



# Master Thesis

Simon Chemnitz-Thomsen

## Reducing the Need for General Anesthesia in Children Undergoing Neuroimaging by Preparation and Motion Correction

Supervisor: Melanie Ganz-Benjaminsen, Associate Professor  
University of Copenhagen, Department of Computer Science

Examiner: Kim Knudsen, Professor  
Technical University of Denmark, Department of Applied Mathematics and  
Computer Science

Date: December 20<sup>th</sup>, 2021

This thesis has been submitted as part of the requirements for the degree  
Master of Science in Statistics (Cand.scient. i statistik)



---

## Abstract

---

Magnetic Resonance Imaging (MRI) is an important medical diagnostic tool. MRI suffers from long acquisition times on the scale of 30-60 minutes, which can lead to motion induced artefacts, especially in pediatric scans. As such, younger children are often scanned under General Anesthesia (GA) to prevent motion during image acquisition. An MR scan with GA requires additional personnel, occupies the scanner for, up to, six times as long as a regular scan and gives rise to concerns about possible health risks related to GA, such as cognitive neurodegeneration. Many methods to reduce GA in pediatric scans are being researched, such as preparation/training with a faux scanner and Motion Correction (MoCo).

Naturally one wishes to determine the effectiveness, of both preparation and MoCo. The aim of this thesis is to evaluate the effect of preparation on anxiety levels in children, particularly if using a mobile phone application beforehand has an influence, and measure the quality of MR images using image quality metrics in an attempt to imitate observer scores as much as possible.

I will demonstrate that preparation/training positively affects children's anxiety levels, as they drop significantly during the course of preparation. When modeling head displacement of the children inside the faux/mock scanner we found that age is a significant predictor indicating that older children move less. The age of the children and whether or not they used an app before the session proved to be significant predictors when modelling the time it took for a child to willingly enter the mock scanner. Indicating that elder children and children who used a mobile phone app beforehand needs less time before entering the mock scanner.

Three mathematical based non-reference metrics, Average Edge Strength (AES), Co-occurrence Entropy (CoEnt) and Tenengrad (TG) were implemented and evaluated on more than 800 MR images, with 500 of the images scored with the golden standard, expert observer assessment. The Spearman correlation between metric scores and observer scores was calculated. This showed that AES had the highest correlation to observer scores and in addition AES had clearer distinction between scores when MoCo was on compared to off. Thus it is recommended to use AES should one choose between AES, TG or CoEnt.



---

## Preface

---

This thesis was written at Neurobiology Research Unit, Copenhagen University Hospital, Rigshospitalet as part of the requirements for the degree Master of Science in Statistics.

The Thesis was formulated in span of four months from August 20<sup>th</sup> 2021 to December 20<sup>th</sup> in accordance with the requirements for a 30 ECTS Master Thesis.

The enclosed thesis was handed in on December 20<sup>th</sup> 2021.

**Supervisor**

Melanie Ganz-Benjaminsen, Associate Professor, University of Copenhagen, Department of Computer Science, Rigshospitalet NRU

**Examiner** Kim Knudsen, Professor

Technical University of Denmark, Department of Applied Mathematics and Computer Science

December 20<sup>th</sup> 2021

A handwritten signature in black ink, appearing to read "Simon Chemnitz". It is enclosed in a thin rectangular border.

Simon Chemnitz-Thomsen



---

## Acknowledgements

---

**Associate Professor Melanie Ganz-Benjaminsen:** Best supervisor one can wish for, giving encouraging and invaluable feedback, advice, enthusiastic involvement in the project and always having time for questions. I am especially thankful for the amazing opportunity to be part of the MoCo group and NRU environment for the last four months and getting a peek behind the curtain of the MRI scanner and procedures.

**Research Assistant Hannah Eichhorn:** For spending countless hours giving feedback and suggestions for my thesis regarding formatting, physics, structure and help with coding and internal systems.

**Post Doc Cheng Teng Ip:** For helping me with statistics of repeated measurements both theoretical and practical implementations in R and SPSS.

**My family:** For support throughout the course of this thesis and with the invaluable feedback and advice regarding formatting, structure and grammar.



---

## Nomenclature

---

The list below describes several symbols and abbreviations that will be used within this thesis

$\delta$  RMS head displacement

$\gamma$  Gyromagnetic Ratio

$\hbar$  Planck Constant

$\omega_0$  Larmor Frequency

$\tau$  Mock scan preparation time

$s$  Spin State, e.g. spin  $\frac{1}{2}$

AES Average Edge Strength

ANOVA Analysis Of Variance

arb'U Arbitrary Unit

CoEnt Co-Occurrence Entropy

CP Cerebral Palsy

EMM Estimated Marginal Means

FID Free Induction Decay

FLAIR Fluid Attenuated Inversion Recovery

fMRI Functional MRI

FOV Field of View

GA General Anesthesia

MoCo Motion Correction

MPRAGE/MPR Magnetization-Prepared Rapid Gradient-Echo

- MRI Magnetic Resonance Imaging  
NMV Net Magnetization Vector  
NRU Neurobiology Research Unit  
pMC Prospective Motion Correction  
PST Phase Stretch Transform  
RF Radio Frequency  
rMC Retrospective Motion Correction  
RRF Rotating Reference Frame  
STAI-CH State-Trait Anxiety Inventory for Children  
TIRM Turbo Inversion Recovery Magnitudes  
TSE Turbo Spin Echo

---

# Contents

---

<b>Nomenclature</b>	<b>VIII</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Background</b>	<b>5</b>
2.1 Clinical Pediatric Neuroimaging . . . . .	5
2.2 Characteristics of MRI . . . . .	6
2.2.1 Nuclear Magnetic Resonance . . . . .	6
2.2.2 Excitation . . . . .	8
2.2.3 Relaxation . . . . .	9
2.2.4 Image Formation . . . . .	11
2.3 Motion Induced Artifacts . . . . .	13
2.4 Motion Correction . . . . .	14
2.4.1 Prospective Motion Correction . . . . .	14
2.4.2 Retrospective Motion Correction . . . . .	15
2.5 MRI & Image Terminology . . . . .	15
2.6 Image Quality Metrics . . . . .	16
2.6.1 Co-Occurrence Entropy . . . . .	16
2.6.2 Average Edge Strength . . . . .	18
2.6.3 Radiographers Visual Assessment . . . . .	20
2.7 Image Processing . . . . .	21
2.8 STAI-CH . . . . .	23
2.9 Statistical Principles . . . . .	23
2.9.1 Notation . . . . .	23
2.9.2 Repeated Measures ANOVA . . . . .	23
2.9.3 General Linear Model . . . . .	25
2.9.4 STAI-CH Example . . . . .	26
2.10 Head Displacement Metrics . . . . .	27
<b>3 Methods</b>	<b>29</b>
3.1 Adult Healthy Controls . . . . .	29
3.1.1 Data . . . . .	29
3.1.2 Analysis . . . . .	30
3.2 Pediatric Healthy Controls . . . . .	32

3.2.1 Data . . . . .	32
3.2.2 Analysis . . . . .	33
3.3 Ongoing Clinical Trial . . . . .	34
3.3.1 Data . . . . .	34
3.3.2 Analysis . . . . .	36
<b>4 Results</b>	<b>37</b>
4.1 Adult Healthy Control . . . . .	37
4.2 Pediatric Healthy Control . . . . .	46
4.3 Pediatric clinical Trial . . . . .	52
<b>5 Discussion</b>	<b>55</b>
5.1 Adult Healthy Control study . . . . .	55
5.1.1 Summary of strengths and weaknesses . . . . .	57
5.2 Pediatric Healthy Control study . . . . .	58
5.2.1 Summary of strengths and weaknesses . . . . .	59
5.3 Pediatric clinical trial . . . . .	60
5.3.1 Summary of strengths and weaknesses . . . . .	60
<b>6 Conclusion &amp; Outlook</b>	<b>63</b>
6.1 Conclusions . . . . .	63
6.2 Outlook . . . . .	64
<b>List of Figures</b>	<b>67</b>
<b>List of Tables</b>	<b>69</b>
<b>Bibliography</b>	<b>71</b>
<b>Appendices</b>	<b>77</b>
<b>A Additional Results</b>	<b>79</b>
A.1 Adult Healthy Control . . . . .	79
A.2 Pediatric Healthy Control . . . . .	82
A.3 Clinical Trial . . . . .	85
<b>B Supplementary Material</b>	<b>87</b>
B.1 Image Representation . . . . .	87
B.2 K-space and Spatial Domain . . . . .	89
B.3 MRI Slices and Sequences . . . . .	90
<b>C STAI-CH</b>	<b>93</b>

# Chapter 1

---

## Introduction

---

Magnetic Resonance Imaging (MRI) is a powerful medical imaging modality which generates 3D images in high detail, especially in soft tissues for instance in the brain. As such MRI is a corner stone in modern medicine and its advancement, particularly in neuroscience [Chang YC, 1999]. MRI is used as a diagnostic test for a variety of conditions such as, brain tumors, epilepsy, cerebral palsy (CP) and more. It is advantageous to other imaging modalities as it has excellent soft tissue contrast and at the same time being non-invasive to the patient, as ionizing radiation is not used. MRI might seem perfect but it is not without pitfalls. A major difficulty of MRI is long acquisition times which will prove a great inconvenience.

Image acquisition times of 30-60 minutes are not unusual, thus MRI is very sensitive to motion. Subjects must lie near perfectly still for longer duration as to not introduce image artifacts. This may prove challenging for younger patients.

In addition, an MR scan can induce anxiety in children as it produces loud noises and has a confined bore, see Figure 1.1 for an image of an MR scanner.

This would often lead to motion induced artifacts such as blurring, ringing and ghosting [Zaitsev et al., 2015]. Hence General Anesthesia (GA) is often used to keep younger children still during a scan. At Rigshospitalet most children under the age of 8-10 years are generally anaesthetised. 1,487 children were scanned with GA from March 2018 to April 2021, 873 of those were cerebrum scans, see Table 2.1 for an overview of how many children got a GA cerebrum scan that last couple of years.

Putting children in general anesthesia introduces new difficulties such as logistical, financial and worst of all, health related risks. Immediate health complications as a result of GA are rare, however studies have demonstrated long term cognitive deficiencies in animals [Stratmann, 2011] and language disabilities in children exposed to anesthesia [Ing et al., 2012]. This naturally leads to rising concerns about possible neuro-degeneration and long term cognitive deficiencies, especially since some children, e.g. with CP or tumors, routinely get multiple scans.

As such Motion Correction (MoCo) has been a research field of great interest with one of the earliest papers published in 1986, where correction of respiratory mo-

tion was studied [Haacke and Patrick, 1986]. Since then a number of techniques and methods have been researched, such as retrospective (rMC) machine learning methods [Duffy et al., 2021, Chatterjee et al., 2020] and prospective (pMC) methods such as continuous update of scanner field of view and selective reacquisition based on motion tracking [Hannah Susanne Eichhorn, 2021, Frost et al., 2019, Slipsager et al., 2019].

Another approach to minimize or avoid motion artifacts is training the patients using a mock scanner. Training can include an introduction to an MR scanner and its various sounds and the experience of lying in a scanner [De Bie et al., 2010]. They showed that training is a viable approach, where children prepared with a mock scanner reduced motion during the real MR scan. By proper preparation 81/90 structural scans were of diagnostic quality and 30/43 functional MRI (fMRI) were of sufficient quality. In addition, [Törnqvist et al., 2015] reported 30/33 trained children had acceptable scans after they were prepared with a story booklet and introduced to the sounds of an MRI scanner.

It is also possible to prepare children with a mobile phone app that introduces them to the MRI scanner the hospital environment and what to expect when getting an MR scan. This has already been successfully implemented at 3 UK institutions [Ashmore et al., 2019]. This approach has also been shown viable by [Runge et al., 2018] where 95% of those receiving proper training (including interacting with an app) had a scan of diagnostic quality while only 45% of the control group had high enough quality. It was estimated that 1,000€ Euro could be saved in financial costs per child who is properly prepared and receive a scan without GA.



**Figure 1.1:** MRI scanner at Rigshospitalet. The confined bore is the circular opening where the patient is placed during scanning. *Photo courtesy of The MoCo project at Rigshospitalet, Copenhagen, Denmark.*

The MoCo project at Neurobiology Research Unit (NRU) Rigshospitalet studies these methods, described above. Specifically prospective-motion correction coupled with state of the art continuous markerless tracking is researched, in addition to preparation with a mock scanner and an app.

It is thus of interest to measure the effect of both motion correction and training with a mock scanner, especially as the project is entering the clinical trial stage (started in October 2021).

I will within this thesis study and examine two distinct methods for evaluating the effectiveness of training and motion correction on MR image quality. I will describe and examine two image quality metrics and their ability to capture image artefacts and how well they correlate to the gold standard of quality metrics namely observer scores given by educated radiographers. The second analysis carried out is a statistical analysis of anxiety levels in children undergoing training/preparation. The intent is to examine changes in anxiety levels and determine what effect, if any, it has when using an app beforehand.

This thesis is partitioned into six chapters including this chapter. In the second chapter a brief background on the physics of MRI, motion artefacts and methods of motion correction. Lastly image quality metrics are outlined, in addition, statistical principles of repeated measures analysis is given, specifically repeated measures ANOVA and the general linear model is touched upon.

The third chapter covers the methods used to conduct the relevant analysis. The chapter starts with a description of the available data that lay the foundation for the analysis carried out in this thesis.

The fourth chapter presents the results obtained by analysing the data outlined in the succeeding chapter. Starting with the metric results of MR images and ending with a statistical analysis of anxiety levels in children undergoing training/preparation.

The results are followed by a discussion in chapter five and the thesis is finally concluded in chapter 6 with a conclusion and outlook.



# Chapter 2

---

## Background

---

In this chapter I will give the background and theory for this thesis starting with a case study of MR imaging children and its difficulties. This is followed by the theoretical background of MRI starting at a fundamental level leading to how images are formed from received nuclear resonance signals. The MRI physics precede the section of motion artefacts and motion correction methods as the basic understanding of physics is needed to grasp why motion is major drawback and eventual methods of correcting it.

Being able to measure image quality at scale is enticing and relevant as any correction method can be evaluated at scale if a satisfactory metric is used. I will describe two image quality metrics in detail namely average edge strength and co-occurrence entropy.

Lastly statistical principles of repeated measurements is briefly described detailing the general linear model and repeated measures ANOVA. A short example is given to familiarize the reader with the notation used.

### 2.1 Clinical Pediatric Neuroimaging

As briefly outlined earlier, long acquisition times are the major drawback of MRI. Naturally, younger children have trouble lying still for the duration of a scan which will produce images of bad quality and in worst case a re-scan is needed. For this reason children at Rigshospitalet are put in General Anesthesia (GA). An excerpt from the Department of Radiology Rigshospitalet can be seen in Table 2.1 which provides an overview of how many children underwent an MR scan with GA for different age groups. As seen in the table hundreds of children are scanned every year with GA with more than 300 children aged 5-12 years the past four years. The Motion Correction group (MoCo-Group) at Neurobiology Research Unit Rigshospitalet (NRU) aims to drastically reduce this number by both preparation and distraction of the children for example by giving a mock scanning session and playing a movie during the clinical scan. Clinical trials for this project started as of October 2021.

**Table 2.1:** Pediatric cerebrum scans under General Anesthesia (GA). An additional 614 children had an MRI under GA although not cerebrum. The table indicates that hundreds of children are scanned every year with GA. The number of children getting a GA scan in the age group 5-12 years old are particularly interesting as they are the main target for the MoCo group, to receive training/preparation in order to get an MR scan without GA.

Year	Age						Total
	0-2	3-4	5-6	7-8	9-10	11-12	
2018(March-December)	93	36	30	15	11	0	185
2019	132	62	43	25	19	15	296
2020	122	56	53	28	15	18	292
2021(January-April)	37	21	20	9	7	6	100
Total	384	175	146	77	52	39	873

A pilot scan and preparation was done on an 8 year old boy prior to the clinical trials. He was scheduled for an MR scan without GA. The child went through the same training and preparation procedures as future participants of the clinical trials, that is, going through a mock scanner (A 1:1 scanner model with no magnet or coils and is visually indistinguishable from a real MR scanner) training session and preparation with an app at home. The training session included playback of various noises the MR scanner would make for different sequences. Starting at lower volume levels and gradually increasing to a volume level that is experienced at a real scan. A movie was played as a method of distraction both during the real scan but also during mock scanner training.

This proved a viable method of distraction as the child was unconcerned about the hospital environment during the real scan and only cared about continuing the movie. The child moved so little that the radiographer decided the scan was of diagnostic quality and forwarded the scans to the appropriate radiologists.

## 2.2 Characteristics of Magnetic Resonance Imaging

A basic understanding of the theory and mechanics of MRI and motion artefacts is needed if one wishes to fix, correct or measures images for motion degradation. In this subsection I briefly outline the necessary theory needed to get this understanding starting with the underlying fundamental physics followed by excitation, relaxation and lastly how an image is formed from measured signals.

### 2.2.1 Nuclear Magnetic Resonance

Atoms and fundamental particles have a magnetic moment proportional to its spin angular momentum,  $\vec{J}$  [Richard Feynman et al., 1969].

Consider a hydrogen atom (single proton, omitting the electron). It has spin and positive charge, thus it will generate a magnetic field and possess a magnetic moment [Lambert and Mazzola, 2003], here denoted by  $\vec{\mu}$ .

The relationship between the magnetic and angular moment is described by:

$$\vec{\mu} = \gamma \vec{J} \quad (2.1)$$

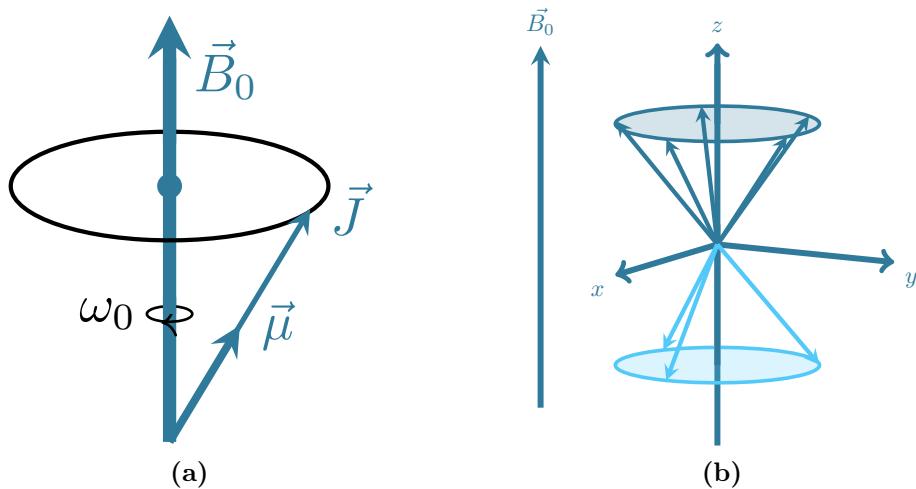
where  $\gamma/2\pi$  denotes the gyromagnetic ratio measured in MHz/Tesla [Dhawan, 2010]. For hydrogen, the gyromagnetic ratio is 42.58 MHz/T.

The nuclear magnetic resonance (NMR) experiment entails to subjecting nuclei, most commonly Hydrogen, to a strong external magnetic field  $\vec{B}_0$ . This will cause the magnetic moments to precess around  $\vec{B}_0$ . The precession happens at a specific frequency,  $\omega_0$ , proportional to the magnetic field strength  $B_0$ , described by the Larmor equation:

$$\omega_0 = \gamma B_0 \quad (2.2)$$

where  $B_0$  is the strength of the magnetic field measured in Tesla. The frequency,  $\omega_0$ , is referred to as the Larmor frequency [Lambert and Mazzola, 2003]. See Figure 2.1 (a) for an illustration of a single atom undergoing precession around a static magnetic field  $\vec{B}_0$ .

The magnetic field affects the energy levels of the nuclei and their magnetic moments align either parallel ( $z\uparrow$ ) or anti-parallel ( $z\downarrow$ ) to the external magnetic field, see Figure 2.1 (b) for an illustration of multiple spin- $\frac{1}{2}$  nuclei undergoing precession. The magnetic moments tend to align parallel compared to anti-parallel. This will lead to a net magnetization,  $\vec{M}$ , along the magnetic field  $\vec{B}_0$  [Lambert and Mazzola, 2003]. The net magnetization vector (NMV) is defined as the average moment per



**Figure 2.1:** (a) Precession of a single atom with angular momentum  $\vec{J}$  and magnetic moment  $\vec{\mu}$ , in the presence of a static magnetic field  $\vec{B}_0$ . The precession frequency is denoted by  $\omega_0$ . (b) Spin- $\frac{1}{2}$  Nuclei aligned in the presence of an external magnetic field  $\vec{B}_0$ . Each nucleus undergoes precession indicated by the two circles. The proportion of nuclei aligning parallel compared to anti-parallel is greatly exaggerated.

unit volume  $V$ :

$$\vec{M} = \frac{1}{V} \sum_{i=1}^N \vec{\mu}_i \quad (2.3)$$

where  $N$  is the total number of magnetic moments [Richard Feynman et al., 1969, Fallis, 2013].

The equilibrium magnetization magnitude,  $M_0$ , at thermal equilibrium is given by:

$$M_0 \approx \frac{N \gamma^2 \hbar^2 s(s+1)}{V 3kT} B_0 = \frac{N \gamma^2 \hbar^2}{V 4kT} B_0, \quad s = \frac{1}{2} \quad (2.4)$$

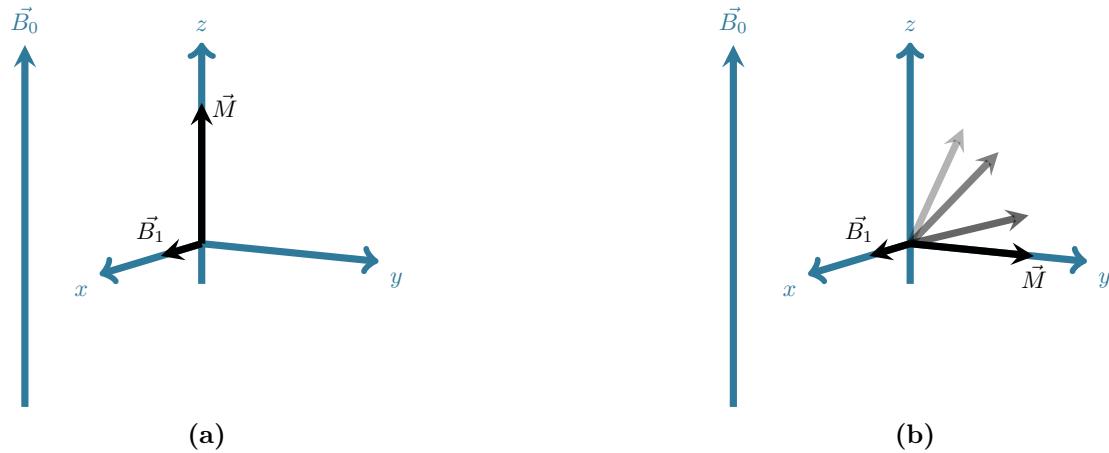
where  $k$  is the Boltzmann's constant,  $T$  the absolute temperature measured in Kelvin,  $\hbar$  the Planck constant and  $s$  the net spin number, which is  $1/2$  in the case of a proton [Fallis, 2013]. For more details on equilibrium magnetization see Chapter 6 in [Fallis, 2013].

Remark: The only term that can be adjusted in order to increase the equilibrium magnetization (and with it the MR signal) is the magnetic field strength  $B_0$ . This drives the desire for higher Tesla scanners.

## 2.2.2 Excitation

The spins are aligned with the magnetic field  $\vec{B}_0$  at equilibrium, thus they do not omit a detectable signal. They need to be excited/tipped in order to detect a signal. We can introduce energy into our spin system to do this. For example by applying a Radio Frequency (RF) pulse  $\vec{B}_1$ . Specifically we need to apply an RF pulse at the Larmor frequency in order to excite the nuclei to exhibit NMR [Dhawan, 2010]. We utilize a rotating reference frame (RRF) to simplify our situation. In the RRF, the coordinate system rotates with the same frequency as  $\vec{B}_1$  (Larmor frequency). The  $\vec{B}_1$  field will look static in the rotating frame. Figure 2.2 shows the NMV as well as the  $\vec{B}_1$  field in the RRF.

Low energy nuclei (spin up  $\uparrow$ ) can receive energy when exposed to the RF pulse  $\vec{B}_1$ . This can cause them to "flip" [Dhawan, 2010]. The force,  $\vec{F}$ , exerted on  $\vec{M}$  is given by  $\vec{F} = \vec{M} \times \vec{B}_1$ , hence  $\vec{M}$  will move away from the z-axis and align with the y-axis as seen on Figure 2.2 where the NMV is shown prior and after application to the  $\vec{B}_1$  field. The NMV will flip 90 degrees if enough energy is applied. This is known as a  $90^\circ$  RF pulse. If even more energy is applied it can flip more, this is the  $180^\circ$  RF pulse.



**Figure 2.2:** Both figures are in the rotating reference frame with an external static magnetic field  $\vec{B}_0$ . (a) Net Magnetization Vector (NMV here denoted  $\vec{M}$ ) prior to application of the  $\vec{B}_1$  field. (b) NMV ( $\vec{M}$ ) after application of the  $\vec{B}_1$  field. The NMV aligns with the y-axis instead of the z-axis.

### 2.2.3 Relaxation

Relaxation happens when the RF pulse stops, the nuclei relaxes and the NMV returns to its equilibrium state. During this phase energy is released and detectable in the radio frequency coils [Dhawan, 2010, Lambert and Mazzola, 2003].

There are mainly two forms of relaxation,  $T_1$  and  $T_2$  relaxation.  $T_1$  relaxation is usually referred to as longitudinal or spin-lattice relaxation, while  $T_2$  is referred to as spin-spin or transverse relaxation.

**Spin-lattice relaxation** is the process of realignment of the magnetic moments with the  $\vec{B}_0$  field, along the z-axis. Thus the relaxation time depends on the magnetic field strength  $B_0$  [Weishaupt and Marincek, 2008].

More specifically spin-lattice relaxation is a result of local magnetization caused by movement of magnetic nuclei [Lambert and Mazzola, 2003]. These local magnetic fields oscillate as the nuclei is in motion. Some spin  $\downarrow$  will transition to spin  $\uparrow$  if the frequency coincide with the Larmor frequency. This generates an initial excess of spin  $\uparrow$  nuclei. Equilibrium ratio of spin  $\uparrow$  and spin  $\downarrow$  is eventually met [Lambert and Mazzola, 2003]. Spin-lattice relaxation is described by the constant,  $T_1$ , in the equation:

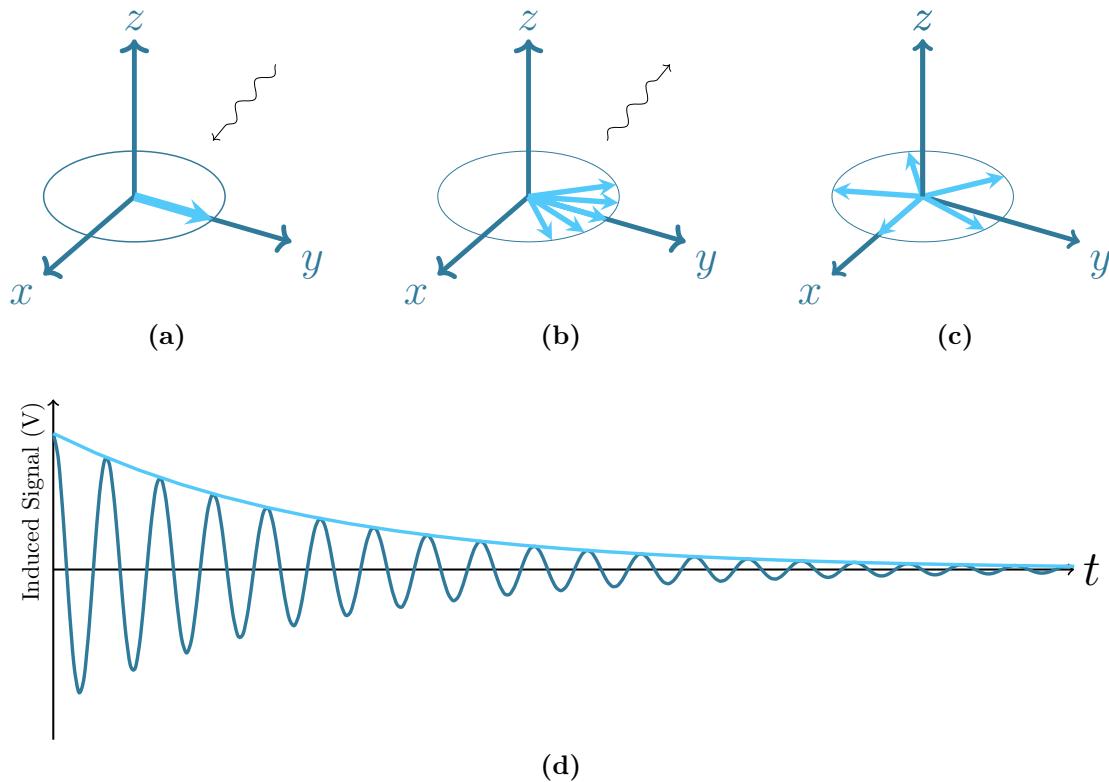
$$M_z(t) = M_0 \left( 1 - \exp \left( -\frac{t}{T_1} \right) \right) \quad t > 0 \quad (2.5)$$

where  $t$  is the time after application of an RF pulse, and  $M_0$  is defined as in equation (2.4) [Fallis, 2013]. The time constant  $T_1$  for water is several seconds while around one second for brain tissue [Talagala and Lowe, 1991]. I want to emphasize that it is the differences in  $T_1$  (and in  $T_2$ ) constants from different tissues that generate the contrasts in an MR image.

**Spin-spin relaxation** refers to the process of two nuclei with opposite spins exchanging spin values (spin  $\uparrow$  to spin  $\downarrow$  and vice versa) [Lambert and Mazzola, 2003]. This will not result in magnetization change along the z-axis. Changing spin states will lead to a phase change, thus dephasing will occur. The detected signal is a damped oscillation often referred to as Free Induction Decay (FID), see Figure 2.3 for an illustration of dephasing and the measured FID signal. Spin-spin relaxation can be described by the time constant  $T_2$ :

$$S(t) \propto \sin(\omega_0 t) \exp\left(-\frac{t}{T_2}\right), \quad t > 0 \quad (2.6)$$

where  $S(t)$  is the signal at time  $t$  and  $\omega_0$  the Larmor frequency [Fallis, 2013]. The sinusoidal expression is omitted in the case of a rotating reference frame. Brain tissue has  $T_2$  relaxation times of around 100 ms while it is several seconds for water [Talagala and Lowe, 1991].

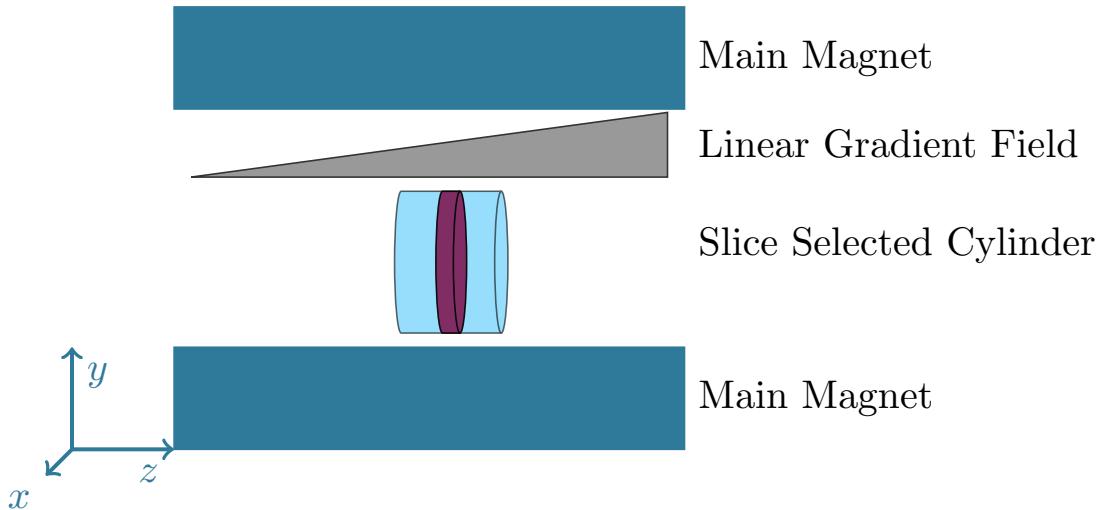


**Figure 2.3:** The transverse magnetic moments in bright blue with a circle indicating precession, in the presence of a static external magnetic field along the z-axis in the same fashion as Figure 2.1, 2.2. (a) Nuclei in a state of phase coherence during excitation of an RF pulse. (b) The transversal magnetic moments will start to dephase after excitation and with it the signal will dampen. (c) The transversal magnetic moments will cancel omitting no detectable signal. (d) The detected signal fades exponentially, on a scale of  $T_2$ , over time,  $t$ , after excitation. This is the damped oscillation referred to as Free Induction Decay (FID). Equation (2.6) describes the dark blue signal while the bright blue represents the signal in the RRF.

## 2.2.4 Image Formation

Suppose we want to perform an MR-scan of a patient's head. The patient is placed in the bore and their protons excited by an RF-pulse at the Larmor frequency and we measure the signal released during relaxation. All the protons in the head will have the same Larmor frequency as the main magnetic field,  $\vec{B}_0$ , is homogeneous. As such the received signal originates from the entire head thus we would have no way of inferring spatial location.

Magnetic gradient field coils are used for selective excitation and subsequently measure signals from a specific slice of the brain. The gradient coils generate additional magnetic fields called gradient fields. This will change the Larmor frequency linearly, recall equation (2.2):  $\omega_0 = \gamma B_0$ . With this we can excite a specific slice using an RF-pulse matching the relevant frequency, while the rest of the head is



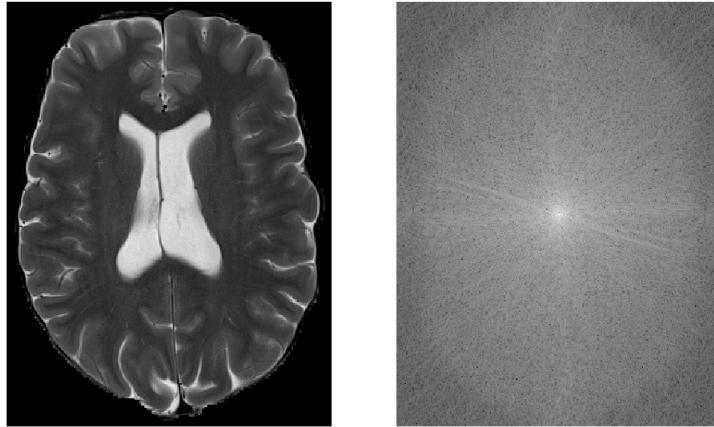
**Figure 2.4:** Gradient slice selection of a cylinder. In gray the gradient magnetic field, in dark blue the main magnet and in light blue a cylinder. A specific slice of the cylinder (in red) is excited when applying an RF-pulse at the relevant frequency. The frequency depends on the position as the gradient field is linear. As such only a specific slice will be excited at a given frequency thus any signal received will originate from the slice. This can be extended with two additional gradient fields along the *x* and *y*-axis. This allows for spatial encoding within the excited slice, which is used to determine where a signal originates from within the slice.

unaffected. This method is referred to as Gradient Slicing see Figure 2.4 for an illustration of slice selecting a cylinder using a linear gradient field.

Phase- and frequency-encoding is used to further infer spatial position, within the slice, of the received signal. This is done by introducing the gradient coils,  $G_x, G_y$  in the *x* respectively *y* direction.

Phase-encoding uses the gradient coil  $G_y$  which is turned on after excitation [Weishaupt and Marincek, 2008]. This will induce a phase shift along the *y-axis*.  $G_x$  is used for frequency encoding along the *x-axis*. For example, the Larmor frequency will increase along the *x-axis*.

Thus each signals spatial location is encoded by its frequency and phase. This leads to image formation in *k*-space (Fourier/Frequency domain) see Figure 2.5 for an illustration of an image in spatial domain and *k*-space.



**Figure 2.5:** On the left an MR image in spatial domain (how we usually see any image, medical or otherwise) and on the right the same image but in k-space (or Fourier/Frequency domain). The center and peripheral of k-space determines the overall shape and details respectively of a given image.

## 2.3 Motion Induced Artifacts

The human body produces different kinds of motion that can influence the MR image. This includes periodic motion such as breathing or heartbeats, sudden movements as coughing, sneezing or swallowing. Some diseases can also induce involuntary movements, essential tremor for example. Lastly there are deliberate movements, e.g., body adjustment.

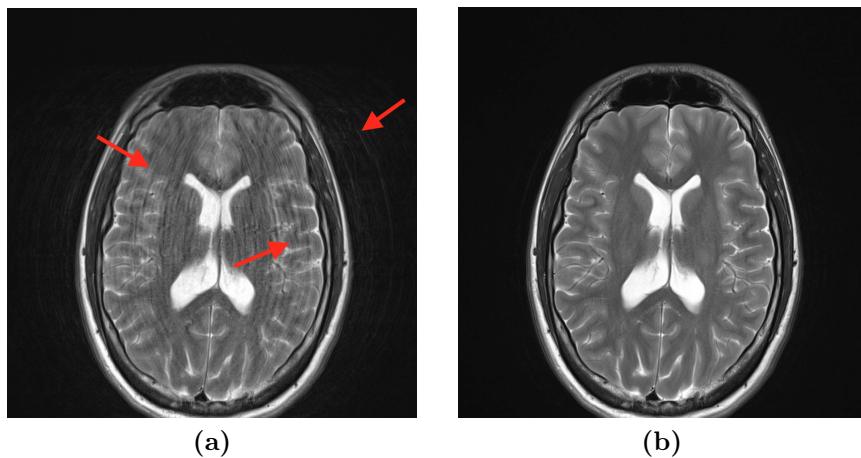
MR images are collected in k-space as opposed to directly in image space. The center of k-space defines the overall shape of an image while the peripheral determines the finer details.

The Fourier transform is applied to transition from frequency- to image-space. Thus every contribution from k-space corresponds to a contribution of a wave to all pixels in image-space [Zaitsev et al., 2015].

Common artifacts include blurring, ringing and ghosting, where the latter is usually caused by periodic movements [Zaitsev et al., 2015, Godenschweger et al., 2016]. See Figure 2.6 for an illustration of artifacts caused by motion where an image with artefacts such as ringing and ghosting is compared to a clean image.

Translation will cause a linear phase ramp of the frequency-space data in the direction of movement, while rotation will lead to an identical rotation in frequency-space data [Godenschweger et al., 2016].

Sampling schemes also has an influence on image artifacts, as image acquisition takes place in frequency-space. For example a "propeller" sampling scheme is less sensitive to motion than its counterparts as it over samples the frequency-space center [Zaitsev et al., 2015]. Although drawbacks of "propeller" scheme is reduced efficiency and sensitivity to gradient timing delays [Godenschweger et al., 2016].



**Figure 2.6:** Two MR images of the same person. During image acquisition of the first image a nodding motion was performed while the latter had no motion during scanning. (a) An MR image with clear blur and ringing artefacts caused by a repeating nodding motion. Examples of the artefacts are annotated by red arrows. (b) An MR image with no apparent motion artefacts.

## 2.4 Motion Correction

As previously mentioned MRI is quite sensitive to motion, if it cannot be avoided motion correction needs to take place. Motion correction (MoCo) can take place during image acquisition, this is called prospective motion correction (pMC), or after image acquisition and is hence called retrospective motion correction (rMC). Motion correction has been a research topic for the past 30 years, with one of the earliest articles published in 1989. Mark Haacke and John Patrick, studied motion correction of MRI related to respiratory motion [Haacke and Patrick, 1986].

### 2.4.1 Prospective Motion Correction

Prospective motion correction is performed during image acquisition, while the subject is in the scanner. The idea of pMC is to continuously update the imaging field of view (FOV) [Slipsager et al., 2019], and can be aided by selective re-acquisition in severe cases of motion. Selective re-acquisition is the process of re-acquiring parts of k-space where motion, during original acquisition, is deemed too severe. A method of motion tracking is essential for both correction methods.

Tracking can be done using a markerless setup, for example using a structured light technique as implemented by *Tracoline* (TracInnovations, Ballerup Denmark) [Frost et al., 2019, Olesen et al., 2012]. As pMC is reliant on motion tracking, a natural limitation could be inadequate quality of the tracking, for example if the tracking system is "tricked" by exaggerated facial expressions.

### 2.4.2 Retrospective Motion Correction

In retrospective motion correction, image acquisition is performed regardless of eventual motion. As a result any modifications to the MR image is done in frequency-space or directly in image-space, after the scan has concluded.

There are several methods of retrospective motion correction. These include machine learning methods such as convolutional neural networks [Duffy et al., 2021] or model based methods [Haskell, 2019].

Retrospective motion correction can, similar to prospective, be aided by methods of tracking.

A basic method of motion correcting rigid body movements, such as translation and rotation, can be achieved using Fourier theory, as object rotation in spatial domain corresponds to a phase shift in frequency/Fourier domain, for more details on this method see Section 4.1 in [Godenschweger et al., 2016].

A limitation of rMC is naturally that only the finalized image is available and thus it relies almost solely on "editing" the image.

## 2.5 MRI & Image Terminology

An MRI volume can be interpreted as a 3 dimensional image and it could, for example, have the shape/size/resolution  $256 \times 256 \times 18$ . This can be viewed as 18 images stacked on top of each other, with each image having size/resolution  $256 \times 256$ . We say that the volume has 18 slices with each slice corresponding to a  $256 \times 256$  image. 2D encoded images usually have a low number of slices, for example 18-26 slices, while 3D encoded sequences have many slices usually 256. A 3D encoded MRI volume could thus have the shape  $256 \times 256 \times 256$  which is particularly more interesting as it allows for three ways to slice the volume (compared to only one for 2D encoded image). The different ways of slicing the volume are: sagittal (from the side), coronal (from the front) and axial (from the top), see Appendix Figure B.6 for an illustration of these different slice views. All images in this thesis are in axial view as 2D encoded images usually can only be viewed in axial view. Let us denote a 3D image by  $V$ . When writing  $V(i, j, k)$  we refer to the intensity value at the location  $(i, j, k)$  that is the value at coordinate  $(i, j)$  for slice  $k$ . This can be any integer in the range  $[0, 255]$  where zero corresponds to black and 255 to white.

As previously outlined, there are two forms of relaxation both leading to different types of images. These images are called  $T_1$  and  $T_2$  weighted images.  $T_1$  sequences will have white matter darker than gray matter and vice versa for  $T_2$ . The specific sequences and protocols are beyond the scope of this thesis but a more detailed explanation can be found in [Hannah Susanne Eichhorn, 2021] who outlines different protocols and sequences in great detail accompanied by well constructed illustrations, however Appendix Figure B.7 has a visualisation of three different sequences.  $T_1$ -weighted MPRAGE and  $T_2$ -weighted TSE are the most important regarding this thesis ( $T_1$ -MPRAGE is needed for a preprocessing pipeline and  $T_2$ -TSE is one of the most used sequences and in all MRI protocols at Rigshospitalet) and the existence of different image sequences should be kept in mind when re-

viewing the results.

When having multiple scans of the same person one would often like the images to be transformed such that all images are aligned. This is referred to as registration and is performed as a preprocessing step. In broad terms, one image is kept fixed and is called the target image and any eventual secondary images are the moving images. These images are moved via translation rotation and scaling such that it aligns with the targets image as much as possible.

## 2.6 Image Quality Metrics

In this section I will give an overview of image quality metrics, explain why good metrics are desired and give examples of metrics used in literature. I describe two mathematical metrics in detail namely average edge strength and co-occurrence entropy which will be evaluated on MR scans. We chose these metrics as they have previously been evaluated on pMC images before with promising results [Pannetier et al., 2016]. The gold standard observer scores are also briefly touched upon including the Likert scale.

The need for image quality metrics arise naturally when, either optimizing or evaluating methods of image processing and correction. Clearly it is needed to assess if the methods are making the image worse or not. Manual expert scoring is time consuming and expensive, hence not viable at a larger scale unlike metrics.

There are mainly two categories of image quality metrics, those with a reference image and those without. A reference image is an image which is considered the ground truth or simply a perfect representation of what the subject image should resemble.

Some of the most common reference image metrics include structural similarity index (SSIM) [Wang et al., 2003, Wang et al., 2004] and peak signal to noise ratio (PSNR) [Topiwala, 1998], although more complex methods have also been proposed, such as perceptual difference models [Miao et al., 2008]. Examples of no reference image metrics includes, average edge strength, co-occurrence entropy [Aksoy et al., 2012, Pannetier et al., 2016, Sciarra et al., 2021] and analysis of noise distribution in the background proposed by [Mortamet et al., 2009]. Model based methods such as deep convolutional neural networks has also been implemented as a method of evaluating the quality of structural brain MR images [Sujit et al., 2019]. Although one could argue that such model based methods lean towards reference metrics and thus they are not completely reference free.

### 2.6.1 Co-Occurrence Entropy

Co-occurrence Entropy (CoEnt) is a Haralick feature based [Robert M Haralick, 1973] metric and as the name suggests it is based on the entropy of the co-occurrence

matrix. CoEnt is given by:

$$CoEnt = - \sum_{p=0}^{255} \sum_{q=0}^{255} \bar{C}(p, q) \log_2 (\bar{C}(p, q)), \quad (2.7)$$

where  $p, q$  are indexed by the different intensity values (for example in a black image the only intensity values are zero and for a white image the only intensity values are 255) in the volumetric image and  $\bar{C}$  is a normalized 2D distribution of the co-occurrence matrix  $C$ .

Remark: One can treat  $\bar{C}(p, q) \log_2(\bar{C}(p, q)) = 0$  in the special case of  $\bar{C}(p, q) = 0$ . That is, the limit of  $x \log_2(x)$  is considered instead, as  $\log_2(0)$  is not well-defined. Indeed the following holds:

$$x \log_2(x) \rightarrow 0 \quad \text{for } x \rightarrow 0, \quad x > 0. \quad (2.8)$$

The curious reader is referred to Example 2.11 in [Ernst Hansen, 2017].

The co-occurrence matrix can be parameterized by  $(\Delta x, \Delta y, \Delta z) \in \mathbb{Z}^3$  through the function:

$$C_{\Delta x, \Delta y, \Delta z}(p, q) = \sum_{(i,j,k)} \begin{cases} 1, & \text{if } V(i, j, k) = p \text{ and } V(i + \Delta x, j + \Delta y, k + \Delta z) = q \\ 0, & \text{otherwise} \end{cases} \quad (2.9)$$

$$= \sum_{(i,j,k)} \mathbf{1}_{\{V(i,j,k)=p\}} \cdot \mathbf{1}_{\{V(i+\Delta x, j+\Delta y, k+\Delta z)=q\}}, \quad (2.10)$$

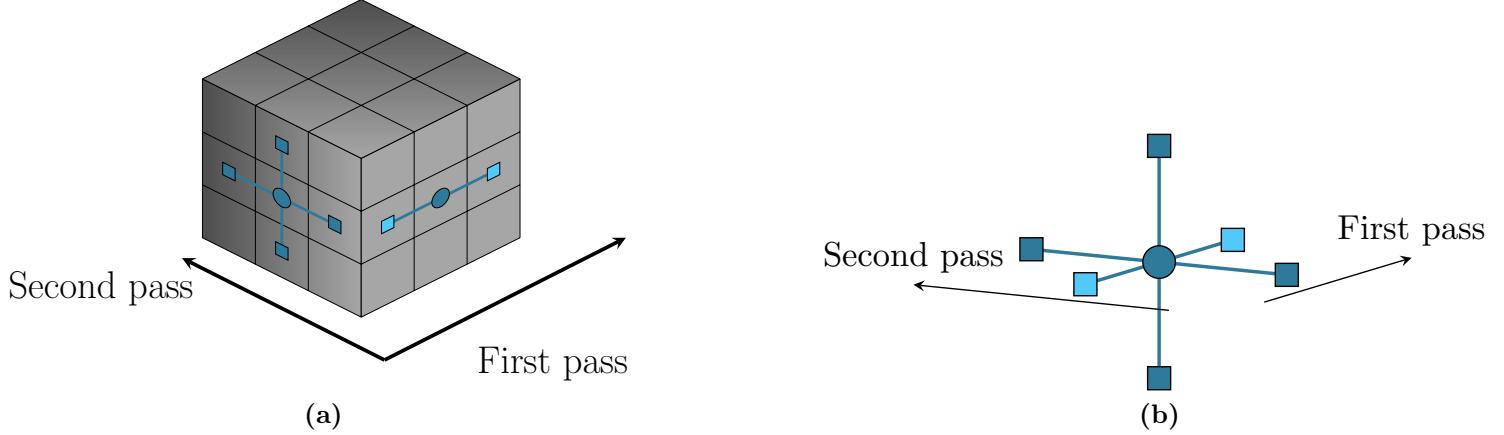
where  $V$  is the volumetric image of interest,  $p$  and  $q$  are intensity values with  $p, q = 0, \dots, 255$ ,  $V(i, j, k)$  the intensity value of  $V$  at coordinate  $(i, j, k)$  and  $\mathbf{1}$  the indicator function [Pannetier et al., 2016].  $\Delta x, \Delta y, \Delta z \in \{-1, 1\}$  corresponds to counting how many of the 6 closest neighbors has intensity value  $q$ . The full co-occurrence matrix for the 6 closest neighbors is then given by:

$$C(p, q) = \sum_{i \in \{-1, 1\}} C_{i,0,0}(p, q) + C_{0,i,0}(p, q) + C_{0,0,i}(p, q), \quad (2.11)$$

For  $p, q = 0, \dots, 255$  Using indicator functions we can rewrite this as:

$$\begin{aligned} C(p, q) &= \underbrace{\sum_i \left( \sum_{(j,k)} \mathbf{1}_{\{V(i,j,k)=p\} \cap \{V(i,j\pm 1,k)=q\}} + \mathbf{1}_{\{V(i,j,k)=p\} \cap \{V(i,j,k\pm 1)=q\}} \right)}_{\text{First pass}} \\ &\quad + \underbrace{\sum_j \left( \sum_{(i,k)} \mathbf{1}_{\{V(i,j,k)=p\} \cap \{V(i\pm 1,j,k)=q\}} \right)}_{\text{Second pass}} \end{aligned} \quad (2.12)$$

This might seem counter-intuitive at first but note the interpretation it leaves. We can write the 3D co-occurrence matrix using two passes of 2D co-occurrence



**Figure 2.7:** Pixel/Voxel of interest represented by a circle, with its neighbors represented by a square. The first pass will count four of the six neighbors while the second pass captures the remaining two. (a) 3-dimensional view of a volumetric image and the 2D procedure highlighted for a single pixel. (b) A voxel of interest, and its six closest neighbors, colored according to which pass they were noted. The dark blue are counted in the first pass and the light blue are counted in the second pass.

calculations. For an illustration of the method see Figure 2.7 where a 3D volume is depicted and the described two passes are visualized for a given slice. Consider, for a fixed value  $i$  the 2D image  $V(i, \cdot, \cdot)$  and assume the function  $\text{com}(\cdot)$  is an optimized function calculating the co-occurrence matrix for a 2D image (i.e. the output is a matrix). The 3D co-occurrence matrix can then be described as:

$$C = \underbrace{\sum_i \text{com}(V(i, \cdot, \cdot))}_{\text{First pass}} + \underbrace{\sum_j \text{com}(V(\cdot, j, \cdot))}_{\text{Second pass}} \quad (2.13)$$

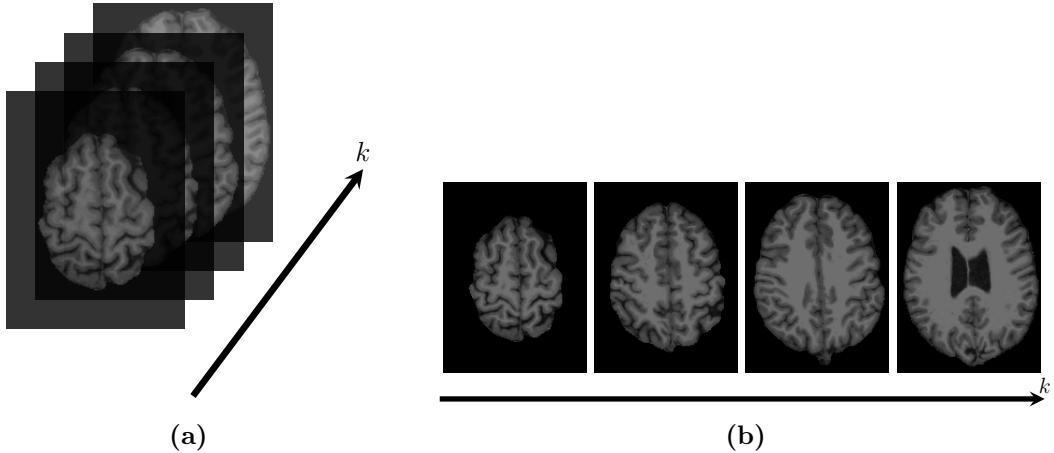
One should take care not to count the same neighboring pixels multiple times, when using this approach. I want to emphasize that for 2D encoded sequences it might not make sense to compare pixel values between slices if the spatial distance between slices is too large.

## 2.6.2 Average Edge Strength

The other metric we examine in detail is Average Edge Strength (AES). It is a gradient based metric and is calculated for each slice  $k$  of the volume image:

$$AES(k) = \frac{\sqrt{\sum_{i,j} E(\mathcal{I}^{(k)})_{i,j} \left[ (G_x(\mathcal{I}^{(k)})_{i,j})^2 + (G_y(\mathcal{I}^{(k)})_{i,j})^2 \right]}}{\sum_{i,j} E(\mathcal{I}^{(k)})_{i,j}} \quad (2.14)$$

where  $\mathcal{I}^{(k)}$  is the  $k'th$  2D image slice e.g.  $V(\cdot, \cdot, k)$  see Figure 2.8 for four different slices of a brain,  $E(\mathcal{I}^{(k)})$  is the binary Canny edge detected image (Canny is



**Figure 2.8:** MRI image slices  $\mathcal{I}^k$  for  $k = 20, 30, 40, 50$ . In (a) the canonical way to picture slices and in (b) a more visually instructive way to show the 4 slices. AES( $k$ ) is calculated for each of the four images and the average is returned.

an edge detection algorithm that thresholds gradient magnitudes of a smoothed image [Canny, 1987]),  $G_x(\mathcal{I}^{(k)})$ ,  $G_y(\mathcal{I}^{(k)})$  are the convolved images of  $I^{(k)}$  with the respective kernels:

$$k_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad k_x = k_y^T \quad (2.15)$$

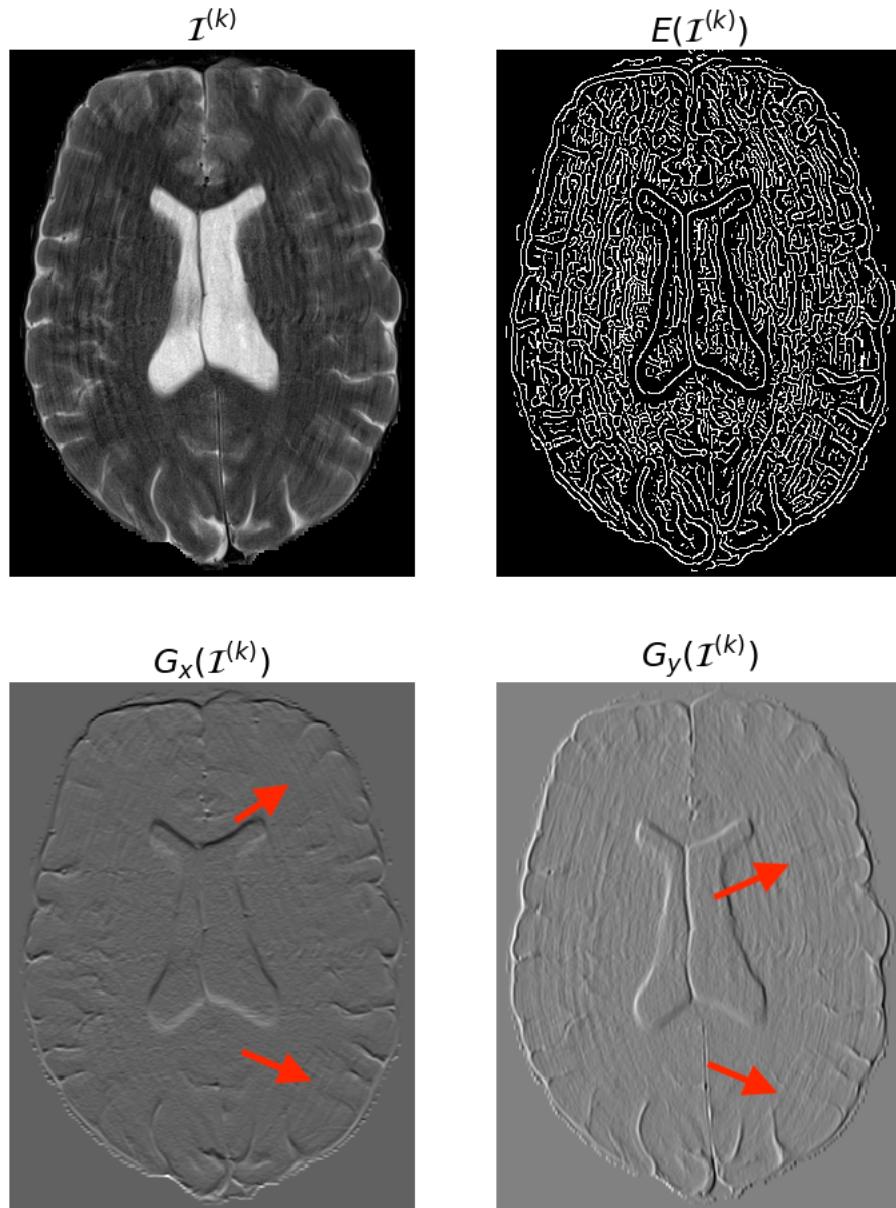
where  $T$  denotes the transpose operator, see Figure 2.9 for an illustration of the different operations, i.e. an application of convolution and Canny on an MR slice. It would seem that convolutions and Canny edge detector captures the ringing in the given picture, this is very desired. The final output is then given by:

$$\frac{1}{K} \sum_{k=1}^K AES(k) \quad (2.16)$$

where  $K$  denotes the total number of slices. The interpretation of  $AES$  is quite nice. We can interpret  $AES$  as the average gradient magnitude along the Canny image, as  $E(\mathcal{I}^{(k)})$  is either zero or one. The numerator is the total gradient strength along the Canny image. The average is then taken over the Canny image and is represented by the denominator as the total edge pixels. The name average comes from the fraction and not the fact that we average over the  $K$  slices.

A lower value of  $AES$  corresponds to higher image blurring.

Remark: AES is inherently modular in the sense that any edge detection can be applied in place of Canny edge detection, e.g. one could use Sobel, Phase Stretch Transform (PST) [Asghari and Jalali, 2015] or a Laplacian based edge detection. This makes AES quite versatile as the edge detection can be specified or adjusted for a given problem.



**Figure 2.9:** Top left, the original image  $I^{(k)}$  (T<sub>2</sub>-weighted TSE), top right the Canny edge detected binary image  $E(I^{(k)})_{i,j}$ , bottom row contains the convolved images as a result of convolution of the original image with the kernels in equation (2.15). It would seem that ringing in the image  $I^{(k)}$  is captured by the convolution operators as annotated with arrows. The Canny edge detected image also seems to capture the ringing, showing overall long strides and ripples.

### 2.6.3 Radiographers Visual Assessment

The gold standard of MRI scoring is the assessment of an experienced radiographer or radiologist. Having extensive knowledge of what different parts of the image should look like and to what extent the quality should be leaves observer scores unparalleled, so far, compared to metrics built from a mathematical foundation. Thus observer scores are often considered the "ground truth". Scoring is usually

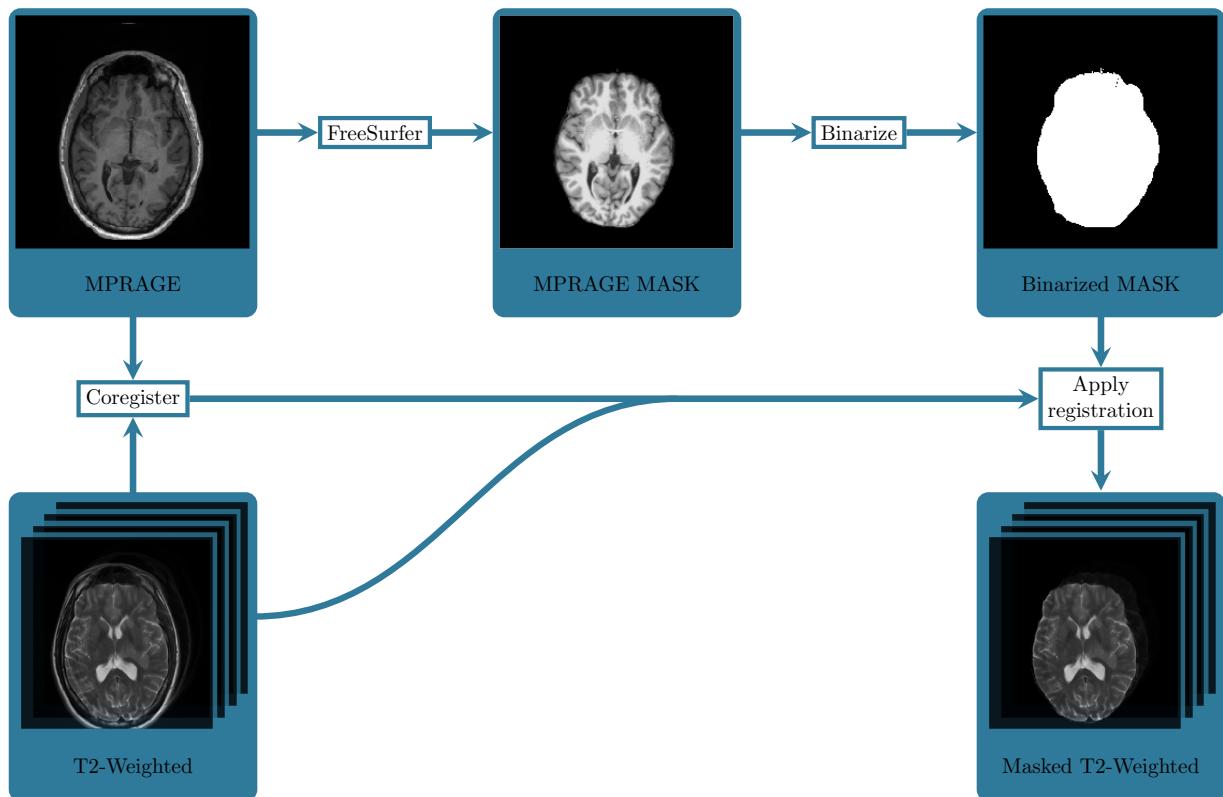
**Table 2.2:** The scale of the observer quality scores are on the Likert scale with 5 being a perfect image and 1 being completely non-diagnostic. Here the scale and what prompts the given score is described for each level.

Observer Score	Prompt
5	Perfect scan
4	Artefacts, however, fully diagnostic
3	Mild/moderate artefacts. Re-scan if any abnormalities
2	Considerable artefacts Re-scan if any abnormalities
1	Completely non-diagnostic

done on the Likert scale [Kecskemeti et al., 2018] where an image of perfect quality is given the score 5 while a completely non-diagnostic image is given the score 1 (this differs from the article where the best score is 4 and the worst score is 0). In Table 2.2 the different scores and what prompts them are described for each level 1-5.

## 2.7 Image Processing

The skull of the scanned subject is often of no significant interest when inspecting or analysing an image, for example when calculating metrics. Thus a process is needed to remove the skull and mask the brain. FreeSurfer [Fischl, 2012] provides this process and as such all scans were processed using FreeSurfer. FreeSurfer is a free software package used for analysing structural neuroimages and provides tools such as skull stripping, segmentation, registration and much more. *recon-all* was used on the T<sub>1</sub>-weighted MPRAGE scan, and the associated brainmask was extracted and binarized. T<sub>2</sub>-TSE and other sequences were registered using *bbregister* [Greve and Fischl, 2009], such that the binarized brain mask is applicable. The T<sub>1</sub> weighted MP-RAGE was registered to the T<sub>2</sub>-weighted scans and the inverse transformation was applied to the binarized brain mask thus generating a mask for the T<sub>2</sub>-weighted images. In Figure 2.10 we see an illustration of this process where a T<sub>1</sub>-MPRAGE is brainmasked, binarized and coregistered to a T<sub>2</sub>-weighted image. In this thesis I only work with brainmasked images as the quality of the background is not of interest.



**Figure 2.10:** FreeSurfer processing pipeline. A T<sub>1</sub>-weighted MPRAGE scan was reconstructed using FreeSurfer's *recon-all*. Any T<sub>2</sub>-weighted scan is then registered to the MPRAGE using FreeSurfer's *bbregister* and the transformation saved. The extracted brainmask is then binarized and the inverse registration-transformation is applied, registering the mask to the relevant T<sub>2</sub>-weighted image. In this thesis I only work with brainmasked images.

## 2.8 STAI-CH

State and Trait Anxiety Inventory for Children (STAI-CH) [Spielberger et al., 1970] is a self reported questionnaire that measures anxiety levels in children. It is a modification of STAI which measures anxiety in adults and has been widely used with more than 14,000 citations [Buela-casal, 2014]. STAI-CH is used to determine children's tendency to experience anxiety (trait) and what state of anxiety they are currently in (state). Examples of state questions are *I feel: calm, upset, uneasy* with responses of the type *Not at all, somewhat, moderately, very much*. Each response is weighted from 1 to 3 with a total of 20 questions. The sum of all responses formulate the anxiety score, with a higher score reflecting a greater state of anxiety. Naturally scores are then in the range of [20,60].

## 2.9 Statistical Principles

In this section I cover the statistical principles needed to analyse data with repeated measurements, for example the results of STAI-CH [Spielberger et al., 1970] questionnaires. The notation used will follow that of [Davis, 2002] although the structure will mostly follow chapter 5 in [Crowder and Hand, 1990].

We start by introducing the notation used followed by repeated measures Analysis Of Variance (ANOVA) and the general linear model.

### 2.9.1 Notation

Consider an experiment involving  $s$  different groups (for example different treatments: drug/placebo) with a total of  $n = \sum_{h=1}^s n_h$  subjects where  $n_h$  denotes the number of subjects associate with group  $h$ . By  $y_{hij}$  denote a repeated measurement of subject  $i$  in group  $h$  at time point  $j$ . Let  $t$  denote the total number of times the measurement is repeated. The  $t \times 1$  bundled observation for subject  $i$  is then  $\mathbf{y}_{hi} = (y_{h1}, \dots, y_{ht})^T$ , for  $h = 1, \dots, s$  and  $i = 1, \dots, n_h$ . The  $h$  subscript is usually omitted to ease notation, we then talk about observations  $\mathbf{y}_i$  for  $i = 1, \dots, n$ . Suppose we have observed  $p$  covariates that are fixed throughout the measurements. We will denote these by the  $p \times 1$  vector  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ ,  $i = 1, \dots, n$ .

### 2.9.2 Repeated Measures ANOVA

The model structure for repeated measures ANOVA is fairly loose and we use this structure to test for the within-effect variability. For example this could be used to examine if there is a significant difference between measured time points or if there is an interaction effect with the repeated measurement, for example treatment or gender.

The model is defined by:

$$y_{hij} = \mu + \gamma_h + \tau_j + (\gamma \cdot \tau)_{hj} + \pi_{i(h)} + e_{hij} \quad (2.17)$$

For  $h = 1, \dots, s$ ,  $i = 1, \dots, n_h$  and  $j = 1, \dots, t$ . Here  $\mu$  is a baseline mean equal for all measurements,  $\gamma$  corresponds to the fixed effect within group  $h$ ,  $\tau_j$  the fixed time effect.  $\pi_{i(h)}$  models the random effects of subject  $i$  in group  $h$  and the  $e_{hij}$ 's are the independent normally distributed error terms. The following constraints apply:

$$\sum_{h=1}^s \gamma_h = 0, \quad \sum_{j=1}^t \tau_j = 0, \quad (2.18)$$

$$\sum_{h=1}^s (\gamma \cdot \tau)_{hj} = 0, \quad \sum_{j=1}^t (\gamma \cdot \tau)_{hj} = 0 \quad (2.19)$$

and lastly:

$$e_{hij} \sim \mathcal{N}(0, \sigma_e^2). \quad (2.20)$$

These formalities are usually swept under the rug when analysing data. Instead the sum of squares and how much variance they describe are of interest. The sum of squares are based on the deviation from different specified means and the ones of interest in the relation to this thesis are giving by:

$$SS_T = n \sum_{j=1}^t (\bar{y}_{..j} - \bar{y}_{...})^2 \quad (2.21)$$

$$SS_R = \sum_{h=1}^s \sum_{i=1}^{n_h} \sum_{j=1}^t (y_{hij} - \bar{y}_{h..} - \bar{y}_{hi.} + \bar{y}_{...})^2 \quad (2.22)$$

where:

$$\bar{y}_{...} = \frac{\sum_{h=1}^s \sum_{i=1}^{n_h} \sum_{j=1}^t y_{hij}}{nt} \quad (2.23)$$

$$\bar{y}_{h..} = \frac{\sum_{i=1}^{n_h} \sum_{j=1}^t y_{hij}}{n_h t} \quad (2.24)$$

$$\bar{y}_{..j} = \frac{\sum_{h=1}^s \sum_{i=1}^{n_h} y_{hij}}{n} \quad (2.25)$$

$$\bar{y}_{h.j} = \frac{\sum_{i=1}^{n_h} y_{hij}}{n_h} \quad (2.26)$$

$$\bar{y}_{hi.} = \frac{\sum_{j=1}^t y_{hij}}{t}. \quad (2.27)$$

Each equation can be interpreted as the mean for, overall, group  $h$ , time  $j$ , group  $h$  at time  $j$  and the  $i$ 'th subject in group  $h$  respectively. Testing the difference between two time points is done with an F-test. The test statistic  $F$  is given by:

$$F = \frac{SS_T / (t - 1)}{SS_R / [(n - s)(t - 1)]} \quad (2.28)$$

hard brackets are used for visual convenience. It should be evaluated in an F-distribution with  $t - 1$  and  $(n - s)(t - 1)$  degrees of freedom.

Similar sum of squares can be defined for different groups and interaction between group and time.

### 2.9.3 General Linear Model

The general linear model can be used to model the repeated measures and to calculate the estimated marginal means of a factor of interest, for example time or group. The general linear model assumes independent normally distributed observations  $\mathbf{y}_i$ ,  $i = 1, \dots, n$ . More specifically:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i \quad (2.29)$$

where  $\mathbf{X}_i$  is a known  $t \times b$  design matrix, with  $b = 1 + (t - 1) + p + (s - 1)$ ,  $\boldsymbol{\beta}$  is a  $b \times 1$  vector of coefficients to be estimated, and  $\boldsymbol{\epsilon}_i$  follows a  $t$ -dimensional Gaussian distribution, with mean zero and unknown covariance matrix  $\Sigma$ , i.e.  $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \Sigma)$  with  $\mathbf{0} = (0, \dots, 0)^T$ .  $\Sigma$  can be interpreted as the within-subject variability. The  $\boldsymbol{\epsilon}_i$ 's are assumed to be independent and hence  $\mathbf{y}_1, \dots, \mathbf{y}_n$  are as well. Having the same matrix  $\Sigma$  for all subjects is called the variance homogeneity assumption and should be checked when analysing data.

This model can be made more general by relaxing the condition that  $b=1+(t-1)+p+(s-1)$ , for example if covariates change over time or by adding within-subject random effects with the addition of  $\mathbf{Z}_i \boldsymbol{\gamma}_i$ . One could also consider  $t_i$  measurements from subject  $i$  instead of  $t$  measurements for all subjects. Although all of this is unnecessary in the context of this thesis.

The estimator given by:

$$\hat{\boldsymbol{\beta}} = \left( \sum_{i=1}^n \mathbf{X}_i^T \hat{\Sigma}^{-1} \mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^n \mathbf{X}_i^T \hat{\Sigma}^{-1} \mathbf{y}_i \right) \quad (2.30)$$

is then a solution to:

$$\sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \Sigma^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \quad (2.31)$$

where  $\hat{\Sigma}$  is the maximum likelihood estimator of  $\Sigma$  and is given by (assuming no missing values):

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^T \quad (2.32)$$

This is clearly a circle argument. The reason for this is that equation (2.30) is actually written with  $\Sigma$  instead of  $\hat{\Sigma}$ , of course under the assumption that  $\Sigma$  is known. In practice estimates of  $\boldsymbol{\beta}$  and  $\Sigma$  are achieved iteratively. This can be done by first calculating  $\hat{\boldsymbol{\beta}}$  in equation (2.30) using some approximation of  $\hat{\Sigma}$ . Using this estimate of  $\hat{\boldsymbol{\beta}}$  to update  $\hat{\Sigma}$  using equation (2.32) and then again using that to update  $\hat{\boldsymbol{\beta}}$  and so on. This is sometimes referred to as a *seesaw* method. Usually equation (2.30) is calculated first, using the sample covariance matrix, then (2.32) is calculated.

One can then perform various test of interest. Suppose we are interested in the linear hypothesis:

$$\mathcal{H} : \quad \mathbf{H} \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad (2.33)$$

where  $\mathbf{H}$  is an  $r \times b$  matrix and  $\boldsymbol{\beta}_0$  is an  $r \times 1$  vector. The test statistic is then given by:

$$T = (\mathbf{H}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T (\mathbf{H}\hat{\Sigma}_{\boldsymbol{\beta}}\mathbf{H})^{-1} (\mathbf{H}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \quad (2.34)$$

where  $\hat{\Sigma}_{\boldsymbol{\beta}}$  is the estimated covariance matrix of  $\hat{\boldsymbol{\beta}}$  [Wald, 1943, Crowder and Hand, 1990]. The test statistic will converge in distribution to a chi-squared distribution:

$$T \xrightarrow{\mathcal{D}} \chi_r^2 \quad (2.35)$$

and hence it is evaluated accordingly [Crowder and Hand, 1990].

We can write a simple hypothesis using the notation as in (2.33) and (2.34). Consider the simple hypothesis given by:

$$\mathcal{H} : \beta_j = \beta_0 \quad (2.36)$$

it can be written as:

$$\mathcal{H} : \mathbf{H}\boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad (2.37)$$

with  $\mathbf{H} = (0, \dots, 0, 1, 0, \dots, 0)$  the  $j$ 'th entry equal to 1. The test statistic is then reduced to:

$$T = (\hat{\beta}_j - \beta_0)^T (\sigma_j^2)^{-1} (\hat{\beta}_j - \beta_0) \quad (2.38)$$

$$= \frac{(\hat{\beta}_j - \beta_0)^2}{\sigma_j^2}. \quad (2.39)$$

Testing of two nested models can be done the usual way with a likelihood ratio test.

#### 2.9.4 STAI-CH Example

Data (see Table 2.3) consists of  $n = 20$  subjects in two groups,  $s = 2$  (Male/Female) who took a state survey before and after training/preparation with a mock scanner. We have  $t = 2$  repeated measurements with  $p = 1$  covariate (age).

A design matrix  $\mathbf{X}_i$  for subject  $i$  could be a  $2 \times 5$  matrix, and  $\boldsymbol{\beta}$  will then be a  $5 \times 1$  vector. They would have the form:

$$\mathbf{X}_i = \begin{pmatrix} 1 & 0 & 0 & g_i & a_i \\ 1 & 1 & 0 & g_i & a_i \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_g \\ \beta_a \end{bmatrix} \quad (2.40)$$

where  $g_i$ , and  $a_i$  denotes the gender and age respectively, of the  $i$ 'th subject.

**Table 2.3:** Exempt of STAI-CH anxiety data, which will be described in greater detail later. There are two repeated measures of state anxiety, before and after preparation/training with a mock scanner, two groups M/F and one covariate, age.

Gender	Age	State Before	State After
<i>M</i>	6	34	24
<i>M</i>	7	28	24
<i>F</i>	5	38	38
<i>M</i>	7	24	22
<i>M</i>	7	55	30
<i>M</i>	10	21	22
<i>F</i>	4	32	29
<i>M</i>	7	23	24
<i>M</i>	7	27	24
<i>F</i>	10	31	28
<i>M</i>	6	26	26

## 2.10 Head Displacement Metrics

Suppose we are tracking a subjects head movement with  $\vec{x}_t$  denoting position of the head at time  $t$ . We could then be interested in the root mean squared (RMS) head displacement here denoted  $\delta$ . Let  $\vec{x}_0$  denote the position of the head at time  $t = 0$  (start of the scan), the RMS head displacement for a given scan is then calculated as:

$$\delta = \sqrt{\frac{1}{N} \sum_{t=0}^N (|\vec{x}_t - \vec{x}_0|)^2}$$

where N denote the total number of motion estimates measured during the scan [Frost et al., 2019].



# Chapter 3

---

## Methods

---

This chapter describes the methods of analysis carried out on the data that is the basis for this thesis. The data analyzed in this thesis consists of three main parts from, an adult healthy control group of a previous study [[Hannah Susanne Eichhorn, 2021](#)], a healthy children population where training effects were tested and lastly, an ongoing clinical trial. Each study is given its own section as the different studies are so dissimilar. A summary and acquisition of data is given for each study followed by the respective analysis performed. Any eventual changes or additions are also mentioned for each study.

### 3.1 Adult Healthy Controls

#### 3.1.1 Data

MR scans from the healthy control group were collected, at Rigshospitalet Copenhagen, from 22 (16 M / 6 F) healthy volunteers ( $23.5 \pm 4.3$  years, mean  $\pm$  standard deviation) over a seven month period in 2020/21. The participants were recruited for the Cimbi [[Knudsen et al., 2016](#)] database and the study protocol was approved by The Ethics Committee of Copenhagen and Frederiksberg, Denmark, with appropriate written consent from each subject.

During the experimental setup, several different sequences were collected from each subject multiple times with and without pMC, for different head motions, leading to more than 800 MR images, see Figure 3.1 for an illustration of image acquisition for a single participant. First participants were instructed to lie as still as possible followed by a scan where they were guided through a nodding motion, while for some sequences an additional shaking motion was performed. A nodding motion was chosen as it has been shown [[Eichhorn and Vascan, 2021](#)] that this motion is the most prevalent for children scanned without GA. The guided motion was planned in such a way that maximum head displacement took place during k-space center acquisition as to introduce as many artifacts throughout the image.

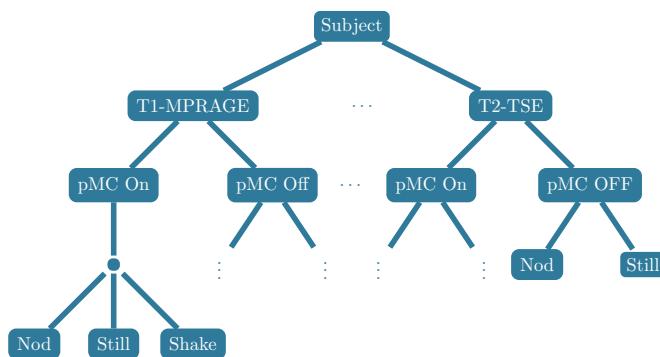
### 3.1.2 Analysis

The MRI data was acquired using a Siemens 3 Tesla Magnetom Prisma (Siemens AG, Erlangen Germany) using a 64-channel head coil. Motion was tracked using the markerless infrared tracking system *Tracoline* (TracInnovations, Ballerup Denmark). Infrared light is projected onto a subjects face and a subsequent 3D point cloud is generated and updated with a frequency of 30Hz. Motion can be estimated, with sub millimeter precession, from the point cloud (When lying as still as possible it can detect the subtle head movements when breathing). Tracoline is mounted on an arm on the MR scanner table itself allowing for ease of access and set up. See Figure 3.2 for an image of the scanning setup including head coil, mirror, headphones and more. A calibration of the tracking system is done with a phantom before each scanning session.

The 800 MRI scans were preprocessed using the FreeSurfer pipeline described in the previous chapter, specifically section 2.7. The processed images, containing only the brain, were cropped such that only non-zero slices remained. This removes the degenerate case of zero division and removes irrelevant information when calculating AES and CoEnt respectively. For CoEnt the images were also binned to 128 intensity levels to follow the procedures of [Pannetier et al., 2016].

A slight modification to the CoEnt metric was made when evaluating on 2D encoded sequences. A 2D co-occurrence matrix and corresponding entropy was calculated for each slice and the average entropy is reported. That is, no intensity comparisons between slices were performed as it does not necessarily make sense when slice thickness is too large. For reference, the slice thickness for 2D encoded sequences are around 4-5 mm while for 3D encoded sequences it is 1 mm.

Please note that comparisons between slices are indirectly done if the entropy is instead calculated on the average co-occurrence matrix. The process of normalizing the matrix before entropy calculations is what inhibits this method of averaging. An additional modification was made to CoEnt both for 3D and 2D encoded se-



**Figure 3.1:** Image acquisition procedure for a single participant/subject. For multiple sequences, a scan was done with pMC on and off while performing one of two motions nodding or laying still (for MPRAGE an additional shaking motion was performed). One can easily see that this quickly leads to many MR images.



**Figure 3.2:** MRI scanner setup. From Left to right: Fiber optics cable leading to the Tracoline camera, Tracoline camera, 64-Channel head coil, mirror which allows the patient to watch a movie while being scanned, MR safe Optoacoustics (Optoacoustics Ltd, Or Yehuda, Israel) headphones where both scanner operator can talk to the patient and play audio from the movie, lastly is the emergency button which sets off an alarm notifying scanner operator of immediate discomfort or need for medical attention. *Photo courtesy of The MoCo project at Rigshospitalet, Copenhagen, Denmark.*

quences. The entry of the co-occurrence matrix corresponding to intensity values  $p = q = 0$  was set to zero, i.e.  $C(0, 0) = 0$  (Recall  $C$  is the co-occurrence matrix while  $\bar{C}$  is the normalised co-occurrence matrix, i.e., normalisation is done after assigning  $C(0, 0)$  to 0.). We allow ourselves to do this as  $C(0, 0)$  corresponds to information of the 3D volume background. This action can be seen as ignoring the background. The TenenGrad (TGrad/TG) [Krotkov, 1989] measure was also evaluated in addition to AES and CoEnt. TGrad is also a reference free image metric which was used in a previous study [Hannah Susanne Eichhorn, 2021].

Mathematical metrics was not the only evaluation of the MR image quality. The gold standard expert-scoring was performed, on roughly 500 images, by two radiography students and one experienced radiologist. The observer scores are on a 5 level Likert scale following [Kecskemeti et al., 2018], with the exception that 5 is given to an image of the highest quality for a perfect scan without artefacts and the lowest quality 1 given if the scan is completely non-diagnostic (Our scale is 1-5 while Kecskemetis scale goes from 0-4). A weighted average of the three scores was then considered with the experienced radiologist having a weight twice of that of a student.

Spearman's correlation was used to compare metric results to the observer scores. The observer scores are subjective as such a degradation in observer scores from 5 to 4 might not lead to the same degradation in image quality as scores degrading from 2 to 1. Thus it is not trivial that the scale should be linear hence we used Spearman's correlation rather than the Pearson correlation.

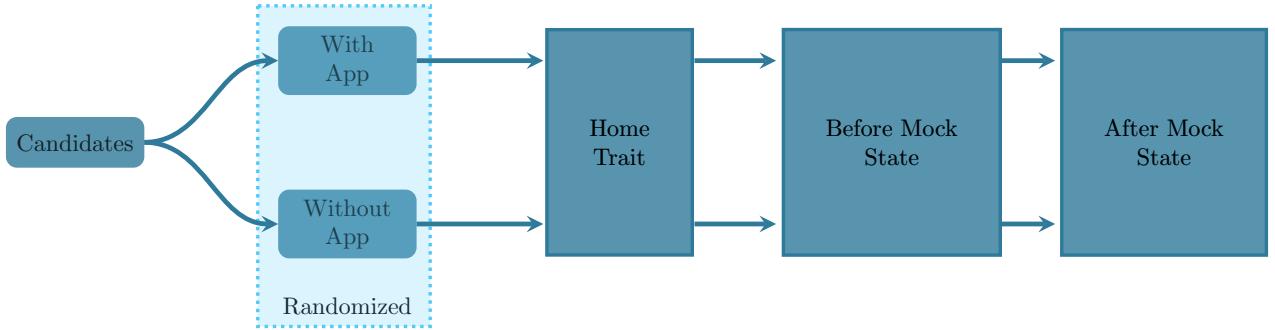
All metric analysis was carried out using Python 3.7 (Python Software Foundation) and all relevant scripts are freely available at GitHub [https://github.com/simonchemnitz/Msc\\_Thesis](https://github.com/simonchemnitz/Msc_Thesis).

## 3.2 Pediatric Healthy Controls

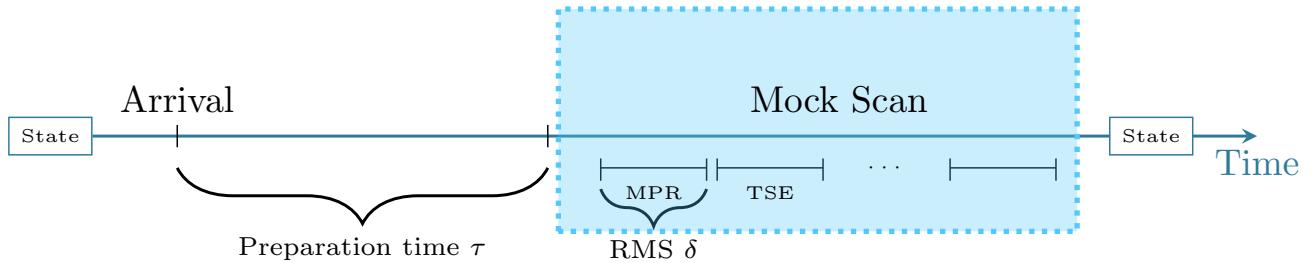
### 3.2.1 Data

The pediatric healthy control group was recruited for a group bachelor project by radiography students from the University College Copenhagen (KP Metropol). In this small study, healthy children were recruited from friends and family of three radiography students. Informed consent was acquired by all legal guardians of the children. Data from this group consists of 20 (11 M / 9 F) repeated measurements of STAI-CH anxiety scores from healthy children ( $7.05 \pm 1.849$  years) volunteers.

Participants were split into two groups, see Figure 3.3, namely those who were and were not instructed to play with an app (*Kom med i scanneren*), which introduces them to the hospital environment such as the reception, waiting room and scanner room. Their anxiety levels were then measured using the self-reported State and Trait Anxiety Inventory for Children (STAI-CH) questionnaire [Spielberger et al., 1970]. Each participant was asked to respond to a Trait survey at home and two State surveys at the hospital, one before the mock training and one after. The preparation time (time before the child enters the mock scanner) and mock scan time was noted during the training session, see Figure 3.4 for a timeline of the mock scanning session. Head motion was continuously tracked using Tracoline while the child was inside the mock scanner. From this the Root Mean Squared (RMS) head displacement can be calculated. In total we have 14 measurements for each participant, six measurements for RMS head displacement for different sequences, three from the STAI-CH questionnaire two from the preparation/scan time and lastly the age, gender and whether they used the app or not.



**Figure 3.3:** Illustration of the experimental design for the healthy control children who received a STAI-CH questionnaire during mock training (recall mock training is where children are put into a mock/faux MRI scanner). The children were randomly chosen to either get preparation with the mobile phone application "*Kom med i scanneren*" or no preparation. They were given a Trait questionnaire at home and two State's before and after mock training (here simply denoted mock).



**Figure 3.4:** Illustration of the mock training session timeline. Children take a state questionnaire before the preparation session. The time from entering the mock scanner room until the child enters the mock scanner is measured and denoted  $\tau$ . Subject head movement is measured for each sequence and the root mean squared head displacement is calculated, for T<sub>1</sub>-MPRAGE, we denote it by  $\delta$ . After the session the children fill out another state survey.

### 3.2.2 Analysis

The statistics outlined in section 2.9 are used to analyse the data from the healthy children. Having two measurements of State anxiety scores for each child gives rise to repeated measures analysis, hence repeated measures ANOVA and a GLM was used to infer results regarding the STAI-CH answers. Specifically the sum of squares was used to determine the within subject effect of time, that is if there are significant differences between anxiety levels before and after the mock training session and if there is a significant interaction with app usage and the anxiety levels before and after. We chose to let gender enter as a covariate as there are not enough observations to test for an interaction between app and gender. This would split the 20 people into four groups of roughly five in each group which is clearly not a large enough sample size. The sum of squares should be evaluated in an F statistic with 15 (20-5) degrees of freedom as we check the effect of time on four of the covariates and between subject effects (time, time\*app, time\*gender, time\*trait)

and time\*age). The only sum of squares that are of interest are time and time\*app as the other factors enters as covariates.

A GLM is also fitted to explain any eventual difference, for example by examining the estimated marginal means or testing parameters if they are different from zero. The GLM is defined for a single subject as:

$$y_i = \beta_0 + \beta_1 \cdot age + \beta_2 \cdot gender + \beta_3 \cdot Trait + \beta_4 \cdot i + \beta_5 \cdot app, \quad i = 0, 1, \quad (3.1)$$

where  $y_i$  is the anxiety score at timepoint  $i$ , with  $i \in \{0, 1\}$  corresponding to before and after training respectively. It is then of interest to estimate the parameters  $\beta_j \in \mathbb{R}, j = 0, \dots, 5$  and test if they are significantly different from zero. This model is chosen as it is fairly simple, (no interactions) this is deliberate as a result from few observations.

Two additional linear models will also be fitted, one to model the Root Mean Squared (RMS) head displacement and one to model the time it took to complete the preparation session (how long it took for the child to enter the mock scanner measured in minutes from when they entered the mock scanner room). That is we fit the models:

$$\delta = \beta_0 + \beta_1 \cdot age + \beta_2 \cdot gender + \beta_3 \cdot Trait + \beta_4 \cdot State + \beta_5 \cdot app \quad (3.2)$$

$$\tau = \beta_0 + \beta_1 \cdot age + \beta_2 \cdot gender + \beta_3 \cdot Trait + \beta_4 \cdot State + \beta_5 \cdot app \quad (3.3)$$

where  $\delta$  is the RMS head displacement measured in millimeters for the first sequence they experience, State is the State anxiety level before the training session,  $\tau$  is measured in minutes and is the time to complete the mock preparation. It would be desirable to see whether or not using the app has an influence on these two measures (RMS,  $\delta$  and time,  $\tau$ ) as it is a core part of the clinical preparation step. The first sequences they experience is chosen as the children are least familiar with the process at the start of the mock scan. It would also be natural to assume that state and trait (the anxiety scores) has an influence, on the two measures (RMS and time), with the reasoning that a more anxious child will require longer preparation time to enter the scanner, as they are less familiar with the entire environment, and move more in the scanner. Once more we do not examine any eventual interactions as we lack a larger sample to consider these interactions.

The repeated measures analysis was done in SPSS (IBM SPSS) and the linear models were fitted in R (The R-Project).

### 3.3 Ongoing Clinical Trial

#### 3.3.1 Data

The ongoing clinical trial aims to drastically reduce the number of GA cerebrum MRI scans in the pediatric population by implementing and validating the aforementioned procedures, such as motion correction and preparation. The children (aged 4-10 years) participating in the study are referred to the experiment by their doctors if an MR scan with GA was already scheduled the experimental protocol

was approved by The Ethics Committee of Copenhagen and Frederiksberg, Denmark with informed written consent from legal guardians of participating children. The parents then has the option to enroll their child in the clinical study given that they do not meet any of the exclusion criteria. Apart from obvious safety reasons, exclusion criteria are needed as it is paramount that the children understands our instructions and the reason why it is necessary for them not to move while being scanned. The criteria are:

- i) Degraded cognitive abilities.
- ii) Mental or physical disabilities which prevents them from lying still.
- iii) Medical condition interfering with MR safety.
- iv) Non-fluent in danish.
- v) Scheduled for contrast fluid MR scan.

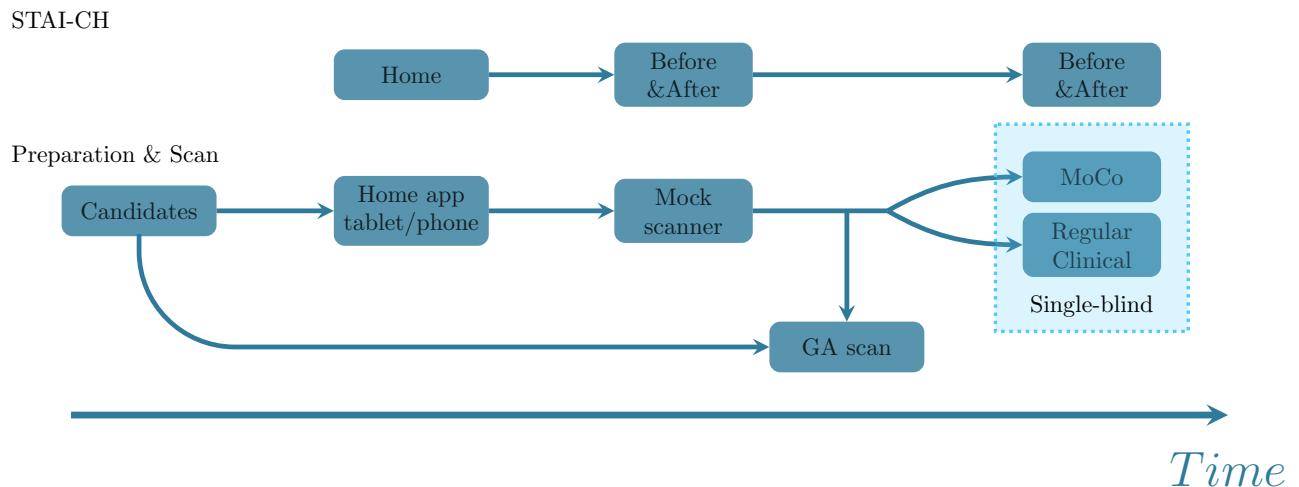
The parents of a participating child are then instructed to show their child the app "*Kom med i scanneren*" and discuss the upcoming experience with their child. The participants are sent a pair of virtual reality goggles that can be used with a phone, which allows the child to "explore" the hospital environment virtually before their physical sessions. This concludes the preparation steps that take place at the child's home.

The child then receives a mock scanning session at the hospital where they can familiarize themselves with the MR scanner. At the mock scan the child will be exposed to the scanning sounds (in most cases the exact noises) they can experience at the real MR scan. The noises are first played at a lower volume gradually increasing to a volume level of a real scan. This allows the children to experience and familiarize themselves with the sounds. The sequence sounds are provided by Siemens themselves. Ideally children would randomly be chosen to either get training or not, although this is not an ethical study and would never be approved by an ethics committee. The clinical experimental design is illustrated in Figure 3.5, starting from the left with the candidates for the clinical study (children who needs to undergo an MRI scan) and what the course of training/preparation is before the real clinical scan with either pMC on or off.

Participating subjects has the option to watch a movie from a selection as a method of distraction, both during the real and mock scanning session. Subject motion is tracked and recorded while in the scanner and during playback of sequence sounds. The child is then scheduled for a real MR scan without GA if he/she can lay adequately still for a longer duration while the sounds are played at a volume level equal to that of the real MR scan.

The initial planned GA scan is kept should motion be deemed too severe at either the mock scan or real scan without GA.

Whether or not pMC is on during the real MR scan is predetermined in a random single-blind setting such that patients, radiographers and radiologist don't know if MoCo was performed or not. Remark: a double blind experiment is not possible as



**Figure 3.5:** Illustration of the experimental design for the clinical trial. Some candidates are excluded from the study for reasons such as mental or physical disabilities, this is illustrated with the direct arrow from candidates to GA scan. At home the patient is asked to play with an app (*Kom med i scanneren*) and answer a Trait questionnaire. They are then introduced to the mock scanner setup where they will familiarise themselves with the scanner environment, scanner and its various sounds. They are either deemed capable or incapable of performing a real scan without GA, this is illustrated by the splitting path going from "Mock scanner" to either a GA scan or to a scan without GA. A single-blind experiment takes place during the MR scan without GA. The patient either gets a regular clinical scan or one with prospective motion correction as illustrated by the branching arrow. A STAI-CH state questionnaire is filled out before and after the mock and real scan.

it can be inferred from sequence protocols if MoCo was on or not as they are coded by the lead researcher and her research assistant.

### 3.3.2 Analysis

MRI scans of the pediatric patients are processed and analyzed in the same manner as the adult healthy control scans as outlined earlier. That is, the images are processed using FreeSurfer to extract a brainmask and the relevant metrics are calculated on the brainmasked images rather than the raw MRI volumes.

# Chapter 4

---

## Results

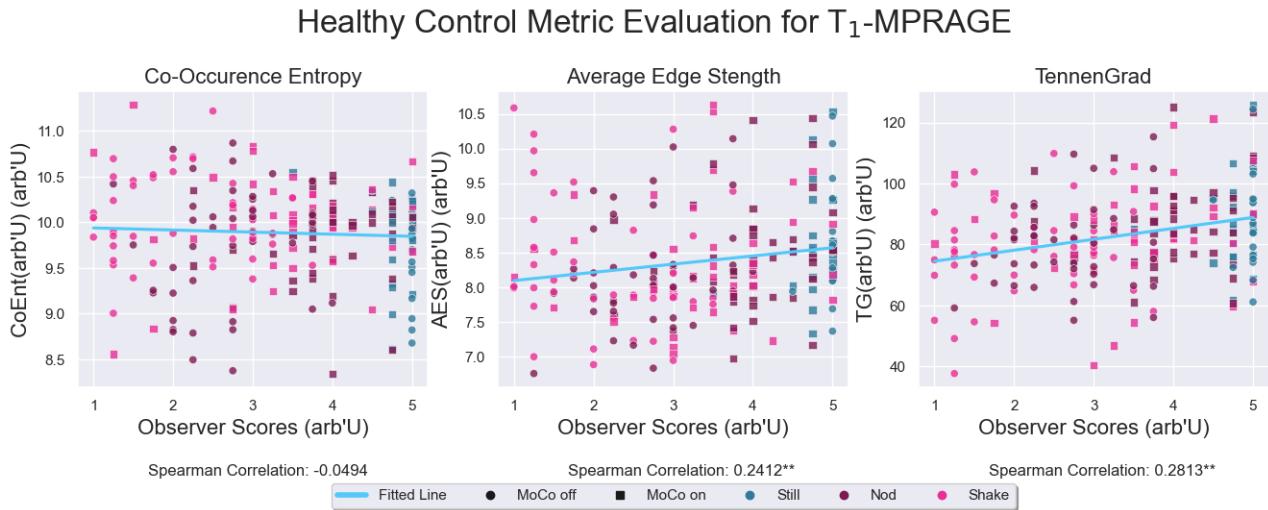
---

In this chapter I will give an overview of the results based on analysis of the data described in the previous chapter. The results of metric analysis of MR images from the adult healthy control is given first where metric scores are compared to observer scores, this is preceded by distributions of metric scores grouped by motion and whether or not prospective motion correction (pMC) was on during acquisition. This is followed by an analysis of the healthy children's data comprised of a statistical analysis of STAI-CH repeated measures where we investigate any association between anxiety levels and gender, age and preparation with an app. Additionally, the effect of app on preparation time and Root Mean Squared (RMS) head displacement is also presented. Finally, preliminary results of the ongoing pediatric patient study are presented.

### 4.1 Adult Healthy Control

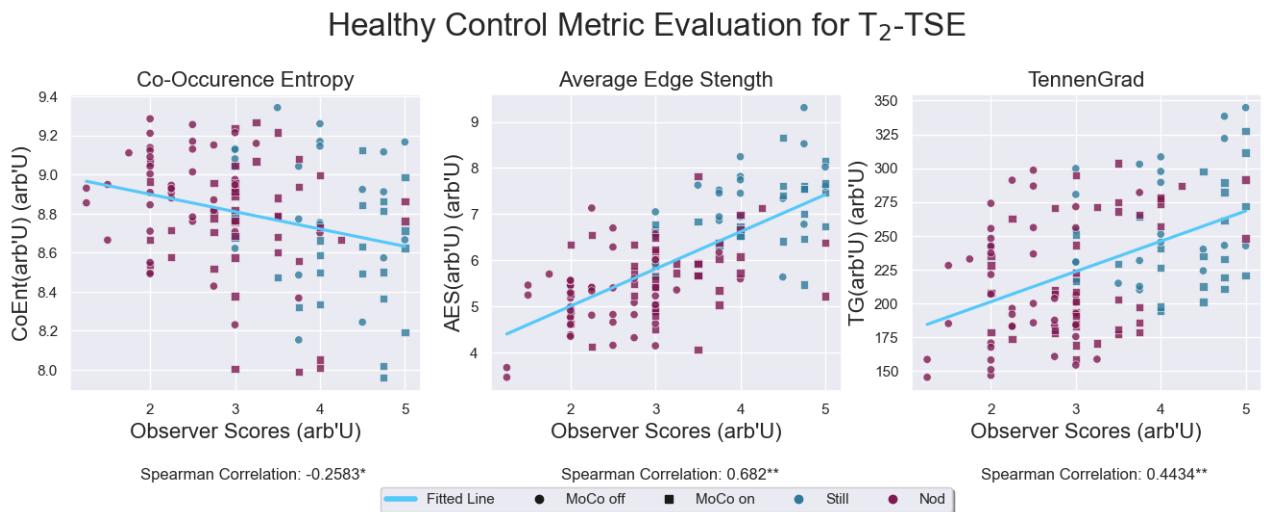
In this section I will evaluate the three metrics (CoEnt, AES and TG), described in the preceding Methods and Background chapters, on the different healthy adult control MR scans. First the correlation between metric and observer scores are examined followed by a quantile description of metric scores via box plots grouped by motion correction (on/off) and expressed motion (nod/still).

Observer scores were available for four of the six original image sequences [[Hannah Susanne Eichhorn, 2021](#)], namely T<sub>1</sub>-MPRAGE, T<sub>1</sub>-TIRM, T<sub>2</sub>-FLAIR and T<sub>2</sub>-TSE. The corresponding metric results were plotted against the weighted average of the observer scores. This is shown in Figures 4.1, 4.2, A.1, A.2 with circles indicating sequences acquired with pMC off and squares with pMC on. Data points are colored according to the movement performed during acquisition (Still, Nod and Shake for MPRAGE). The relevant Spearman correlation is shown below each figure with the corresponding significance level annotated with stars. A linear regression line is plotted to quickly assess the correlation between the given metric and observer scores, although a linear fit does not necessarily explain the data optimally. Note that the scale of the metrics are different hence it is not possible to directly compare the values, but rather, only how much they correlate with observer scores, see

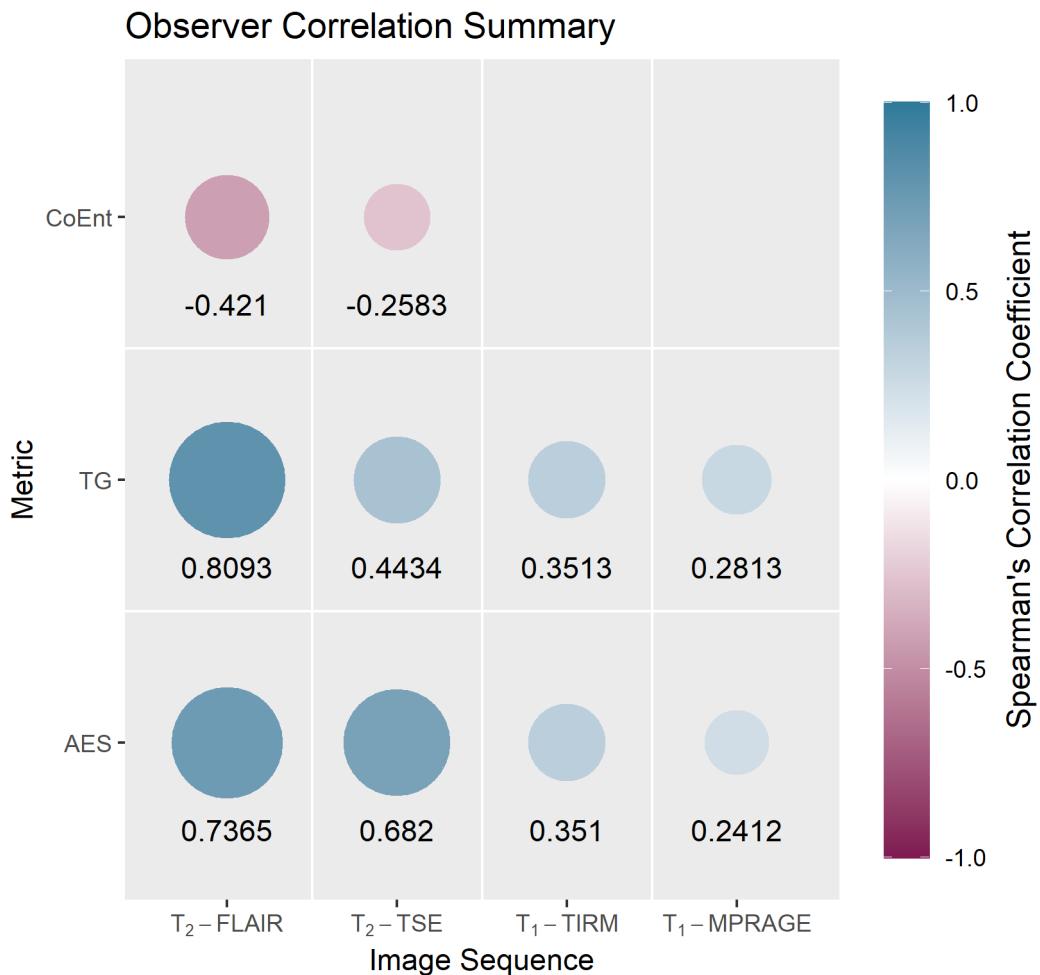


**Figure 4.1:** Correlations between average observer score and the corresponding metric evaluated score from CoEnt, AES and TG for T<sub>1</sub>-weighted MPRAGE scans without reacquisition. Data is colored according to motion and the style of the marker represent whether pMC was on or off with a square denoting pMC scans. For each metric the associated Spearman correlation is reported below the relevant figure and the significance level is denoted either \* or \*\* representing a p-value  $p < 0.05$  and  $p < 0.01$  respectively. A regression line is plotted to easily see the general trend. AES and TG show significant correlation while CoEnt does not.

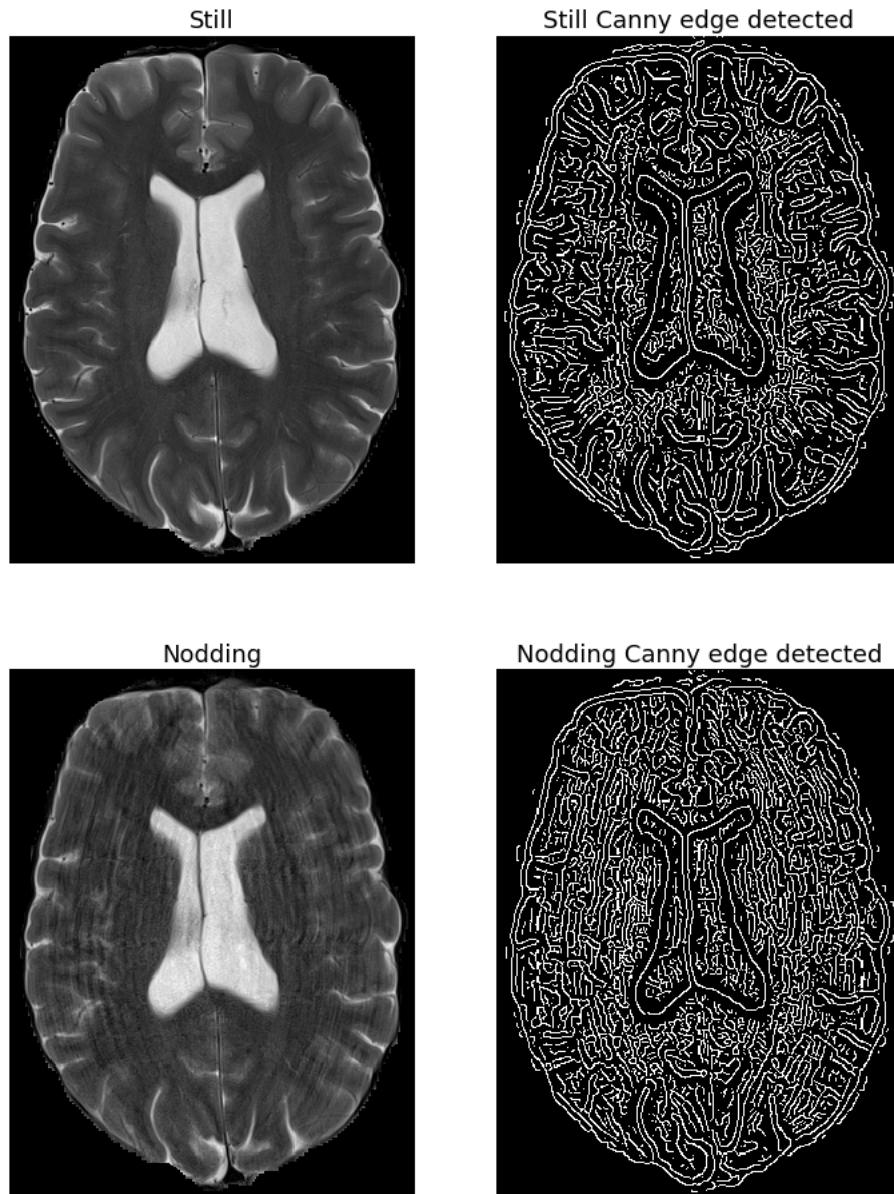
Figure 4.3 for clear overview of the significant Spearman correlations. The gradient metrics (TG and AES) correlate well with observer scores for the sequences T<sub>2</sub>-TSE and T<sub>2</sub>-Flair. On the contrary CoEnt performs relatively worse throughout the different sequences. The observer correlation is poor for T<sub>1</sub>-MPRAGE jointly for all three metrics which might be explained by relatively few differences between nodding and still images. This difference can be seen in Figure 4.6, where an illustration of two T<sub>1</sub>-MPRAGE images with one image acquired while a nodding motion was performed and another while lying still is shown. A T<sub>2</sub>-TSE image with nodding has more prominent ringing as previously seen in Figure 2.6 and we also observe this as Canny edge detector and convolutions are able to detect such ringing as seen in Figure 4.4 and 4.5, which might explain the higher correlation. It is essential that both the edge detection and the convolutions detect ringing for the metric to be useful. Two additional methods of edge detection for AES were evaluated, namely a Laplacian threshold based edge detection and a Phase Stretch Transform [Asghari and Jalali, 2015]. These two additional edge detectors showed sub par performance levels when compared to AES with Canny and TG across all four sequences and some even worse than CoEnt. With either non significant correlation or inconsistent correlation for example being negatively correlated for one sequence but positively correlated for another.



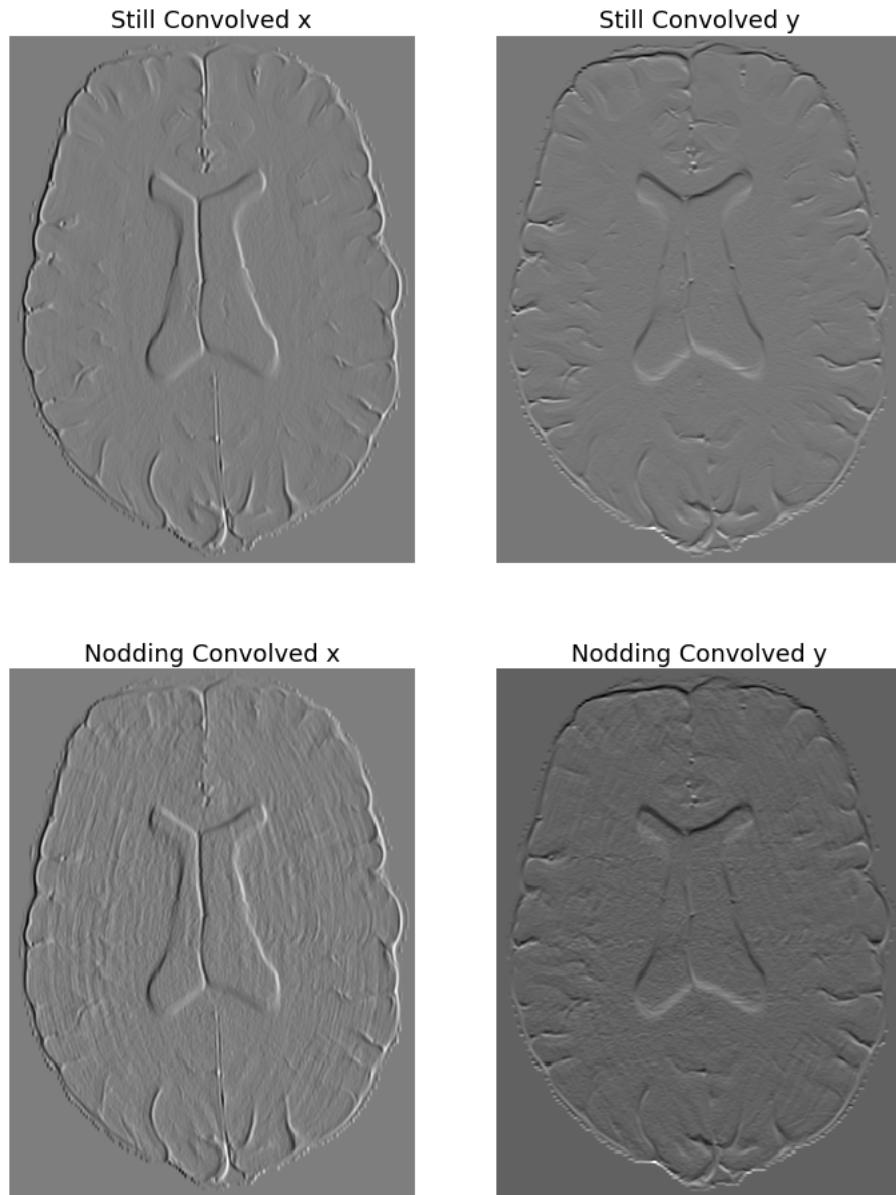
**Figure 4.2:** Correlations between average observer score and the corresponding metric evaluated score from CoEnt, AES and TG for T<sub>2</sub>-weighted TSE scans without reacquisition. Data is colored according to motion and the style of the marker represent whether pMC was on or off with a square denoting pMC scans. For each metric the associated Spearman correlation is reported below the relevant figure and the significance level is denoted either \* or \*\* representing a p-value  $p < 0.05$  and  $p < 0.01$  respectively. A regression line is plotted to easily see the general trend. AES, TG and CoEnt show significant correlation although for CoEnt the correlation is negative.



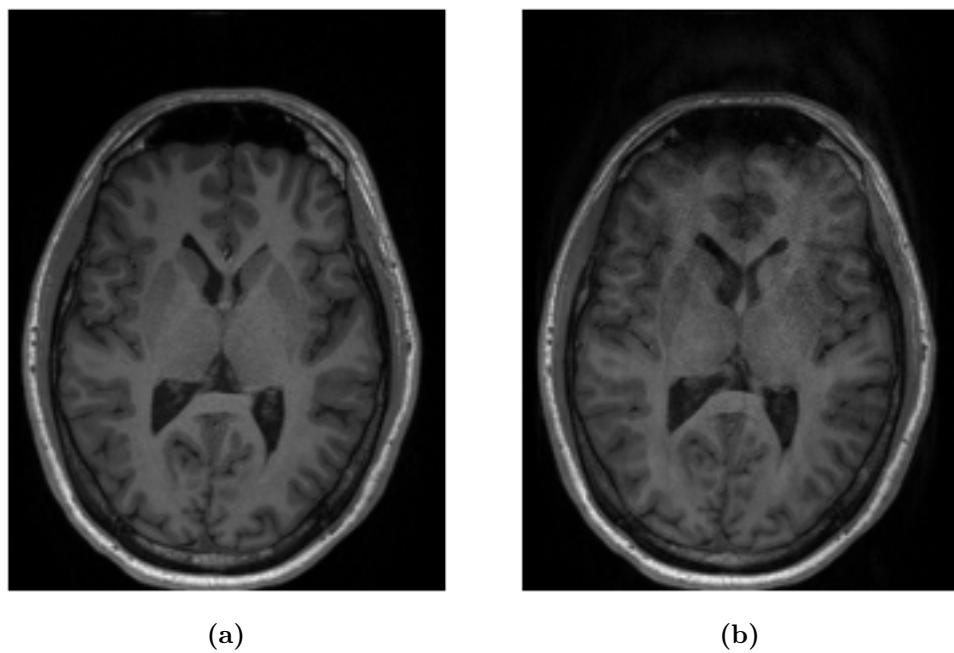
**Figure 4.3:** Spearman correlation coefficients of metric scores for each image sequence. Note that TG and AES have similar results with AES slightly better for T<sub>2</sub>-TSE while CoEnt correlation lacks considerably.



**Figure 4.4:** Original scan (T<sub>2</sub>-weighted TSE) on left with the corresponding Canny edge detected image on the right. Images in the bottom row originates from a scan where a nodding motion was performed during image acquisition and the top row where the subject was as still as possible. Canny is successful in detecting the severe case of ringing which can be seen by the fewer number of breaks in the detected lines for the nodding image relative to that of the still image.



**Figure 4.5:** Convolved T<sub>2</sub>-weighted TSE slices from Figure 4.4. Bottom row corresponds to slices from the nodding scan and top of the still scan. Images in the first column are convolved in the x direction and the second in the y direction. Ringing is visible in both directions for the nodding scan with the x direction being the most prominent. No ringing is visible for the corresponding still scan.

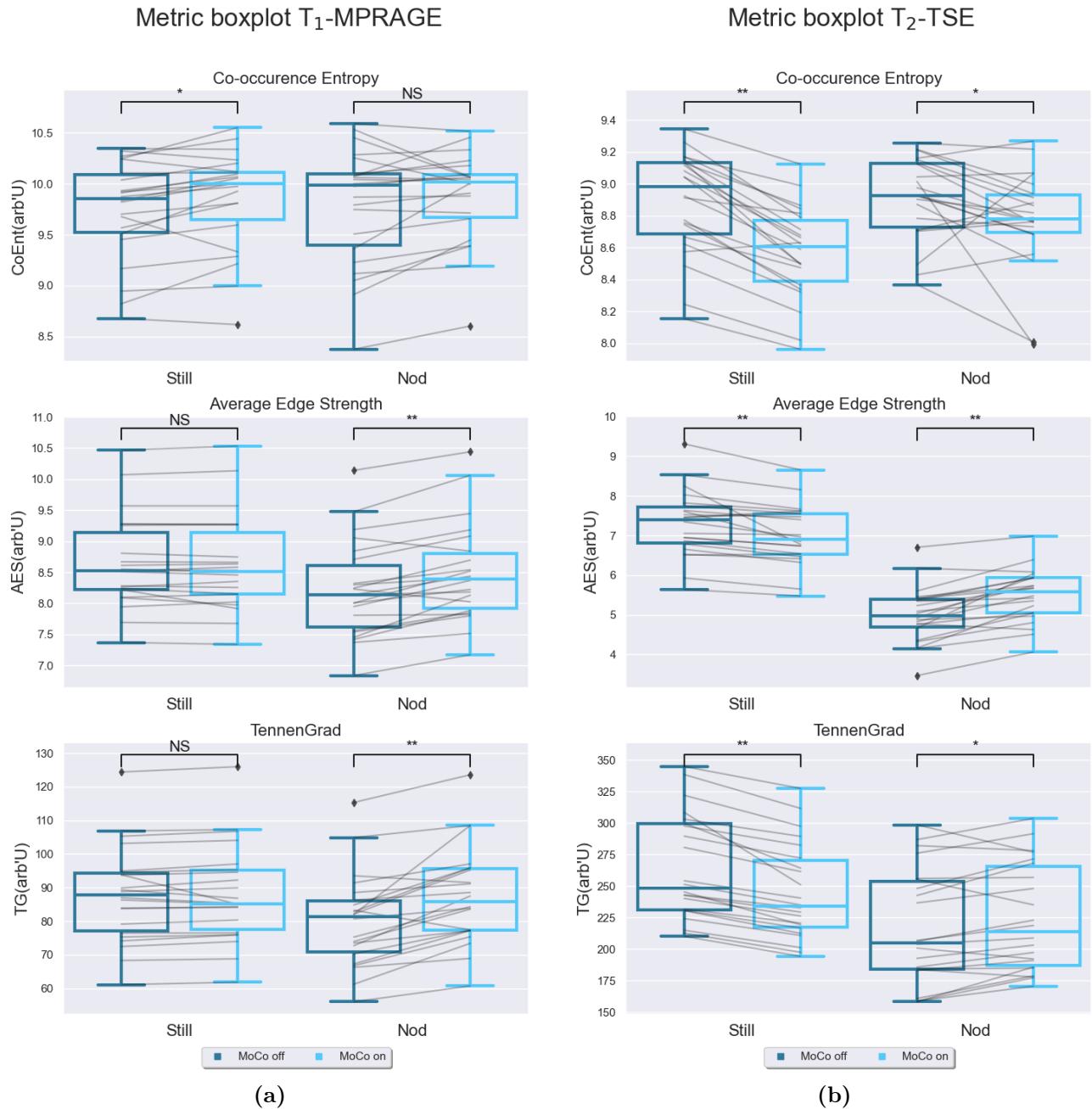


**Figure 4.6:** T<sub>1</sub>-weighted MPRAGE from the same person. Visually they are almost the same with some blurring of the image to the right. (a) Image was acquired when lying still. (b) Image was acquired with a head nodding motion. This might explain the poor correlation of metrics for MPRAGE scans. See Figure B.8 for a visualization of an MPRAGE image with a shaking motion. This renders the image completely non diagnostic.

Box plots of the metric scores were also created to see how the metric scores distribute when pMC is on, respectively off, for the two motions still/nodding. This is shown in Figure 4.7 for the MPRAGE and TSE sequences and Figure A.3 for  $T_2^*$  and  $T_2$ -FLAIR. We compare the metric scores using Wilcoxon rank test [Wilcoxon, 2012] as we have paired data (the same subjects performed 4 scans with pMC on/off while nodding/ laying still.). The p-values are adjusted for multiple tests using Benjamini–Hochberg false discovery correction [Benjamini and Hochberg, 1995] and noted with either \*, \*\* or NS for p-values less than 0.05, less than 0.01 and greater than 0.05 respectively. Connecting lines between individual subjects are drawn in gray. There were no significant differences when pMC on/off for the  $T_1$ -MPRAGE scans when the subject is lying still, except when considering the CoEnt which increased. The difference in metric scores was highly significant ( $p < 0.01$ ) when a nodding motion was made, with a higher metric score when pMC was on, which is the expected outcome. Contrary to AES and TG, CoEnt showed no significant difference when pMC was on/off, when a nodding motion was performed.

The results are somewhat different for the  $T_2$ -TSE scan. A significant difference ( $p < 0.01$ ) in metric scores was found for all three metrics when considering pMC on/off for subjects lying still, however the the difference is more noticeable for CoEnt and TG compared to AES. All of the metrics showed a significant difference when a nodding motion was performed. CoEnt seemed to drop in scores when pMC is on which is inconsistent when comparing to  $T_1$ -MPRAGE, i.e., CoEnt does not consistently increase or decrease with image quality. There is a more noticeable difference in AES scores when pMC is on with nodding motion compared to the other metrics. There is little distinction between pMC on/off for  $T_2^*$  and  $T_2$ -FLAIR.

So far it would seem that AES exhibits higher correlations with observer scores and provides a clearer distinction for nodding scans compared to TG and CoEnt. It is apparent that both Canny edge detector and the convolution operators capture any eventual ringing. The correlation to observer scores are bad for all metrics when focusing on  $T_1$ -MPRAGE scans which might be explained by the relatively few differences in a nodding and a still scan. The artefacts for  $T_1$ -MPRAGE are not obvious.



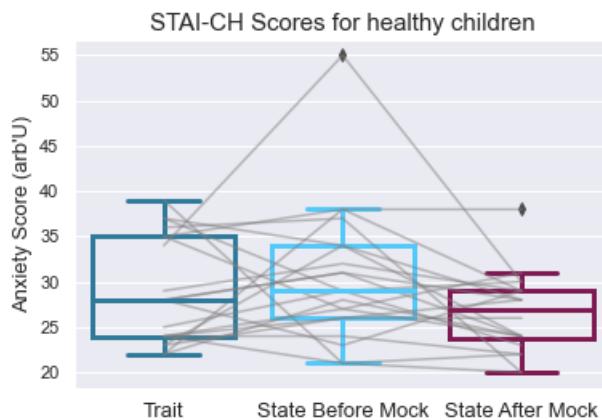
**Figure 4.7:** Quartile distribution of metric scores grouped by motion correction (off and on colored dark blue and bright blue respectively) and head movement(still/nodding) for three metrics (CoEnt, AES and TG in descending position). Paired data are connected with gray lines and the significance level of the difference is denoted either *NS*, \* or \*\* representing a not significant p-value  $p \geq 0.05$ , a p-value  $p < 0.05$  and  $p < 0.01$  respectively. All p-values have been adjusted for multiple testing (a) Metric results for T<sub>1</sub>-weighted MPRAGE. (b) Metric results for T<sub>2</sub>-weighted TSE.

## 4.2 Pediatric Healthy Control

In this section I will determine the statistical significance of training and preparation when measuring anxiety levels in children participating in the healthy control study outlined in section 3.2. In addition the effect of the mobile phone app and other variables on RMS head displacement and preparation time is also presented. The 20 children (9 F, 11 M) were split into two groups, those who played with the provided mobile phone app before training and those who did not. A Trait questionnaire was given at home and two State questionnaires at the mock scanner training (before and after mock scanning). During the scanning their movement was tracked and recorded using Tracoline.

I fitted three different models, a repeated measures ANOVA, a GLM to assess the main effect of the training, i.e., if there is a significant difference of anxiety levels before and after mock scanning session and what the effect might be. The last two models are again linear models to explore any significant effects of the measured variables on the RMS head displacement and the time it took to complete the preparation session (The time before getting the child into the mock scanner). In Table 4.1 a summary of the scale variables are given grouped by gender in the first columns and grouped by usage of app in the last columns. One can take particular note of the apparent difference when comparing with app to without app and no apparent difference when comparing boys to girls. In Figure 4.8 we show the distribution of anxiety scores of Trait, and State before/after with connecting lines for each participant. Here, we observe the drop of anxiety scores going from before mock training to after training. Finally, a distribution of preparation time, Trait score and State scores are given in Figure 4.9 grouped by those who used the app and those who did not. In this we can note the clear difference in preparation time, this is extremely desired.

The STAI-CH scores were analysed using SPSS (IBM SPSS Statistics), specifically a repeated measures general linear model was fitted. In Table 4.2 we see the analysis of the sum of squares for both the within subjects effect of time (within each subject is there a difference at the two time points before/after) and the between subjects effect of using the app (are there differences between subjects who get the app and those who do not get the app). We see the within subjects main effect is significant at the 0.05 significance level, as an F-statistic of 6.985 evaluated in an  $F(1, 15)$  distribution leads to a p-value of 0.018. We also see the between subjects effect of using the app is not significant with an F-statistic of 2.376, which leads to a p-value of 0.14 when evaluated in a  $F(1, 15)$  distribution. This implies that the Timepoint (Before/After) is significant on a acceptance level of 0.05, in other words there is a significant difference of means for the two measured anxiety levels (before and after training). But this only specifies that there is a difference, not if it drops or increases (we would expect it to drop). Hence, the GLM model described in equation (3.1) is used to examine the Estimated Marginal Means (EMM) in order to determine if the anxiety levels drops or increases. The pairwise comparison of the means (30.550 vs 26.450) show again a significant difference at the 0.05 level with a p-value of  $p = 0.012$ . This shows that anxiety levels drop significantly during



**Figure 4.8:** STAI-CH scores for the healthy children. Connecting lines for each subject is colored grey. The reader should note the drop in anxiety scores after mock training, this is expected.

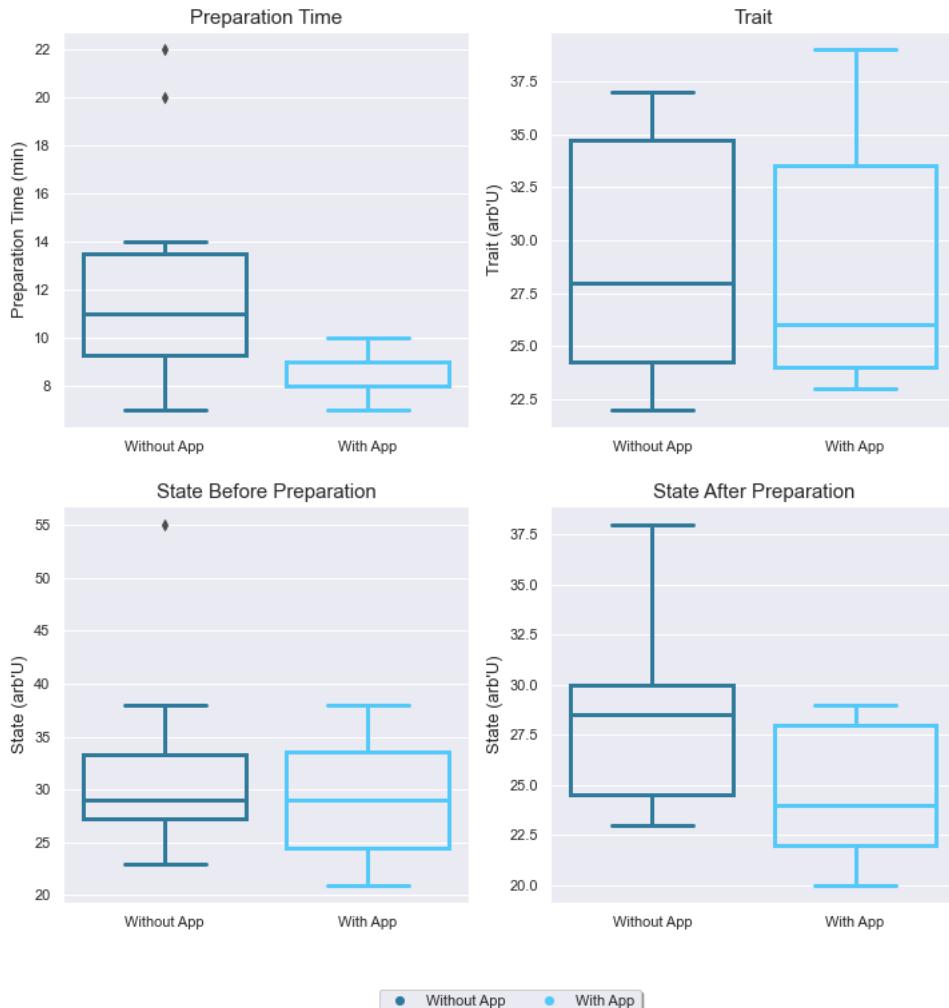
the course of the training/preparation (which is as expected). See Table 4.3 for the EMM of the two timepoints (before/after) and Table 4.4 for the pairwise comparison of these EMM.

The pairwise comparison can also be done using a GLM regression in "long format". That is each observation is the anxiety scores and an additional predictor value indicating if the score is before or after preparation. Such calculations are performed in R (R-project) and lead to the same mean difference, see Table A.1. In this Table the p-values is different from Table 4.4 as calculations are done slightly different, although they should agree on the size of the difference which they do. With more observations the p-values would differ less. Model diagnostics of this model can be found in the Appendix in Figures A.4, A.5. Generally, the normality assumption holds and the residual seem to be uniformly distributed around zero. Levene's test for homogeneity [Levene et al., 1960] shows that we can assume variance homogeneity, see Table A.2 for more details on results of Levene's test.

Turning our attention to the additional two linear models, we see that age is a significant predictor when modelling the RMS head displacements and using the app is significant predictor when modelling the time it took before the child to enter the mock scanner. This could indicate that older children better understand the importance of lying still and that using the app helps them familiarize themselves with the scanner faster than those who did not use the app. In Tables 4.5 and 4.6 the estimated coefficients  $\beta$  and the corresponding standard error and p-value for the test of the hypothesis  $\beta = 0$  is given for the RMS and prep-time model respectively. Model diagnostics for these two models can be found in the Appendix Figure A.6, A.7 where we see that the normality assumption is reasonable. There are some patterns which can be found in the residuals that might indicate an interaction should be considered but the small sample size could also give misleading results.

**Table 4.1:** Summary table of pediatric healthy control grouped by gender in the first two columns and by usage of app in the last two columns. It can be seen that, when grouping by gender, there are no noteworthy differences of the various variables (age,Trait etc.) however it seems there is a difference in state, and preparation time when grouping by those who used and didn't use the app. This is interesting and significance of this difference is desired.

Variable	Female, N=9	Male, N=11	With App, N = 10	Without App, N = 10
age				
—Mean	6.66	7.36	7.3	6.8
—Range	[5,9]	[4,10]	[4,10]	[4,10]
Trait				
—Mean	28	29.55	28.5	29.2
—Range	[22,37]	[22,39]	[23,39]	[22,37]
State Before Mock				
—Mean	31.33	29.91	29.1	32
—Range	[21,38]	[21,55]	[21,38]	[23,55]
State After Mock				
—Mean	26.66	26.27	24.6	28.3
—Range	[20,38]	[22,31]	[20,29]	[23,38]
Preparation Time				
—Mean	10.8	10	8.3	12.5
—Range	[7,20]	[7,22]	[7,10]	[7,22]



**Figure 4.9:** Boxplot of the variables preparation time, trait, state (before/after) grouped by those who used the app and those who did not. There are no visible differences in Trait scores, this is expected as using the app should not have an influence on their tendency to anxiety. There is a clear and noteworthy difference in preparation time, which is expected.

**Table 4.2:** Sum of squares ANOVA table for anxiety levels for the healthy control children. Top of the table is the ANOVA for the within subject effects of time and the bottom for the between subjects effect. The tests of interest are highlighted in gray. Significant p-values are denoted in boldface with \*,  $\times$  denotes interaction. The rows for age, Trait and gender are not of interest as they enter as covariates. We see that there is a significant difference of anxiety scores before and after training with a p-value of 0.018 of Timepoint, however it does not state whether it drops or increases, only that there is a difference. We do not see a significant effect of app on anxiety levels as seen by the p-value of 0.144 however it is relatively close to being significant and it might be a problem with the sample size.

Source	Sum of Squares	df	F	p
Timepoint	142.716	1	6.985	<b>0.018*</b>
Timepoint $\times$ app	0.00	1	0.000	0.998
Timepoint $\times$ gender	28.22	1	1.381	0.258
Timepoint $\times$ age	97.62	1	4.778	<b>0.045*</b>
Timepoint $\times$ Trait	154.52	1	7.562	<b>0.015*</b>
Residual Error	306.48	20-5		
<hr/>				
Intercept	177.784	1	3.196	0.094
app	132.207	1	2.376	0.144
gender	34.766	1	0.625	0.442
age	41.477	1	0.746	0.401
Trait	28.149	1	0.506	0.488
Residual Error	834.495	20-5		

**Table 4.3:** Estimated marginal means of anxiety scores obtained from the model described in equation (3.1). The model is evaluated for Trait = 28.85, age = 7.05, gender = 0.45, app = 0.5. The mean of the two time points is different with the mean after preparation being lower.

Timepoint	Mean	Std.Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Before Session	30.550	1.706	26.914	34.186
After Session	26.450	0.945	24.435	28.465

**Table 4.4:** Pairwise comparisons of the estimated marginal means from Table 4.3. The p-value is Bonferroni adjusted [Jafari and Ansari-Pour, 2019] for multiple comparisons. The mean difference, 4.1, is significant at the 0.05 level. This can also be seen as zero is not in the 95% confidence interval. This can be interpreted as a significant drop in anxiety scores during the course of preparation/training.

Timepoint(I)	Timepoint(J)	Mean Difference (I-J)	95% CI			
			Std.Error	Sig	Lower	Upper
Before Session	After Session	4.100	1.429	<b>0.012*</b>	1.053	7.147
After Session	Before Session	-4.100	1.429	<b>0.012*</b>	-7.147	-1.053

**Table 4.5:** Regression Coefficients of RMS linear model. Age is significant on the 0.05 level with a negative estimated coefficient. This can be interpreted as older children move less. The other parameters are not significant on the 0.05 level hence we cannot reject that they are zero.

Variable	$\beta$	SE <sup>1</sup>	95% CI <sup>1</sup>	p-value
age	-1.00	0.42	-1.9, -0.06	<b>0.038*</b>
gender	-0.95	1.20	-3.6, 1.7	0.4
app	-0.59	1.19	-3.2, 2.1	0.6
Trait	0.14	0.13	-0.15, 0.44	0.3
Before_Mock	-0.06	0.09	-0.25, 0.14	0.5

<sup>1</sup>SE = Standard Error, CI = Confidence Interval

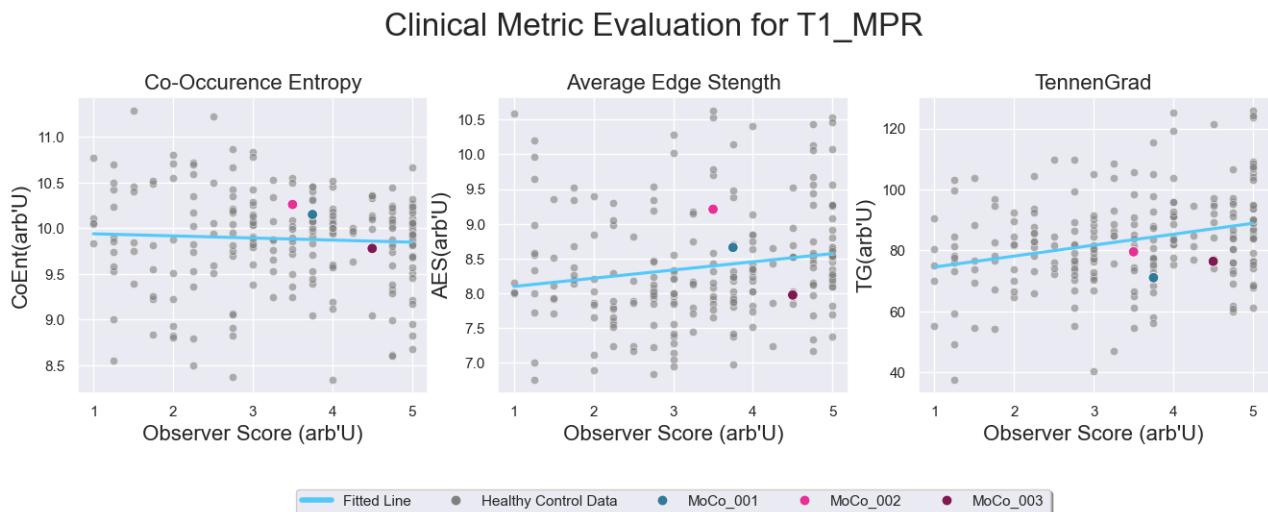
**Table 4.6:** Regression Coefficients of prep model. Note that both age and app is significant with negative estimated coefficients. The interpretation is that the older you are the less preparation time you need to enter the mock scanner and if you used the app you need less preparation time as well. We cannot reject that the remaining parameters are zero.

Variable	$\beta$	SE <sup>1</sup>	95% CI <sup>1</sup>	p-value
age	-1.40	0.45	-2.4, -0.41	<b>0.008*</b>
gender	-0.11	1.48	-3.3, 3.1	>0.9
app	-3.30	1.42	-6.4, -0.27	<b>0.035*</b>
Trait	-0.15	0.14	-0.45, 0.15	0.3
Before_Mock	0.10	0.10	-0.12, 0.32	0.4

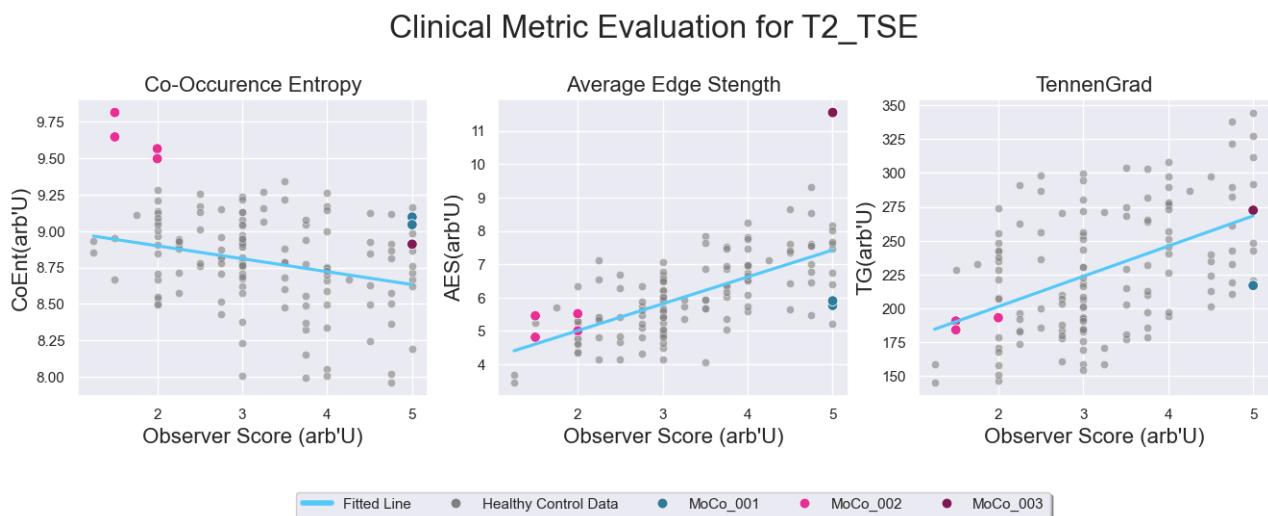
<sup>1</sup>SE = Standard Error, CI = Confidence Interval

## 4.3 Pediatric clinical Trial

Metric correlations of the pediatric clinical scans can be seen in Figure 4.11 where we see AES and TG scores compared to those of the healthy control. Each participant is colored differently to easily compare the scores. The AES and TG scores for the second participant were lower than that of the others. This patient did not have a scan of diagnostic quality and needs an MRI with GA, however the AES score for the first patient is almost as close (only 0.5 larger than that of the second patient). The metric scores for T<sub>2</sub>-FLAIR is plotted in Figure A.8 where the difference between subject 002 and 003 is very low.



**Figure 4.10:** Metric evaluation of the clinical scans of the three participants MoCo\_001, 2 and 3 in dark blue, red and bright red, overlaid the adult healthy control results in gray.



**Figure 4.11:** Metric evaluation of the clinical scans of the three participants MoCo\_001, 2 and 3 in dark blue, red and bright red, overlaid the adult healthy control results in gray. For AES there is almost no distinction between the third and second participant with a difference of only 0.5. The second participant had to get an MRI with GA.



# Chapter 5

---

## Discussion

---

In this chapter I will discuss the results from the preceding chapter. I start by giving a brief overview of the studied metrics and the subsequent results in the healthy adult followed by limitations. The same procedure is followed for the analysis of the healthy pediatric population and clinical trial, where a brief overview of the analysis is given, its implications and lastly the limitations and challenges.

### 5.1 Adult Healthy Control study

In this thesis we studied and evaluated two image quality metrics [Pannetier et al., 2016], a gradient based metric and a Haralick feature based metric. The gradient metric, Average Edge Strength (AES) averages the gradient magnitude of a single image slice, but only for edges that are detected by a given edge detection algorithm, in this thesis Canny edge detector [Canny, 1987]. This is calculated for each slice in a volume and the mean score across the different slices is reported. The idea is that sharp images will have larger gradients than blurry images which will lead to a higher value for AES.

The concept for the Haralick based Co-occurrence Entropy (CoEnt) is that we compare pixel or voxel intensity values across an image. For each pixel we look at the closest neighboring pixels and note how often two intensity values are neighbors. An image with many structures will produce a dense co-occurrence matrix and subsequently high entropy. In addition we compared the metric scores of the aforementioned metrics to those of the Tenengrad (TG) measure [Krotkov, 1989]. We chose to specifically look at these metrics as they are reference-free metrics and they have previously shown promising results for pMC MRI scans [Pannetier et al., 2016].

The three metrics, AES, CoEnt and TG, were evaluated on more than 500 MRI volumes, distributed on a total of 4 different MRI sequences, which were scored by two radiography students and one experienced radiologist. The Spearman correlation between the observer and metric scores was examined. The metrics were also evaluated on 300 additional images that were not scored by observers. The distribution of the metric scores was also examined by boxplots, where the scores

were grouped by motion (still/nod) and MoCo (on/off) to inspect if the scores drop, increase or remain equal. A Wilcoxon rank correlation was used to check if there was a significant difference between having MoCo (on/off).

Turning our attention to the examination of the observer score correlation, we see that CoEnt is by far the worst metric of the three. However, all metrics performed poorly on T<sub>1</sub>-weighted MPRAGE scans when compared to other sequences. This might be explained by subtle differences between an image with and one without artefacts as seen in Figure 4.6. The artefacts are very subtle relative to artefacts found for example in a T<sub>2</sub>-Weighted TSE as seen in figure 2.6. All three metrics collectively had the highest correlation on T<sub>2</sub>-weighted images specifically T<sub>2</sub>-FLAIR had the highest followed by T<sub>2</sub>-TSE. TG and AES had comparable correlations on T<sub>1</sub>-MPRAGE/TIRM and T<sub>2</sub>-FLAIR, however AES has a relatively higher correlation on T<sub>2</sub>-TSE compared to TG. In Figure 4.3 we give a clear overview of the significant Spearman correlations for each image sequence (T<sub>1</sub>-MPRAGE, T<sub>1</sub>-STIR, T<sub>2</sub>-TSE, T<sub>2</sub>-FLAIR) and each metric (AES, TG and CoEnt).

We see both AES and TG showing no significant difference in metric scores for T<sub>1</sub>-MPRAGE scans when comparing pMC on/off when lying still however there is a significant increase in CoEnt scores for this sequence. The non significance is to be expected as pMC does not deteriorate image quality when lying still (if there is no movement there is no update of FOV and no re-acquisition). Both AES and TG show a significant increase in scores for the nodding motion when pMC is on compared to off. CoEnt shows no significant differences.

For T<sub>2</sub>-weighted scans we see a different pattern, and specifically for T<sub>2</sub>-TSE. All metrics show a significant difference when having pMC on compared to off for the still scans. This could be explained by noisy motion tracking which would lead to a disruptive change in FOV. I want to emphasize that even though the images were somewhat degraded, that they were still of fully diagnostic quality (4 or 5 on the Likert scale). Both AES and TG show a significant increase in scores when nodding and pMC is on compared to off. Although the two distributions are more distinguishable for AES compared to TG. This could be explained by its more complicated machinery, such as the Canny edge being better at capturing the artefacts, and thus leading to lower scores for the scans without pMC.

Based on our study the gradient metrics AES and TG are the metrics that correlate the most with observer scores, while CoEnt falls considerably behind and has shown repeatedly that it is an inconsistent metric with a lack of correlation with observer scores and an almost uniform distribution of values regardless of the respective observer scores.

If one should choose between AES and TG, AES is recommended as it correlates well with observer scores (save for T<sub>1</sub>-Weighted MPRAGE) and it specifically outperforms TG for T<sub>2</sub>-Weighted TSE scans. This is important as the T<sub>2</sub>-Weighted TSE is a widely used image sequence for diagnostic purposes and is included in all of Rigshospitalet's sequence protocols. Although one should note that in order for AES to capture eventual artefacts it is required that both convolutions and the Canny edge detector are able to capture the artefacts, i.e. if just one of the operations cannot detect ringing then AES will not be able to detect ringing. As such one

should take care that gradient metrics make sense with regards to what artefacts one wishes to characterize. Gradient methods will inherently be worse at capturing blurring as a blurred image usually does not contain many sharp detectable edges. Still, AES is preferred due to its inherent modularity as any binary edge detector can be used in place of a Canny edge detector. One could either tune or use an edge detector suited for a specific problem and then use AES. Across literature [[Aksoy et al., 2012](#), [Pannetier et al., 2016](#), [Sciarra et al., 2021](#)] AES has only been evaluated using Canny edge detector. It is not obvious that other methods of edge detection has been evaluated with Canny subsequently being the best, as none mention this. A different edge detector, Phase Stretch Transform (PST) [[Asghari and Jalali, 2015](#)] was briefly examined however the results proved to be sub par. One should note that there are many parameters to tune when considering the PST, as such the wrong parameters might have been chosen when evaluating the metric. Using AES with different edge detectors is interesting and should definitely be investigated further.

Another interesting idea is to further build on CoEnt to get comparable results. We only looked at the six closest neighbors of a voxel, yet one could just as easily argue that looking at the 26 closest neighbors is valid (corresponding to the closest diagonal voxels in addition to six closest neighbors, i.e. the 3x3x3 cube encapsulating the voxel of interest). Extending the co-occurrence matrix might also be a viable approach as shown by [[Kovalev et al., 2001](#)], where the co-occurrence matrix is extended to *multisort* co-occurrence matrices by considering, not only the intensity values but also image gradients and mutual orientations. It would be interesting to see if a derived *multisort* CoEnt would lead to a better performance. We have to note that [[Kovalev et al., 2001](#)] uses the matrices for classification and segmentation purposes; this was also the original applications of the Haralick features [[Robert M Haralick, 1973](#)].

### 5.1.1 Summary of strengths and weaknesses

Being able to fast, objectively and algorithmically determine the quality of an MR image or volume is a difficult task and very much still an open problem. The desire for high correlation to the gold standard observer scores further complicates this task. A single metric that can accurately measure multiple artefacts (e.g. blurring, ringing and ghosting) does not currently exist and it is not obvious that one such metric does exist. It might be possible to consider some weighted average of a selection of metrics although this gives rise to the problem of choosing the weights. In addition, observer scores are given as a whole taking multiple factors into account and a given image might score differently from each observer. The rankings of the observer scores can also be discussed, as an image of fully diagnostic quality does not exclude it from having artefacts although the purpose of image metrics should, arguably, be to measure, at a large scale, if MR images are of diagnostic quality and not necessarily if they contain any artefacts.

Most articles feature reference metrics rather than non-reference metrics and the performance of metrics are often shown on images which are artificially distorted rather than actual clinical scans [[Chow and Paramesran, 2016](#)]. Reference metrics

are superior to non-reference metrics when addressing correlation to observer scores although only in the special case when a reference is available. But a reference image is not available in a clinical setting, which to some extent, renders such metrics useless. A natural question could be if it was possible to create a faux reference image based on high quality scans, be it multiple or a single one of acceptable quality. Another student briefly examined this but it lead to problems with registration and thus creating a faux reference image is not a possibility at the moment but should be further investigated.

There is a sea of image metrics to choose from, requiring that it is non-reference and has shown acceptable performance in literature on real MR data rather than artificially distorted data or natural images, drastically reduces the number of options to go with. One might suggest the use of machine learning or model based methods instead as they have shown good results [Sujit et al., 2019]. Such model based methods are dependent on the availability of MR scans with the various artefacts that one could encounter during clinical scans (arguable this renders them as reference image metrics). Such a large amount of MRI volumes, with relevant artefacts, being accessible is not trivial. There are many pools of MR data although the volumes could originate from different scanners (General Electric, Phillips or Cannon) whom might have different magnetization level (Tesla) or an application of a smoothing filter before relaying the image (Phillips). This could train a model that is not optimal for our specific goals and equipment and could render it useless, although it would be interesting to examine such model based methods in greater detail, for example by training a model on a pool of MR images and evaluating on our healthy control images.

## 5.2 Pediatric Healthy Control study

The mock scanner setup and app described in chapter 3 is not cheap thus, it is of great interest to show their effectiveness at helping children undergoing neuroimaging. One study to examine the effectiveness was of a healthy children controls where anxiety levels of 20 healthy children was measured twice, before and after a mock scanning session, using the self reported questionnaire STAI-CH [Spielberger et al., 1970]. A repeated measures ANOVA (examination of the sum of squares) was performed to determine if the means at the two timepoints (before/after) were significantly different. The analysis showed that there is a significant difference between the two timepoints with an F statistic of 6.985, which leads to a p-value of 0.018 when evaluated in a  $F(1, 15)$  distribution, See Table 4.2 for more details. Furthermore, a GLM was fitted with the intent to determine if the anxiety scores increase or decrease during the mock scanning session. The model is used to calculate the EMM which contains the relevant information regarding the difference. A mean difference of 4.1 with a standard deviation of 1.429 was found significant at the 0.05 level, see Table 4.3 for more details. In other words the anxiety levels drop significantly during the mock scanning session. The between-subjects effect of using the app was not found significant with an F statistic of 2.376 and a p-value of 0.144.

The findings of the repeated measures ANOVA are exactly the results we had expected, with the possible exception of non significance of the effect of using the app. A drop of anxiety levels indicates that the children become more comfortable with the environment and scanner which is the entire reasoning for having the mock scanner setup and sessions. It would have been desirable to have the effect of app (whether or not they used the app prior to mock session) be significant with lower anxiety score for those who used it compared to those who did not. A reason for the non significance this might be the low sample size ( $n=20$ ) as it is nearly significant ( $p=0.144$ ).

Two additional linear models were fitted to describe the root mean squared (RMS) head displacement,  $\delta$ , and the time it took to enter the mock scanner,  $\tau$ . The idea is that more anxious children and those who did not use the app would move more and require longer time before they would enter the mock scanner. We also see interesting results looking at these two additional models. We saw that age is a significant predictor when focusing on the RMS head displacement. The interpretation of the estimated parameter (see Table 4.5) is that older children move less than younger children. This result backs the natural assumption that younger children might have more trouble lying still for extended periods of time.

It was also found that using the app was significant at reducing the time needed before entering the mock scanner with a coefficient estimate of -3.3 and an associated p-value of 0.035. This is interpreted as an average decrease of around three minutes when using the app compared to not using the app. Initially this would be deemed quite a low difference when preparation times take around 10 minutes, however it should be noted that one participant who did not receive the app was not willing to enter the mock scanner at all. In addition another participant had to stop the scan and thus the entire scanning session did not finish. Age was also significant when modelling the preparation time which can be interpreted as, the older children more easily adjust/familiarize themselves with the environment, which makes sense.

### 5.2.1 Summary of strengths and weaknesses

Our results indicate that the training session is helping the children to get more comfortable with the hospital environment as shown by the repeated measures analysis and the effect of the the app reduces the time it takes before the child enters the mock scanner. This is the exact reason why the mock scanner training takes place and why the app is offered to the children, which is why these results highly relevant when considering the clinical trial. One of the major limitations for this study is the low sample size. 20 participants is not a lot, especially, when we further divide them into two groups (those with and those without the app). Another complication of the study is the fact that STAI-CH is a self reported questionnaire, hence some answers might be exaggerated or understated in order to answer what their parents want them to, rather than how they truly feel. In addition the answers are dependent on the parents interpretation of the various questions as the children have limited comprehension of some words/phrases used in the questions and naturally the children also has limited reading capabilities. This interpretation is not necessarily equal across all parents of participating children.

Some researchers have looked into supporting the written questions with additional pictograms where the children are asked to pick the picture that they relate to the most, for example by choosing which of four faces (sad, happy, surprised, scared) they relate to the most [Nilsson et al., 2012], although the study did not confirm if it is usable for children under the age of 7.

The app did have a significant effect on preparation time, however some parents of the younger children noted that the children were too easily distracted from the app and wanted to do something else. Perhaps it might be of use to have an option for the younger children 5-7 years old's and one for the elder children aged 8-10. In summary, the studied validated the use of the mobile phone app and preparation/training with the mock scanner.

## 5.3 Pediatric clinical trial

The children participating in the clinical trial got the same preparation as the children in the healthy control study, that is using an app and mock scanner training. They underwent a clinical scan with either pMC on/off if it was deemed that they did not move too much during mock scanner training. The metric results are similar to that of the healthy controls which is to be expected. The lowest metric scores were measured on the second subject and his scan was deemed to be not of diagnostic quality. Of the three participating children his was the only scan with pMC on.

There are two distinct possible reasons subject 002 did not have a scan of diagnostic quality. The first main reason would naturally be that subject motion was too severe even for pMC thus leading to a scan of poor quality. A second possibility is that the scan was not necessarily degraded due to motion of the head, but rather exaggerated facial expressions like blinking and squeezing of the face. Tracoline tracks motion based on light projected on the face, as such facial contusions like excessive hard blinking could lead to Tracoline being "tricked" thinking there is a head motion and in turn updating the scanner FOV incorrectly. Subject 002 was tired at time of scanning but was determined and focused on watching the movie that the child expressed exaggerated facial expressions, like the ones described above, in order to continue the movie. The blame is not entirely on wrong tracking with Tracoline, as the mock scan for this participant also showed motion to such a degree that it was discussed if the MRI scan without GA should be cancelled or not. TracInnovation Aps are aware of such possible mistracks and are looking into possible solutions. Currently we do not know if tracking was tricked or if it was head motion that cause the bad quality.

### 5.3.1 Summary of strengths and weaknesses

It is difficult to draw strong conclusions from four subjects, however three of the participants had a fully diagnostic scan. It is not known with certainty if it was head motion or tricked tracking that was the source of the poor scan quality for

subject 002. All parents of expressed gratitude for the mock training session and the course of the trial. The parents of the child with a non diagnostic scan, were so happy with the mock training that they asked for a second training session even though the next MR scan is with GA.

Unfortunately an amendment to the clinical protocol was not approved in time leading to a lack of STAI-CH repeated measurements data of the pediatric patients. Low participation rates further complicated the course of the experiment. As of December 20<sup>th</sup> only four (only three analyzed in this thesis as a result of time constraints) children participated in the study in the course of four months. This is staggeringly low and unexpected to all parties, especially as one hundred patients received a cerebrum scan in the course of four months (January to April) prior to the start of this thesis. Many attempts were done to encourage doctors to refer children to the study, such as coffee mugs with advertisements on them and further meetings with the pediatricians. One group of doctors has shown a lot of enthusiasm towards the projects namely the oncology pediatricians. Scans of oncology patients are done with contrast fluid injected intravenous which adds an additional layer of complication to the scan. A catheter (needle) needs to be inserted in the vein which is painful and especially distressing if the child is afraid of needles. The injection of contrast fluid can also be uncomfortable, nurses both report a cold feeling, a warm feeling or no feeling when injected with contrast fluid. It has also been discussed if the intravenous catheter can be inserted at a different location in order to separate the two experiences as much as possible. Being able to help this group could make a big impact on many patients especially as brain tumors are the second most common form of cancer, for children, accounting for roughly 27% of pediatric oncology patients [Steward and Wild, 2014] and they undergo multiple scan sessions. In summary, we have too few participants (so far) to draw any strong conclusion, however the study looks promising as three out of the four has a fully diagnostic scan.



# Chapter 6

---

## Conclusion & Outlook

---

In this chapter, I will give an overview of what conclusions can be drawn from this thesis. The chapter is finished with an outlook of what topics can or should be further investigated to expand the analysis or experiments performed in relation to this thesis.

### 6.1 Conclusions

Generally, motion sensitivity is the main drawback of MRI, which can lead to image artefacts hence children are often put in GA should they undergo neuroimaging. In this thesis three reference-free image quality metrics, Average Edge Strength (AES), Co-occurrence Entropy (CoEnt) and Tenengrad (TG) were examined with the intent to measure image artefacts and correlate metric scores with observer scores. The effect of training/preparation on anxiety levels in children participating in a healthy control study was also studied using statistical methods such as repeated measures ANOVA and linear models. Lastly, image metrics were evaluated on clinical scans from the MoCo project's clinical trial.

The metric analysis on more than 800 MR images from 22 healthy adults showed that CoEnt was the worst performing metric of all three metrics (AES, CoEnt, TG) with inconsistent or no correlation to observer scores. CoEnt also showed inconsistency when grouping scores by those with and without prospective motion correction. Observer correlations of AES and TG are comparable with consistent results where AES had slightly higher correlation on most sequences. All metrics had poor correlation on T<sub>1</sub>-MPRAGE and highest correlations for T<sub>2</sub>-weighted images. AES had a higher correlation on T<sub>2</sub>-TSE compared to the other metrics. AES might be better as it is a more advanced function utilizing both image gradients and edge detection. The edge detection in AES intrinsically modular, meaning any binary edge detector can be used, which might be exploitable for a given problem. This leads to the recommendation and conclusion that AES should be preferred over TG and CoEnt, should one choose between these reference free metrics.

The effect of training/preparation on anxiety levels in children participating in a healthy control study was also examined. We found that anxiety scores drop significantly during the course of training. Additionally, we found that using a mobile phone app at home significantly reduced the time it took before the child entered the mock scanner, supporting the relevance and effectiveness of training/preparation. Age was found to be a significant factor when modelling both the root mean squared head displacement and the preparation time. These results are very fascinating as it shows exactly what we had hoped for and it is in accordance with studies recently reported the literature. However I want to emphasize that this exact kind of study (analysis of anxiety levels during the course of training/preparation) has not been done before.

The main goal of the MoCo project is to prepare children for their MRI examination, to such an extent that it renders GA redundant. This thesis shows a positive effect of training and preparation on anxiety levels and it is now up to the clinical trial to determine how many scans are of diagnostic quality and if there is a significant difference in quality for those getting a scan with pMC, for example by evaluating AES on the images. So far three out of four children in the clinical trial had a fully diagnostic scan without GA. At current time it is unknown if the quality of the non diagnostic scan was low as a consequence of head motion or tricked motion tracking. It should also be mentioned that the subject was already on the boundary of not getting an MR scan without GA.

## 6.2 Outlook

The search for optimal non reference image quality metrics is still an open research field with many possibilities and challenges ahead. It is of great interest to expand the metric analysis by further studying:

- Different edge detectors for AES, such as Phase Stretch Transform (PST). To the best of my knowledge, there are no studies which determine the effectiveness of edge detectors different from Canny. With more than 4 parameters to tune, it might be possible to tune PST such that it is applicable in our situation.
- Generalize CoEnt to contain *multi-sort* matrices like gradient matrices. CoEnt could also be extended to count the diagonal neighbors. One could also count intensity values of pixel with not just a distance of 1 to the target pixel but a distance of 2 corresponding to summing over  $i = -2, -1, 1, 2$  in equation (2.11).
- Faux reference images such that reference metrics can be applied. Problems with registration was reported by another student who briefly looked at this. This problem should be addressed first and then one can determine the effectiveness of a reference metric, for example by evaluating on the adult healthy controls. One could evaluate on images different from the still scans without pMC and examine how faux reference metric scores correlate to the "real"

reference metric scores where the still scans without pMC are used as a reference.

- Model based methods trained on pooled MR data. One could train a model on many of the pooled MR data sources such as Cimbi and UK Biobank and then evaluate on the 500 observer scored images as testing to validate if this approach is even viable. Naturally one needs MR scans with some form of associated image score. Differences in image sequences and scanners should be addressed first followed by differences in image scoring.

To expand the MoCo project it might be of interest to have an app option for the 8-10 year old children and another option for the younger children aged 5-7 years, as parents of the younger children reported that their children were easily bored or distracted when using the app. Naturally, having more data would greatly increase the quality of the performed analysis, this also applies to the healthy pediatric controls. It would also be of interest to analyse the STAI-CH scores of the patients participating in the clinical trial and compare them with the pediatric healthy control scores. One could examine if there are differences in scores and how they evolve over time including before and after the real MR scan.

The prospect of scanning children without GA looks promising and might be a possibility in the near future. Having the preparation protocol implemented as a core part of pediatric neuroimaging could drastically reduce the number of children having to undergo GA. This would alleviate the need for anesthesiologists and their time could be spent elsewhere. More pediatric patients could be scanned as only eight GA scans can be performed in a given week. Most importantly, the course of treatment would be greatly improved with reduced health risks for those who need it the most.



---

## List of Figures

---

1.1	MRI Scanner at Rigshospitalet . . . . .	2
2.1	Precession . . . . .	7
2.2	NVM before and after RF pulse . . . . .	9
2.3	$T_2$ relaxtion and FID . . . . .	11
2.4	Gradient Slice selection . . . . .	12
2.5	Spatial domain and k-space of an MR image . . . . .	13
2.6	MRI slice with artifacts compared to clean slice . . . . .	14
2.7	Co-Occurrence Matrix . . . . .	18
2.8	MRI slice with artifacts compared to clean slice . . . . .	19
2.9	Image slice with canny, and convolved images. . . . .	20
2.10	MRI Registration . . . . .	22
3.1	Image acquisition for adult healthy control study . . . . .	30
3.2	MRI Scanner Setup at Rigshospitalet . . . . .	31
3.3	STAI-CH Experimental design . . . . .	33
3.4	Timeline of the Mock training session . . . . .	33
3.5	Clinical Experimental design . . . . .	36
4.1	$T_1$ -MPRAGE Correlation plot . . . . .	38
4.2	$T_2$ -TSE Correlation plot . . . . .	39
4.3	Spearman correlation coefficients of metric scores . . . . .	40
4.4	Edge detected $T_2$ -TSE . . . . .	41
4.5	Convolved $T_2$ -TSE . . . . .	42
4.6	$T_1$ -MPRAGE nodding and still . . . . .	43
4.7	Metric distribution for $T_1$ -MPRAGE and $T_2$ -TSE . . . . .	45
4.8	Distribution of Anxiety Scores . . . . .	47
4.9	Distribution of STAI-CH variables . . . . .	49
4.10	Clinical Metric Results $T_2$ -TSE . . . . .	52
4.11	Clinical Metric Results $T_2$ -TSE . . . . .	53
A.1	$T_2$ -FLAIR Correlation plot . . . . .	79
A.2	$T_1$ -TIRM Correlation plot . . . . .	80
A.3	Metric distribution for $T_2$ -FLAIR and $T_2^*$ -STAR . . . . .	81

A.4	Model diagnostics for the repeated measure GLM . . . . .	82
A.5	QQ-plot of STAI-CH State scores . . . . .	83
A.6	Model diagnostics for the RMS linear model . . . . .	84
A.7	Model diagnostics for the preparation time linear model . . . . .	84
A.8	Clinical Metric Result T <sub>2</sub> -FLAIR . . . . .	85
B.1	Matrix visualised as image . . . . .	88
B.2	Box blurring Convolution . . . . .	89
B.3	Gradient Convolution . . . . .	89
B.4	Spatial and Fourier/frequency domain of a natural image . . . . .	90
B.5	Spatial and Fourier/frequency domain of a wavy image . . . . .	90
B.6	MRI axial, sagittal, coronal and render view . . . . .	91
B.7	Slice view of different MR sequences . . . . .	92
B.8	T <sub>1</sub> -MPRAGE still, nod and shake motion . . . . .	92
C.1	Example of STAI-CH State . . . . .	93
C.2	Example of STAI-CH Trait . . . . .	93

---

## List of Tables

---

2.1	Cerebrum GA scans of children . . . . .	6
2.2	Likert scale of expert observer scores . . . . .	21
2.3	Exempt of STAI-CH data . . . . .	27
4.1	Pediatric HC Summary Table . . . . .	48
4.2	Sum of squares ANOVA table . . . . .	50
4.3	Estimated Marginal Means . . . . .	50
4.4	Pairwise Comparison of estimated marginal means . . . . .	51
4.5	RMS Regression Coefficients . . . . .	51
4.6	Preparation Time Regression Coefficients . . . . .	51
A.1	EMM from regression model . . . . .	82
A.2	Levene's test of variance homogeneity . . . . .	83



---

## Bibliography

---

- [Aksoy et al., 2012] Aksoy, M., Forman, C., Straka, M., Cukur, T., Hornegger, J., and Bammer, R. (2012). Hybrid Prospective and Retrospective Head Motion Correction to Mitigate Cross-Calibration Errors. *Magnetic Resonance in Medicine*, 67(5):1237–1251.
- [Asghari and Jalali, 2015] Asghari, M. H. and Jalali, B. (2015). Edge Detection in Digital Images Using Dispersive Phase Stretch Transform. *International Journal of Biomedical Imaging*, 2015.
- [Ashmore et al., 2019] Ashmore, J., Di Pietro, J., Williams, K., Stokes, E., Symons, A., Smith, M., Clegg, L., and McGrath, C. (2019). A free virtual reality experience to prepare pediatric patients for magnetic resonance imaging: Cross-sectional questionnaire study. *JMIR Pediatrics and Parenting*, 2(1):1–10.
- [Benjamini and Hochberg, 1995] Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*, 57(1):289–300.
- [Buela-casal, 2014] Buela-casal, A. G.-r. G. (2014). Meta-Analysis of Group Comparison and Meta-Analyis of Reliability Generalization of the State-Trait Anxiety Inventory Questionnaire. *Spanish Magazine of Public Health*, 88:101–112.
- [Canny, 1987] Canny, J. (1987). A Computational Approach to Edge Detection. pages 184–203. Morgan Kaufmann, San Francisco (CA).
- [Chang YC, 1999] Chang YC, Huang KM, C. J. S. C. (1999). Impact of Magnetic Resonance Imaging on the advancement of medicine. *Journal of The Formosan Medical Association*, 99(11):740–748.
- [Chatterjee et al., 2020] Chatterjee, S., Sciarra, A., Dünnwald, M., Oeltze-Jafra, S., Nürnberg, A., and Speck, O. (2020). Retrospective Motion Correction of MR Images using Prior-Assisted Deep Learning. (NeurIPS).
- [Chow and Paramesran, 2016] Chow, L. S. and Paramesran, R. (2016). Review of medical image quality assessment. *Biomedical Signal Processing and Control*, 27(November 2019):145–154.

- [Crowder and Hand, 1990] Crowder, M. J. and Hand, D. J. (1990). *Analysis of repeated measures*. Chapman and Hall, first edition.
- [Davis, 2002] Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. Springer.
- [De Bie et al., 2010] De Bie, H. M., Boersma, M., Wattjes, M. P., Adriaanse, S., Vermeulen, R. J., Oostrom, K. J., Huisman, J., Veltman, D. J., and Delemarre-Van De Waal, H. A. (2010). Preparing children with a mock scanner training protocol results in high quality structural and functional MRI scans. *European Journal of Pediatrics*, 169(9):1079–1085.
- [Dhawan, 2010] Dhawan, A. P. (2010). *Medical Image Analysis*. Wiley, 2 edition.
- [Duffy et al., 2021] Duffy, B. A., Zhao, L., Sepehrband, F., Min, J., Wang, D. J., Shi, Y., Toga, A. W., and Kim, H. (2021). Retrospective motion artifact correction of structural MRI images using deep learning improves the quality of cortical surface reconstructions. *NeuroImage*, 230(January):117756.
- [Eichhorn and Vascan, 2021] Eichhorn, H. S. and Vascan, A.-V. (2021). Characterisation of children ' s head motion for Magnetic Resonance Imaging with and without general anaesthesia. 1(December):1–10.
- [Ernst Hansen, 2017] Ernst Hansen, Søren Eilers, T. G. M. (2017). *Indledende Matematisk Analyse*. Institut For Matematiske Fag Københavns Universitet, 8 edition.
- [Fallis, 2013] Fallis, A. (2013). *MRI Physical Principles and Sequence Design*. Wiley Blackwell, 2 edition.
- [Fischl, 2012] Fischl, B. (2012). FreeSurfer. *NeuroImage*, 62(2):774–781.
- [Frost et al., 2019] Frost, R., Wighton, P., Karahanoglu, F. I., Robertson, R. L., Grant, P. E., Fischl, B., Tisdall, M. D., and van der Kouwe, A. (2019). Markerless high-frequency prospective motion correction for neuroanatomical MRI. *Magnetic Resonance in Medicine*, 82(1):126–144.
- [Godenschweger et al., 2016] Godenschweger, F., Kägelein, U., Stucht, D., Yarach, U., Sciarra, A., Yakupov, R., Lüsebrink, F., Schulze, P., and Speck, O. (2016). Motion correction in MRI of the brain. *Physics in Medicine and Biology*, 61(5):R32–R56.
- [Greve and Fischl, 2009] Greve, D. N. and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48:63–72.
- [Haacke and Patrick, 1986] Haacke, E. and Patrick, J. L. (1986). Reducing motion artifacts in two-dimensional Fourier transform imaging. *Magnetic Resonance Imaging*, 4(4).

- [Hannah Susanne Eichhorn, 2021] Hannah Susanne Eichhorn (2021). *Comparison of prospective and retrospective motion correction for Magnetic Resonance Imaging of the brain (Msc Thesis)*.
- [Haskell, 2019] Haskell, M. W. (2019). *Retrospective Motion Correction for Magnetic Resonance Imaging*. PhD thesis, Harvard.
- [Ing et al., 2012] Ing, C., DiMaggio, C., Whitehouse, A., Hegarty, M. K., Brady, J., Von Ungern-Sternberg, B. S., Davidson, A., Wood, A. J., Li, G., and Sun, L. S. (2012). Long-term differences in language and cognitive function after childhood exposure to anesthesia. *Pediatrics*, 130(3).
- [Jafari and Ansari-Pour, 2019] Jafari, M. and Ansari-Pour, N. (2019). Why, when and how to adjust your P values? *Cell Journal*, 20(4):604–607.
- [Kecskemeti et al., 2018] Kecskemeti, S., Samsonov, A., Velikina, J., Field, A. S., Turski, P., Rowley, H., Lainhart, J. E., and Alexander, A. L. (2018). Robust motion correction strategy for structural MRI in unsedated children demonstrated with three-dimensional radial MPnRAGE. *Radiology*, 289(2):509–516.
- [Knudsen et al., 2016] Knudsen, G. M., Jensen, P. S., Erritzoe, D., Baaré, W. F., Ettrup, A., Fisher, P. M., Gillings, N., Hansen, H. D., Hansen, L. K., Hasselbalch, S. G., Hennigsson, S., Herth, M. M., Holst, K. K., Iversen, P., Kessing, L. V., Macoveanu, J., Madsen, K. S., Mortensen, E. L., Nielsen, F. , Paulson, O. B., Siebner, H. R., Stenbæk, D. S., Svarer, C., Jernigan, T. L., Strother, S. C., and Frokjaer, V. G. (2016). The Center for Integrated Molecular Brain Imaging (Cimbi) database. *NeuroImage*, 124:1213–1219.
- [Kovalev et al., 2001] Kovalev, V. A., Kruggel, F., Gertz, H.-j., and Cramon, D. Y. V. (2001). Three-Dimensional Texture Analysis of MRI Brain Datasets. 20(5):424–433.
- [Krotkov, 1989] Krotkov, E. P. (1989). *Active Computer Vision by Cooperative Focus and Stereo*. Number 8. Springer-Verlag.
- [Lambert and Mazzola, 2003] Lambert, J. B. and Mazzola, E. P. (2003). *Nuclear Magnetic Resonance Spectroscopy An Introduction to Principles, Applications, and Experimental Methods*, volume 129. Pearson Prentice Hall, 1 edition.
- [Levene et al., 1960] Levene, H., Olkin, I., and Hotelling, H. (1960). Robust tests for equality of variances. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, page 278–292. Stanford University Press.
- [Miao et al., 2008] Miao, J., Wong, W. C. K., and Wilson, D. L. (2008). Weighted perceptual difference model (case-PDM) for MR image quality evaluation. *Medical Imaging 2008: Image Perception, Observer Performance, and Technology Assessment*, 6917(March):69170L.

- [Mortamet et al., 2009] Mortamet, B., Bernstein, M. A., Jack Jr, C. R., Gunter, J. L., Ward, C., Britson, P. J., Meuli, R., Thiran, J.-P., and Krueger, G. (2009). Automatic quality assessment in structural brain magnetic resonance imaging and the Alzheimer's Disease Neuroimaging Initiative \*. *Magnetic Resonance Imaging*, 26(2):365–372.
- [Nilsson et al., 2012] Nilsson, S., Buchholz, M., and Thunberg, G. (2012). Assessing Children's Anxiety Using the Modified Short State-Trait Anxiety Inventory and Talking Mats: A Pilot Study. *Nursing Research and Practice*, 2012:1–7.
- [Olesen et al., 2012] Olesen, O. V., Paulsen, R. R., Højgaard, L., Roed, B., and Larsen, R. (2012). Motion tracking for medical imaging: A nonvisible structured light tracking approach. *IEEE Transactions on Medical Imaging*, 31(1):79–87.
- [Pannetier et al., 2016] Pannetier, N., Stavrinos, T., Ng, P., Herbst, M., Zaitsev, M., Young1, K., Matson, G., and Schuff, N. (2016). Quantitative framework for prospective motion correction evaluation. *Magnetic Resonance in Medicine*, 75(2):810–816.
- [Richard Feynman et al., 1969] Richard Feynman, Matthew Sands, and Robert B. Leighton (1969). *The Feynman Lectures on Physics, Volume II Chapter 35*. California Institute of Technology, Addison-Wesley.
- [Robert M Haralick, 1973] Robert M Haralick, K. Shanmugam, I. D. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, 6:610–621.
- [Runge et al., 2018] Runge, S. B., Christensen, N. L., Jensen, K., and Jensen, I. E. (2018). Children centered care: Minimizing the need for anesthesia with a multi-faceted concept for MRI in children aged 4–6. *European Journal of Radiology*, 107(August 2018):183–187.
- [Sciarra et al., 2021] Sciarra, A., Mattern, H., Yakupov, R., Chatterjee, S., Stucht, D., Oeltze-Jafra, S., Godenschweger, F., and Speck, O. (2021). Quantitative evaluation of prospective motion correction in healthy subjects at 7T MRI. *Magnetic Resonance in Medicine*, (July):1–12.
- [Slipsager et al., 2019] Slipsager, J. M., Ellegaard, A. H., Glimberg, S. L., Paulsen, R. R., Dylan Tisdall, M., Wighton, P., Van Der Kouwe, A., Marner, L., Henriksen, O. M., Law, I., and Olesen, O. V. (2019). Markerless motion tracking and correction for PET, MRI, and simultaneous PET/MRI. *PLoS ONE*, 14(4).
- [Spielberger et al., 1970] Spielberger, C., Gorsuch, R., and Lushene, R. (1970). *STAI manual for the state-trait anxiety inventory. Self-Evaluation Questionnaire*.
- [Steward and Wild, 2014] Steward, B. and Wild, C., editors (2014). *World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015*. Internation Agency for Research on Cancer.

- [Stratmann, 2011] Stratmann, G. (2011). Neurotoxicity of anesthetic drugs in the developing brain. *Anesthesia and Analgesia*, 113(5):1170–1179.
- [Sujit et al., 2019] Sujit, S. J., Gabr, R. E., Coronado, I., Robinson, M., Datta, S., and Narayana, P. A. (2019). Automated Image Quality Evaluation of Structural Brain Magnetic Resonance Images using Deep Convolutional Neural Networks. *2018 9th Cairo International Biomedical Engineering Conference, CIBEC 2018 - Proceedings*, (December):33–36.
- [Szeliski, 2021] Szeliski, R. (2021). *Computer Vision: Algorithms and Applications*. Springer, 2nd edition.
- [Talagala and Lowe, 1991] Talagala, S. L. and Lowe, I. J. (1991). Introduction to magnetic resonance imaging. *Concepts in Magnetic Resonance*, 3(3):145–159.
- [Topiwala, 1998] Topiwala, P. N. (1998). Wavelet Image and Video Compression. chapter 4.6, page 79. Kluwer Academic Publishers.
- [Törnqvist et al., 2015] Törnqvist, E., Måansson, , and Hallström, I. (2015). Children having magnetic resonance imaging: A preparatory storybook and audio/visual media are preferable to anesthesia or deep sedation. *Journal of Child Health Care*, 19(3):359–369.
- [Wald, 1943] Wald, A. (1943). Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of the American Mathematical Society*, 54(3):426.
- [Wang et al., 2004] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.
- [Wang et al., 2003] Wang, Z., Simoncelli, E. P., and Bovik, A. C. (2003). Multi-scale structural similarity for image quality assessment. *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, 2(January 2014):1398–1402.
- [Weishaupt and Marincek, 2008] Weishaupt, D. and Marincek, B. (2008). *How Does MRI Work*. 2 edition.
- [Wilcoxon, 2012] Wilcoxon, F. (2012). Individual Comparisons by Ranking Methods. *American Cyanamid Co.*, 1(6):80–83.
- [Zaitsev et al., 2015] Zaitsev, M., Maclarens, J., and Herbst, M. (2015). Motion artifacts in MRI: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, 42(4):887–901.



# Appendices

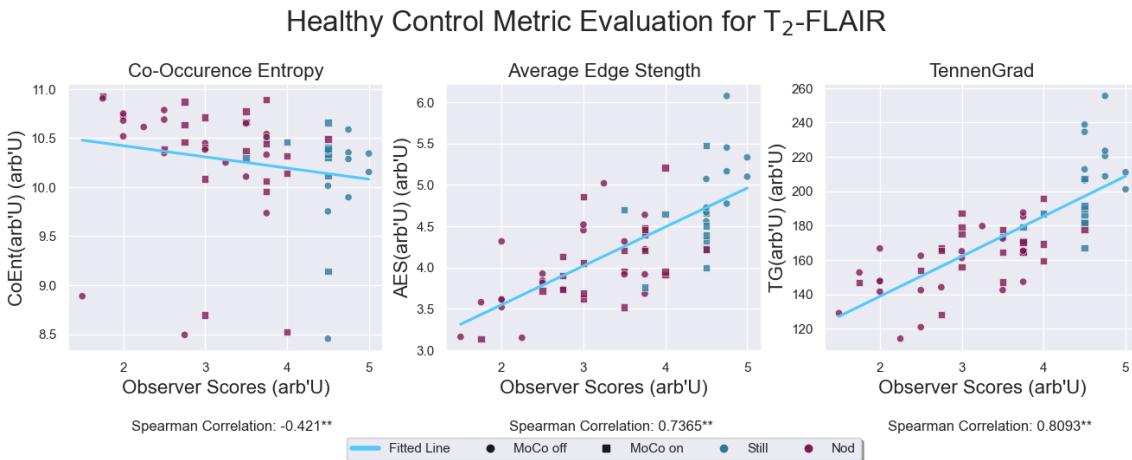


# Chapter A

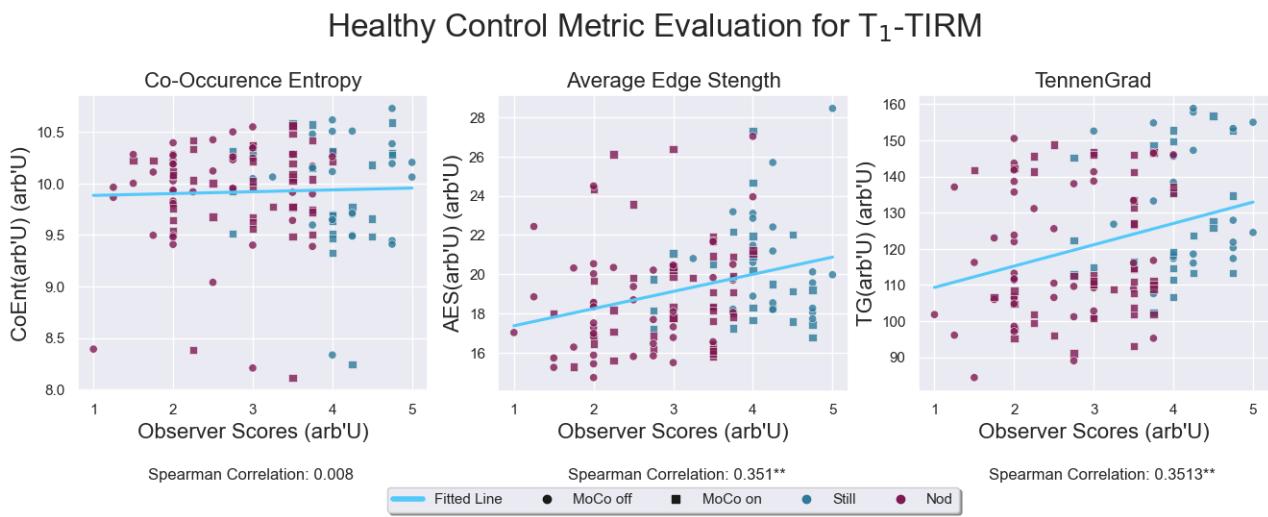
## Additional Results

### A.1 Adult Healthy Control

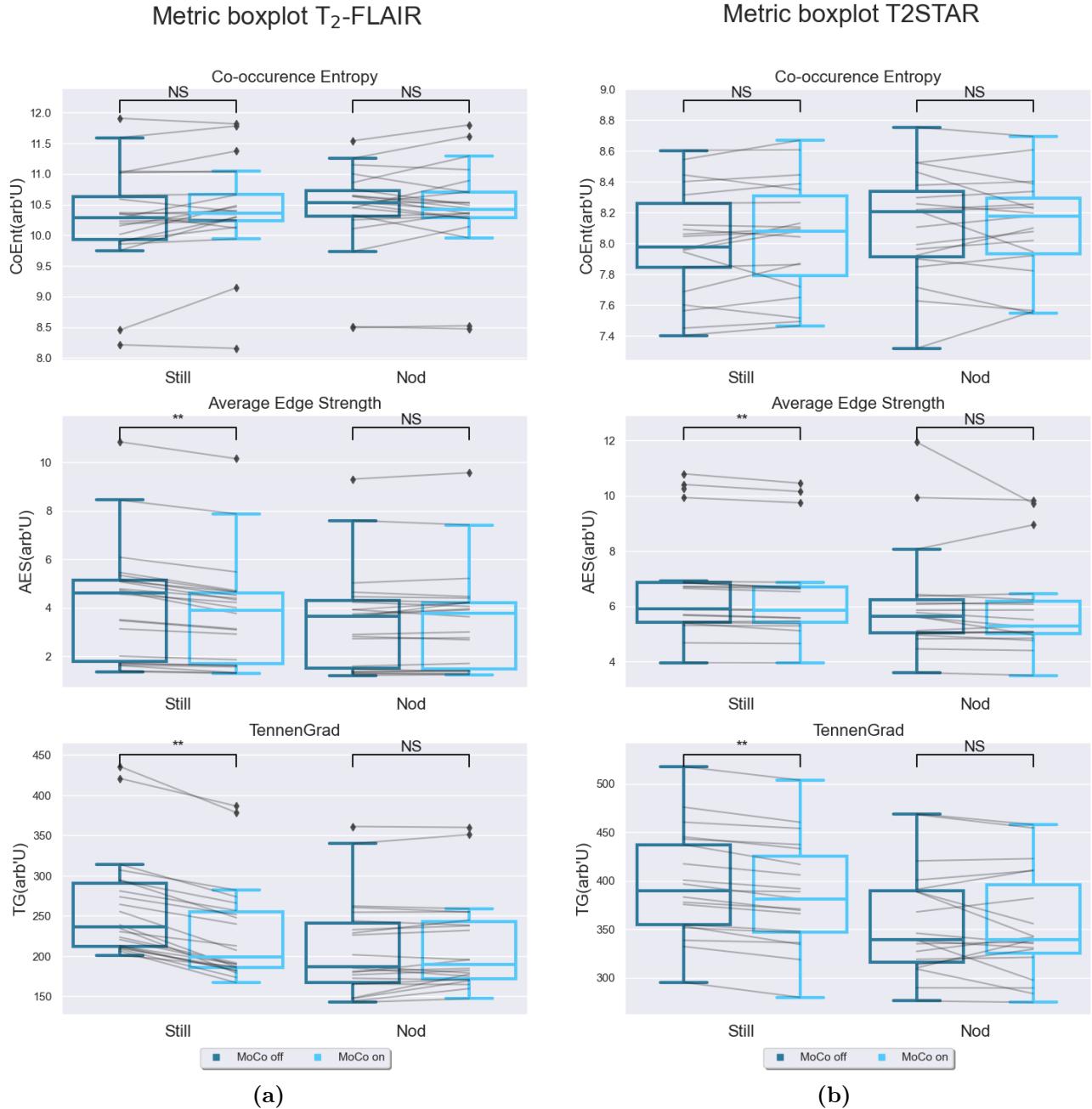
In this section we have additional metric Results



**Figure A.1:** Correlations between average observer score and the corresponding metric evaluated score from CoEnt, AES and TG for T2-weighted FLAIR scans without reacquisition. Data is colored according to motion and the style of the marker represent whether pMC was on or off with a square denoting pMC scans. For each metric the associated Spearman correlation and its p-value is reported below the relevant figure. A p-value less than 0.001 is annotated with \*\*, one less than 0.05 with \*. A regression line is plotted to easily see the general trend. AES, TG and CoEnt show significant correlation although for CoEnt the correlation is negative.



**Figure A.2:** Correlations between average observer score and the corresponding metric evaluated score from CoEnt, AES and TG for T1-weighted TIRM scans without reacquisition. Data is colored according to motion and the style of the marker represent whether pMC was on or off with a square denoting pMC scans. For each metric the associated Spearman correlation is reported below the relevant figure. A regression line is plotted to easily see the general trend. AES and TG show significant correlation while CoEnt does not.



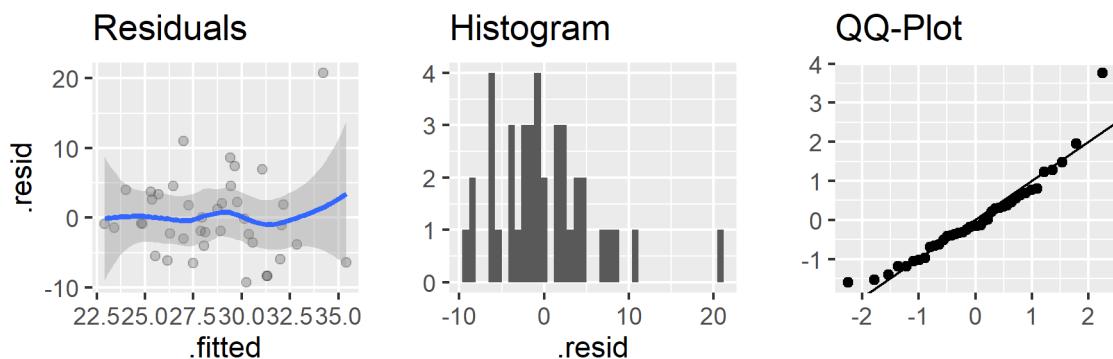
**Figure A.3:** Quartile distribution of metric scores grouped by motion correction (off and on colored dark blue and bright blue respectively) and head movement(still/nodding) for three metrics (CoEnt, AES and TG in descending position). Paired data are connected with gray lines and the significance level of the difference is denoted either NS, \* or \*\* representing a not significant pvalue  $p \geq 0.05$ , a pvalue  $p < 0.05$  and  $p < 0.01$  respectively. (a) Metric results for T2-Weighted FLAIR. (b) Metric results for T2\*.

## A.2 Pediatric Healthy Control

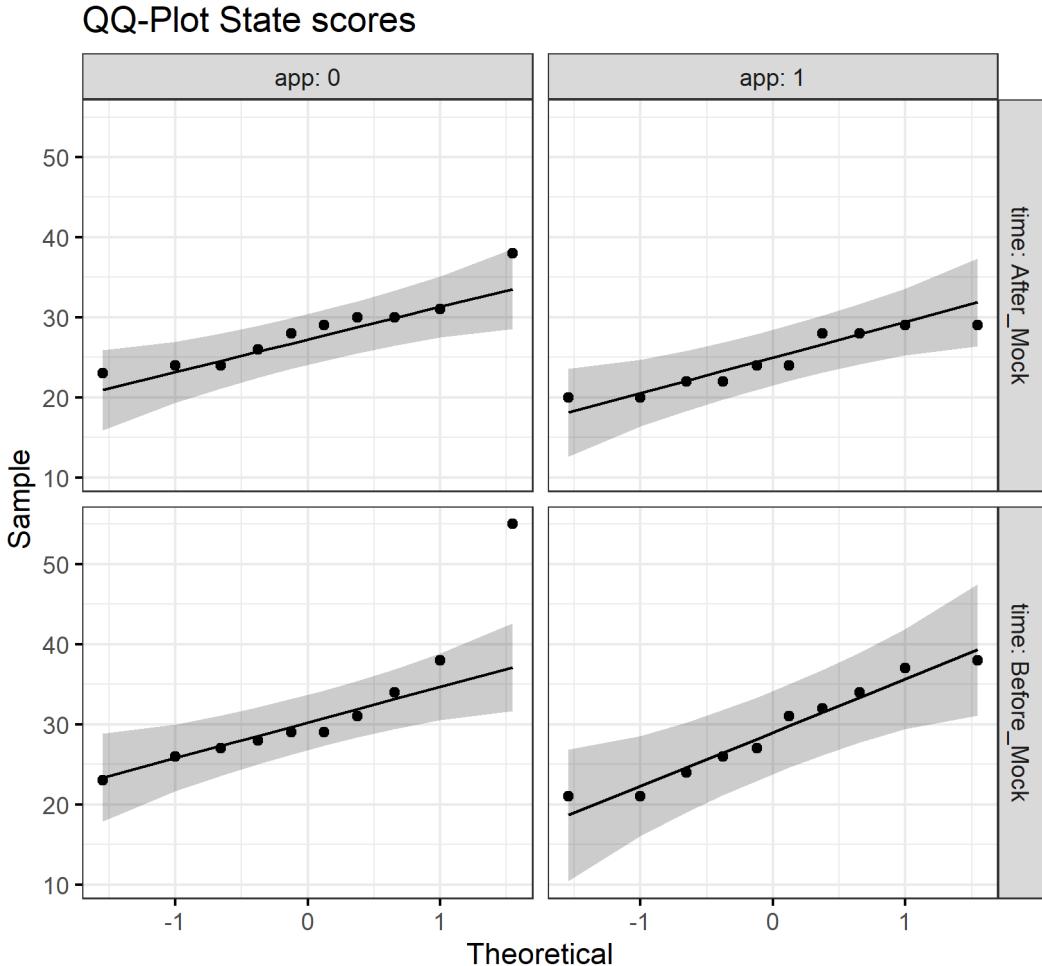
**Table A.1:** Estimated marginal means from a regression model. The mean difference is shown in the Beta column and row Before\_Mock. This is also significant. Difference in p-value is explained by slightly different methods of calculations of the two programs (SPSS, R).

Characteristic	Beta	SE <sup>1</sup>	95% CI <sup>1</sup>	p-value
time				
After_Mock	—	—	—	
Before_Mock	4.1	1.98	0.08, 8.1	<b>0.046</b>
age	0.64	0.619	-0.62, 1.9	0.3
Trait	0.16	0.190	-0.22, 0.55	0.4
gender	2.0	2.10	-2.3, 6.2	0.4
app	-3.7	2.01	-7.8, 0.39	0.074
<sup>1</sup> SE = Standard Error, CI = Confidence Interval				

### A.2.1 Model Diagnostics



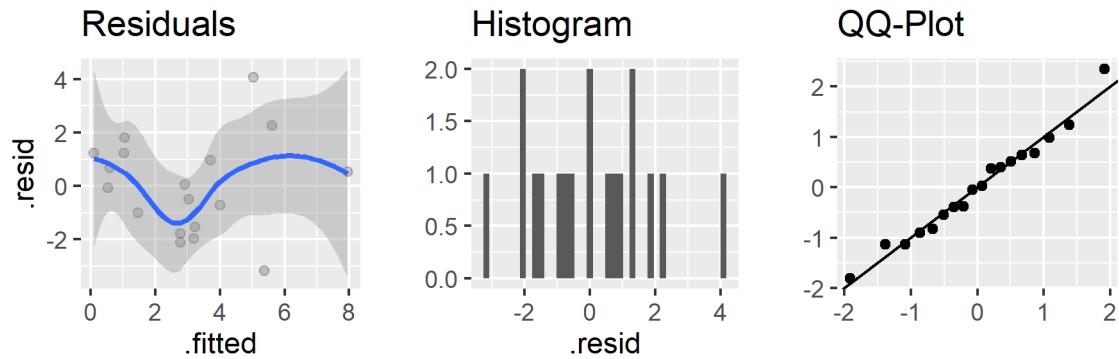
**Figure A.4:** Model diagnostics for the repeated measure GLM. Residuals looks good (uniformly distributed around 0) and the normality assumptions looks reasonable when observing the QQ-plot.



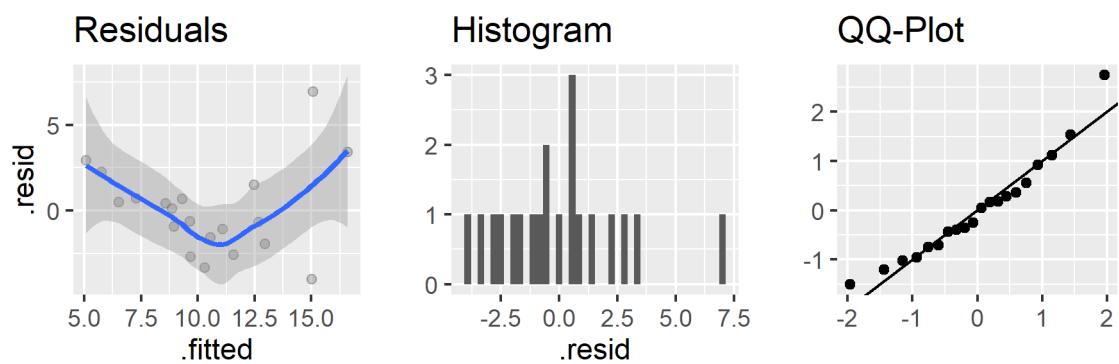
**Figure A.5:** QQ-plot of the STAI-CH State scores, the scores should all be normally distributed (lie, somewhat, on the straight line which they do) for each level of (before, after)  $\times$  (with app, without app). The normality assumption holds.

**Table A.2:** Levene's test of variance homogeneity [Levene et al., 1960]. Tests the null hypothesis that the error variance of the dependent variable is equal across groups. As the p-value is not significant, we cannot reject the hypothesis thus we can assume variance homogeneity.

	F	df1	df2	p-value
Before Mock	0.780	1	18	0.389
After Mock	0.000	1	18	0.993

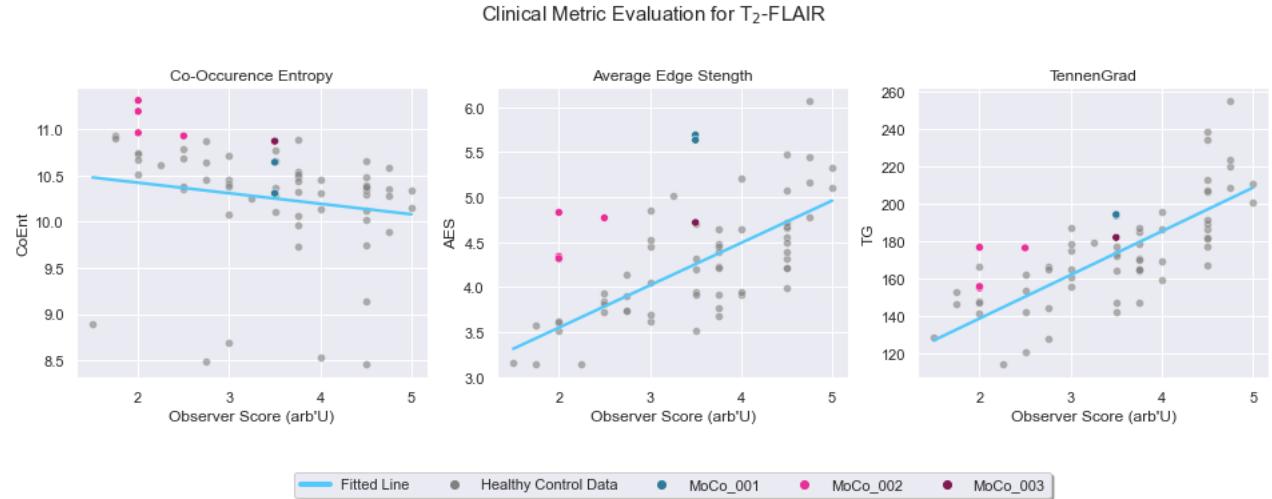


**Figure A.6:** Model diagnostics for the RMS linear model. It seems the normality assumption holds when looking at the qqplot however looking at the residuals it would seem that linearity does not hold and perhaps an interactions should be included.



**Figure A.7:** Model diagnostics for the preparation time linear model. It seems the normality assumption holds when looking at the qqplot however looking at the residuals it would seem that linearity does not hold and perhaps an interactions should be included.

### A.3 Clinical Trial



**Figure A.8:** Metric evaluation of the clinical scans of the three participants MoCo\_001, 2 and 3 in dark blue, red and bright red, overlaid the adult healthy control results in gray. For AES there is almost no distinction between the third and second participant. The second participant had to get an MRI with GA. The AES scores seem higher than for the adult healthy controls.



# Chapter B

---

## Supplementary Material

---

In this chapter I will briefly outline the basics of how computers represent images and some of the core background. This includes how an image is represented by a matrix and the spatial domain and K-space of images. Only grey scale images will be considered as MRI can only produce grey scale volumes. The presented theory and examples are built on [Szeliski, 2021] see <https://szeliski.org/Book/> for greater details. Different MR images are also visualized such as different sequences and slice axes.

### B.1 Image Representation

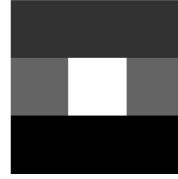
A grey scale image can be represented by a matrix who's dimension corresponds to the resolution of the image. Each entry of the matrix is an integer in the range  $[0,255]$  and the integer values are referred to as intensity values. A black pixel will have the intensity value of 0 while a white pixel will have an intensity value of 255. This leaves  $2^8 = 256$  different intensity values.

#### Example

Consider the  $3 \times 3$  matrix  $M \in [0, 255]^{3 \times 3}$

$$M = \begin{bmatrix} 50 & 50 & 50 \\ 100 & 255 & 100 \\ 0 & 0 & 0 \end{bmatrix}. \quad (\text{B.1})$$

This can be visualized as an image  $\mathcal{I}$  by mapping the values  $[0,255]$  to intensities of pixels. The bottom row will be black, the center white while the top row will be grey, as seen below in Figure B.1. Any image on a computer is represented and stored in a matrix like above, with the only difference being the shape of the matrix. An image  $\mathcal{I}$  with "HD" resolution would be represented by a matrix  $M \in [0, 255]^{1080 \times 1920}$  with 1920 columns and 1080 rows.



**Figure B.1:** The matrix  $M$  visualised as an image.

One can also interpret an image  $\mathcal{I}$  as a function  $f : \mathbb{N}_0 \times \mathbb{N}_0 \rightarrow [0, 255]$ . With this we can define convolution with a kernel/filter  $\omega$  by:

$$g(x, y) = \sum_{dx=-a}^a \sum_{dy=-b}^b \omega(dx, dy) \cdot f(x + dx, y + dy) \quad (\text{B.2})$$

where  $a, b$  denotes the shape of the kernel/filter and  $\cdot$  represents regular multiplication ( $f(x + dx, y + dy)$ ,  $\omega(dx, dy)$  are just regular natural numbers). Note that  $a, b$  can be of a different size than the image itself, often kernels are of size  $3 \times 3$ . Blurring and edge detection is built on the convolution operator as will be shown below.

### Example

The identity kernel is given by the matrix:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (\text{B.3})$$

As the name suggest nothing will happen if an image is convolved with the identity kernel.

The box blurring kernel of size 3 is given by:

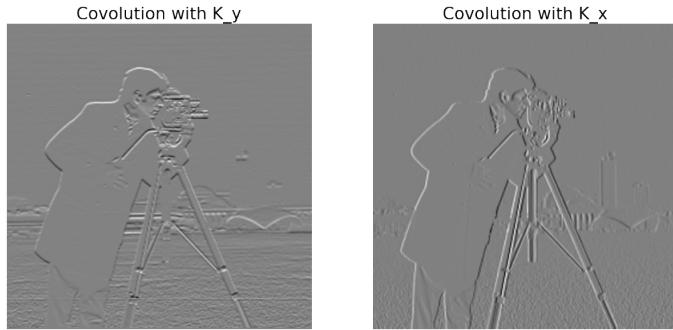
$$\frac{1}{3^2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \quad (\text{B.4})$$

This kernel will blur an image. If one instead convolves with a blurring kernel of say size  $15 \times 15$  then the image would get more blurry as seen in Figure B.2 where we blur an image with three different box blurr kernels. Images can also be "differentiated" by applying gradient kernels. We can differentiate with respect to  $y$  by convolving with the kernel

$$k_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad (\text{B.5})$$



**Figure B.2:** Convolution with box blurring kernel of kernel size k=3,9,15, Note how the blurring increases with kernel size



**Figure B.3:** Convolutions of an image with the kernel described in equation (B.5) on the left. On the right the same image convolved with the transposed kernel. Note the different lines that are detected. Horizontal lines are detected with the kernel  $k_y$  and vertical with  $k_x$ .

taking inspiration from the equation

$$\frac{f(x+h) - f(x)}{h} \quad (\text{B.6})$$

for some differentiable function  $f$ . The gradient in the  $x$  direction is the transposed of  $k_y$  that is  $k_x = k_y^T$ . Convoluting with these gives: These derivatives can be used to detect edges in an image as an edge will have a larger gradient. The reason that  $k_y$  detects horizontal lines is that the pixels in the top row of the kernel are compared to those of the bottom row. If the pixel values are equal, there is no edge and the sum of values would be zero. However if they are different it will lead to a value different from zero. This is the value we see in Figure B.3

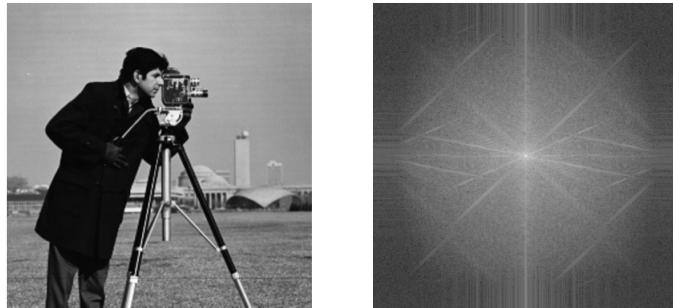
## B.2 K-space and Spatial Domain

It is well known that an integrable function  $f : \mathbb{R} \rightarrow \mathbb{C}$  has the Fourier representation:

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) \exp(2\pi i x \xi) d\xi, \quad \forall x \in \mathbb{R}. \quad (\text{B.7})$$

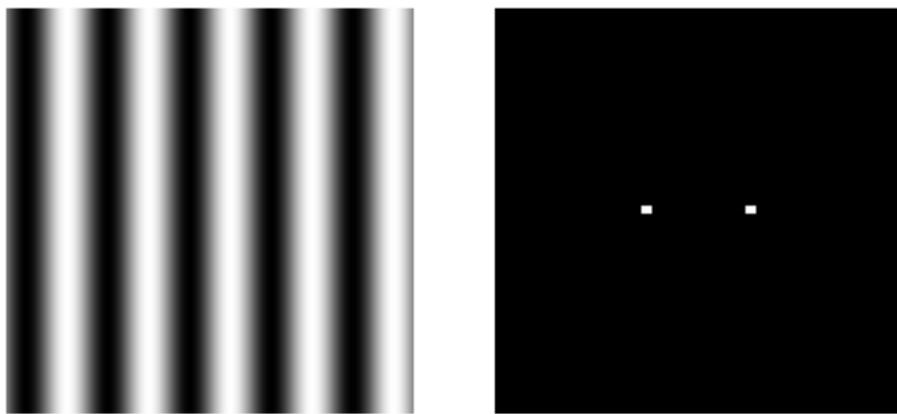
In loose terms we can write the function  $f$  with a bunch of wave functions. The same can be done for an image using the discrete 2D Fourier transform, recall that we can treat an image as a discrete function. This can be visualized as well as

seen in Figure B.4 where we on the left have the normal image in what we call spatial domain and on the right we have the image in Fourier domain or k-space. In Fourier space each pixel corresponds to a wave which can be seen if we take



**Figure B.4:** A natural image on the left in spatial domain and the corresponding image in frequency domain on the right.

the Fourier transform of a wavy image as seen in Figure B.5 We say the image is

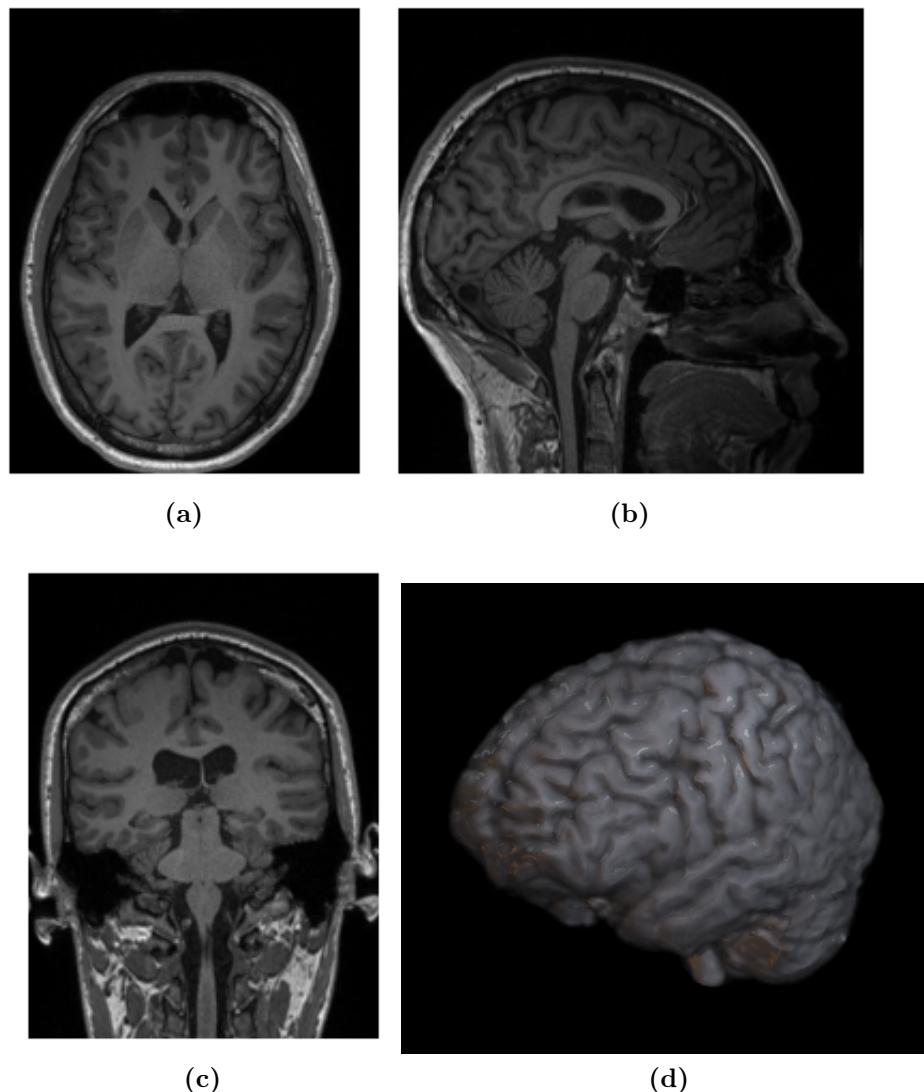


**Figure B.5:** On the left a wavy image in spatial domain with the corresponding image in frequency domain on the right.

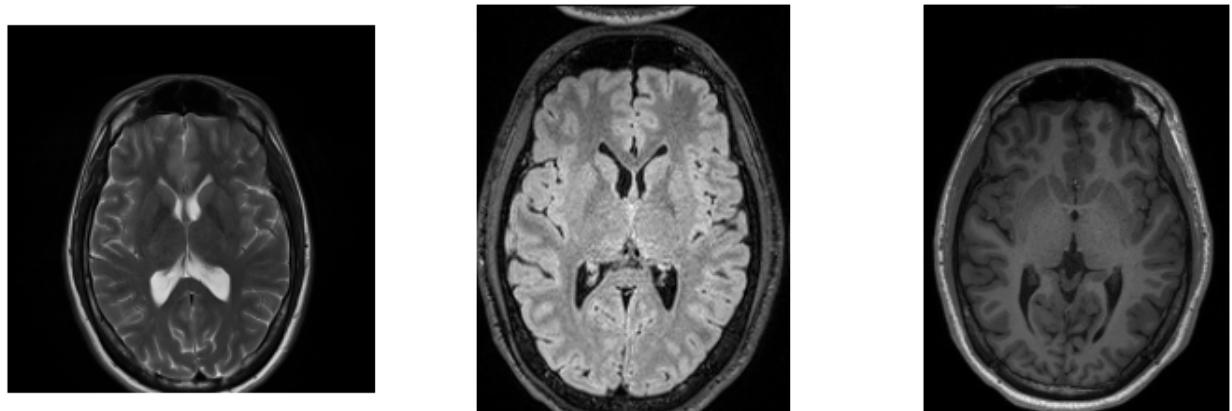
acquired in spatial domain when you take a picture with your phone or camera however MR images are "taken" in k-space and the image we see is the inverse Fourier transformed k-space image.

### B.3 MRI Slices and Sequences

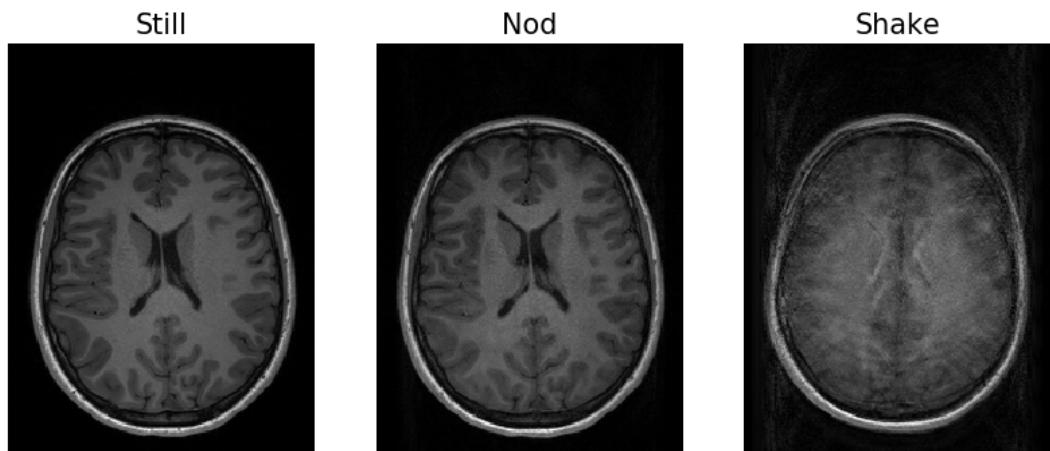
As mentioned earlier, 3D encoded images can be viewed from multiple axis namely, axial, sagittal and coronal. In addition we can also render a 3D object if we desire. Axial view is used throughout the thesis but in Figure B.6 the additional three slicing views are visualised for a T<sub>1</sub>-MPRAGE. Different sequences lead to images with different types of contrast which is visualized in Figure B.7 where we compare T<sub>2</sub>-TSE, T<sub>2</sub>-FLAIR and T<sub>1</sub>-MPRAGE, note that different parts of the brain are bright for the different sequences.



**Figure B.6:** T1-MPRAGE in (a) axial, (b) sagittal, (c) coronal and (d) render view



**Figure B.7:** Three different sequences visualized. On the left T<sub>2</sub>-TSE, in the middle T<sub>2</sub>-FLAIR, and on the right T<sub>1</sub>-MPRAGE. Note how each image has different parts of the brain highlighted.



**Figure B.8:** Three T<sub>1</sub>-MPRAGE scans from the same subject, first lying still, followed by a scan with nodding motion and one with a shaking motion. There are subtle differences for the nodding and still scans, however the scan with a shaking motion is completely of non diagnostic quality.

## Chapter C

---

### STAI-CH

---

Examples of STAI-CH questions

For copyright reasons only five example questions are allowed per questionnaire.

1. Jeg føler mig..... meget rolig       rolig       ikke rolig
2. Jeg føler mig..... meget ude af den       ude af den       ikke ude af den
3. Jeg føler mig..... meget godt tilpas       godt tilpas       ikke godt tilpas
4. Jeg føler mig..... meget nervøs       nervøs       ikke nervøs
5. Jeg føler mig..... meget anspændt       anspændt       ikke anspændt

**Figure C.1:** Five examples of STAI-CH State questions

1. Jeg er bekymret for at lave fejl..... nærmest aldrig       nogle gange       ofte
2. Jeg har lyst til at græde..... nærmest aldrig       nogle gange       ofte
3. Jeg føler mig ked af det..... nærmest aldrig       nogle gange       ofte
4. Jeg har svært ved at bestemme mig..... nærmest aldrig       nogle gange       ofte
5. Jeg har svært ved at se mine problemer i øjnene  nærmest aldrig       nogle gange       ofte

**Figure C.2:** Five examples of STAI-CH Trait questions