**Karolinska
Institutet**

Institutionen för klinisk neurovetenskap

Psykologprogrammet, termin 10

Huvudämne: Psykologi

Examensarbete (D-nivå) i psykologi (2PS026), 30 poäng

Höst-/vårterminen 2018/2019

Investigating rapid effects of motor training on the brain's gray matter volume

A magnetic resonance imaging study with repeated testing

Arian Jafari & Bo Jenner

Supervisor: Dr Kristoffer Månsson, Centrum för psykiatriforskning, Institutionen för klinisk neurovetenskap, Karolinska Institutet

Examiner: Professor John Axelsson, Institutionen för klinisk neurovetenskap, Karolinska Institutet

Institutionen för klinisk neurovetenskap

Psykologprogrammet, termin 10

Huvudämne: Psykologi

Examensarbete (D-nivå) i psykologi (2PS026), 30 poäng

Höst-/vårterminen 2018/2019

Investigating rapid effects of motor training on the brain's gray matter volume

A magnetic resonance imaging study with repeated testing

Sammanfattning/Abstract

Nyliga MRI-studier har visat att träning kan leda till en synbar ökning i grå substansvolym (GMV) i uppgiftsrelevanta hjärnregioner inom timmar eller till och med minuter. Denna studie undersökte snabb plasticitet i kontexten av motorisk träning, med temporala upplösningar på två minuter och ungefär en timme. Tjugofem friska deltagare övade på att utföra en komplex fingerfärdighetsuppgift (FTT) inuti en magnetresonanskamera. Tidigare forskning på FTT användes för att *a priori* definiera ett område i primära motorkortex (PMC_{ROI}). GMV i PMC_{ROI} analyserades vid tre tidpunkter: början, mitten och slutet av scanning. Vi fann en ökning i GMV under vila jämfört med träning i slutet av scanning, vilket indikerar en möjlig fördröjd träningseffekt. *Post hoc*-analyser av det sensomotoriska nätverket resulterade i jämförbara effekter som försvann vid korrigering. Inga signifikanta skillnader i GMV under träning i PMC_{ROI} hittades mellan de tre tidpunkterna. Såvitt vi vet har inga tidigare studier rapporterat förändringar i GMV på så här kort tid (120 sekunder). Framtida forskning bör fortsätta undersökandet med hög temporal upplösning för att utforska en potentiell fördröjning i den snabba effekten av träning på GMV.

Nyckelord: complex finger tapping, grå substansvolym, motorisk träning, snabb strukturell förändring, voxelbaserad morfometri

Recent MRI studies have shown that training can cause apparent gray matter volume (GMV) increases in task-relevant brain regions within hours or even minutes. The present study examined rapid plasticity in the context of motor training, at temporal resolutions of two minutes as well as roughly one hour. Twenty-five healthy volunteers practised performing a complex finger tapping task (FTT) while inside an MRI scanner. Previous research on FTT was used to define *a priori* region of interest PMC_{ROI} in the primary motor cortex. GMV in PMC_{ROI} was analyzed during rest and training at three time points: beginning, middle, and end of scan. We found an increase in GMV during rest compared to training at end of scan, indicating a potential delayed training effect. *Post hoc* analyses of the sensorimotor network resulted in comparable effects which did not survive correction. No significant differences in training GMV between the three time points were found in PMC_{ROI}. To our knowledge, no previous study has reported changes in GMV in such a short time period (120 seconds). Future research should continue investigations with high temporal resolution to explore a potential delay in rapid effects of training on GMV.

Keywords: complex finger tapping, gray matter volume, motor training, rapid structural change, voxel-based morphometry

Investigating rapid effects of motor training on the brain's gray matter volume

A magnetic resonance imaging study with repeated testing

Arian Jafari & Bo Jenner

Although it is often said that the adult brain was considered to be anatomically static a good bit into the 20th century, much work has been done to document the dynamic properties of brain cells since the beginning of the 1900s (Stahnisch, 2003). Inquiries into the relationship between neural structure and learning are older still, and animal experiments on the subject are documented as far back as 1783. In the 1960s more substantial support was gathered for the idea that the brain is altered anatomically as a result of experience, but by the turn of the millennium the evidence for these processes in humans was still sparse (Rosenzweig, 2002). This largely changed with a seminal study which showed that London taxi and bus drivers had a greater volume of gray matter in mid-posterior hippocampus compared to controls, an area associated with spatial knowledge (Maguire et al., 2000). Volume further correlated with the length of their driving careers. The authors suggested that extensive training in spatial navigation leads to increased gray matter volume.

This discovery had not been possible without magnetic resonance imaging (MRI). The first such images were produced in the 1970s, and since then MRI has developed into one of the most important tools available to document the structure and function of the brain (Huettel, Song, & McCarthy, 2009). The technology utilizes the behavior of magnetic particles, specifically hydrogen in most cases. Hydrogen atoms are the most abundant in the body, and because their proportions vary between different types of tissue, measuring hydrogen density allows for identification of the tissue. In the case of brain scans this makes MRI suitable for differentiating between, for example, gray matter (largely glia and neuronal cell bodies) and white (mainly myelinated axons). It has the added benefits of being non-invasive and not associated with any dangers, with the exception of cases where participants have metal in their bodies or are sensitive to loud noise (Huettel et al., 2009).

Principles of Magnetic Resonance Imaging

MRI depends largely on the interaction between the magnetism of different objects, such as protons. Not all atomic nuclei can be observed by the scanner, but hydrogen meets all the requirements and has the potential for nuclear magnetic resonance. By subjecting them to the static field of the scanner, the magnetic moments of the nuclei – their magnetism – are aligned. Alignment means that nuclei take one of two positions: parallel, which requires less energy and is thus the general rule, and anti-parallel. If more energy is supplied, more of them will flip into the anti-parallel state (Huettel et al., 2009).

This is what is called excitation, and it has two purposes. One is the process that follows, relaxation, in which the energy absorbed by nuclei during excitation is released, providing the scanner with a signal. The other is to change the direction of the magnetization to be perpendicular to the static field, which is necessary for the signal to be picked up by the scanner. The strength of the signal is heavily dependent on the magnetization remaining in this direction. (Huettel et al., 2009).

Contrast imaging. The rate at which this energy is released, the relaxation time, is called T1 and varies between types of tissue. This means that it is possible to identify e.g. gray and white matter from the amount of energy in the respective area. So called T1-weighted images are

created by repeating excitation at the time when the difference of energy release between the tissue types is the greatest, resulting in the biggest contrast in the finished image (Huettel et al., 2009).

In order to be able to localize the signal, it is necessary to complement the static field with *gradient fields*. These have the effect of altering the strength of the magnetism along three axes (x, y, and z). The first of these makes it possible to choose one slice of the brain to examine, and the two others alter the characteristics of the signal so that its origin is associated with a unique variation (Huettel et al., 2009).

Blood-oxygen-level dependent contrast. MRI scanners are also used for registering blood flow. Neural activity requires oxygen to be sustained, which is steadily supplied by the vascular system. Blood flow is therefore measured as a proxy of neural firing and constitute the basis of functional MRI (fMRI). While oxygenated blood is not magnetic, deoxygenated blood attracts weakly to an external magnetic field. The latter fact means that deoxygenated blood disturbs the processes which keep the magnetic field perpendicular to the static field, causing the signal to decrease. The amount of oxygen in the blood can therefore be calculated from the degree of signal loss, and blood-oxygen-level dependent (BOLD) contrast images be created (Huettel et al., 2009).

Neuroplasticity

Since the study on London taxi drivers, studies making use of MRI have continued to document the effects of experience on brain structure, as measured by the volume of gray or white matter in different regions. Several cross-sectional studies confirm that gray matter volume correlates positively with the number of years that an individual has been doing something for, such as using a language or practicing mathematics (Aydin et al., 2007; Carreiras et al., 2009; Mechelli et al., 2004). These are examples of neural plasticity, meaning alterations in the brain's anatomy or connectivity following new experiences. Functional plasticity may consist of changes in electrophysiological properties and neuronal responsiveness, whereas structural plasticity reflects changes in gray matter volume and the structure of white matter (Debowska et al., 2016).

Experimentally these changes can be induced by training – that is, repeated practice of an ability with the object of improving at it – and a large body of literature today demonstrates the effects when performed over months (e.g. Draganski et al., 2004; Hölzel et al., 2012; Månsson et al., 2016) and weeks (e.g. Landi, Baguear, & Della-Maggiore, 2011; Wenger, Kühn, et al., 2017).

How much time is required for structural plasticity to occur is of increasing interest, as changes are progressively detected more quickly. Changes in gray matter volume of the visual cortex have been observed after less than two hours of practicing new color names over three days (Kwok et al., 2011). White matter changes in the fornix have also been observed, with diffusion tensor imaging, after two hours of playing a car racing game (Hofstetter, Tavor, Tzur Moryosef, & Assaf, 2013). Practicing a balancing task for only 45 minutes lead to increased cortical thickness in the primary motor cortex (Taubert, Mehnert, Pleger, & Villringer, 2016), and stimulating the index finger for the same duration increased gray matter volume in the parietal operculum (Schmidt-Wilcke et al., 2018). Most surprisingly, an increase in visual cortex gray matter volume was recently induced in under five minutes, simply by displaying complex images to participants (Månsson, Cortes, Manzouri, Hau, & Fischer, 2018).

There is evidently a relationship between proficiency and gray matter volume, but its precise definition is not known. When it comes to motor ability, it seems like it is not enough

simply to use a faculty, as opposed to training it, to produce such changes. One study controlled for the possibility that the observed increase in cortical thickness from motor task training might in part be a result of mere muscular activity, but could find no such effect (Taubert et al., 2016). As for correlations between the two measures, one study reported both negative and positive correlations between two-point discrimination ability and gray matter volume at baseline. The latter of these remained after the intervention, whereas the former disappeared (Schmidt-Wilcke et al., 2018). Another study found no correlation between gray matter concentration and rate of learning, but the increase in gray matter concentration correlated with speed of relearning one year later (Landi et al., 2011).

Often the pattern has appeared to be that skill acquisition is related to a corresponding structural enlargement, but more recent findings indicate that expansion is not linear to the development of the skill. For example, volume increase associated with non-dominant hand practice was observed after four weeks, but was no longer discernible three weeks later, in spite of continued proficiency improvements (Wenger, Kühn, et al., 2017). An expansion–renormalization model has been suggested where initial gray matter expansion is followed by a renormalization of the volume (Wenger, Brozzoli, Lindenberger, & Lövdén, 2017). In this view, brain areas grow larger as a result of increased recruitment, possibly due to such processes as dendritic branching and glial swelling. Once more efficient pathways have been established, the size of the area is more or less restored. Only rarely is an increase of neurons a part of this process, as indications of adult neurogenesis has only been indicated in the dentate gyrus, the olfactory bulb, and the striatum (Wenger, Brozzoli, et al., 2017), and there is ongoing debate over whether it exists at all (e.g. Snyder, 2019).

Motor Skill Learning

The finger tapping task (FTT) has been used to quantitatively measure motor control ability dating as far back as the 19th century (Jobbágy, Harcos, Karoly, & Fazekas, 2005), and variations of it have been used to study clinical populations, such as individuals with Korsakoff's syndrome, Alzheimer's disease, and Parkinson's disease (Barut, Kiziltan, Gelir, & Köktürk, 2013), as well as healthy individuals such as professional musicians (Jäncke, Schlaug, & Steinmetz, 1997). In its most basic form, the exercise is simply to tap a finger against a surface repeatedly (Reitan & Wolfson, 1993), but several variations exist. A more complex version was used already in early fMRI studies of motor skill learning (Karni et al., 1995, 1998), and has also found widespread use in research on the effects of sleep on motor-learning consolidation. A meta-analysis of that field which looked at 34 studies described the task as most commonly involving a sequence of five or more steps, and consisting of tapping the index finger through little finger against either the thumb or a keyboard. Typically, the subject alternates between blocks of performance and rest, and is later tested after having slept, with number of correct sequences as primary measure (Pan & Rickard, 2015). This type of FTT has also been used in investigations of bilateral activation of premotor and primary motor regions (Horenstein, Lowe, Koenig, & Phillips, 2009) and mental training (Nyberg, Eriksson, Larsson, & Marklund, 2006), among other areas.

The complex-sequence, multi-finger version of the FTT has several advantages. It is easy to learn and to practice, and its difficulty can easily be adjusted to the participants' skill level. The risk of floor and ceiling effects can be controlled by adding or removing elements, making the sequence more or less challenging to master. It can also be performed while the participant is

inside the MRI scanner and recorded with the use of a response pad, allowing for simultaneous observation of the brain, and for comparisons between neurological and behavioral data.

While different variations in the design of FTT have been observed to correlate with different neural activation patterns, all variations generally produce increased activity in the primary sensorimotor cortex, the supplementary motor area, the basal ganglia and the cerebellum (Witt, Laird, & Meyerand, 2008). In addition to the primary sensorimotor cortex being central to the motor system, these aforementioned areas have all been observed to play important roles in different aspects of motor action: the supplementary motor area with generating movements in the absence of explicit sensory cues, the basal ganglia with gating initiation of movement, and the cerebellum with coordinating sensory and motor information (Purves et al., 2013).

Much of the evidence from studies of motor skill learning highlights the primary motor cortex (PMC) as an area of interest. The potential for structural change in the area is demonstrated by studies that find an effector-specific increase in PMC volume after balance training (Taubert et al., 2016) and bilateral PMC expansion from practicing handwriting (Wenger, Kühn, et al., 2017). It is not obvious that improvement in motor task performance primarily depends on motor areas. Learning to juggle over the course of three months only resulted in expansions in the mid-temporal area and posterior intraparietal sulcus, areas more related to visual-motion information (Draganski et al., 2004), and subsequent studies on juggling have likewise found no structural changes in the motor cortex (Dayan & Cohen, 2011). Direct evidence of where to look for structural effects of finger tapping is scarce. Increases in cortical thickness have been found in the pre- and postcentral gyri as well as the dorsolateral prefrontal cortex, after participants practiced a finger tapping task daily for four weeks (Sale et al., 2017). Looking for white-matter changes in the same dataset, increased fractional anisotropy was found in the right caudate nucleus, and decreased mean diffusivity in the left nucleus accumbens (Reid, Sale, Cunnington, Mattingley, & Rose, 2017). One study used an FTT in combination with visual stimulation to examine the influence of cerebral blood flow (CBF) on contrast imaging, but no training was involved and the details of the task design were omitted (Ge et al., 2017). Apparent structural changes were nonetheless found bilaterally in the basal ganglia and visual cortex.

There is all the more functional data related to finger tapping (Witt et al., 2008), but although neural activity indicates where to look for structural change, the two are not the same. A study using a paradigm that involved tracking a target on a screen with a joystick predicted increase in gray matter concentration based on BOLD data collected with the same paradigm. Structural changes were expected in the anterior cerebellum, contralateral PMC, and posterior parietal cortex, but were only found in the PMC (Landi et al., 2011). Based on the neural activity data available, however, the PMC is the most appropriate area to investigate with the complex-sequence, multi-finger FTT paradigm. The earliest evidence shows that practicing a finger tapping sequence for three weeks leads to a larger activation of the PMC contralaterally, compared to performing an untrained sequence (Karni et al., 1995), and more recent studies confirm the connection (Horenstein et al., 2009; Nyberg et al., 2006).

Given that FTTs are commonly used in investigations of motor behavior it is appropriate to continue examinations of their impact on brain anatomy, especially since the short-term effects of practicing it are unknown. This will also further the documentation of rapid structural effects of learning, a field which is in its infancy, and allow for additional analyses of the relationship between proficiency and brain volume.

Aims

The present study aimed to extend the current literature by examining rapid plasticity in the context of motor training, with the added novel approach of investigating plasticity in very short intervals. More specifically, this study aimed to help us answer four questions: (1) Can structural change measured by gray matter volume (GMV) be observed within an hour of complex motor sequence training? (2) Can change in GMV be observed already within minutes of training? (3) To what extent does change in GMV over the period of an hour, if observed, correlate with improved finger tapping performance? (4) If a difference in GMV between adjacent blocks of training and rest can be measured within minutes, how does the size of that difference develop over time?

Method

Participants

Demographics. Twenty-five healthy ($\bar{x}_{\text{Self-rated health-7}} = 6.3, s = .7$), right-handed adults (thirteen [52%] female) between 19 and 32 years of age ($\bar{x}_{\text{age}} = 25.8, s = 3.3$) participated in the study. Nineteen participants were current university students, five employed full time, and one specified their occupation as “other”.

Recruitment. Participants were recruited through physical bulletin boards, online advertisement, and word of mouth. An advertisement was designed, asking for healthy volunteers between ages 20 and 30 interested in participating in brain research. By recruiting young adults, we aimed to ensure a high enough ability to improve continuously at a motor task, as well as acceptable homogeneity in the same respect. The advertisement described the study as one where participants would lie in an MRI scanner for a period of time slightly longer than an hour while being tasked with trying to become as skilled as possible at performing a task with their hands. Volunteers were promised a gift card worth 300 SEK as well as a high resolution 3D image of their brain as compensation for their participation. Physical copies of the advertisement were posted on various bulletin boards at the Karolinska Institutet campus in Solna, Södertörn University’s campus in Huddinge, and Stockholm University’s campus in Stockholm. The advertisement was also shared on Facebook in the authors’ feeds as well as in two student union related groups, one for psychology students at Karolinska Institutet and one for IT students at Stockholm University.

The advertisement included a URL to a web page containing a consent form and a questionnaire, where participants were asked to answer questions relevant to the screening process. The online questionnaire contained questions about age, sex, height, weight, level of education, employment status, handedness (Edinburgh Handedness Inventory [EHI]; Oldfield, 1971), self-rated health (SRH-7; Eriksson, Undén, & Elofsson, 2001), vision impairments, hearing impairments, recent hand or arm injuries, experience playing musical instruments, experience playing video games, and fluency in Swedish. The questionnaire also asked questions relating to MRI safety, including questions on surgical history, kidney impairments, fear of confined spaces (claustrophobia), tattoos, presence of pacemakers or similar implants, pregnancy or breastfeeding status, metal splinter in the eyes, and metal implants.

Screening. Screening of participants was done in two phases. In the first phase, some participants were excluded based on their answers to the sign-up questionnaire. Participants were excluded if they fulfilled any of the following criteria: being below the minimum age of 19 or above the maximum age of 34, not being right-handed (defined as scoring below +71 on the EHI (consistent with previously developed criteria for categorical coding

(Deep-Soboslay et al., 2010; Dragovic, 2004), needing vision-correcting glasses and not having access to contact lenses, poor self-rated health (i.e. rating their overall health as three or lower on the seven point SRH-7), not being fluent in Swedish, having recently experienced hand or arm injury, or answering affirmatively to any of the MRI safety questions.

In the second phase of screening participants performed a short computerized test. The purpose of this screening test was to exclude participants who showed signs of being too proficient at the FTT at baseline. This was done to minimize the risk of participants reaching a peak in performance too quickly during the main trial. The screening test was identical to one training block in the main scanner test, but with a duration of 60 seconds and the correct numerical sequence being 3 2 4 1. Participants were blinded to the criteria for screening as a precaution against the risk of affecting test performance. Screening tests were conducted at Karolinska Institutet's campus in Solna and at Karolinska University Hospital in Huddinge. Participants were contacted through the channels they had chosen upon signing up (email and/or SMS). Participant screening test scores (i.e. the total number of sequences correctly performed during one minute) were standardized with z-transformation. Participants whose z-scores were above +0.8 were excluded for above mentioned performance reasons.

Materials and Apparatus

Pilot tests were conducted using a laptop computer and the libre software Notepad++ (2016), using keys 1–4 of the keyboard as the input device and a mobile phone as the stopwatch for timing the blocks. The software script used for displaying information to the participants and registering their input was developed in E-Prime 2.0 (Psychology Software Tools, 2015). To ensure that the script was handling user input correctly several rounds of testing were performed during its development. Screening and pre-scan tests were performed on a 15.6" laptop computer, once again using keys 1–4 for user input.

MRI data was acquired at Karolinska University Hospital in Huddinge, using a Siemens whole-body clinical MRI 3 Tesla scanner (Magnetom Prisma-fit, Siemens Medical Solutions, Erlangen, Germany) with a gradient system of 8 Gauss/cm and with a compact 64-channel phased-array receiving head coil. Structural images for GMV analyses were produced with a 119 seconds long T1-weighted imaging protocol with TE = 3.46, TR = 1730, and TI = 900ms. Isotropic voxel size was set to 1mm, with a matrix size of 256x256x192 and a flip angle of 7°. The input device used for the scanner test consisted of two MR compatible input devices with two buttons, taped together to form one controller with one horizontal row of four buttons.

Procedure

Finger tapping task. For the present study, we chose to use a complex sequence multi-finger tapping task, similar in design to the one used by Walker, Brakefield, Morgan, Hobson, & Stickgold (2002). This variant of the finger tapping task requires subjects to repeatedly input a specific sequence of numbers, each number corresponding to a button on a mechanical input device, using four fingers on their non-dominant hand. The trial consists of several identical blocks. The sequence of numbers to be input remains the same for the duration of the full trial.

The final version of the trial design used in the present study used blocks of 120 seconds of either training (i.e. finger tapping) or resting. The subject's task during training blocks was to accurately input the six-number sequence 1 4 2 3 1 3 as many times as possible. The primary outcome measure was the total number of sequences input correctly during each training block. The outcome and specific numerical sequence were chosen in order to be consistent with the

literature (see Nettersheim, Hallschmid, Born, & Diekelmann, 2015; Pan & Rickard, 2015; Sale et al., 2017) while also satisfying our criteria described above. The correct sequence was displayed to the participants at all times during training in order to avoid encumbering working memory. After each training block, participants received visual feedback describing the total number of sequences correctly input, as well as the total amount of performed key presses, during the previous block.

Pre-scan test. Upon arrival at the location of the camera test, participants were welcomed and instructed according to a standardized protocol. First, participants were directed to a test station where they performed a pre-scan test. This test was identical to the screening test, both with regards to instruction and task design, apart from the correct numerical sequence being 1 3 4 2. The purpose of this pre-scan test was to ensure that participants were fully comfortable with the instructions of the task they were to perform in the scanner, in order to minimize the risk of potential confusion and subsequent loss of quality of performance data. Participants were then given verbal instructions for the scanner test.

Scanner test. Before being placed in the scanner, participants were told that the task in the scanner would be nearly identical to the task in the two tests they had already performed but with a different number sequence, that they would be constantly either performing the task or resting in two minute blocks, and that their goal for the duration of the test was to become as skilled as possible at performing the task. Participants were told to keep their eyes open during training as well as during rest.

While in the scanner, participants viewed a digital monitor presenting them with instructions. Before each block, participants were visually instructed to either prepare for training or to move their hand from the input device and rest. During rest, the screen was black. Participants were given verbal encouragement halfway through the trial, validating them on the strenuous nature of the task and encouraging them to keep trying to improve as much as possible for the remainder of the trial.

Block design. Each test block consisted of the subject either resting or training (figure 1) while T1-weighted images were acquired. Block order was identical for all participants. The ordering of subject activity during two consecutive blocks at the same time point (i.e. training–rest or rest–training) was reversed at some time points in order to control for the possibility of an order effect. Participants continued training and resting intermittently between time points. On average, participants spent 73 minutes inside the scanner ($s = 202$ seconds).

Figure 1
Design of the scanner test

Note. Each block represents 120 seconds of either training (T) or rest (R). Pairs of blocks were used in analyses of GMV at three time points (A, B, and C). Resting data at time point B (B_R) was not used in any analyses and is shown for illustrative purposes only. Subjects continued training and resting intermittently between the different time points.

Data Reduction and Statistical Analyses

Finger tapping performance. Participant finger tapping performance was analyzed at three time points (figure 1); beginning of scan (A_T), middle of scan (B_T), and end of scan (C_T). Performance was measured by total number of sequences correctly input during the specified training block. To investigate whether there was a difference in performance between the three time points, we performed a one-way repeated measures ANOVA. The Shapiro-Wilk test of normality showed that FTT data could be assumed to be approximately normally distributed at all three time points ($W_{AT} = .922, p = .057$; $W_{BT} = .988, p = .987$; $W_{CT} = .975, p = .770$). Mauchly's Test of Sphericity returned significant results (Mauchly's $W = .555, p = .001$), indicating that sphericity (i.e. equal variances of differences in performance between all time points) could not be assumed. To counter the risk of inflated F-ratios, we applied the conservative Greenhouse-Geisser correction of Mauchly's W ($\epsilon = .692$) to the results. Pairwise comparisons to check differences in performance between all three possible combinations of time points were performed using paired samples t -tests. Significance level of statistical analyses was set to $p < .05$.

Magnetic resonance images.

Voxel-based morphometry. Pre-processing of MRI T1-weighted images was done using voxel-based morphometry (VBM) in order to allow for voxel-wise comparison of the distribution of gray matter in specific brain regions. VBM consists of several steps, including mapping of the spatial data onto a reference brain atlas, classification of tissue into gray matter, white matter, cerebrospinal fluid and bone, as well as smoothing of gray-matter segments and bias correction (Ashburner & Friston, 2005). VBM was performed using Computational Analysis Toolbox (Gaser & Dahnke, 2019) for the Statistical Parametric Mapping software (The Functional Imaging Laboratory Methods Group, 2014) using default settings.

Region of interest selection. We defined an *a priori* region of interest (ROI) located within the PMC for analyses of GMV (PMC_{ROI}), consisting of a sphere with a 10 mm radius. Coordinates 40, -16, 54 (MNI x, y, z) have previously been observed as peak coordinates for activation patterns in motor cortex of the right hemisphere during left handed complex sequence finger tapping in a sample of young, right-handed, healthy subjects (Nyberg et al., 2006), and were thus used to define the epicenter of the ROI in the current study.

Post hoc analyses were conducted on a set of six different regions suggested to form the sensorimotor network (figure 2) (Shirer, Ryali, Rykhlevskaia, Menon, & Greicius, 2012). The six regions are (1) left pre- and postcentral gyrus (lPPG), (2) right pre- and postcentral gyrus (rPPG), (3) right supplementary motor area (rSMA), (4) left thalamus (lThalamus), (5) bilateral cerebellar lobules IV, V, and VI (4-6Cb), and (6) right thalamus (rThalamus). Mean GMV for each region was calculated by applying NIfTI files representing each region (Shirer et al., 2012, Supplementary materials) as masks to our structural images, and calculating the mean of mean GMV values of each individual point in the region.

Magnetic resonance imaging analyses. Training GMV in PMC_{ROI} was analyzed at three time points – beginning of scan (A_T), middle of scan (B_T), and end of scan (C_T) (see table 1 for descriptives). Potential differences in GMV were analyzed using one-way repeated measures ANOVA, using paired samples t -tests for pairwise comparisons when required. Difference in training GMV and rest GMV within the same time point was analyzed twice, once at beginning of scan and once at end of scan (i.e. A_T-A_R and C_T-C_R), using paired samples t -tests. Deltas from these two comparisons were then compared to test whether difference in training GMV and

rest GMV differed between the two time points, again using paired samples t -test. Significance level of statistical analyses was set to $p < .05$.

The Shapiro-Wilk normality test showed that normality could not be assumed ($W_{\text{GMV-AT}} = .899$, $p = .013$; $W_{\text{GMV-BT}} = .895$, $p = .01$; $W_{\text{GMV-CT}} = .881$, $p = .005$), which led us to opt for the Friedman test as a non-parametric option to repeated measures ANOVA.

Figure 2

Masks of the sensorimotor network

Note. Areas of the sensorimotor network as defined by Shirer et al. (2012), analyzed in *post hoc* tests of GMV. ROIs are color coded as follows: left pre- and postcentral gyrus, green = right pre- and postcentral gyrus, yellow = right supplementary motor area, orange = left thalamus, blue = bilateral cerebellar lobules IV, V, and VI, purple = right thalamus.

Post hoc tests. After our analyses of GMV difference between baseline and end of scanning in PMC_{ROI} yielded no significant results, we decided to investigate further potential regions of interest to see whether our findings would generalize to GMV in other brain areas.

We conducted the same tests of GMV on the six regions in the sensorimotor network. In order to perform a repeated measures ANOVA comparing mean training GMV at the three time points A_T , B_T , and C_T , we first verified that the necessary assumptions were fulfilled. Shapiro-

Wilk tests of normality showed that normality could be assumed for GMV at all three time points across all six regions. Mauchly's Test of Sphericity indicated that sphericity could be assumed for all six regions ($W_{\text{IPPG}} = .93, p = .41$; $W_{\text{rPPG}} = .83, p = .11$; $W_{\text{rSMA}} = .96, p = .62$; $W_{\text{IThalamus}} = .87, p = .19$; $W_{4-6Cb} = .94, p = .46$; $W_{\text{rThalamus}} = .95, p = .59$). To account for the risk of committing type I errors due to multiple comparisons, while not sacrificing too much statistical power, false discovery rate (FDR) corrected critical values were calculated using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995).

Correlations between performance and magnetic resonance imaging data. To prepare data for testing whether changes in performance correlated with changes in GMV, new delta measures were constructed for each participant. Three new data points describing change in training GMV between time points were created for each participant – (1) difference between beginning of scan and end of scan (i.e. $A_T - C_T$), (2) difference between beginning of scan and middle of scan (i.e. $A_T - B_T$), and (3) difference between middle of scan and end of scan (i.e. $B_T - C_T$). The same procedure was then used to create another three data points, this time measuring each participant's change in performance. In the case of significant differences being observed in both performance and GMV, three different Pearson correlation coefficients (r) would be calculated, one for each pair of time points. Calculation of confidence intervals and hypothesis testing ($p < .05$) would then be carried out using the Fisher transformation.

Results

Finger Tapping Performance

Repeated measures ANOVA showed a significant improvement in number of completed finger tapping sequences between time points ($F(1.38, 33.21) = 137.42, p < .001, \eta_p^2 = .85$). Pairwise comparisons of FTT performance at the three time points showed significant differences between all three possible pairs ($\bar{x}_{\text{FTT-A}} = 28.08, s = 9.32$; $\bar{x}_{\text{FTT-B}} = 49.56, s = 12.12$; $\bar{x}_{\text{FTT-C}} = 60.84, s = 15.91$; $\bar{x}_{\text{FTT-AC}} = -32.76, p < .001$; $\bar{x}_{\text{FTT-AB}} = -21.48, p < .001$; $\bar{x}_{\text{FTT-BC}} = -11.28, p < .001$), with participants on average improving their performance consistently over the duration of the test. Reported p values of pairwise comparisons are adjusted for multiple comparisons using the Bonferroni correction.

Magnetic Resonance Images

A Friedman test was conducted to compare training between beginning (A_T), middle (B_T), and end (C_T) of scan. This test did not reveal a significant difference between training GMV at any of the three time points ($\chi^2(2) = 2.96, p = .23$).

Differences in training GMV and resting GMV were compared within time points A and C respectively (see table 1 for descriptive statistics), using a paired samples t -test. We found a significant difference between training and resting at C ($t(24) = 2.60, p = .02$, Cohen's $d = .52$) indicating that GMV at C was higher during rest than during training. We found no significant difference at time point A ($t(24) = -1.17, p = .25$, Cohen's $d = .23$). A paired samples t -test comparing the difference between training GMV and rest GMV at A ($A_T - A_R$) to the difference between training GMV and rest GMV at C ($C_T - C_R$) found no significant effect of time on GMV difference ($t(24) = .64, p = .53$).

Table 1
Descriptive statistics of gray matter volume in PMC_{ROI}

	Mean GMV	<i>s</i>
A _T	.3777	.0479
A _R	.3794	.0466
B _T	.3784	.0474
C _R	.3789	.0462
C _T	.3760	.0482

Note. Means and standard deviations (*s*) for gray matter volume in PMC_{ROI} for all analyzed combinations of time point and activity – training at beginning of scan (A_T), resting at beginning of scan (A_R), training at middle of scan (B_T), resting at end of scan (C_R), and training at end of scan (C_T).

Post hoc repeated measures ANOVA tests of the six regions in the sensorimotor network showed no significant differences between training GMV at any of the three time points (table 2).

Table 2
Results of post hoc repeated measures ANOVAs

	<i>F</i> (2, 48)	<i>p</i>	η_p^2
Left PPG	1.02	.37	.04
Right PPG	.11	.89	.01
Right SMA	3.08	.06	.11
Left thalamus	.16	.86	.01
4-6Cb	3.05	.06	.11
Right thalamus	2.12	.13	.08

Note. Tests compared mean training GMV in ROIs in the sensorimotor network (Left/right pre- and postcentral gyri [PPG], right supplementary motor area [SMA], bilateral cerebellar lobules IV, V, and VI [4-6Cb]) at beginning of scan (A_T), middle of scan (B_T), and end of scan (C_T). Reported for each region are *F* values, *p* values, and effect sizes as measured by partial eta squared (η_p^2).

We then compared GMV during training and rest at two time points (A and C) for each of the six ROIs in the sensorimotor network using paired samples *t*-tests (table 3). Initially, we found significant differences for four of the tests – GMV was higher during rest than during training in lPPG and rPPG at A, and in lPPG and rThalamus at C. However, no initially significant *post hoc* results survived FDR correction.

Table 3

Results of post hoc pairwise comparisons of gray matter volume in the sensorimotor network

Comparison	GMV difference	<i>s</i>	<i>t</i> (24)	<i>p</i>	FDR _{critical}	<i>d</i>
IPPG _{AT-AR}	-.0030	.006	-2.64	.014*	.008	.53
IPPG _{CR-CT}	.0039	.009	2.24	.035*	.017	.45
rPPG _{AT-AR}	-.0025	.006	-2.26	.034*	.013	.45
rPPG _{CR-CT}	.0025	.009	1.40	.176	.029	.28
rSMA _{AT-AR}	-.0032	.008	-2.03	.054	.021	.41
rSMA _{CR-CT}	.0009	.013	.33	.745	.050	.07
lThalamus _{AT-AR}	-.0022	.032	-.35	.731	.046	.07
lThalamus _{CR-CT}	.0041	.032	.63	.536	.033	.13
4-6Cb _{AT-AR}	.0006	.008	.39	.703	.038	.08
4-6Cb _{CR-CT}	-.0020	.007	-1.42	.170	.025	.28
rThalamus _{AT-AR}	-.0013	.018	-.37	.716	.042	.07
rThalamus _{CR-CT}	.0116	.020	2.93	.007*	.004	.59

Note. Reported are differences in mean GMV between training and rest for six regions in the sensorimotor network, reported for time points A and C, obtained from performing *post hoc* paired samples *t*-tests. Also reported are standard deviations (*s*), *t* values, *p* values, critical values after using false discovery rate (FDR) correction, and effect sizes as measured by Cohen's *d*. Note that training preceded rest at time point A, and vice versa at C. Analyzed regions are left pre- and postcentral gyrus (IPPG), right pre- and postcentral gyrus (rPPG), right supplementary motor area (rSMA), left thalamus (lThalamus), bilateral cerebellar lobules IV, V, and VI (4-6Cb), and right thalamus (rThalamus). Asterisks (*) highlight *p* values in comparisons that showed initially significant effects at an alpha level of .05. None of these effects survived FDR correction.

Correlations Between Performance and Magnetic Resonance Imaging Data

Since comparisons of GMV between the three time points did not result in any observations of a significant effect of time, no correlational analyses were conducted.

Discussion

The aim of this study was to contribute to the literature on rapid structural changes, specifically in the area of motor learning, by capturing structural brain images while subjects practiced performing a complex finger tapping task over a period of time averaging 73 minutes. Specifically, our first aim was to investigate whether structural change measured by GMV could be observed over the full period of approximately one hour of motor training. We report no difference in GMV in the *a priori* defined region PMC_{ROI} between beginning of scan, middle of scan, and end of scan. Exploratory analyses of the sensorimotor network revealed initially

significant changes, that however did not survive FDR correction, in the right SMA and the cerebellum. Secondly, we aimed to explore whether we, through training, could induce change in GMV within a matter of minutes. In support of this, we report larger GMV during rest as compared to training at end of scanning for the *a priori* region in the PMC. *Post hoc* analyses of the entire sensorimotor network found some initially significant effects also indicating larger GMV during rest in bilateral pre- and postcentral gyri and right thalamus than during training at end of scan. However, these effects did not survive FDR correction. Thirdly, we wanted to investigate the relationship between improved finger tapping performance and GMV increase over an hour. While we did observe an overall increase in performance, the lack of observed difference in training GMV over the course of the entire trial led us to not conduct correlational analyses. Finally, we aimed to explore the development of differences in GMV between adjacent blocks (i.e. training vs rest) over time. Our results showed no significant difference in GMV delta between beginning of scan and end of scan.

Primary Findings

We observed an interesting pattern of GMV changes between adjacent blocks of training and rest at the end of the scanning period, significant in PMC_{ROI} and initially significant (i.e. pre-correction) in several other regions in the sensorimotor network, with medium effect sizes (Cohen, 1988). Specifically, GMV in these regions was at the end of the trial higher during rest than during the succeeding training block. These findings stand in contrast with the findings of a previous study that found rapid bidirectional morphological changes in less than 263 seconds when participants viewed pictures, relative to resting (Månsson et al., 2018). Another similarly contrasting observation is that of apparent structural change in both the basal ganglia and the visual cortex during FTT over roughly 270 seconds (Ge et al., 2017). One possible explanation for the difference between these studies and the current one is that the transient effects on volume of engaging a brain region do not appear instantly, but with a delay of up to a few minutes. This delay would likely be short enough to not affect the results of the previously mentioned studies (i.e. no longer than approximately four minutes), but long enough to shift our observations of training-induced GMV increase into the succeeding resting blocks. Consequently, such a delay could mean that our estimations of GMV during training blocks have resulted in observations of smaller effects of GMV change over time than if corresponding blocks of rest (i.e. A_R, B_R and C_R) had been compared, since GMV during resting blocks would likely retain the effects of preceding training blocks. It is therefore possible that the difference in results between that study and the present one, i.e. the lack of GMV increase during training blocks, can be partly attributed to the shorter image acquisition times (i.e. 120 seconds) in the current study. Another possibility is that sensorimotor systems and visual systems differ regarding the nature of short term structural change. Perhaps visual stimuli are a more time efficient catalysts for rapid apparent structural plasticity in the visual network than is motor training for corresponding sensorimotor plasticity. Both above mentioned hypotheses could also explain the lack of observed difference between training GMV and rest GMV at baseline – if rapid plasticity is generally delayed or if motor training-induced plasticity specifically is slower than that induced by visual stimuli, our baseline measurement of training GMV would be unaffected by training simply due to its nature as the first training block. Future experimental research on rapid plasticity should investigate this hypothetical delay in order to obtain a more accurate understanding of training-induced changes in brain structure.

The absence of any lasting or cumulative changes in the PMC, given the presence of a learning effect, is surprising considering previous findings on motor-task-related structural change. This is especially so because these have been made in similar time frames (Taubert et al., 2016) and with a similar task (Wenger, Kühn, et al., 2017), although the latter over weeks. To some degree it also conflicts with the fact that another study found structural effects using a finger tapping task, but this change was, importantly, measured after four weeks of training (Sale et al., 2017).

Methodological Considerations

Development and pilot testing of the finger tapping task. When designing the finger tapping task for this study, several important factors regarding the task's difficulty had to be considered. The task needed to provide participants an opportunity to continually improve their finger tapping performance during the entire duration of scanning. In other words, the task had to not be difficult to the extent that participants would struggle to begin measurably improving (i.e. producing a floor effect), nor so simple that participants would reach a performance point where further improvement would be halted (i.e. producing a ceiling effect). This difficulty level was adjusted by changing the length of the number sequence. We also had to carefully choose a duration for each trial block. Shorter trials would lead to less fatigue but also, due to a shorter total practice time, likely yield less overall performance improvement, as well as allowing for less total MR image acquisition.

In order to find a sequence length and block duration that fulfilled these requirements, we conducted a series of pilot tests outside of the scanner, using a laptop computer and a text editor to imitate the task's conditions. Five rounds of pilot tests were conducted in total. Performance scores were calculated for each trial, as measured by the product of the number of correct sequences and the ratio of correct to incorrect key presses. The first four trials were performed on the authors themselves. We ran two trials with a sequence length of four numbers (using the sequences 1 3 4 2 and 1 3 2 4, one for each trial) with block durations of one minute and two and a half minutes respectively, and two trials with a sequence length of six (using the sequences 1 3 4 2 4 3 and 1 3 2 4 2 1) with the same variation in block durations. Tests were conducted using the left hand (both of the authors are right handed), with the number 1 corresponding to the little finger and 4 corresponding to the index finger. The results of these four trials led us to conclude that the combination of a longer sequence and a longer block duration were most likely to fulfill our above mentioned criteria regarding difficulty and room for performance improvement. To account for the possibility of the learning curve varying between subjects depending on their baseline finger dexterity, we then conducted one additional pilot test on one external subject not participating in the main study. The purpose of this fifth pilot test was to test whether we could see continual improvement throughout the entire test duration in a subject who had less experience utilizing computer keyboards than the authors. Results from this fifth pilot test confirmed that these test conditions would likely be suitable for a broad range of subjects.

It is possible that changes to the block duration, sequence length, or correct sequence would have led to us observing different results. Therefore, care should be put into the design of the specifics of the finger tapping task when conducting future research on motor-training-induced rapid GMV increase.

Region of interest selection. In the current study, we only analyzed one *a priori* brain region. By limiting our analyses to one region we run the risk of missing areas in which training leads to a significant increase in measured GMV. Because the ROIs used in the *post hoc* analyses

are rather large and are only represented by a single average of all volume data belonging to it, there is the risk that effects that would appear large with a more precise delineation are heavily diminished. Accurately predicting suitable targets for analysis using the current literature is a challenging task. According to an automated meta-analysis, 75 studies exist that all report coordinates relevant for finger tapping tasks (Neurosynth, 2019; see Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011 for methodology). However, there is a high variance among these studies with regards to both the population studied and the type of finger tapping task used, making the results less applicable for a study like the present one. Another meta-analysis reported several coordinates for activation foci for variants of the finger tapping task in a more granular way derived from a total of 38 published articles, 15 of which specifically studied multi-finger sequence variants of the task (Witt et al., 2008). These studies all utilized right-handed finger tapping as opposed to the present study which had subjects using their non-dominant left hand, but the activation patterns reported might still hold potential to be utilized as a starting point for directing future FTT-related GMV analyses. One approach is to use regression models to find regions where performance improvement predicts GMV increase (Schmidt-Wilcke et al., 2018). This data-driven approach has the benefit of avoiding the problems associated with manual ROI selection, while the lack of *a priori* defined ROIs leads to losing the opportunity to formulate ROI-specific hypotheses. Yet another option is to do what our original intention was with the present study (see Appendix), and define ROIs by analyzing BOLD data collected within the same scanning session. This approach has the benefit of avoiding the challenges of defining ROI(s) from the literature, thus reducing the risk of selecting the wrong region for tests of GMV. One potential challenge of the BOLD signal approach stems from the fact that MRI scanners cannot produce both structural and functional images simultaneously. Functional activation patterns might themselves change with training, making the choice of timing of the BOLD sequence an important one. However, due to its potential accuracy, this approach holds promise for future studies of training effects on rapid GMV increase.

Statistical power. The present study is limited in its statistical power. Our sample size was relatively small, likely contributing to several significant effects not surviving correction despite moderate effect sizes. Due to a combination of time constraints and limited knowledge regarding expected effect sizes we did not perform a power calculation while planning data collection. We encourage researchers to calculate power *a priori* as well as use larger samples for future studies, ensuring that they will be able to more confidently answer questions regarding the nature of rapid GMV increase in the sensorimotor network.

Brain Regions and Finger Tapping

While our observation of a positive effect of practice on finger tapping performance supports the use of the FTT paradigm for studying motor skill learning, one possible explanation of the lack of lasting GMV change in the PMC is that training did not challenge participants' motor ability enough, and that performance therefore never or rarely exceeded the constraints of the existing neural structures. Alternatively, it may be that other regions were strained more heavily than those common to most motor behavior, comparable to results in studies on juggling which have mainly found alterations in areas relevant to visual motion (Dayan & Cohen, 2011). For example, the contralateral ventral premotor cortex and ipsilateral posterior cerebellum are differentially activated in finger tapping tasks that are self-paced, compared to externally paced by visual or auditory stimuli. The premotor cortex is central in preparing motor behavior based on sensory information, not least when the behavior is sequential, and with adjusting its

frequency and complexity (Witt et al., 2008). Given the simplicity of the task, and that its efficient execution may be largely dependent on how the movements are planned (e.g. according to a rhythm), it is conceivable that the demands on this area of the brain are higher than on those that are responsible for the mechanical operation.

Additional areas have been identified that are not part of the sensorimotor network, but that are implicated in relevant versions of the finger tapping task, such as the basal ganglia and parts of the parietal cortex (Witt et al., 2008). The latter is involved in the production and execution of complex and sequential motor behavior. More specifically, the inferior parietal lobe is active in the encoding of sequence-specific information and the integration of sensorimotor processes. The contributions of parietal areas are strongly attested to in visually and auditorily driven movements, but the inferior parts are also observed with internally driven motor behavior, possibly tasked with directing attention to somatosensory feedback (Witt et al., 2008). These findings suggest that analyses of GMV may be warranted in areas outside of the sensorimotor network as defined in the current study.

The above hypothesis of unexamined areas does not account for the observed decrease in GMV between rest and training at the last time point. Overall there was no difference in volume between baseline and end, but in light of this finding it is worth noting the trajectory of GMV during training, which is an increase at time point B and a subsequent decrease at point C (both of them non-significant). This development also appears necessary for a significant change to appear in the last measurement but not register as significant when compared to the baseline value. The expansion–renormalization model predicts the trajectory suggested here. The observations it is based on generally indicate that the pattern occurs over days or weeks, but its principles could conceivably apply on a smaller temporal scale: over-recruitment of cells, followed by selection of the most efficient ones and stabilization (Wenger, Brozzoli, et al., 2017). The fact that the BOLD signal typically evolves in the same way over a session of training a motor task (Dayan & Cohen, 2011) lends some support to the hypothesis that these changes can occur in a shorter space of time. A previous finding runs counter to this, where balance training resulted in a linear increase on structural measures (Taubert et al., 2016). This task might have presented a different kind of challenge to the brain than the FTT and differ in its effects because of that, but the observations of the current study should be noted as conflicting with a model of linear structural development, whether they are considered as absence of change or suggestive of fluctuations.

Voxel-Based Morphometry and Mechanisms of Rapid Plasticity

To our knowledge, this is the first time that differences between two VBM-based measurements of GMV are reported with as little as two minutes separating measurements. There is an ongoing discussion about what factors underlie results such as these. Contrast images are limited in what they can say about the depicted tissue, and it is not certain what kind of microstructural changes are indicated when differences are found, or whether these reflect structural changes at all. As mentioned earlier, neurogenesis is an unlikely contributor, at least outside of the hippocampus (Zatorre, Fields, & Johansen-Berg, 2012). Astrocytes and microglia are able to divide and expand their surfaces as a consequence of new experiences, and these functions – gliogenesis – could be responsible for some part of a registered gray matter increase (Wenger, Brozzoli, et al., 2017). Vasculature, synaptogenesis and dendritic branching are also implicated (Zatorre et al., 2012). Based on our observations of apparent rapid structural

development, it is possible that VBM is a noisier method for measurement of gray matter than was previously thought.

Although structural change is in all likelihood the result of several processes (Tardif et al., 2016; Wenger, Brozzoli, et al., 2017; Zatorre et al., 2012), rapid developments complicate matters further. The chemical reactions behind dendritic branching have been observed within minutes of an intervention (Bosch et al., 2014), but the formation of synapses takes closer to an hour (Schmidt-Wilcke et al., 2018). With a timeframe of roughly two minutes, the most likely culprit may be regional blood flow as its relaxation time is close to that of gray matter (Franklin et al., 2013). Reductions were found in VBM gray matter measures 110 minutes after administration of a single dose of baclofen, which partially overlapped with reductions in CBF (Franklin et al., 2013). The study concluded that actual decreases of GMV were unlikely and that CBF interferes with volumetric evaluations of T1-weighted images. Comparable effects were found after consuming 200mg of caffeine and, separately, visual-sensorimotor stimulation (Ge et al., 2017). Caffeine intake resulted in apparent reductions in gray matter density in several brain regions, roughly 30 minutes later. These all disappeared when the observed differences in CBF were regressed out. Performing a self-paced bimanual FTT while watching a shifting checkerboard pattern lead to apparent increases in gray matter density in the visual cortex and basal ganglia. As with the caffeine, the former change disappeared entirely when CBF was regressed out, while the effect in the basal ganglia was reduced but remained significant (Ge et al., 2017). The authors suggest that because there is no reason to expect structural changes following interventions such as these, studies investigating short-term changes should regularly control for CBF. The study on balance training previously discussed raised the same concern, and consequently used pulsed arterial spin labeling to control for it (Taubert et al., 2016).

Conclusions

Much remains to be learned about rapid plasticity, and this study suggests that its continued documentation may be more than a question of relating skill practice with evidence of cortical expansion. The changes in morphological measures found in this study are some of the quickest so far, but because further fluctuations are hinted at it may be necessary to develop more granular models to account for what they indicate and how they relate to changes in proficiency. Future studies should continue to document these changes with the temporal resolution used in the current study, and consider what factors may create non-linear relationships between performance and structural measures, or why they might not appear at all where they are expected. In the same vein, we also believe that the possible delay between motor training and observable GMV changes should be further elucidated.

Ethical Considerations

It is important to consider participant safety when performing experimental research. This study was approved by the regional ethics board in Stockholm. MRI is considered non-invasive, and exposure to magnetic fields of the magnitude used in the present study is considered safe, assuming that safety protocols are followed and persons belonging to at-risk groups (e.g. individuals with pacemakers) are excluded. Participants were asked MRI safety questions at two time points – both upon signing up for the study and upon arrival at the scanner location – to minimize the risk of including in the study an at-risk participant based on a mistake in their answering process.

Mitigation of potentially life-threatening risks notwithstanding, some aspects of the experience of lying in an MRI scanner for an extended duration of time can still be disturbing. While potential participants with self-reported claustrophobia were excluded, it is possible that some included participants found the experience of the scanner's relatively tightly confined space to be anxiety provoking. MRI scanners also produce high-volume noise (Sheppard, Chen, & Salvi, 2018) which is a common cause of participant discomfort (McJury & Shellock, 2000), and has even been reported to cause hearing loss (Mollasadeghi et al., 2013). All participants were warned upon arrival that the machine would produce noise that some find disturbing. Additionally, participants were scanned with at least two separate layers of hearing protection to protect against the risk of hearing loss. Finally, participants were provided with a device with which they could stop the scan at any time regardless of reason. These things considered, one still has to keep in mind the possibility of participants high in discomfort being reluctant to go against the test leader's authority and risk what they might perceive as ruining the trial to stop the test.

While studies of clinical populations are generally more sensitive with regards to risks to personal integrity, the safe handling of participant data is an important matter to consider also in non-clinical studies. Our procedure to minimize the risks associated with handling of participant data was as follows. Upon signing up, each participant was assigned a study ID consisting of a randomized string of three numbers and three letters (e.g. 123ABC). Participant data that was gathered was subsequently tied only to their study ID and not their real identity, with one exception. In order to utilize the MR scanner, participants had to be registered in the Karolinska University Hospital's MR patient system by name and Swedish personal identity number (*personnummer*). Thus, potential data security risks associated with the hospital's internal handling of patient data were beyond our control. The connection between study IDs and participant identities (i.e. email addresses and/or phone numbers) was stored on a password-protected device at a secure location. All non-personal communication with participants was carried out through email, phone, and/or SMS, and did not include any sensitive data. On a handful of occasions, consecutively booked participants briefly ran into each other at the scanning location. This created the risk of deanonymizing participants who did not wish for their participation in the study to be known to anyone outside of the research team. We reason that the risk of harm in such a situation is low, due to the researched population being non-clinical and healthy.

To our knowledge, there is at this time little potential harm that can be inflicted upon an individual simply by having access to their neuroimaging data. Still, research on using imaging data to gather knowledge on a person (e.g. accurately predicting an individual's age [Cole et al., 2017]) is advancing. It is not unthinkable that this kind of research will in the future lead to the development of applications where individuals' neuroimaging data can be used against them in potentially harmful ways, such as for purposes of commercial advertising or political propaganda. When participants receive an image of their brain over email as part of compensation for their participation in the study, the image will end up in the digital hands of their email provider, making personal integrity difficult to enforce. The risks associated with this can be partially mitigated by encouraging participants to download the images to store safely and then delete the email itself.

This kind of research is costly. The purchase, operation, and maintenance of an MRI scanner and associated equipment all have sizeable effects on both university and hospital budgets, and it is therefore important that they are used for purposes of corresponding relevance.

In other words, MRI research directly competes with other, potentially more societally beneficial MRI applications. Scanning for this project was largely done during weekends and evenings, when the camera often is not otherwise occupied, somewhat mitigating this drawback. Additionally, some data collection took place during time slots reserved specifically for MRI research, further reducing the risk of negative impact on clinical use. However, even within the area of research exist questions of differing societal relevance. Weighing the potential benefits of basic research against those of clinical research is a formidable task. It is up to the individual(s) to whom the time slots are allotted to decide how they are best used, knowing that if one question is being examined, several others are not.

Acknowledgements

The scope of this study has extended far beyond the abilities of two master's students. Its design and execution are heavily indebted to our supervisor, Kristoffer Månsson, from whom we have learnt enough to ensure gray matter increases globally, and who has stood us by with support and feedback throughout the project. It also could never have been done without the programming and data processing of Amirhossein Manzouri, or the physics expertise and generosity to offer us his camera hours of Tie-Qiang Li, both of whom have contributed invaluable to the discussions shaping this study. Finally, we would like to thank our good friend Shervin Bukhari who was the one to suggest that we use a complex finger tapping task as an MR-compatible way to measure effects of motor learning in real time.

Reference List

- Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, 26, 839–851.
<https://doi.org/10.1016/j.neuroimage.2005.02.018>
- Aydin, K., Okur, O. O., Unal, Z., Ozturk, C., Yilmaz, S., Ucar, A., ... Agayev, A. (2007). Increased gray matter density in the parietal cortex of mathematicians: A voxel-based morphometry study. *American Journal of Neuroradiology*, 28, 1859–1864.
<https://doi.org/10.3174/ajnr.a0696>
- Barut, Ç., Kiziltan, E., Gelir, E., & Köktürk, F. (2013). Advanced analysis of finger-tapping performance: A preliminary study. *Balkan Medical Journal*, 30, 167–171.
<https://doi.org/10.5152/balkanmedj.2012.105>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300. Retrieved from
<http://www.jstor.org.proxy.kib.ki.se/stable/2346101>
- Bosch, M., Castro, J., Saneyoshi, T., Matsuno, H., Sur, M., & Hayashi, Y. (2014). Structural and molecular remodeling of dendritic spine substructures during long-term potentiation. *Neuron*, 82, 444–459. <https://doi.org/10.1016/j.neuron.2014.03.021>
- Carreiras, M., Seghier, M. L., Baquero, S., Estévez, A., Lozano, A., Devlin, J. T., & Price, C. J. (2009). An anatomical signature for literacy. *Nature*, 461, 983–986.
<https://doi.org/10.1038/nature08461>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences, second edition. Statistical Power Analysis for the Behavioral Sciences* (2nd ed.).
- Cole, J. H., Poudel, R. P. K., Tsagkrasoulis, D., Caan, M. W. A., Steves, C., Spector, T. D., & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, 163, 115–124.

- <https://doi.org/10.1016/j.neuroimage.2017.07.059>
- Dayan, E., & Cohen, L. G. (2011). Neuroplasticity subserving motor skill learning. *Neuron*, 72(3), 443–454. <https://doi.org/10.1016/j.neuron.2011.10.008>
- Debowska, W., Wolak, T., Nowicka, A., Kozak, A., Szwed, M., & Kossut, M. (2016). Functional and structural neuroplasticity induced by short-term tactile training based on braille reading. *Frontiers in Neuroscience*, 10. <https://doi.org/10.3389/fnins.2016.00460>
- Deep-Soboslay, A., Hyde, T. M., Callicott, J. P., Lener, M. S., Verchinski, B. A., Apud, J. A., ... Elvevåg, B. (2010). Handedness, heritability, neurocognition and brain asymmetry in schizophrenia. *Brain*, 133, 3113–3122. <https://doi.org/10.1093/brain/awq160>
- Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Changes in grey matter induced by training. *Nature*, 427, 311–312. <https://doi.org/10.1038/427311a>
- Dragovic, M. (2004). Categorization and validation of handedness using latent class analysis. *Acta Neuropsychiatrica*, 16(4), 212–218. <https://doi.org/10.1111/j.0924-2708.2004.00087.x>
- Eriksson, I., Undén, A., & Elofsson, S. (2001). Self-rated health. Comparisons between three different measures. Results from a population study. *International Journal of Epidemiology*, 30, 326–333. <https://doi.org/10.1093/ije/30.2.326>
- Franklin, T. R., Wang, Z., Shin, J., Jagannathan, K., Suh, J. J., Detre, J. A., ... Childress, A. R. (2013). A VBM study demonstrating “apparent” effects of a single dose of medication on T1-weighted MRIs. *Brain Structure and Function*, 218, 97–104. <https://doi.org/10.1007/s00429-012-0385-6>
- Gaser, C., & Dahnke, R. (2019). CAT12 (Computational Anatomy Toolbox for SPM) (Version 12) [Desktop application]. Retrieved from <http://www.neuro.uni-jena.de/cat/>
- Ge, Q., Peng, W., Zhang, J., Weng, X., Zhang, Y., Liu, T., ... Wang, Z. (2017). Short-term apparent brain tissue changes are contributed by cerebral blood flow alterations. *PLoS ONE*, 12, 1–12. <https://doi.org/10.1371/journal.pone.0182182>
- Hofstetter, S., Tavor, I., Tzur Moryosef, S., & Assaf, Y. (2013). Short-term learning induces white matter plasticity in the fornix. *Journal of Neuroscience*, 33, 12844–12850. <https://doi.org/10.1523/JNEUROSCI.4520-12.2013>
- Horenstein, C., Lowe, M. J., Koenig, K. A., & Phillips, M. D. (2009). Comparison of unilateral and bilateral complex finger tapping-related activation in premotor and primary motor cortex. *Human Brain Mapping*, 30, 1397–1412. <https://doi.org/10.1002/hbm.20610>
- Huettel, S. a., Song, A. W., & McCarthy, G. (2009). *Functional magnetic resonance imaging* (2nd ed.). Sunderland, MA: Sinauer Associates.
- Hölzel, B. K., Carmody, J., Vangel, M., Congleton, C., Yerramsetti, S. M., Gard, T., & Lazara, S. W. (2012). Mindfulness practice leads to increases in regional brain gray matter density. *Psychiatry Res.*, 191, 36–43. <https://doi.org/10.1016/j.psychresns.2010.08.006> Mindfulness
- Jobbágy, Á., Harcos, P., Karoly, R., & Fazekas, G. (2005). Analysis of finger-tapping movement. *Journal of Neuroscience Methods*, 141, 29–39. <https://doi.org/10.1016/j.jneumeth.2004.05.009>
- Jäncke, L., Schlaug, G., & Steinmetz, H. (1997). Hand skill asymmetry in professional musicians. *Brain and Cognition*, 34, 424–432. <https://doi.org/10.1006/brcg.1997.0922>
- Karni, A., Meyer, G., Jezard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1995). Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature*, 377, 155–158. <https://doi.org/10.1038/377155a0>
- Karni, A., Meyer, G., Rey-Hipolito, C., Jezard, P., Adams, M. M., Turner, R., & Ungerleider, L.

- G. (1998). The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 861–8. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9448252>
- Kwok, V., Niu, Z., Kay, P., Zhou, K., Mo, L., Jin, Z., ... Tan, L. H. (2011). Learning new color names produces rapid increase in gray matter in the intact adult human cortex. *Proceedings of the National Academy of Sciences*, 108, 6686–6688. <https://doi.org/10.1073/pnas.1103217108>
- Landi, S. M., Baguear, F., & Della-Maggiore, V. (2011). One week of motor adaptation induces structural changes in primary motor cortex that predict long-term memory one year later. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 31, 11808–13. <https://doi.org/10.1523/JNEUROSCI.2253-11.2011>
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, 97, 4398–4403. <https://doi.org/10.1073/pnas.070039597>
- McJury, M., & Shellock, F. G. (2000). Auditory noise associated with MR procedures: A review. *Journal of Magnetic Resonance Imaging*, 12, 37–45. [https://doi.org/10.1002/1522-2586\(200007\)12:1<37::AID-JMRI5>3.0.CO;2-I](https://doi.org/10.1002/1522-2586(200007)12:1<37::AID-JMRI5>3.0.CO;2-I)
- Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Structural plasticity in the bilingual brain. *Nature*, 431, 757–757. <https://doi.org/10.1038/431757a>
- Mollasadeghi, A., Mehrparvar, A. H., Atighechi, S., Davari, M. H., Shokouh, P., Mostaghaci, M., & Bahaloo, M. (2013). Sensorineural hearing loss after magnetic resonance imaging. *Case Reports in Radiology*, 2013, 1–3. <https://doi.org/10.1155/2013/510258>
- Månsson, K. N. T., Cortes, D. S., Manzouri, A., Hau, S., & Fischer, H. (2018). Viewing pictures triggers rapid morphological enlargement in the human visual cortex. *BioRxiv*, 1–8. <https://doi.org/10.1101/408658>
- Månsson, K. N. T., Salami, A., Frick, A., Carlbring, P., Andersson, G., Furmark, T., & Boraxbekk, C.-J. (2016). Neuroplasticity in response to cognitive behavior therapy for social anxiety disorder. *Translational Psychiatry*, 6, e727–e727. <https://doi.org/10.1038/tp.2015.218>
- Nettersheim, A., Hallschmid, M., Born, J., & Diekelmann, S. (2015). The role of sleep in motor sequence consolidation: Stabilization rather than enhancement. *Journal of Neuroscience*, 35, 6696–6702. <https://doi.org/10.1523/JNEUROSCI.1236-14.2015>
- Neurosynth. (2019). Finger tapping - an automated meta-analysis of 75 studies. Retrieved May 12, 2019, from [http://neurosynth.org/analyses/terms/finger tapping/](http://neurosynth.org/analyses/terms/finger%20tapping/)
- Notepad++ team. (2016). Notepad++. [Windows application]. Retrieved from <https://notepad-plus-plus.org/>
- Nyberg, L., Eriksson, J., Larsson, A., & Marklund, P. (2006). Learning by doing versus learning by thinking: An fMRI study of motor and mental training. *Neuropsychologia*, 44, 711–717. <https://doi.org/10.1016/j.neuropsychologia.2005.08.006>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Pan, S. C., & Rickard, T. C. (2015). Sleep and motor learning: Is there room for consolidation? *Psychological Bulletin*, 141, 812–834. <https://doi.org/10.1037/bul0000009>

- Psychology Software Tools, I. (2015). E-Prime (Version 2.0.10.356). [Windows application]. Retrieved from <https://www.pstnet.com>
- Purves, D., Cabeza, R., Huettel, S. A., LaBar, K. S., Platt, M. L., & Woldorff, M. M. (2013). Motor systems: The organization of action. In *Principles of cognitive neuroscience* (2nd ed., pp. 131–166). Sunderland, MA: Sinauer Associates.
- Reid, L. B., Sale, M. V., Cunnington, R., Mattingley, J. B., & Rose, S. E. (2017). Brain changes following four weeks of unimanual motor training: Evidence from fMRI-guided diffusion MRI tractography. *Human Brain Mapping, 38*, 4302–4312. <https://doi.org/10.1002/hbm.23514>
- Reitan, R. M., & Wolfson, D. (1993). *The Halstead–Reitan neuropsychological test battery: Theory and clinical interpretation* (2nd ed.). Tucson, AZ: Neuropsychology Press.
- Rosenzweig, M. R. (2002). Aspects of the search for neural mechanisms of memory. *Annual Review of Psychology, 47*, 1–32. <https://doi.org/10.1146/annurev.psych.47.1.1>
- Sale, M. V., Reid, L. B., Cocchi, L., Pagnozzi, A. M., Rose, S. E., & Mattingley, J. B. (2017). Brain changes following four weeks of unimanual motor training: Evidence from behavior, neural stimulation, cortical thickness, and functional MRI. *Human Brain Mapping, 38*, 4773–4787. <https://doi.org/10.1002/hbm.23710>
- Schmidt-Wilcke, T., Wulms, N., Heba, S., Pleger, B., Puts, N. A., Glaubit, B., ... Dinse, H. R. (2018). Structural changes in brain morphology induced by brief periods of repetitive sensory stimulation. *NeuroImage, 165*, 148–157. <https://doi.org/10.1016/j.neuroimage.2017.10.016>
- Sheppard, A., Chen, Y., & Salvi, R. (2018). MRI Noise and Hearing Loss. *The Hearing Journal, 71*, 30. <https://doi.org/10.1097/01.HJ.0000532395.75558.2d>
- Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., & Greicius, M. D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex, 22*, 158–165. <https://doi.org/10.1093/cercor/bhr099>
- Snyder, J. S. (2019). Recalibrating the relevance of adult neurogenesis. *Trends in Neurosciences, 42*, 164–178. <https://doi.org/10.1016/j.tins.2018.12.001>
- Stahnisch, F. W. (2003). Making the brain plastic: Early neuroanatomical staining techniques and the pursuit of structural plasticity, 1910–1970. *Journal of the History of the Neurosciences, 12*, 413–435. <https://doi.org/10.1076/jhin.12.4.413.27917>
- Tardif, C. L., Gauthier, C. J., Steele, C. J., Bazin, P. L., Schäfer, A., Schaefer, A., ... Villringer, A. (2016). Advanced MRI techniques to improve our understanding of experience-induced neuroplasticity. *NeuroImage, 131*, 55–72. <https://doi.org/10.1016/j.neuroimage.2015.08.047>
- Tassopoulos, M. (n.d.). *ICA guidelines*. Unpublished document.
- Taubert, M., Mehnert, J., Pleger, B., & Villringer, A. (2016). Rapid and specific gray matter changes in M1 induced by balance training. *NeuroImage, 133*, 399–407. <https://doi.org/10.1016/j.neuroimage.2016.03.017>
- The Functional Imaging Laboratory Methods Group. (2014). SPM12 (Statistical Parametric Mapping) (Version 12) [Desktop application]. Retrieved from <http://www.fil.ion.ucl.ac.uk/spm>
- Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with sleep makes perfect: Sleep-dependent motor skill learning. *Neuron, 35*, 205–211. Retrieved from https://journals-scholarsportal-info.myaccess.library.utoronto.ca/pdf/08966273/v35i0001/205_pwsmp.xml
- Wenger, E., Brozzoli, C., Lindenberger, U., & Lövdén, M. (2017). Expansion and

- renormalization of human brain structure during skill acquisition. *Trends in Cognitive Sciences*, 21, 930–939. <https://doi.org/10.1016/j.tics.2017.09.008>
- Wenger, E., Kühn, S., Verrel, J., Mårtensson, J., Bodammer, N. C., Lindenberger, U., & Lövdén, M. (2017). Repeated structural imaging reveals nonlinear progression of experience-dependent volume changes in human motor cortex. *Cerebral Cortex*, 27, 2911–2925. <https://doi.org/10.1093/cercor/bhw141>
- Witt, S. T., Laird, A. R., & Meyerand, M. E. (2008). Functional neuroimaging correlates of finger-tapping task variations: An ALE meta-analysis. *NeuroImage*, 42, 343–356. <https://doi.org/10.1016/j.neuroimage.2008.04.025>
- Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8, 665–670. <https://doi.org/10.1038/nmeth.1635>
- Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature Neuroscience*, 15, 528–536. <https://doi.org/10.1038/nn.3045>

Appendix: Independent Component Analysis and Manual fMRI Denoising

When conducting BOLD fMRI research, it is important to differentiate between actual neural-related signal and noise generated by other factors such as subject body movement and autonomous physiological factors (Bright, Tench, & Murphy, 2017). This differentiation can be done by using software that attempts to automatically separate noise from legitimate BOLD signal. In other cases, one might use a semi-automated process, where an algorithm is trained to be able to differentiate between signal and noise. This requires some initial hand classification of fMRI images, the results of which are then fed into the algorithm in order to help it learn how to classify future images. However, neither automated nor semi-automated processes for denoising is as effective at separating signal from noise as hand classification of ICA components. While hand classification is the gold standard due to its high accuracy, it is both time consuming and requires expertise (Griffanti et al., 2017).

Initially, our methodological aim was to utilize manual denoising due to reasons described above. In order to be able to accurately and reliably classify ICA components, the authors underwent denoising training. First, the authors studied material and guidelines for manual classification (Garrett et al., 2014, Supplementary materials; Griffanti et al., 2017; Kelly Jr. et al., 2010; Tassopoulos, n.d.) and attended a one day denoising workshop. They then proceeded by independently classifying components in a sample containing data from 9 BOLD fMRI trials unrelated to the present study. This data was then checked against a database of correct responses, ensuring that the authors achieved at least 80% correct decisions for each subject. After independent classification, corner cases were reviewed in a joint discussion between the authors and their supervisor. The authors then used the same classification procedure when performing manual denoising on the data set intended to be used for the present study. However, ICA components from our trials proved to contain an unusually high amount of noise, possibly due to our short MR sequence durations. Thus, our ICA components were deemed unusable for the purpose of this study, which instead led us to utilize the existing literature for defining a region of interest.