Detailing reliability estimation of the individual working brain by varying spatial and temporal resolution in accelerated echo planar MR imaging
Für meine Frau
# Table of Contents

1. Introduction ........................................................................................................ p. 1
2. Material and methods ............................................................................................... p. 4
   2.1. Subjects ........................................................................................................... p. 4
   2.2. Task ............................................................................................................... p. 4
   2.3. Procedure ....................................................................................................... p. 5
   2.4. Imaging protocols ............................................................................................ p. 5
   2.5. fMRI preprocessing ......................................................................................... p. 6
   2.6. Processing of anatomical data and cortex based alignment (CBA) .......... p. 7
   2.7. Behavioral data ............................................................................................... p. 9
   2.8. Data plausibility check with GLM ................................................................. p. 9
   2.9. Fast Fourier transform (FFT) ......................................................................... p. 10
   2.10. Test-retest reliability analysis ...................................................................... p. 12
   2.11. Conjunction of significant reliability analysis ............................................ p. 13
   2.12. Fixed and gliding GLM brain activity mask analyses ................................... p. 14
3. Results ................................................................................................................... p. 16
   3.1. Behavioural data (reaction time analyses) .................................................... p. 16
   3.2. Data plausibility check (GLM results) ............................................................ p. 16
   3.3. FFT .............................................................................................................. p. 18
   3.4. Test-retest reliability analysis ...................................................................... p. 18
   3.5. Conjunction of reliability analysis ............................................................... p. 25
   3.6. Fixed and gliding GLM brain activity mask analyses ................................... p. 28
4. Discussion .............................................................................................................. p. 34
5. Conclusions ............................................................................................................ p. 38
6. References ............................................................................................................. p. 39
7. Summaries .............................................................................................................. p. 45
   7.1. Summary (English version) ........................................................................... p. 45
   7.2. Summary (German version) .......................................................................... p. 46
8. Abbreviations ......................................................................................................... p. 47
9. Appendix ............................................................................................................... p. 49
   9.1. List of figures ................................................................................................. p. 49
   9.2. List of tables ................................................................................................ p. 51
   9.3. „Erklärung § 5 Abs. 1 zur Datenaufbewahrung“ ........................................ p. 52
   9.4. „Erklärung über den Eigenanteil“ ................................................................. p. 53
   9.5. „Danksagung“ .............................................................................................. p. 56
   9.6. Curriculum vitae ............................................................................................ p. 57
1. Introduction

The human brain is a complex oscillating system that generates specific neural oscillations which are related to neural functions. With fMRI these neural oscillations can be studied when converting the BOLD time course into the frequency domain respectively the fMRI frequency spectrum. The properties of the fMRI frequency spectrum rely on the chosen fMRI acquisition parameters. Fast repetition time (TR) would yield a broad fMRI frequency spectrum that represents fast fMRI frequencies. Large voxel size would yield a good signal-to-noise ratio (SNR) of the fMRI signal while spatial resolution is reduced. In fMRI studies it is crucial to balance the competing constraints of these acquisition parameters. Consequently, voxel sizes ranging from $2^3 - 3^3$ mm$^3$ and repetition times (TR) of approximately 2000 ms are used (Turner and Jones, 2003). The exact impact of voxel size and TR on the reproducibility and validity of fMRI frequency spectra remains under investigation.

Several approaches have been pursued to investigate test-retest reliability in general. One approach focuses on intraclass correlation (ICC; Shrout and Fleiss, 1979) and was originally developed to assess the consistency, or conformity, of subjective judgments made by human observers. ICCs are suitable for many research questions in the context of reliability, when the conditions are appropriate and the proper ICC model is chosen (McGraw & Wong, 1996). Analyzing fMRI reliability with ICC was first introduced by Specht et al., 2003 and has gained some acceptance among fMRI researchers since that time. While this approach focused on conventional general linear model analysis, the ICC was also adopted to estimate reliability of specific components of the whole fMRI frequency spectrum (Zuo et al., 2010a). While ICCs in fMRI were first applied to group level studies ICCs can be applied in a straightforward way to assess intra-individual differences in test-retest reliability in fMRI (Plichta et al., 2012, Gorgolewski et al., 2013).

So far the test-retest reliability of brain activity (Bennett and Miller, 2010) and brain connectivity (Braun et al., 2012; Chen et al., 2008; Damoiseaux et al.,
2006; Deuker et al., 2009; Schuyler et al., 2010; Shehzad et al., 2009; Telesford et al., 2010; Wang et al., 2011; Zuo et al., 2010b) has been widely assessed. Almost all fMRI reliability studies were conducted with conventional single-shot echo planar imaging (EPI) sequences. Single-shot EPI sequences achieve whole brain coverage by balancing the competing constraints of spatial resolution and temporal resolution leading to voxel sizes around $3^3 \text{ mm}^3$ and TR around 2000 ms (Turner and Jones, 2003). Test-retest reliability of these conventional acquisition parameters is relevant for conventional fMRI but many new fMRI methods require smaller voxel sizes and faster TR. In particular, it is currently well accepted that multi voxel pattern analysis (MVPA) is ideally performed on spatially unsmoothed high spatial resolution fMRI data (Norman et al., 2006). Furthermore, it has been shown that Granger causality is optimally detected with TR of approximately 1000 ms (Goebel et al., 2003). Small distance connectivity is detected at frequencies around 0.3 Hz which would require a TR of approximately 1000 ms (Salvador et al., 2005). Finally, highly exotic sequences are chosen when fMRI is combined with EEG, which indeed lead to TR around 330 ms (Scheeringa et al., 2011). Remarkably enough, very little is known about the test-retest reliability of these unconventional sequence types with small voxel size and fast TR.

Problems with temporal or spatial resolution are partly overcome when partial parallel imaging (PPI) techniques are used. In MVPA whole brain coverage at a given TR could sometimes only be achieved with PPI. PPI is an acceleration technique that affords TR reduction and helps to reduce image distortions in EPI commonly used for fMRI. The use of PPI can be difficult because it shows less favorable SNR when compared to non-PPI sequences. Nevertheless, the number of observations of brain activity is increased with PPI when compared to conventional EPI sequences. Remarkably enough, to date only one study investigated the test-retest reliability of PPI in a BOLD imaging context (Krinzinger et al., 2011).

When discussing test-retest reliability in fMRI, researchers should always consider SNR and temporal SNR (TSNR). SNR and TSNR decrease proportionally with voxel size (Triantafyllou et al., 2005). Consequently, BOLD
imaging with a small voxel size can be disadvantageous indeed. By contrast, an increase in TR through reduction of imaging flip angle is less disadvantageous. In principle optimal TSNR is found for the Ernst angle (Ernst and Anderson, 1966; Helms et al., 2008; Ye et al., 2010), but reduction of imaging flip angle well below the Ernst angle results in negligible loss in TSNR (Gonzalez-Castillo et al., 2011).

In this study we investigated how changes in voxel size and TR affect the test-retest reliability of the fMRI frequency spectrum by means of the ICC. As discussed above modern imaging methods require that the whole fMRI frequency spectrum is scanned with sufficient test-retest reliability. Furthermore, it has been shown that fMRI frequencies above the low frequency oscillation band (LFO; 0.01-0.1 Hz) contain meaningful neural information (Salvador et al., 2005; Zuo et al., 2010a). Consequently, we explicitly focused on the test-retest reliability of the entire fMRI frequency spectrum by means of Fourier coefficients. Fourier coefficients are elements of the fast Fourier transform (FFT) and are therefore at the base of every calculation of frequency bands. We focused on spatially unsmoothed data sets because they are a technical requirement for high resolution data analysis approaches. We investigated reliability at the individual subject level which may have implications for clinical applications. Furthermore, we completed our analysis on a whole-brain, vertex-wise basis. As one of the strictest criteria for reliability, this approach provides a global measure of concordance that indicates how effectively the fMRI frequency spectrum is represented in each test-retest pairing (Bennett and Miller, 2010).
2. Materials and methods

2.1. Subjects

Twenty-five healthy right-handed male subjects were included in the study (mean age 27.6; range 20-44 years). All subjects were free of neurological or psychiatric disorders and were without a history of alcohol or drug abuse. None of the subjects took medications known to alter brain activity. From each subject we obtained written informed consent prior to the study in accordance with the Declaration of Helsinki.

2.2. Task

A Stroop interference task (Stroop, 1935, see McLeod, 1991, for overview) was chosen because this paradigm is well studied and is known to invoke widespread activation in the brain (Laird et al., 2005; Roberts et al., 2008; Nee et al., 2007). The Stroop task consisted of a direction and a color condition. In all conditions a yellow or a blue arrow were presented that pointed to the left or the right hand side. During the direction condition subjects had to press the left or the right index finger congruent with the direction of the arrow (“normal condition”). During the color condition subjects had to press the left index finger for a yellow arrow or the right index finger for a blue arrow, irrespective of the direction of the arrow (“Stroop condition”). Congruent stimuli were defined as yellow arrows that pointed to the left hand side and blue arrows that pointed to the right hand side. Incongruent stimuli were defined as yellow arrows that pointed to the right hand side and blue arrows that pointed to the left hand side. Presentation frequency of arrows was balanced for congruency and for incongruency as well as for left respectively right pointing direction.
2.3. Procedure

Each subject underwent two scanning sessions that are further referred to test and retest sessions. The measurement protocol for both sessions was identical. The time interval between sessions was 81.0 ± 46.6 (SD) min. Within this time subjects were asked to relax outside of the scanner. The test and the retest session both consisted of six fMRI scans. At the end of the test session an anatomical scan was acquired. The task for each scan was the same and the scans varied only in the fMRI acquisition protocol (see below). Each fMRI scan consisted of four blocks, where the color and the direction conditions alternated always starting with the direction block. Each block consisted of 36 trials and lasted for 32.4 sec. The trial consisted of the stimulus presentation with a duration of 0.5 seconds and an inter-stimulus interval of 0.4 seconds when only a fixation cross was presented. The trials with the yellow/blue arrow pointing to the left/right were randomized within the blocks. Prior to each block an instruction screen was presented for 3.5 seconds informing the subject about the next condition followed by a 2.5 seconds empty interval. After each block a baseline interval for 12 seconds was inserted. During the baseline only a fixation cross was presented. At the end of the scan another baseline interval of 20 seconds was included. Total duration of one scan including the instructions and the baseline was 221.6 seconds.

2.4. Imaging protocols

Functional and structural images were acquired on a 3-Tesla Scanner (Tim Trio, Siemens Medical Systems, Erlangen, Germany), equipped with a 12-channel head coil. Six different single-shot echo planar imaging (EPI) protocols which varied in temporal and spatial resolutions (e.g. TR = 1250 ms or 1600 ms or 2400 ms and voxel size = 2.5³ mm³ or 3.5³ mm³ or 4.5³ mm³) were designed (Table 1). Acquisition acceleration was applied in all protocols using parallel imaging with an acceleration factor R = 2 and 24 reference lines together with a generalized autocalibrating partially parallel acquisitions (GRAPPA)
reconstruction algorithm (Griswold et al., 2002). Images were acquired in interleaved slice order using a product EPI sequence (TE: 30 ms, distance factor: 10 %, receiver bandwidth: 2298 Hz/Px, echo spacing: 0.52 ms). The optimal flip angles were selected to be equal to the Ernst angle (Ernst and Anderson, 1966; Helms et al., 2008; Ye et al., 2010). Phase encoding was placed along the anterior to posterior direction. At the end of the test session, high spatial resolution T1-weighted images were acquired, using a magnetization prepared rapid gradient echo protocol (MPRAGE, 1³ mm³, matrix size = 256 x 240, TR = 2300 ms, TE = 2.98 ms, TI = 900 ms, flip angle = 9°, slices = 160, FOV = 256 x 240 mm²). In addition, four 2D fast low-angle shot T1-weighted data sets were acquired - two within the first session and two within the second one (2D-FLASH, 0.43 x 0.43 x 4 mm³, matrix size = 512 x 448, TR = 220 ms, TE = 2.46 ms, flip angle = 70°, slices = 25, FOV = 192 x 220 mm²). They were further used for co-registration purposes.

<table>
<thead>
<tr>
<th>Voxel size (mm³)</th>
<th>TR (ms)</th>
<th>TE (ms)</th>
<th>Flip angle (deg)</th>
<th>Slices</th>
<th>FOV (mm²)</th>
<th>Matrix size</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5 x 4.5 x 4.5</td>
<td>2400</td>
<td>30</td>
<td>80</td>
<td>23</td>
<td>288 x 288</td>
<td>320 x 320</td>
</tr>
<tr>
<td>4.5 x 4.5 x 4.5</td>
<td>1600</td>
<td>30</td>
<td>72</td>
<td>23</td>
<td>288 x 288</td>
<td>320 x 320</td>
</tr>
<tr>
<td>4.5 x 4.5 x 4.5</td>
<td>1250</td>
<td>30</td>
<td>68</td>
<td>23</td>
<td>288 x 288</td>
<td>320 x 320</td>
</tr>
<tr>
<td>3.5 x 3.5 x 3.5</td>
<td>2400</td>
<td>30</td>
<td>80</td>
<td>30</td>
<td>224 x 224</td>
<td>384 x 384</td>
</tr>
<tr>
<td>3.5 x 3.5 x 3.5</td>
<td>1600</td>
<td>30</td>
<td>72</td>
<td>30</td>
<td>224 x 224</td>
<td>384 x 384</td>
</tr>
<tr>
<td>2.5 x 2.5 x 2.5</td>
<td>2400</td>
<td>30</td>
<td>80</td>
<td>42</td>
<td>220 x 220</td>
<td>616 x 616</td>
</tr>
</tbody>
</table>

2.5. fMRI preprocessing

The obtained data were further processed using BrainVoyager QX 2.6 software (Brain Innovation, Maastricht, Netherlands). The first two images of each functional run were discarded to avoid T1 saturation effects. For preprocessing purposes slice scan time correction, linear trend removal and 3D motion
correction were applied. The results of 3D motion correction were checked visually. In a single scan of one subject the head motion estimates were larger than the voxel size of the EPI protocol used. Nevertheless, we did not exclude these data because the beta weights obtained at the data plausibility check (see below) for this session did not differ from the beta weights of the remaining five sessions. We did not apply any temporal filter since we aimed at tracing the entire fMRI frequency spectrum. It has been demonstrated that all frequencies within the fMRI frequency spectrum are of potential interest (Buzsáki and Draguhn, 2004; Penttonen et al., 2003; Salvador et al., 2005). Frequencies around 0.01 Hz are believed to link small and big world connectivity, while frequency bands around 0.3 Hz link small world connectivity (Salvador et al., 2005). We did not apply spatial smoothing to assure equal conditions for the six scanning protocols because different voxel sizes would call for different smoothing kernels (Weibull et al., 2008). Furthermore, spatial smoothing is known to obscure effects of interest (Kriegeskorte et al., 2006; Fransson et al., 2002; White et al., 2001). We resampled all volume time-courses to $1^3 \text{mm}^3$ to assure comparability of six scanning protocols. We used nearest neighbor interpolation, since we intended to conserve the original spatial properties of the data as much as possible.

2.6. Processing of anatomical data and cortex based alignment (CBA)

Individual MPRAGE T1-weighted datasets underwent manual Talairach transformation. The transformation was performed in two steps. As a first step, we manually defined the anterior and posterior commissure, which was used as a starting point for an AC-PC plane transformation with trilinear interpolation. As a second step, we manually defined the extreme points of the cerebrum and transformed it into standard space with trilinear interpolation. The coregistration of the fMRI images consisted of two steps. In a first step, 2D-FLASH T1-weighted images acquired within the sessions were co-registered to the MPRAGE T1-weighted image using NeuroElf software (www.neuroelf.com). In a second step, the fMRI images were co-registered to the 2D-FLASH T1-
weighted images. The 2D-FLASH T1-weighted images revealed a better anatomical contrast than the conventionally used first EPI volume of the fMRI images leading to more sufficient coregistration results (Gartus et al., 2007; Koten et al., 2009). The results of the final co-registration were visually inspected but no irregularities were found. Finally, the obtained transformation matrices were further applied to the functional data and volumetric time course (.vtc) data in Talairach space were created.

We restricted our analysis to grey matter voxels. For this purpose cortex based alignment (CBA; Fischl et al., 1999) is a promising technique for four reasons. First, CBA leads to a mask that only considers activity within grey matter. This is of particular importance, because fMRI frequency spectra of grey matter represent more meaningful neural activity than those of white matter (Zuo et al., 2010a). It is obvious that the most accurate masks are obtained at the individual subject level. Second, CBA results in superior overlap between subjects, although it does not rely on spatial smoothing, which is a weighted spatial averaging procedure. Spatial averaging procedures have two main pitfalls for fMRI data (Fransson et al., 2002; White et al., 2001). A Gaussian filter of $9^3$ mm$^3$ - as required for larger voxel sizes - would exceed cortical thickness and therefore introduce irrelevant signal originating from white matter, blood vessels or spinal fluid. A Gaussian filter of $5^3$ mm$^3$ - as required for small voxel sizes - would mitigate high spatial resolution of the image. Finally, CBA is known to coregistrate the cytoarchitectonic borders of the cortex more accurately as compared to volume based coregistration methods (Fischl et al., 2008). Furthermore it has been shown that CBA-based fMRI analyses provide more robust results even in conventional GLM analyses especially without smoothing of the functional data (Goebel et al., 2006). Finally, our ICC model requires spatially independent data that are only obtained in unsmoothed datasets (Telesford et al., 2010).

To perform CBA we carried out brain segmentations with half-automated procedures implemented in BrainVoyager QX 2.6. The results of the segmentations were visually inspected and manually corrected if necessary. The cortices were reconstructed to 3D surfaces (150 iterations, smoothing force...
0.07) and inflated to spheres. To standardize the alignment procedure, individual spheres were sampled from a standard sphere with 40962 vertices and individual curvature maps were computed. These curvature maps were used to run an extended multi-scale version of the CBA procedure with maximization of curvature overlap and creation of individual CBA transformation matrices (Fischl et al., 1999; Goebel et al., 2006). The curvature maps were further used as masks to extract mesh time courses (.mtc) from the .vtc data. Based on the individual CBA transformation matrices these .mtc data were further aligned between brains. As such we achieved optimal correspondence between the anatomies of the individual subjects. At the end we obtained the time courses of 40962 vertices for each subjects, so that the vertices were anatomically optimal correspondent between subjects. In the following, we will refer to a voxel, when we discuss fMRI resolution, whereas we will refer to a vertex, when we discuss statistical analyses and specific positions on the cerebral surface.

2.7. Behavioral data

Reaction times of both test and retest session were recorded and averaged per subject. The mean RT per subject was subjected to a paired \(t\)-test for a within-subject comparison of task conditions. In addition, an intraclass correlation analysis (ICC, Shrout and Fleiss, 1979) was performed over mean response time per subject to inform about the test-retest reliability of the behavioral data.

2.8. Data plausibility check with GLM

To assure that our data are plausible we performed a second level GLM analysis. Each condition of the experiment was modeled as a separate predictor and as predictors of no interest we added motion parameters. These design matrices were subjected to a RFX GLM analysis on the data. To this end, we obtained twelve GLMs - one for each of two sessions of six scanning protocols.
Furthermore we applied two contrasts for each GLM obtained, such as (color condition + direction condition) vs. baseline condition as well as color condition vs. direction condition. The \(t\)-maps obtained from each GLM were averaged across GLMs and projected onto the brain surfaces. These brain surfaces were furthermore used to compute GLM brain activation masks in order to validate the Fourier-based measure (see below).

2.9. Fast Fourier transform (FFT)

We performed a vertex-wise fast Fourier transform (FFT) over time-courses for the test and retest session using Matlab 2012b (MathWorks, Natick, Massachusetts, USA). The highest detectable frequency of the power spectrum is given by the Nyquist frequency that is a direct function of TR. In our case the fMRI frequency spectrum of the fastest protocol was almost two times broader than the fMRI frequency spectrum of the slowest protocol (Fig. 1). Thus the number of Fourier coefficients entered into the computation of the ICC was larger for protocols with faster TR (TR = 1250 ms and TR = 1600 ms, both 129 Fourier coefficients) as compared to protocols with slower TR (TR = 2400 ms, 65 Fourier coefficients). The number of Fourier coefficients of protocols with fast TR did not differ due to the nature of the FFT algorithm which only considers powers of two. For the FFT we explicitly did not analyze separately the direction and the color condition, because this would lead to fragmentation of the time-course. Fragmentation of the time-course would mitigate block related frequencies and especially LFO, which are better estimated over longer time intervals. We computed the grand mean time-course across subjects and the related fMRI frequency spectrum for each scanning protocol to illustrate the analysis procedure (Fig. 1B).
Fig. 1. Time courses and fMRI frequency spectra obtained with six different EPI protocols including three voxel sizes and three TR for a ROI in superior parietal cortex. (A) Effect of voxel size and TR on the time course for a randomly selected subject. Prominent intraindividual differences between session A and B are apparent for time courses and fMRI frequency spectra. (B) Effect of voxel size and TR on the mean time course obtained from \( n = 25 \) subjects. Intraindividual differences level out after averaging. Resulting fMRI frequency spectra revealed two main peaks representing the task frequency (0.017 Hz) and the second harmonic (0.034 Hz) which served as a plausibility check. BOLD spectra are broader for faster sampling rates due to the Nyquist frequency.

2.10. Test-retest reliability analysis

The Fourier coefficients derived from the FFT analysis were used to estimate reliability by means of intraclass correlation (ICC; Shrout and Fleiss, 1979; Specht et al., 2003). ICCs were computed separately for all vertices of every subject using the following equation (cf. Caceres et al., 2009):

\[
ICC = \frac{BMS - EMS}{BMS + (k - 1)EMS}
\]

For every vertex, a data matrix was created with Fourier coefficients as “targets” and \( k = 2 \) sessions as “judges”. The within-target mean square (EMS) is defined as the mean sums of squares obtained for the session factor divided by the respective degrees of freedom. The between-targets mean square (BMS) is defined as the mean sums of squares obtained for the target factor divided by the degrees of freedom. Finally, a total number of 40962 ICCs per subject was obtained corresponding to the number of vertices per subject. Although slightly unconventional, first level ICC estimates are accepted within the imaging community (Plichta et al., 2012; Gorgolewski et al., 2013).

For every subject six ICC maps were computed, one for each scanning protocol. Subsequently, the ICC estimates were transformed using the Fisher \( z' \)-transformation to better approximate a normal distribution of the ICC coefficients. The transformed correlations were averaged across subjects and standard deviations were computed, followed by a \( t \)-test for departure from zero at the group level. Finally, mean correlations and their standard deviations were back-transformed to ICCs and projected onto cortical maps. Three histograms
where generated, which show the distribution of mean ICC, the distribution of the standard deviation of the ICC and the distribution of the significance of mean ICC expressed as \( t \)-values.

In order to assess significant differences in reliability between different voxel sizes and TRs we computed vertex-wise second-level paired \( t \)-tests on the Fisher z'-transformed ICCs. Subsequently, the vertex-wise mean difference in reliability between two scanning protocols was accessed. In principle fifteen contrasts are possible for six scanning conditions but due to limitations of space, we only computed contrasts as a function of either voxel size or TR. Within this context, the three scanning protocols with TR of 2400 ms were tested against each other as well as the three protocols with voxel size of \( 4.5^3 \text{mm}^3 \). Resulting maps were corrected for multiple testing with a cluster threshold estimation procedure (Hagler et al., 2006).

In order to compute so called radio curves and to visualize the increase in the number of significantly reproducible vertices at different significance thresholds, we first computed the total number of vertices. For this analysis we took the whole brain into account leading to a total number of 81924 vertices because of 40962 vertices per hemisphere. All vertices were averaged across subjects. The three scanning protocols with TR of 2400 ms and the three protocols with voxel size of \( 4.5^3 \text{mm}^3 \) were analyzed separately. Then we computed the relative increase in reproducibility by means of ratios, deliberately choosing the smallest voxel size of \( 2.5^3 \text{mm}^3 \) and the slowest TR of 2400 ms as a reference value.

2.11. Conjunction of significant reliability analysis

One might argue that fMRI frequency spectra in specific parts of the brain are always detected reliably independent of voxel size or TR, whereas in other parts of the brain reliability might profit from specific acquisition parameters. We have tried to assess the effect of voxel size and TR by means of conjunction and exclusion analysis. As described above mean significance of reliability of brain activity was estimated using the \( t \)-test statistic for deviations from zero. The
presence of these $t$-maps opens the door for the commonly used conjunction analysis (Nichols et al., 2005). All vertices that showed a significance value above the Bonferroni corrected significance threshold ($t_{24} = 6.71, \alpha = 3.051 \times 10^{-7}$) were taken into account. The conjunction analysis was applied in a systematic and hierarchical way and grand conjunction maps were computed. First we assessed the effect of voxel size. For this analysis all data with voxel sizes of $2.5^3$ mm$^3$, $3.5^3$ mm$^3$ and $4.5^3$ mm$^3$ with a TR of 2400 ms were subjected to a conjunction analysis. This conjunction analysis revealed, which parts of the brain are reliable independent of the applied voxel size. In a next step we computed the conjunction of maps with a voxel size of $3.5^3$ mm$^3$ and $4.5^3$ mm$^3$ with a TR of 2400 ms. Finally we identified areas that were reliable exclusively with the $4.5^3$ mm$^3$ sequence. For the effect of TR a similar approach was followed. In this analysis all data with TR of 2400 ms, 1600 ms and 1250 ms and a voxel size of $4.5^3$ mm$^3$ were subjected to a conjunction analysis. In a next step we computed the conjunction of maps with a TR of 1600 ms and 1250 ms and a voxel size of $4.5^3$ mm$^3$. Finally the effect of the remaining sequences with a TR of 1250 ms was assessed.

### 2.12. Fixed and gliding GLM brain activity mask analyses

In order to validate the Fourier-based reliability measure we assessed how the Fourier derived ICC measures are related to brain activity. It has been shown that time courses of fMRI data are more reliable in areas activated by the paradigm than in areas that are not activated (Gorgolewski et al., 2013). We decided to investigate this property with two kinds of GLM brain activity masks. In a first step, we assessed the number of reliably detected vertices in relation to the sequence under study within a so called fixed GLM brain activity mask. In a second step, we wanted to provide external validation of our newly introduced Fourier derived ICC measure with a so called gliding GLM brain activity mask. We assumed that brain areas of higher brain activity go along with higher reliability of brain oscillations independent of the sequence in use.
For the fixed GLM brain activity mask we performed a classical conjunction analysis over all twelve GLMs obtained from the six sequences per session (Nichols et al., 2005). We used a rather liberal threshold of 5% uncorrected ($t_{24} = \pm 2.064$). For the activated area we took those vertices where all twelve GLMs showed activation above the positive threshold. For the deactivated area we took those vertices where all twelve GLMs showed deactivation below the negative threshold. All vertices with $t$-values between the positive and the negative threshold were assigned to the not (de-)activated area, respectively the area of non-systematic brain activation.

Three histograms were computed, which showed the distributions of mean ICC in areas of activation, deactivation and in the area that was neither activated nor deactivated consistently. Six more histograms were computed, which showed the distributions of standard deviations of the ICCs and the distributions of the significance of mean ICC expressed as $t$-values in the areas described above.

For the gliding GLM brain activity mask we performed a similar conjunction analysis over all twelve GLMs with a gliding $t$-value threshold ranging between 0.25 and 5 in steps of 0.25. Three areas were defined including areas of activation, areas of deactivation and areas of non-systematic brain activation. We computed plots that show the relation between the gliding $t$-statistics of the three areas and measures of reliability. The measures of reliability were obtained by averaging the ICC estimates within the confines of the gliding brain activity maps over all vertices. Furthermore plots of mean standard deviation and mean statistical significance of the ICC were obtained.
3. Results

3.1. Behavioural data (reaction time analyses)

Subjects performed the experiment with a mean accuracy of 94 percent (sd = 4.5 percent). One subject was excluded from the study because he did not complete the experiment. The ICC analysis of response behavior revealed a value of ICC = 0.87. The mean response times were 390.1 ms for session 1 and 376.6 ms for session 2. A paired t-test revealed that a significant statistical difference was present (p < 0.001).

3.2. Data plausibility check (GLM results)

The conventional second level GLM analysis showed that brain activation patterns related to the color condition and the direction condition were similar (Fig. 2A). There was an overlap in finger action related areas including parts of motor area 1 (M1), ventral premotor cortex, supplementary motor area (SMA) and intraparietal sulcus (IPS). In addition, we found an overlap in visual processing areas including primary visual cortex and in the posterior and frontal eye fields. The contrast color condition vs. direction condition revealed a large set of activated brain areas including anterior cingulate cortex (ACC), parietal cortex and dorso-lateral premotor cortex (Fig. 2B). All these areas have been linked to the Stroop effect in previous studies (Laird et al., 2005; Roberts et al., 2008; Nee et al., 2007).
Fig. 2. Second level GLM contrasts estimated from six EPI protocols measured in two sessions which were used for a data plausibility check. (A) Beta-weights from the 6 by 2 measurements were averaged and contrasted against the resting baseline. (B) Beta-weights from the 6 by 2 measurements were averaged and the Stroop contrast color condition vs. direction condition was computed.
3.3. FFT

The fMRI frequency spectra of both sessions for all scanning protocols as computed from the grand mean time-course across subjects revealed a consistent peak at the task frequency (0.017 Hz) and its second harmonic (0.034 Hz; Fig. 1B). These observations are in line with former investigations of task induced fMRI frequency spectra (Bandettini et al., 1993).

3.4. Test-retest reliability analysis

The general finding is that test-retest reliability of fMRI frequency spectra at the individual subject’s level was affected by voxel size and by TR. The prominent effect was an increase in reliability of fMRI oscillations when larger voxel sizes and faster TR were applied. Effects of varying voxel sizes were stronger than effects of varying TR.

A qualitative neuroanatomical assessment of the mean reliability of fMRI spectra maps (Fig. 3A) revealed the following picture. The fMRI protocol with the smallest voxel size (2.5 mm$^3$) led to reliable detection of fMRI frequency spectra in occipital brain areas including cuneus and precuneus. When protocols with the next larger voxel size (3.5 mm$^3$) were applied, reliable fMRI frequency spectra were also detected in parietal cortex and SMA. For protocols with the largest voxel size (4.5 mm$^3$) a considerable increment in detection of fMRI frequency spectra was found for medial as well as for lateral aspects of the cortex. Slight increments in detection of reliable BOLD oscillations were found for protocols with faster TR. Areas with high mean reproducibility of fMRI frequency spectra seemed to overlap areas with high standard deviation of mean reproducibility of fMRI frequency spectra (Fig. 3B). The increment in standard deviation is substantially larger for the step from 2.5 mm$^3$ to 3.5 mm$^3$ than for the step from 3.5 mm$^3$ to 4.5 mm$^3$. Application of faster TR seems to decrease the standard deviation in frontal brain areas and on the medial hemisphere.
Significance of mean reproducibility of fMRI frequency spectra increased when larger voxel sizes were applied (Fig. 4). For all applied voxel sizes brain areas with high mean reproducibility of fMRI frequency spectra showed high significance, despite the simultaneous increase in standard deviation. Furthermore, application of faster TR seemed to slightly increase observed significance.

The effects of voxel size and TR on the reproducibility of the fMRI frequency spectra were characterized by the following descriptive statistics. Distributions of mean ICC (Fig. 5A) revealed three subgroups of distributions organized...
around the three voxel sizes (2.5³ mm³ / 2400 ms: mean ICC = 0.206; 3.5³ mm³ / 2400 ms: mean ICC = 0.294; 4.5³ mm³ / 2400 ms: mean ICC = 0.367; 3.5³ mm³ / 1600 ms: mean ICC = 0.302; 4.5³ mm³ / 1600 ms: mean ICC = 0.372; 4.5³ mm³ / 1250 ms: mean ICC = 0.383). Within each subgroup a faster TR led to higher reliability. Distributions of the standard deviations showed that substantial individual differences in reproducibility existed (Fig. 5B). These differences were nearly independent of the applied protocol, although voxel size 2.5³ mm³ revealed smaller standard deviations. Finally, distributions of significance of mean ICC correlations expressed as t-values revealed a pattern similar to the distributions of mean ICC (Fig. 5C).
Fig. 5. (A) Distributions of mean reproducibility of fMRI frequency spectra: Each of the three subgroups represents one voxel size; within each subgroup EPI protocols with faster TR reveal higher reproducibility. (B) Distributions of mean standard deviation of reproducibility: Large variability within EPI protocols exists whereas differences between EPI protocols are rather small. (C) Distributions of significance of reproducibility expressed as $t$-values. In accordance with B and C, larger voxel sizes and faster TR reveal higher significance values.

The differences in reliability between the different sequences were tested for significance using paired $t$-tests. The large voxel size reliability contrast $4.5^3$ mm$^3$ vs. $2.5^3$ mm$^3$ revealed large brain areas excluding inferior-temporal and
insular areas ($p < 0.01$) (Fig. 6A). The more narrow reliability contrast $3.5^3 \text{ mm}^3$ vs. $2.5^3 \text{ mm}^3$ revealed frontal, parieto-occipital and temporal areas as well as medial aspects of cortex ($p < 0.01$) (Fig. 6B). The other more narrow reliability contrast $4.5^3 \text{ mm}^3$ vs. $3.5^3 \text{ mm}^3$ revealed mainly areas of deactivation, especially for the precuneus, which were small compared to the other contrasts ($p < 0.01$) (Fig. 6C). Reliability contrasts between the different TR did not reveal significant effects.

The absolute and relative number of vertices that exhibit significant reliability at increasing $t$-values substantially differed between the scanning protocols (Fig. 7). Both larger voxels (Fig. 7A) and faster TR (Fig. 7B) led to an improved detection of reliable vertices, when the significance level of reproducibility thresholds was stricter. The empirical distributions illustrating the voxel size effect (Fig. 7A, top) showed prominent differences, while this was less the case for the TR effect (Fig. 7B, top). The ratio curves related to the voxel size effect (Fig. 7A, bottom) started to diverge at $t$-value = 7 with a maximum divergence at $t$-value = 14. At a $t$-value = 14 the protocol with voxel size = $3.5^3 \text{ mm}^3$ detected approximately 20 times more vertices when compared to the protocol with voxel size = $2.5^3 \text{ mm}^3$. At the same $t$-value = 14 the protocol with voxel size = $4.5^3 \text{ mm}^3$ detected approximately 200 times more vertices when compared to the protocol with voxel size = $2.5^3 \text{ mm}^3$. The ratio curves related to the TR effect (Fig. 7B, bottom) started to diverge between $t$-value = 7 and $t$-value = 8 suggesting that this threshold is critical for both voxel size and TR effects. The ratio curves of the TR effect showed their maxima at $t$-value = 15. At this $t$-value the protocol with TR = 1600 ms detected 2.5 times more vertices when compared to the protocol with TR = 2400 ms while the 1250 ms protocol detected 3.5 times more vertices.
Fig. 6. Significant differences in reproducibility maps expressed as t-values estimated at the vertex level for the voxel size effect assessed at a TR of 2400 ms. (A) The large reliability contrast $4.5^3 \text{mm}^3$ vs. $2.5^3 \text{mm}^3$ revealed large brain areas, excluding parahippocampal and insular areas. (B) The more narrow reliability contrast $3.5^3 \text{mm}^3$ vs. $2.5^3 \text{mm}^3$ revealed frontal areas and parts of posterior cingulate cortex and parietal cortex as well as medial aspects of cortex. (C) The other more narrow reliability contrast revealed mainly areas of deactivation, especially ventral precuneus.

Fig. 7. Absolute and relative increase in the number of vertices showing significant reproducibility of fMRI frequency spectra. (A) Larger voxel sizes led to more significantly reproducible vertices, which is underlined by substantial increase factors computed from the reproducibility of the smallest voxel size as a reference value. (B) Faster TR led to a higher absolute number of significantly reproducible vertices, which is underlined by increase factors computed from the reproducibility of the slowest TR as a reference value.

3.5. Conjunction of reliability analysis

The grand conjunction over all voxel sizes at a TR = 2400 ms revealed reliable fMRI frequency spectra in parieto-occipital cortex (Fig. 8). Only for the right hemisphere we found reliable fMRI frequency spectra in activated motor areas covering parts of BA 4 and possibly BA 6 as well as activated somatosensory areas covering parts of BA 1. Both hemispheres revealed reliable fMRI frequency spectra in somatosensory cortex corresponding to BA 3 and BA 3a. Furthermore, reliable fMRI frequency spectra were found in visual cortex extending into the IPS and the posterior parietal lobe. Cortical areas of
deactivation were reliably traced, including precuneus, lateral parietal cortex and medial prefrontal cortex. Conjunction over voxel sizes $3.5^3 \text{ mm}^3$ and $4.5^3 \text{ mm}^3$ at a TR of 2400 ms revealed reliable fMRI frequency spectra in extended additional frontal areas extending from premotor cortex into frontal eye fields as well as function carrying systems in central sulcus and middle temporal areas. The largest voxel size $= 4.5^3 \text{ mm}^3$ at a TR of 2400 ms exclusively revealed reliable fMRI frequency spectra in SMA, pre-SMA and cingulate motor area (CMA). In addition, reliable fMRI frequency spectra were found in superior parietal cortex, anterior and posterior cingulate cortex and in insular cortex. Finally, we found no systematic voxel size effects on the reliability of fMRI frequency spectra in ventral parts of the cortex. Parahippocampal gyrus and distinct parts of insular cortex did not reveal increasing reliability due to larger voxel sizes.

In order to identify areas with systematic TR effects on the reliability of fMRI frequency spectra we performed a similar consistency analysis taking into account all protocols with a voxel size of $4.5^3 \text{ mm}^3$ (Fig. 9). We only found TR effects at very high $t$-values ($t = 10$) and large cortical areas did not reveal any systematic TR effect. The grand conjunction over all TR revealed reliable fMRI frequency spectra in the precuneus, lateral parietal cortex, and parts of lateral prefrontal cortex. The conjunction over 1600 ms and 1250 ms led to additional fMRI frequency spectra detection in ventral premotor cortex and inferior parietal cortex. The fastest TR of 1250 ms exclusively revealed reliable fMRI frequency spectra in SMA, M1 and CMA.
**Fig. 8.** Conjunction analysis maps investigating systematic voxel size effects on the reproducibility of fMRI frequency spectra.

**Fig. 9.** Conjunction analysis maps investigating systematic TR effects on the reproducibility of fMRI frequency spectra.
3.6. Fixed and gliding GLM brain activity mask analyses

The results reported in the previous section imply that an increase in reliability due to different scanning protocols is both a global and regional phenomenon. In this section we show that mean test-retest reliability of fMRI frequency spectra was higher in areas of task induced brain activation and deactivation. These areas were visualized by GLM derived brain activity masks (Fig. 10A, Fig. 11A). Both masks revealed regions which are commonly activated in focussed attention and working memory tasks, like IPS and the frontal eye fields (Fox et al., 2005). Due to finger movement, also motor related areas like parts of pre-motor cortex and SMA were activated. In contrast, both masks revealed regions of deactivation, like posterior cingulate cortex, precuneus, lateral parietal cortex and medial prefrontal cortex.

For the fixed brain activity mask nine histograms were computed (Fig. 10B-D). The distributions of mean ICC within the fixed brain activation mask showed a similar pattern as compared to the distributions of the mean ICC of the whole brain (Fig. 5A). Mean ICCs were higher in activated and deactivated areas than in areas of no systematic activity. Standard deviation and significance of reproducibility of fMRI frequency spectra appeared to be similar when compared to the whole brain although both seemed to be higher in areas of activation and deactivation but not in areas lacking any systematic activity.

The previously reported increase in mean ICC, in standard deviation and in the significance of reproducibility within the fixed brain activity mask was even more pronounced when the same measures were inspected within the gliding brain activity mask. In this case mean ICC, standard deviation and significance of reliability increased quite substantially, when the gliding activation and deactivation masks were set at more conservative $t$-thresholds (Fig. 11B-D). This was not the case in areas lacking any systematic activity. Large voxel sizes showed better reliability than small voxel sizes and fast TR was slightly better when compared to slow TR. The differences in reliability between voxel sizes remained well conserved for all three states of brain activity.

In order to elucidate explanatory power of the analysis procedure, we quantified the effects of voxel size and TR on the reproducibility of fMRI frequency
spectra. For this purpose we added all vertices in areas of activation, deactivation and in areas of no systematic activity (Fig. 11E). More vertices of the brain surface appeared to be deactivated than activated.
Fixed brain activity mask

A

B

distribution of mean ICC

C

sd of ICC estimate

D

significance of ICC estimate

- 2.5 x 2.5 x 2.5 mm³ / 2400 ms
- 3.5 x 3.5 x 3.5 mm³ / 1600 ms
- 4.5 x 4.5 x 4.5 mm³ / 1250 ms
- 3.5 x 3.5 x 3.5 mm³ / 2400 ms
- 4.5 x 4.5 x 4.5 mm³ / 1600 ms
- 4.5 x 4.5 x 4.5 mm³ / 2400 ms

3.6 x 3.5 x 3.5 mm³ / 2400 ms

- 2.064 ≤ t ≤ 2.064
- t ≤ 2.064
- t ≥ -2.064

df = 24
Fig. 10. (A) Fixed brain activity mask as obtained from a second level GLM conjunction analysis over all EPI protocols measured at two sessions and the related distributions of reliability estimates. (B) Distribution of the mean ICC estimates within the three masks obtained from the maps depicted in Fig. 3A. (C) Distribution of the standard deviation of the ICC estimates within the three masks obtained from the maps depicted in Fig. 3B. (D) Significance of ICC estimates within the three masks obtained from the maps depicted in Fig. 4.
**Fig. 11.** Gliding brain activity mask and corresponding plots of relation between gliding activation mask and measures of reproducibility. (A) Gliding brain activity mask as obtained from a conventional second level GLM conjunction analysis over all EPI protocols measured at two sessions. It revealed areas of activation and deactivation at eight statistical thresholds. The area of non-activation is not shown for visualization purposes. (B) Plots of the grand mean ICC estimates within the three masks obtained from the maps depicted in Fig. 3A. For the non-activated area ‘mask t-value’ refers to all t-values in a window of the depicted t-value and its negative equivalent (e.g. 1.25* refers to all t-values between -1.25 and 1.25). (C) Plots of the mean standard deviation of ICC estimates within the three masks obtained from the maps depicted in Fig. 3B. (D) Plots of mean significance of ICC estimates within the three masks obtained from the maps depicted in Fig. 4. (E) The number of vertices within the three masks. For the activated area ‘mask t-value’ refers to the actually depicted t-value; for the deactivated area ‘mask t-value’ refers to the negative equivalent of the depicted t-value. For the non-activated area ‘mask t-value’ refers to all t-values in a window of the depicted t-value and its negative equivalent.
4. Discussion

In this study we evaluated the test-retest reliability of fMRI frequency spectra by varying voxel size and TR in a sample of 25 male subjects. As a major finding there are prominent effects of voxel size on test-retest reliability of fMRI frequency spectra (Fig 3A). While these effects are mainly explainable by an increase in the temporal signal-to-noise ratio (TSNR) due to changes in voxel size (Triantafyllou et al., 2005, Ball et al., 2012), we evaluated the exact effects on the test-retest reliability of fMRI frequency spectra (Fig. 5A-C). Improvements due to increase in voxel size appear mainly in cortical areas activated by the paradigm and are therefore not physiological noise (Fig. 11B). These improvements were present in test-retest reliability and in statistical significance. There are, indeed, several task-relevant cortical areas with high statistical significance irrespective of the chosen voxel size (Fig. 8). These areas include task relevant areas such as the IPS and primary motor cortex as well as large parts of the default network including angular gyrus and precuneus. It should be mentioned that statistical significance tends to be high even in areas with poor test-retest reliability which was the case for 2.5³ mm³ voxels where ICC values rarely exceeded 0.3. These high resolution images are quite interesting from a psychological point of view but have to be applied with care. The direct contrasts of test-retest reliability between the three voxel sizes emphasized this point (Fig. 6A-C). Much more neural task-relevant information is gained when voxel size is increased. For this reason we advise to use large voxel sizes when the focus of research is on the cerebral cortex. One might argue that large voxels of 4.5³ mm³ show poor spatial resolution. But it should be mentioned that it is current practice to smooth 3³ mm³ voxels with a smoothing kernel of 6³ mm³ up to 9³ mm³ (Ball et al., 2012; Hagberg et al., 2012). This leads to much lower spatial resolution than assumed and, moreover, obscures effects of interest (Fransson et al., 2002, White et al., 2001). Within this context we would advise that alignment in cortex oriented studies with large voxel sizes is ideally performed with cortex based alignment.
that does not rely on spatial smoothing (Goebel et al., 2006). More reliable fMRI will be achieved with a non-smoothed analysis technique and especially in a clinical environment routine usage of smoothing procedures is critical (Geissler et al., 2005). In contrast, usage of large voxel sizes is certainly limited, thus TSNR reaches a plateau at voxel sizes larger than $4.5^3 \text{mm}^3$ (Triantafyllou et al., 2005). Finally, large voxels are of course not ideal for studies that focus on subcortical structures that might suffer from partial volume effects (Weibull et al., 2008).

The TR effect on test-retest reliability was less prominent than the voxel size effect but not less interesting. Imaging flip angles decreased when TR gets faster, which led to a slight increase in test-retest reliability (Fig. 5A). Fast TR goes along with a high number of observations which probably could account for this observation. As a confounding factor in our design one could mention that we did not keep constant TR to study the flip angle effect in isolation. But this would lead to an uneconomical experimental design and the advantages of flip angle reduction – a faster TR – would be diminished. In our case comparability of different sequences was assured by computing the relation of TR and flip angle by means of the Ernst equation (Ernst and Anderson, 1966; Helms et al., 2008; Ye et al., 2010). It has been shown that the reduction of imaging flip angle well below the Ernst angle results in negligible loss in TSNR (Gonzales-Castillo et al., 2011). This suggests that sequences with very fast TR might show fair reliability (average whole brain ICC > 0.4). Thus future studies at 3T might shed light on fMRI ultra-fast sequences. Recently one study showed that temporal ICA is feasible with TR of no less than 300 ms (Goebel et al., 2012). This is potentially very attractive because fast TR opens a window on the dynamically changing aspects of cognition. In addition fast TR allows research in higher frequency domains which are cognitively relevant (Salvador et al., 2005; Zuo et al., 2010a).

Our ICC measure takes into account the Fourier coefficients of the fMRI frequency spectrum. Fourier coefficients represent specific frequencies within the whole frequency spectrum and with averaging the Fourier coefficients one can compute frequency bands. Thus the Fourier coefficients ground modern
fMRI analysis techniques. We assessed reliability of the fMRI frequency spectrum as a whole to get one ICC value for every fMRI frequency spectrum. Our previous discussion suggests that this measure captures cognitively relevant information. Areas of high activation and deactivation show a relatively high test-retest reliability of task related oscillations (cf. Fig. 11). This indeed suggests that Fourier coefficient based test-retest reliability research is relevant for resting state research as well as for task related research. A major drawback of the Fourier coefficient based analysis method is that it cannot distinguish between activation and deactivation. Thus ideally Fourier coefficient analysis methods are combined with brain activity maps.

Neural oscillations are present during all states of mind and are a general property of brain function (Vincent et al., 2007). They can be assessed with fMRI and for test-retest reliability a main confounding factor is voxel size most likely due to its large impact on TSNR (Triantafyllou et al., 2005, Ball et al., 2012). When investigating the data on the individual level, there are subjects that exhibit higher test-retest reliability than other subjects. We only can speculate why these individual differences in test-retest reliability are present and one possible reason is that there are differences in individual TSNR. These differences on the individual level are independent of fMRI acquisition parameters. At this time it is by far not standard to assess physiological parameters like heart beat, respiration rate and changes in volume of respiration parallel to the fMRI scanning process. These factors are known to introduce irrelevant variation to the fMRI signal (Krüger and Glover, 2001; Birn et al., 2006). Our study also lacks application of these methods, although it was proposed that only 5 percent of the unwanted variance in the fMRI signal are explainable by heart beat, respiration rate and changes in volume of respiration. (Petridou et al., 2009). Further studies may concentrate on the relation between intra-individual differences in the fMRI signal and physiological parameters, while also other parameters like hemoglobin concentration should be addressed. This would shed light on how to maximize individual TSNR in order to improve individual test-retest reliability of fMRI. Furthermore, it has been shown that the quality of high resolution imaging at 3T is substantially improved
when larger numbers of observations are made (Murphy et al., 2007). But depending on the TSNR and the effect size of the experiment, high resolution fMRI studies might take more time than practically feasible. Future studies may concentrate on search of methods to improve TSNR of high resolution fMRI to gain reliable insight in the changing dynamics of the brain at high spatial resolution.
5. Conclusions

Test-retest reliability of temporally and spatially unfiltered fMRI frequency spectra increases as a function of voxel size and decreases as a function of TR, when twofold accelerated EPI/GRAPPA protocols are applied. This is also true when reliability measures are restricted to task activated or deactivated brain areas for all subjects. Higher test-retest-reliability tends to be found in regions either activated or deactivated independent of the scanning protocol used in the study. Increase in reliability found for larger voxel sizes goes along with increase in standard deviations of individual reliability measures which is primarily because of interindividual differences in TSNR. Test-retest reliability of partial parallel imaging in fMRI seems to be promising.
6. References


symbolic arithmetic in typically developing children aged between 6 and 12 years. Dev Neuropsychol 36, 721–740.


7. Summaries

7.1. English summary

Little is known about the effects of voxel size and repetition time (TR) on the test-retest reliability of fMRI frequency spectra. Here we assessed these effects by means of a new Fourier based intraclass correlation (ICC) measure at the individual subject level. Experimental data were obtained from a Stroop task in a sample of \( n=25 \) male subjects. We found large inter-individual differences in reliability independent of voxel size and TR, primarily because of inter-individual differences in temporal signal-to-noise ratio (TSNR). Whole brain ICC roughly increased linearly with voxel size (\( 2.5^3 \text{ mm}^3 = 0.2; 3.5^3 \text{ mm}^3 = 0.3; 4.5^3 \text{ mm}^3 = 0.4 \)). Additionally, we found a subtle but significant increase in reliability for faster TR. Furthermore, reliability of fMRI frequency spectra was higher in activated and deactivated brain areas for both voxel size and TR when compared to areas of no task induced activation. The increase in reliability found for larger voxel sizes is rather global. However, when a consistency analysis was applied, we identified task relevant brain areas in which fMRI frequency spectra were exclusively assessable with larger voxel sizes. Furthermore, as we applied a parallel imaging technique (GRAPPA), we could corroborate considerations which suggest parallel imaging as a reliable technology in the context of fMRI.
7.2. German summary

In der funktionellen Magnetresonanztomographie (fMRT) ist der Effekt von Änderungen der MRT-Aquisitonsparameter Voxelgröße und Bildwiederholungsrate auf die Test-Retestreliabilität von fMRT-Frequenzspektren nahezu unbekannt. In dieser Studie wurde dieser Effekt mittels eines eigens dafür entwickelten Analyseverfahrens erfasst, welches auf Intraklassen-Korrelation (ICC) beruht. Das angewendete Verfahren ermöglicht eine Reliabilitätsanalyse pro Individuum, welche anhand experimenteller Daten einer Stroop-Aufgabenstellung in einer Gruppe von $n=25$ männlichen Probanden durchgeführt wurde. Die Analyse ergab große interindividuelle Unterschiede der Test-Retestreliabilität unabhängig von Voxelgröße und Bildwiederholungsrate, in erster Linie aufgrund von interindividuellen Unterschieden des zeitlichen Signal-zu-Rausch Verhältnisses (TSNR). Über der gesamten Hirnoberfläche kam es zu einem annähernd linearen Anstieg des ICC mit Zunahme der Voxelgröße ($2.5^3$ mm$^3 = 0.2; 3.5^3$ mm$^3 = 0.3; 4.5^3$ mm$^3 = 0.4$). Zusätzlich kam es zu einem kleinen aber statistisch signifikanten Anstieg des ICC bei Verkürzung der Bildwiederholungsrate. Unabhängig von Voxelgröße und Bildwiederholungsrate lag zudem eine höhere Test-Retestreliabilität der fMRT-Frequenzspektren in aktivierten und deaktivierten Hirnarealen vor als in Hirnarealen, die nicht durch das Paradigma angesprochen wurden. Die Zunahme der Test-Retestreliabilität von fMRT-Frequenzspektren durch eine Vergrößerung der Voxelgröße ist eher globaler Natur und umfasst den gesamten Kortex. Mithilfe von Konsistenzanalysen konnten jedoch für das Paradigma relevante Hirnareale identifiziert werden, die ausschließlich mittels großer Voxelgrößen reliabel erfasst werden konnten. Da in dieser Studie eine parallele Bildgebungsmethode (GRAPPA) verwendet wurde, konnten zusätzlich Annahmen gestützt werden, dass parallele Bildgebungsmethoden auch in Bezug auf fMRT reliabel sind.
## 8. Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>anterior cingulate cortex</td>
</tr>
<tr>
<td>AC-PC</td>
<td>anterior commissure - posterior commissure line</td>
</tr>
<tr>
<td>BA</td>
<td>Brodmann area</td>
</tr>
<tr>
<td>BOLD</td>
<td>blood oxygenation level dependent</td>
</tr>
<tr>
<td>BMS</td>
<td>between-target mean square</td>
</tr>
<tr>
<td>CBA</td>
<td>cortex based alignment</td>
</tr>
<tr>
<td>CMA</td>
<td>cingulate motor area</td>
</tr>
<tr>
<td>EEG</td>
<td>electroencephalogram</td>
</tr>
<tr>
<td>EMS</td>
<td>within-target mean square</td>
</tr>
<tr>
<td>EPI</td>
<td>echo planar imaging</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional magnetic resonance imaging</td>
</tr>
<tr>
<td>FFT</td>
<td>fast Fourier transform</td>
</tr>
<tr>
<td>FOV</td>
<td>field of view</td>
</tr>
<tr>
<td>GLM</td>
<td>general linear model</td>
</tr>
<tr>
<td>GRAPPA</td>
<td>generalized autocalibrating partially parallel acquisitions</td>
</tr>
<tr>
<td>ICC</td>
<td>intraclass correlation coefficient</td>
</tr>
<tr>
<td>IPS</td>
<td>intraparietal sulcus</td>
</tr>
<tr>
<td>LFO</td>
<td>low frequency oscillations</td>
</tr>
<tr>
<td>M1</td>
<td>motor area 1</td>
</tr>
<tr>
<td>MPRAGE</td>
<td>magnetization-prepared rapid gradient-echo</td>
</tr>
<tr>
<td>MVPA</td>
<td>multi voxel pattern analysis</td>
</tr>
<tr>
<td>.mtc</td>
<td>mesh time course (file format)</td>
</tr>
<tr>
<td>PPI</td>
<td>partial parallel imaging</td>
</tr>
<tr>
<td>Pre-SMA</td>
<td>pre-supplementary motor area</td>
</tr>
<tr>
<td>RFX GLM</td>
<td>random-effects general linear model</td>
</tr>
<tr>
<td>SD</td>
<td>standard deviation</td>
</tr>
<tr>
<td>SMA</td>
<td>supplementary motor area</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>TE</td>
<td>echo time</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>TR</td>
<td>repetition time</td>
</tr>
<tr>
<td>TSNR</td>
<td>temporal signal-to-noise ratio</td>
</tr>
<tr>
<td>.vtc</td>
<td>vertex time course (file format)</td>
</tr>
<tr>
<td>2D-FLASH</td>
<td>2D fast low-angle shot</td>
</tr>
</tbody>
</table>
9. Appendix

9.1. List of figures

Figure 1. (p. 11): Time courses and fMRI frequency spectra obtained with six different EPI protocols including three voxel sizes and three TR for a ROI in superior parietal cortex.

Figure 2. (p. 17): Second level GLM contrasts estimated from six EPI protocols measured in two sessions which were used for a data plausibility check.

Figure 3. (p. 19): Mean reproducibility of power spectra maps obtained from ICC measures at the individual level and the corresponding mean standard deviation maps ($n = 25$).

Figure 4. (p. 20): Lateral and medial hemispheres depicting the significance of reproducibility expressed as t-values for the six EPI protocols obtained from ICC measures at the level of the subject ($n = 25$).

Figure 5. (p. 22): (A) Distributions of mean reproducibility of fMRI frequency spectra. (B) Distributions of mean standard deviation of reproducibility. (C) Distributions of significance of reproducibility expressed as t-values.

Figure 6. (p. 24): Significant differences in reproducibility maps expressed as t-values estimated at the vertex level for the voxel size effect assessed at a TR of 2400 ms.

Figure 7. (p. 25): Absolute and relative increase in the number of vertices showing significant reproducibility of fMRI frequency spectra.
Figure 8. (p. 27): Conjunction analysis maps investigating systematic voxel size effects on the reproducibility of fMRI frequency spectra.

Figure 9. (p. 27): Conjunction analysis maps investigating systematic TR effects on the reproducibility of fMRI frequency spectra.

Figure 10. (p. 30): Fixed brain activity mask as obtained from a second level GLM conjunction analysis over all EPI protocols measured at two sessions and the related distributions of reliability estimates.

Figure 11. (p. 32): Gliding brain activity mask and corresponding plots of relation between gliding activation mask and measures of reproducibility.
9.2. *List of tables*

**Table 1.** (p. 6): Different single-shot EPI protocols, including three spatial resolutions and three temporal resolutions.
9.3. „Erklärung § 5 Abs. 1 zur Datenaufbewahrung“

Erklärung § 5 Abs. 1 zur Datenaufbewahrung

Hiermit erkläre ich, dass die dieser Dissertation zu Grunde liegenden Originaldaten in der Brain Imaging Facility des IZKF der RWTH Aachen, Universitätsklinikum RWTH Aachen, Pauwelsstraße 30, 52074 Aachen

hinterlegt sind.
9.4. „Erklärung über den Eigenanteil“

Eidesstattliche Erklärung gemäß § 5 Abs. (1) und § 11 Abs. (3) 12. der Promotionsordnung


Die Durchführung der experimentellen Datenanalyse sowie statistische Auswertungen und fMRT-Datenverarbeitungsschritte, wie im Folgenden dargelegt:


Aktueller Stand der Patentanmeldung (02.10.2015): Recherche durch das DPMA nach § 43 Patentgesetz.

Des Weiteren von mir durchgeführt wurden die folgenden Arbeitsschritte:

Bei der Durchführung der Arbeit hatte ich folgende Hilfestellungen, die in der Danksagung angegeben sind:

A. Univ. Prof. Dr. Klaus Willmes-von Hinckeldey, LFG Neuropsychologie, RWTH Aachen: Studiendesign, Korrektur des Manuskriptes

B: Dr. Jan Willem Koten, LFG Neuropsychologie, RWTH Aachen: Studiendesign, Unterstützung bei der statistischen Analyse und bei der Programmierung, Korrektur des Manuskriptes

C. Bianca Molnar, Doktorandin, LFG Neuropsychologie, RWTH Aachen: Probandenbetreuung und Beaufsichtigung der experimentellen Messungen

D. Dipl. Ing. André Schüppen, IZKF, RWTH Aachen: Unterstützung bei der Programmierung mittels Matlab

E. Prof. Dr. rer. nat. Thoralf Niendorf, Max-Delbrück Center for Molecular Medicine, Berlin: Technische Unterstützung bezüglich MRT-Parallel-Imaging

F. Dr. nat. med. Mikhail Zvyagintsev, MD, BSc, IZKF, RWTH Aachen, Unterstützung bei der Erstellung des Manuskriptes

G: Dr. Jochen Weber, Columbia University, New York, USA: Programmierung der MRT-Koregistrierung von T1-MPRAGE und T1-2D-FLASH

Hannes Schwenke

Als Betreuer der obigen Dissertation bestätige ich die Angaben von Hannes Schwenke

Univ.-Prof. Dr. Klaus Willmes-von Hinckeldey
9.5. Danksagung

9.6. Curriculum vitae

Hannes Schwenke
Arzt
geboren am 04.08.1985 in Anklam

ÄRZTLICHE TÄTIGKEIT UND STUDIUM

Seit Juni 2014  Assistenzarzt am Institut für Neuroradiologie
               UK-SH Campus Lübeck
April 2013 – Mai 2014 Assistenzarzt an der Klinik für Neurologie
                         UK-SH Campus Lübeck
Dezember 2012  Approbation an der RWTH Aachen
Oktober 2006   Beginn des Studiums der Humanmedizin, RWTH Aachen

ZIVILDIENST

2005 - 2006   Klinik für Psychiatrie und Psychotherapie
              Universitätsklinikum Rostock

SCHULAUSBILDUNG

2005         Abitur, CJD Christophorus-Gymnasium Rostock

Lübeck, den 02.10.2015