Mapping the hand, foot and face representations in the primary motor cortex – Retest reliability of neuronavigated TMS versus functional MRI

Carolin Weiss a,c,⁎,1, Charlotte Nettekoven b,1, Anne K. Rehme b, Volker Neuschmelting a, Andrea Eisenbeis a, Roland Goldbrunner a, Christian Grefkes b,c

a University of Cologne, Department of Neurology, 50924 Cologne, Germany
b Max Planck Institute for Neurological Research, 50931 Cologne, Germany
c University of Cologne, Department of Neurosurgery, 50924 Cologne, Germany

⁎ Corresponding author at: Klinik für Allgemeine Neurochirurgie, Uniklinik Köln, Bettenhaus Eb. 11, Kerpener Str. 62, 50931 Cologne, Germany. Fax: +49 221 478 6025.
1 Both authors contributed equally (shared first-authorship).

Article history:
Accepted 19 October 2012
Available online 29 October 2012

Keywords:
Motor cortex
Test–retest
Reliability
Intraclass correlation
fMRI
TMS

Abstract

Introduction: Functional magnetic resonance imaging (fMRI) is a frequently used non-invasive mapping technique for investigating the human motor system. Recently, neuronavigated transcranial magnetic stimulation (nTMS) has been established as an alternative approach. We here compared the test–retest reliability of both mapping techniques with regard to the cortical representations of the hand, leg, face and tongue areas.

Methods: Ten healthy subjects were examined three times (intervals: 3–5 days/21–35 days) with fMRI and nTMS. Motor-evoked potentials were recorded from the abductor pollicis brevis, plantaris, mentalis and the tongue muscles. The same muscles were activated in an fMRI motor task. Euclidean distances (ED) between hotspots and centers of gravity (CoG) were computed for the respective somatotopic representations. Furthermore, spatial reliability was tested by intersession overlap volumes (OV) and voxel-wise intraclass correlations (ICC).

Results: Feasibility of fMRI was 100% for all body parts and sessions. In contrast, nTMS was feasible in all sessions and subjects only for the hand area, while mappings of the foot (90%), face (70%) and tongue representations (40%) remained incomplete in several subjects due to technical constraints and co-stimulation artifacts. On average, the mean ED of the hotspots was better for fMRI (6.2±1.1 mm) compared to nTMS (10.8±1.9 mm) while stability of CoG was similar for both methods. Peak voxel reliability (ICC) was high for both methods (>0.8), and there was no influence of inter-session intervals. In contrast, the reliability of mapping the spatial extent of the hand, foot, lips and tongue representations was poor to moderate for both fMRI and nTMS (OVs and ICC<50%). Especially nTMS mappings of the face and tongue areas yielded poor reliability estimates.

Conclusion: Both methods are highly reliable when mapping the core region of a given target muscle, especially for the hand representation area. In contrast, mapping the spatial extent of a cortical representation area was only little reliable for both nTMS and fMRI. In summary, fMRI was better suited when mapping motor representations of the head, while nTMS showed equal reliability for mapping the hand and foot representation areas. Hence, both methods may well complement each other.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Functional magnetic resonance imaging (fMRI) and navigated transcranial magnetic stimulation (nTMS) are currently the most frequently used techniques to investigate the motor cortex in vivo. Several studies dealing with motor processing such as motor cortex plasticity (Hermes et al., 2012; Simoes et al., 2012) and neurorehabilitation (Stinear, 2010; Wang et al., 2011), motor skill learning (Dayan and Cohen, 2011; Simões et al., 2012) and neurorehabilitation (Stinear, 2010; Hamzéi et al., 2012) and the anatomy of motor representation (Munzert et al., 2009; Stoddle and Schmahmann, 2010) have furthered our understanding of the neural correlates underlying motor action. Both fMRI and nTMS have also been applied for clinical purposes (Holody et al., 2011; Najib et al., 2011), e.g., in post-stroke rehabilitation (Grefkes and Fink, 2011; Schlaug et al., 2011; Stinear, 2010) and...
presurgical brain mapping (Najib et al., 2011; Ruohonen and Karhu, 2010; Tiedman et al., 2009; Wengenroth et al., 2011). High test–retest reliability is mandatory for achieving valid results. This is particularly true when brain mapping techniques are used in clinical applications with strong impact on neurological functions like, e.g., preoperative mapping of motor cortex representations to achieve total tumor resection (Gorlia et al., 2008; Lacroix et al., 2001) while preserving neurological functions (Duffau et al., 2005; Pouratian and Bookheimer, 2010).

However, both fMRI and TMS are only indirect measures of neuronal representations. Previous studies suggested that fMRI of the motor system has only a poor to moderate test–retest reliability. For example, intraclass correlation coefficients (ICC) for hand movements range from 0.37 (Kong et al., 2007) to 0.82 (Bosnell et al., 2008). Test–retest reliability also depends on the parameter that is taken into account. For example, while the peak voxel reliability of fMRI clusters seems to be high, spatial overlaps of fMRI activation clusters were reported to have poor reproducibility across sessions (approximately 30%, Bennett and Miller, 2010). Furthermore, reliability of fMRI motor mappings is also influenced by the body part that is mapped, yielding rather poor values for the cortical representations of the feet (0.21) and mouth (0.14) (Havel et al., 2006). Recently, neuronavigated transcranial magnetic stimulation (nTMS) has attracted rising attention as an alternative motor mapping method which is also readily applicable in the clinical routine: based on structural MRI scans and continuous detection of the position of the subject’s head, motor-evoked potentials (MEPs) at peripheral muscles are systematically induced upon precise nTMS of M1 (Basser, 1994; Basser and Roth, 2000; Ilmoniemi et al., 1999; Picht et al., 2009; Picht et al., 2011a, 2011b). Hence, nTMS mappings may localize anatomical targets more accurately than standard non-navigated procedures (Ahdab et al., 2010; Julkunen et al., 2009). Non-navigated TMS mapping studies reported inter-session reliability scores between 0.69 and 0.86 for the representation of different hand muscles (Malcolm et al., 2006), which corresponds to moderate to high test–retest reliability. Similar ICCs were reported for the “motor hotspot” of the anterior tibialis muscle (Cacchio et al., 2009; Cacchio et al., 2011). However, data on the reliability of neuronavigated TMS are scarce, particularly for mappings of different body parts. Furthermore, mappings of the face representation, which are of particular interest for clinical mappings in patients with lesions in the opercular region (e.g., tumors), have not been examined so far. Finally, data directly comparing reliability between nTMS and fMRI mappings are scarce as well.

Therefore, we designed a serial, combined fMRI–nTMS study to investigate the test–retest reliability of both approaches in a group of healthy subjects. A particular feature of our study was that we considered not only “standard” hand motor mappings but also included mappings of the foot, face and tongue representations. We compared inter-session differences in TMS hotspots or fMRI local maxima, center of mass, and RMT latencies within the following ranges for the respective group of muscles: 17–27 ms for APB, 36–50 ms for PM, and 7–15 ms for MM as well as for MEP recordings of the LT (Rodel et al., 1999, 2001, 2003; Saisanen et al., 2008). The order of stimulation (APB, PM, MM, LT) was randomized across subjects. Data were acquired from the dominant (i.e. left) hemisphere. Stimulation intensity was adjusted to 110% resting motor threshold (RMT) of the respective muscle. The RMT was defined as the minimum stimulus intensity capable of inducing MEPs greater than 50 μV peak-peak amplitude in at least 5 out of 10 consecutive trials upon stimulation of the motor ‘hotspot’ at rest (Julkunen et al., 2009; Rossini et al., 1994). The hotspot was defined as the cortical stimulation site at which coil positioning, orientation and tilt yielded the highest MEP amplitude.

Material and methods

Subjects and study design

We examined 10 healthy, right-handed subjects (5 males, mean age 28.3 years, range 24–34 years; handedness was verified according to the Edinburgh Handedness Inventory, Oldfield (1971)). All subjects were recruited from the data base of the Max Planck Institute for Neurological Research, Cologne, Germany. The study was according to the declaration of Helsinki (1969, last revision 2008) and approved by the local ethics committee. Written informed consent was obtained from all subjects. All subjects participated in three sessions: (i) session 1, (ii) session 2 (3–5 days after session 1), and (iii) session 3 (3–5 weeks after session 2). At each session, motor cortex mappings were performed by means of fMRI and nTMS in a randomized sequence for each subject.

nTMS technique

nTMS was conducted using the eXimia NBS system (version 3.2.2, Nexstim Ltd., Finland) and a figure-of-eight-shaped stimulation coil. The subjects were comfortably seated in an adjustable armchair with head-rest. For head tracking, we used the standard tracking unit which was fixed to the subjects’ front using an elastic strap and additional tapes. The head of the subject was coregistered with the corresponding high-resolution anatomical MRI which had been acquired from each subject before entering the study. Coregistration was achieved using three main anatomical landmarks (nasion and crus helicis of each ear) and nine additional points along the whole extent of the skull. This procedure yields a coregistration error of less than 2 mm (root mean square difference between the coregistered anatomical landmarks estimated by the neuronavigation software). Motor-evoked potentials (MEPs) were recorded using surface electrodes (Ambu Neuroline, Bad Nauheim, Germany) mounted above the abductor pollicis brevis muscle (APB), the plantaris muscle (PM), the mentalis muscle (MM) and the anterior lateral tongue muscles (LT). The PM is less frequently used in nTMS studies compared to, e.g., the tibialis anterior muscle (Cacchio et al., 2009, 2011). However, our fMRI pilot scans showed that toe flexions (engaging the PM) evoked less movement artifacts in head-motion sensitive regions (e.g., ventricles, CSF interface) compared to the ankle movements (engaging the tibialis anterior muscle).

To avoid false positive registrations of the nTMS-induced MEPs (e.g., due to direct stimulation of the facial nerve), we only accepted latencies within the following ranges for the respective group of muscles: 17–27 ms for APB, 36–50 ms for PM, and 7–15 ms for MM as well as for MEP recordings of the LT (Muellbacher et al., 2001; Rodel et al., 1999, 2001, 2003; Saisanen et al., 2008). The order of stimulation (APB, PM, MM, LT) was randomized across subjects. Data were acquired from the dominant (i.e. left) hemisphere. Stimulation intensity was adjusted to 110% resting motor threshold (RMT) of the respective muscle. The RMT was defined as the minimum stimulus intensity capable of inducing MEPs greater than 50 μV peak-to-peak amplitude in at least 5 out of 10 consecutive trials upon stimulation of the motor ‘hotspot’ at rest (Julkunen et al., 2009; Rossini et al., 1994): The hotspot was defined as the cortical stimulation site at which coil positioning, orientation and tilt yielded the highest MEP amplitude.

The anatomical position of the hotspot was established by performing a coarse mapping around the anatomical landmarks of the respective motor cortex representation area (Yousry et al., 1997; Penfield and Rasmussen, 1950): (i) The “hand knob formation” for ABP mapping, (ii) the cortex close to the interhemispheric fissure for mapping of the foot representation, and (iii) the frontal operculum for mapping the lips and tongue area. The coil orientation was kept stable (according to its orientation at the hotspot) during the whole mapping
procedure. For mapping of the hand area, the coil position was usually perpendicular to the course of the central sulcus. For mapping of the foot area, the coil position was usually perpendicular to the course of the interhemispheric fissure (medialateral current). For mapping of the face and tongue areas, the coil orientation was more variable but usually posterior–anterior.

For each muscle representation, 150–250 pulses (depending on the size of the respective representations) were applied using a stimulation grid projected onto the head representation and visualized by the neuronavigation software. The spacing between the grid nodes was 5 mm. Note that for all mappings, coil orientation was independent from the grid. Three pulses were applied per stimulation position. The outer margin of a given functional area was determined by two adjacent negative spots, i.e., the position where no MEP could be elicited in three consecutive stimulation trials.

FMRI acquisition

The fMRI experiment was performed on a 3 T Scanner (Trio, Siemens, Erlangen, Germany) using a block design. The fMRI task was designed to elicit activations associated with movements of the same muscles as used for nTMS (see NTMS technique). Hence, the subjects were asked to perform rhythmic, stereotypic movements with the thumb (horizontal abductions activating the APB), the foot (flexions of all toes activating the PM), the lips (pursing activating the MM) and the tongue (horizontal movements with slightly opened jaw activating intrinsic tongue muscles). The subjects were instructed by means of text cues (“hand”, “foot”, “lips”, and “tongue”) displayed on a video screen which was visible through a mirror attached to the MR head coil. After 1 s, the instruction text was replaced by an empty circle which started to blink in red color at a frequency of 1.5 Hz, thereby pacing the respective movements. After 20 s the blinking circle was replaced by a black screen indicating the end of the respective block of trials. The duration of this resting condition was 15 s. The order of conditions (movements of different body parts) was randomized. The whole fMRI experiment lasted approximately 11 min. Subjects were trained for correct task performance immediately before each session and advised to avoid any additional movements. Performance of the subjects in the scanner was controlled by video monitoring. Motor performance in terms of movement frequency was very stable across sessions in each subject due to the relative simplicity of the tasks. Furthermore, there were no unintended/mirror movements in any of the subjects during the resting baselines or in the task conditions.

We used a gradient-echo planar imaging (EPI) sequence sensitive to detect blood-oxygenation level dependent (BOLD) changes in tissue contrast using the following imaging parameters: repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, flip angle = 90°, voxel size 3.4 × 3.4 × 3.4 mm³, field of view (FOV) 220 × 220 mm, 32 slices (whole brain) and 320 volumes. High resolution T1 volumes were acquired using a magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence with the following parameters: TR = 2250 ms, TE = 3.93 ms, FOV = 256 mm, 176 sagittal slices, and voxel size = 1.0 × 1.0 × 1.0 mm³. In fMRI, the signal-to-noise ratio, and hence stability of the data, are also influenced by the acquisition plane (Gustard et al., 2001). In the present study, slice orientation was axial, slightly tilted backwards so that the orientation of the slices was approximately perpendicular to the course of the central sulcus.

FMRI preprocessing

The MRI volumes were processed using the Statistical Parametric Mapping software package (SPM 8; Wellcome Department of Imaging Neuroscience, London, UK, http://www.fil.ion.ucl.ac.uk) implemented in Matlab (version 2011, The MathWorks Inc., MA, USA). The first three images of a session were excluded from further analyses (“dummy images”). The remaining EPI of the three fMRI sessions were realigned to the first volume of the respective session by affine registration using a two-pass procedure, and finally registered to the first image of the first session. The realigned EPVs were then coregistered with the corresponding high resolution T1 volume. For group comparisons, all volumes were spatially normalized to MNI (Montreal Neurological Institute) space employing the unified segmentation approach (Ashburner and Friston, 2005). Finally, the EPI volumes were spatially smoothed using an isotropic Gaussian kernel of 8 mm full width at half maximum (FWHM). A smaller smoothing kernel might have been better for achieving a higher spatial resolution, but would have been at the cost of the signal-to-noise ratio (Turner et al., 1998; Webbula et al., 2008). We hence decided to use the default SPM8 parameters.

FMRI statistics

The experimental conditions were modeled using boxcar stimulus functions convolved with a canonical hemodynamic response function. The time series in each voxel were high-pass filtered at 1/128 Hz. A one-way General Linear Model (GLM) was applied for identification of significantly activated voxels. Head movement estimates were used as confound regressors. Note that the statistical analysis was performed twice, i.e., on the non-normalised EPVs (for computing differences in the coordinates of local maxima, centers of gravity and cluster overlaps; see below) and on the normalized dataset which was warped to MNI space (for computations of voxel-wise ICCs). For single subject analyses, voxels were considered significantly activated when passing a threshold of T = −4.59 (p < 0.05; family-wise error (FWE) corrected on the voxel level). As the objective of the current study was to analyze test–retest reliability of motor cortex mappings (and not the reliability of whole brain activation patterns), we restricted the fMRI analysis to the activation cluster yielding the highest BOLD activity in left precentral gyrus, thereby adopting a similar approach as in nTMS where mappings were restricted to the precentral region. Accordingly, we identified the fMRI peak activation cluster in the precentral gyrus in every subject and session (cluster forming threshold: p < 0.05, FWE corrected on the voxel level) and removed all other voxels from the respective SPM(T) map (Fig. 1).

For the group analysis, masked images underwent a full-factorial analysis of variance (ANOVA) using a random effects model (p < 0.05; FWE corrected on the voxel level) (Friston et al., 1999, 2002). Note that this masking procedure guaranteed that only activated voxels were considered for ICC computations, thereby preventing the inclusion of voxels without activity in any of the three sessions (for which no ICC can be computed due to zero in the denominator, see ICC formula, Section Statistical reliability measures).

NTMS data processing

MEPs with amplitudes > 50 μV and latencies within the limits as defined above (see NTMS technique) were used for further analyses. Mean MEP amplitudes were calculated for each stimulation node and session. A three-dimensional map of the averaged MEPs was reconstructed based on the individual T1 MNI coordinates (as provided by the neuronavigation software) for each stimulation session using a Matlab script and the SPM8 software package. As it is unknown at which position, superficial or deep, TMS-induced neuronal excitation occurs (Fox et al., 2004; Salinas et al., 2009; Thielscher and Wichmann, 2009), all TMS data were projected onto the cortical surface of the individual 3D head model as provided by the “peeling tool” of the neuronavigation software (Diekhoff et al., 2011). For group analyses, all individual MRI volumes (and coregistered MEP maps) were spatially normalized to MNI space using SPM8 (Ashburner and Friston, 2005).
Definition of the fMRI Region of Interest

Fig. 1. Definition of the fMRI region of interest. The activation cluster hosting the voxel with the highest effect size (“peak activation”) on the precentral gyrus was identified in the original SPM(T) map (1). Then, a binary mask was applied to isolate that cluster from the rest of the brain (2). This masked activation cluster (3) was used for computing centers of gravity, inter-session volume overlaps, and intra-class correlations. The picture on the right side shows the final rendering of an fMRI activation cluster on the left M1 (hand motor) region, the anatomical location with highest fMRI parameter estimates (“local maximum”; marked by “○”) and the location of the center of gravity (CoG; marked by “□”).

Statistical reliability measures

Euclidean distances

The between-session differences between the coordinates of the nTMS hotspots or the fMRI local activation maxima were evaluated by computing Euclidean distances in 3D single subject space using the following formula (Deza, 2009):

$$ED = \sqrt{(x_1-x_2)^2 + (y_1-y_2)^2 + (z_1-z_2)^2}$$

with $x_{1,2}, y_{1,2}, z_{1,2}$ representing the respective coordinates.

The “nTMS hotspots” were defined as the coordinates yielding the highest averaged MEP amplitudes (nTMS) while the “local activation maximum” represents the coordinate with the highest t-value (fMRI) in a given functional area (Fig. 1). We furthermore computed the CoG for each nTMS and fMRI map. While the nTMS hotspots or the fMRI peak voxels represent the site of maximal regional excitability/activity in a region, the CoG also considers the spatial distribution of excitability and activity in the respective motor map. The CoG (Fig. 1) was computed based on the averaged MEP amplitudes or BOLD-signal estimates weighted by the highest MEP amplitude or fMRI intensity obtained for the respective motor map (Diekhoff et al., 2011; Wassermann et al., 1992) using the following formula:

$$CoG = \left[ \frac{\sum_j a_j (x_j y_j)}{a_{j,\max}} \right]$$

with $a_j$ being the mean amplitude at position $x_j$ or $y_j$ (Classen et al., 1998a).

Spatial overlap volumes

For each subject and modality (nTMS and fMRI), the spatial overlap (percentage of voxels) of the first session with sessions 2 and/or 3 (overlap volumes, OV) was calculated based on binarized image volumes. All computations were performed in 3D single-subject space.

Intraclass correlation coefficients

For analyzing the reliability of the respective maps, voxel-wise ICCs (McGraw and Wong, 1996; McGraw KO, 1996; Shrout and Fleiss, 1979; Specht et al., 2003) were calculated from the spatially normalised TMS and fMRI maps using the following formula:

$$ICC = \frac{MSB-MSW}{MSB + (k-1)MSW}$$

Here, MSB represents the mean sum of squares between the subjects, MSW the mean sum of squares within subjects, and $k$ the number of sessions, i.e.,

$$MSB = \sum_{j=1}^{n} \sum_{i=1}^{k} \left( x_{ij} - \bar{x}_j \right)^2 / (n-1)$$

and

$$MSW = \sum_{i=1}^{n} \sum_{j=1}^{k} \left( x_{ij} - \bar{x}_j \right)^2 / (n(k-1))$$

The data used for computing the mean sums of squares were the MEP amplitudes or the fMRI t-values of each voxel, respectively.

The ICC ranges from 0 to 1. ICC values below 0.5 are usually considered to reflect poor reliability, 0.5–0.75 moderate reliability and ICC > 0.75 high reliability (Portney and Watkins, 2000).

Mixed long-term analysis for nTMS mappings of the lips and the tongue

For a number of subjects, nTMS data of the face and tongue area could not be assessed in all three sessions due to technical/biological constraints. This problem often occurred in session 1 or session 2 (i.e., the short-term comparisons). In order to increase the sample size (and hence statistical power) for the long-term reliability estimates, we considered the comparison “session 2 vs. session 3” in those subjects who were lacking data for the comparison “session 1 vs. session 3”. Note that the inter-session intervals between sessions 2 and 3 (3–5 weeks inter-session period) are very similar to those between sessions 1 and 3 (i.e. 3½–5½ weeks inter-session period). We will refer to this analysis as “mixed long-term analysis” or sessions “1/2–3” whenever used throughout the manuscript.

Statistical tests

Paired t-tests were used to test for significant differences between the variables of interest (PASW Statistics 18.0, SPSS Inc., Chicago, IL, USA). Correction for multiple comparisons was achieved by using the false-discovery-rate (FDR) approach (Benjamini and Hochberg, 1995).

Results

Feasibility

All subjects accomplished all three fMRI sessions without any problem. The SPM realignment procedure showed that the between-EPI displacements were < 1.1 mm in x, y, and z for each movement condition.
output intensity in one subject. MM mapping was feasible in 80% of the subjects, but in only 40% of the subjects in each and every session. Likewise, LT mapping was feasible in 90% of the subjects, but in only 70% of the subjects in all three sessions. Accordingly, short-term comparison could be obtained only from \( n = 4 \) for the lip region and \( n = 7 \) for the tongue area. In contrast, long-term results (spaced by an interval of 3–5½ weeks, either session 1 vs. session 3 or session 2 vs. session 3) could be obtained from \( n = 8 \) subjects for the lips representation and \( n = 9 \) for the tongue area (see Suppl. Figs. 1, 2, 5 and 7). The limited reproducibility of MM or LT mappings across nTMS sessions was caused by co-stimulation of jaw muscles (Guggisberg et al., 2001) and direct stimulation effects of facial nerve fibres passing through the stimulated frontolateral area, previously described as ‘S responses’ (Dubach et al., 2004). In these subjects, stimulation thresholds of facial nerve fibres were lower than those of the cortical representations, leading to false positive MEPs in target muscles as evident from very short latencies (in the range of 3–6 ms).

**Congruency between nTMS and fMRI mappings**

When directly comparing the hotspot coordinates of the nTMS mappings with the corresponding fMRI local maxima, we found an average “mismatch” (EDs) between 12.9 mm (thumb/APB, session 1) and 19.9 mm (tongue/LT, session 2; \( p < 0.05 \)). The EDs between nTMS CoGs and fMRI CoGs was smaller compared to the distance between nTMS hotspots and fMRI local maxima, and ranged between 10.5 mm (thumb/APB, session 2) and 19.37 mm (tongue/LT, session 2). There was no significant difference between the three sessions, neither for hotspots nor CoGs (Table 3). Paired t-test revealed that distances between nTMS and fMRI coordinates were smaller for the thumb/APB representation compared to the foot/PM representation (\( p < 0.05 \), FDR corrected) or the lip/MM representation (\( p = 0.1 \) FDR corrected). In summary, nTMS and fMRI mappings yielded similar but not identical information on the anatomical position of the “core region” of a given body part, especially with respect to the foot representation.

**Reliability of the nTMS hotspots and fMRI local activation maxima**

Averaged across body parts and sessions, the mean EDs were \( 10.8 \pm 1.9 \) mm for the nTMS hotspots and \( 6.2 \pm 1.1 \) mm for the fMRI activation maxima. Paired t-tests revealed that nTMS hotspots were significantly less reliable than the local fMRI activation maxima (\( p < 0.05 \)). Separate analyses of the respective body parts (all FDR corrected) revealed that this difference was significant for the tongue (LT; result significant for \( n = 7 \) [3 sessions] and for the mixed long-term analysis \( n = 9 \) [2 sessions]) and foot (PM) representation (\( p < 0.05 \)) while the thumb representation showed a statistical trend for smaller EDs in the fMRI sessions (\( p = 0.095 \), FDR corrected) (Fig. 2a). ICC values for the \( x \), \( y \) and \( z \)-coordinates showed a good reliability of the hotspot coordinates acquired by both fMRI (ICC = 0.879 ± 0.099) and nTMS measures (ICC = 0.892 ± 0.122) (Table 1). There was no effect of inter-session interval (1–2, 1–3, and 2–3) on EDs, neither for nTMS hotspots (Fig. 2c) nor for local fMRI maxima (Fig. 2b and Suppl. Figs. 1 and 3).

**Reliability of the nTMS CoGs and fMRI CoGs**

Mean EDs of the CoGs averaged across all body parts and sessions were \( 6.7 \pm 0.7 \) mm for nTMS and \( 6.8 \pm 2.0 \) mm for fMRI (no statistical difference, \( p = 0.69 \)). There were no significant differences between body parts (Fig. 3a). Averaged across all sessions and body parts, the reliability of CoG \( x \), \( y \), and \( z \)-coordinates assessed by the respective ICC values turned out to be high for both nTMS (0.95 ± 0.06) and fMRI (0.86 ± 0.17). There was no significant difference between both methods. For nTMS, CoG EDs were not significantly different across
sessions when considering each body part separately. This result was reproduced when considering the mixed long-term analysis (MM: n = 8; LT: n = 9; see Suppl. Figs. 2 and 4). Likewise, ICC values of nTMS and fMRI CoGs were similar for short-term and long-term intersession intervals (Figs. 3bc).

Spatial reliability — overlap volumes

The mean overlap volumes (OV) of the respective regions of interest (ROIs) between all three sessions were 32 ± 4% for fMRI and 24 ± 4% for nTMS (p = 0.21). In the subjects for which data could be assessed in all three sessions, the average spatial overlap between sessions 1 and 2 (“short-term”) was not significantly different compared to the average “long-term” (sessions 1–3) overlap for both approaches (fMRI: 45 ± 7% short-term vs. 46 ± 10% long-term; nTMS: 44 ± 6% short-term vs. 54 ± 6% long-term). The results of the mixed long-term analysis were within the same range (nTMS OV mixed long-term: 53% ± 6%). There was no difference between OVs for different body parts across inter-session intervals nor between different body part representation areas after correcting for multiple comparisons (Fig. 4; Suppl. Figs. 5 and 6). At uncorrected values (P < 0.05), we found higher spatial overlaps for the thumb and foot representations of the nTMS mappings compared to the fMRI intersection volumes. However, even those slightly higher overlap volumes remained below 50%. Therefore, in summary, test–retest reliability in terms of OVs has to be considered poor for both nTMS and fMRI irrespective of which area had been mapped (usually <50% for any body part).

Spatial reliability — intraclass correlation coefficients

The mean ICC for all three sessions computed over all voxels and regions yielded poor between-session reliability estimates for both nTMS (0.30 ± 0.09) and fMRI (0.48 ± 0.09) with average higher ICC values for fMRI (p < 0.05) (Table 2). The effect of time (“short-term”, “long-term”) was not significant for fMRI ( ICCs for sessions 1–2: 0.51 ± 0.06; sessions 1–3: 0.40 ± 0.09) while nTMS yielded significantly higher ICCs of the motor maps for the short-term comparison (0.32 ± 0.05) compared to the long-term interval (0.17 ± 0.07). When testing for differences between methods for different body parts we found significantly higher mean ICCs for fMRI mappings of the hand, lips and tongue compared to nTMS (p < 0.05, FDR corrected for all comparisons). This finding was confirmed when considering the mixed long-term analysis for the fMRI data. However, the fMRI and nTMS ICC maps also demonstrated that within a given representation area some voxels showed high ICC values (> 0.75). The anatomical location of these voxels was close to the fMRI local maxima and the nTMS hotspots. Therefore, although average ICCs indicated poor reliability for nTMS/fMRI clusters in terms of average voxel intensity (reflecting MEP amplitudes or fMRI parameter estimates), between-session reliability at the “inner core” of the somatotopic representation was high for both fMRI and nTMS (Fig. 5). Here, the number of voxels with ICCs > 0.75 was usually higher for fMRI, especially for the tongue area (233 voxels = 24.2% of the cluster) compared to nTMS (3 voxels = 5.7%). All these data were confirmed when comparing ICC values for the mixed long-term analysis (MM: n = 8; LT: n = 9; see Suppl. Fig. 4).

Discussion

We here compared the test–retest reliability of fMRI and nTMS with respect to motor cortex mappings of four different body parts at three different time points in a cohort of healthy subjects. The results show that both approaches reached high reliability values for localizing the “core” representations of a given body part. However, nTMS was often

<table>
<thead>
<tr>
<th>Spatial direction</th>
<th>Motor area (muscle)</th>
<th>Hotspot values</th>
<th>CoG values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>fMRI</td>
<td>nTMS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ICC</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>Thumb (APB)</td>
<td>0.724 ± 0.059</td>
<td>0.823 ± 0.092</td>
</tr>
<tr>
<td></td>
<td>Foot (PM)</td>
<td>0.716 ± 0.056</td>
<td>0.914 ± 0.079</td>
</tr>
<tr>
<td></td>
<td>Lips (MM)</td>
<td>0.858 ± 0.078</td>
<td>0.900 ± 0.096</td>
</tr>
<tr>
<td></td>
<td>Tongue (LT)</td>
<td>0.762 ± 0.082</td>
<td>0.350 ± 0.094</td>
</tr>
<tr>
<td>Y</td>
<td>Thumb (APB)</td>
<td>0.957 ± 0.092</td>
<td>0.889 ± 0.094</td>
</tr>
<tr>
<td></td>
<td>Foot (PM)</td>
<td>0.831 ± 0.092</td>
<td>0.890 ± 0.098</td>
</tr>
<tr>
<td></td>
<td>Lips (MM)</td>
<td>0.968 ± 0.096</td>
<td>0.903 ± 0.094</td>
</tr>
<tr>
<td></td>
<td>Tongue (LT)</td>
<td>0.893 ± 0.084</td>
<td>0.876 ± 0.097</td>
</tr>
<tr>
<td>Z</td>
<td>Thumb (APB)</td>
<td>0.957 ± 0.086</td>
<td>0.969 ± 0.089</td>
</tr>
<tr>
<td></td>
<td>Foot (PM)</td>
<td>0.966 ± 0.066</td>
<td>0.937 ± 0.086</td>
</tr>
<tr>
<td></td>
<td>Lips (MM)</td>
<td>0.971 ± 0.067</td>
<td>0.976 ± 0.088</td>
</tr>
<tr>
<td></td>
<td>Tongue (LT)</td>
<td>0.923 ± 0.076</td>
<td>0.920 ± 0.087</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.879 ± 0.082</td>
<td>0.862 ± 0.095</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>0.099 ± 0.068</td>
<td>0.122 ± 0.066</td>
</tr>
</tbody>
</table>

Table 1

Intraclass correlation coefficients of fMRI and nTMS hotspots and CoGs. fMRI ICC values were similar for hotspots and CoGs; by contrast, CoGs of nTMS measures were more reliable than the respective hotspots. The reliability values of fMRI and nTMS hotspot coordinates were within a similar range; however, depth (z) was rather unsteady upon nTMS sessions.

Table 2

Between-session intraclass correlation coefficients of fMRI and nTMS. The intraclass correlation coefficients (ICCs) of each voxel were calculated for each mapping method separately and include all three sessions. A significant difference towards higher reproducibility of the hand, the lips and the tongue areas was observed for fMRI as compared to nTMS experiments (marked with asterisks, P < 0.05, FDR corrected).

<table>
<thead>
<tr>
<th>Body part (muscle)</th>
<th>Hotspot/local activation maximum</th>
<th>ED (CoGs) Mean</th>
<th>ED (hotspot/local activation maximum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lips (MM)</td>
<td>Session 1</td>
<td>15.96 ± 5.61</td>
<td>18.87 ± 2.90</td>
</tr>
<tr>
<td></td>
<td>Session 2</td>
<td>15.16 ± 1.69</td>
<td>15.73 ± 3.30</td>
</tr>
<tr>
<td></td>
<td>Session 3</td>
<td>12.41 ± 1.69</td>
<td>18.67 ± 2.84</td>
</tr>
<tr>
<td></td>
<td>Mean ± SEM</td>
<td>14.51 ± 4.08</td>
<td>17.76 ± 1.85</td>
</tr>
<tr>
<td>Feet (PM)</td>
<td>Session 1</td>
<td>15.04 ± 3.07</td>
<td>17.59 ± 3.13</td>
</tr>
<tr>
<td></td>
<td>Session 2</td>
<td>13.90 ± 2.50</td>
<td>18.08 ± 2.88</td>
</tr>
<tr>
<td></td>
<td>Session 3</td>
<td>17.19 ± 2.19</td>
<td>18.74 ± 2.20</td>
</tr>
<tr>
<td></td>
<td>Mean ± SEM</td>
<td>15.37 ± 2.55</td>
<td>18.14 ± 2.66</td>
</tr>
<tr>
<td>Thumb (APB)</td>
<td>Session 1</td>
<td>10.84 ± 1.06</td>
<td>12.86 ± 1.35</td>
</tr>
<tr>
<td></td>
<td>Session 2</td>
<td>10.51 ± 1.22</td>
<td>12.98 ± 2.32</td>
</tr>
<tr>
<td></td>
<td>Session 3</td>
<td>11.35 ± 2.52</td>
<td>14.74 ± 2.24</td>
</tr>
<tr>
<td></td>
<td>Mean ± SEM</td>
<td>10.9 ± 1.67</td>
<td>13.53 ± 1.97</td>
</tr>
<tr>
<td>Tongue (LT)</td>
<td>Session 1</td>
<td>17.78 ± 3.37</td>
<td>16.41 ± 1.43</td>
</tr>
<tr>
<td></td>
<td>Session 2</td>
<td>19.37 ± 3.02</td>
<td>19.97 ± 1.80</td>
</tr>
<tr>
<td></td>
<td>Session 3</td>
<td>13.53 ± 2.00</td>
<td>19.11 ± 3.90</td>
</tr>
<tr>
<td></td>
<td>Mean ± SEM</td>
<td>16.89 ± 3.08</td>
<td>18.5 ± 3.11</td>
</tr>
</tbody>
</table>

Table 3

Euclidean distances between local fMRI activation maxima/nTMS hotspots and CoGs computed for different target muscles across methods. Centers of gravity: the Euclidean 3D distances between the nTMS hotspots and fMRI local maxima of the thumb/APB were significantly smaller than those of the foot/PM representation (p < 0.05, FDR corrected).
not feasible when mapping the motor representation of the face and tongue. Furthermore, mapping the spatial extent in terms of overlap volumes or voxel-wise ICCs was not very reliable neither in fMRI nor in nTMS. Interestingly, TMS hotspot and fMRI maxima were robust across short-term and long-term inter-session intervals. However, the fMRI activation maxima were significantly more reliable compared to the nTMS hotspots. Likewise, fMRI was superior in mapping the tongue and face representations.

Feasibility and technical aspects

FMRI — feasibility and signal variance

FMRI is a widely used, non-invasive method to indirectly localize neural activity associated with motor, sensory or cognitive tasks with a millimetre resolution. However, major disadvantages of this technique are its poor signal-to-noise ratio, and the rather low temporal resolution resulting from the delayed hemodynamic response following a neuronal event (Logothetis, 2000). Furthermore, factors like thermal noise and systemic noise of the MRI scanner as well as physiological noise caused by movements of the brain (e.g., due to cardiac or respiratory cycles) or the entire head (e.g., task-related head movements) influence the fMRI signal (Kruger and Glover, 2001). In the current study, head movements inside the scanner did not exceed 1.1 mm between scans, hence reaching excellent values in our group of healthy subjects. However, patient populations may show stronger head movements, thereby reducing signal quality and hence reliability (Seto et al., 2001). The signal-to-noise ratio, and hence stability of the data, are also influenced by the acquisition plane. In the present study, we used an “oblique axial” orientation, i.e., perpendicular to the central sulcus. This acquisition plane has been shown to yield high signal detection accuracy at 3 T (Gustard et al., 2001).

In addition to these technical factors, neuropsychological factors such as attention, alertness and novelty of stimuli are potentially relevant for the reliability of fMRI analyses (Peyron et al., 1999; Poldrack et al., 1999; Rostami et al., 2009; Sterr et al., 2007). We sought to control for learning effects by using a relatively simple motor task which was extensively trained prior to each fMRI session, enabling stable task performance between sessions. However, we cannot exclude that slight variations in other factors like movement acceleration or force and non-motor factors such as vigilance might have influenced inter-session variability and, hence, test–retest reliability of the fMRI data.

nTMS — feasibility and variance of MEPs

In contrast to fMRI, nTMS has an excellent temporal resolution within the range of a few milliseconds. The spatial resolution of the electric field induced by TMS strongly depends on anatomical factors and tissue conductivities (Fox et al., 2004; Thielacker and Wichmann, 2009). Furthermore, the reliability of nTMS measures is influenced by the registration mismatch (in this study <2 mm), orientation and tilt of the stimulation coil (Mills et al., 1992; Miranda et al., 1997; Pascual-Leone et al., 1994). Importantly, realistic head models such as tissue-segmented MR images with realistic anatomical features like gyri and sulci demonstrated that the electric field forms a complex pattern onto the folded cerebral cortex with secondary fields of different magnitudes occurring especially at tissue boundaries (Amanian et al., 1998; Salinas et al., 2009). Therefore, variations in MEP responses across sessions as observed in the present study might also result from slight differences in coil orientation during stimulation. This putative confound is inherent to TMS as the complex...
Fig. 5. Between-session intraclass correlation coefficient (ICC) maps of fMRI and nTMS experiments. Intraclass correlation coefficients (ICCs) calculated for each voxel of the respective functional fMRI and nTMS experiments and spatially normalized to a standard brain (Brain 152, MNI, Canada). The heat map illustrates the ICC of the respective fMRI t-value or nTMS MEP amplitudes (ICC = 0.0: dark violet–ICC = 1.0: red). For further illustration, the upper right window of each composite figure shows only the voxels amounting to highly reliable ICC values greater than 0.75. On average, fMRI yielded higher ICC scores compared to nTMS. Especially ICCs for the tongue and lip representations were significantly smaller for nTMS compared to fMRI.
relationship between anatomy, electric field generation, and tissue excitation makes it currently impossible to stimulate each grid position with its optimal coil orientation in an acceptable period of time. However, while absolute MEP values strongly depend on coil orientation, relative measures like the CoG position and even the hotspot position remain relatively stable when repeating the mapping with a different but constant coil orientation (Diekhoff et al., 2011).

In principle, another source of the inter-session variance might be slight shifts of the brain in relation to the skull, depending on the position of the subject's head. However, head positions were quite similar for each of the "nTMS mapping sessions owing to a neck rest in a standard position. Likewise, the position of the head was nearly identical in all fMRI sessions. Furthermore, in young subjects the anatomical structures attached to the brain (arachnoidal adhesions, falx and tentorium) leave only little room for brain shifts in a closed skull. Therefore, brain shifts between sessions represent — if at all — only a minor source of variance that might have influenced test–retest reliability in the current study.

**Difficulties in nTMS mapping of regions in the M1 head representation**

Difficulties in the TMS mappings of the cortical face and tongue representations arose from directly stimulating peripheral facial nerve fibers situated underneath the TMS coil (Dubach et al., 2004) and the unpleasant stimulation of jaw muscles (e.g. temporal muscle) mediated by trigeminal affereces (Guggisberg et al., 2001). As in some subjects the thresholds for eliciting short latency MEPs were below the cortical stimulation thresholds (Dubach et al., 2004), hotspots and RMTs could not be determined in all cases. Thus, the number of subjects with all three MM and LT mappings completed was low. The additional mixed long-term analysis with a larger sample size including those subjects in which at least two mappings were completed MM/LT mappings did not reveal better reliability values compared to the smaller sample size. Therefore, our data suggest that nTMS mappings of the face and the tongue area generally have rather low test–retest reliability, at least for test–retest intervals longer than 3 weeks.

**Mismatch between nTMS and fMRI coordinates**

The Euclidean distances between fMRI and nTMS coordinates (hotspots/local maxima and CoGs) found in the present study are within a similar range to what has been published before (Classen et al., 1998b; Diekhoff et al., 2011; Herwig et al., 2002; Lotze et al., 2003). For example, Diekhoff et al. (2011) reported a difference between APB hotspots and the local fMRI maxima for a thumb abduction task of on average 10.5±3.0 mm which is close to the spatial difference found in the present study (13.5±1.9 mm). Such differences might result from the fMRI motor task (which usually activates several groups of muscles) and the type of fMRI sequence (Diekhoff et al., 2011). Furthermore, intrinsic properties of the TMS mapping technique also contribute to differences between methods, i.e., poor specificity in stimulation depth for nTMS as well as complex inhomogeneities of the electric field induced by the TMS pulse (Salinas et al., 2009) and direct stimulation effects on axons in the white matter (Fox et al., 2004). The higher consistency of fMRI and nTMS CoGs results compared to local fMRI activation maxima / nTMS hotspots in our study, once again, leads to the suggestion that averaging increases the stability of nTMS and fMRI mappings.

**Reliability of hotspots and CoGs**

The hotspots and local activation maxima were highly reliable across all three sessions for both fMRI and nTMS. When directly compared to each other, fMRI yielded significantly smaller between-session variance than the nTMS experiments. Compared to a previous publication that reported mean hotspot deviations of 8.9±4.6 mm for three mapping sessions of the extensor digitorum communis muscle performed within 7–14 days (Wolf et al., 2004), our findings showed slightly higher inter-session variance for mappings of the APB hotspots (deviation of 10.3±1.5 mm). However, the ICC obtained in the present study for both the nTMS hotspot coordinates and fMRI local maxima reached high scores (ICC > 0.75). Reliability of the CoGs in terms of mean inter-session ED was similar for nTMS and fMRI. The high CoG reliability found for nTMS is in line with results published for non-navigated TMS (4.1±0.4 mm and 6.8±0.3 mm) (Uy et al., 2002; Wolf et al., 2004). Likewise, CoG results of the fMRI hand motor task were similar to what have been reported in previous publications (2±3 mm to 4±2 mm; Alkadhi et al., 2002; Marshall et al., 2004; Mattay et al., 1996). Also the between-session differences of the fMRI CoGs obtained for mapping the tongue representation resemble those reported in the literature (7.6±1.6 mm (Alkadhi et al., 2002)).

We found a higher stability of nTMS CoGs compared to nTMS hotspot coordinates. The higher variability of the hotspot coordinate might result from the fact that the identification of the hotspot is based on a considerably lower number of MEPS than those used for computing the CoG (which considers all MEPS acquired for a given region, i.e., 150–250). Averaging seems to be especially important for TMS as a rather low reliability has been reported for single MEP amplitudes with ICC values from 0.01 to 0.34 (Jung et al., 2010; Livingston and Ingersoll, 2008). The variability of single MEPS is strongly influenced by trial-to-trial fluctuations in, e.g., alertness (Gerloff et al., 1998; Ziemann et al., 1996) or involuntary precontraction of the muscle (Hess, 1996; Ortu et al., 2008). Differences in fMRI CoGs might result from compliance and training effects and from artificial yet inconstant activation caused by major blood vessels (Luh et al., 2000). Moreover, the reliability of fMRI CoGs strongly depends on the statistical threshold (Fesl et al., 2008). In this study, we used a “conservative” threshold of p<0.05, family-wise error corrected on the voxel level. Overall, ICC calculations suggested that the reproducibility of the CoGs is high for both nTMS and fMRI (ICC > 0.75). In contrast to our initial hypothesis (Bennett and Miller, 2010), the reliability of hotspots and CoG coordinates was not influenced by inter-session intervals, respectively.

In summary, we found nearly equal between-session reproducibility for fMRI local activation maxima and corresponding CoGs. In contrast, reliability of nTMS CoGs was significantly better as compared to hotspots. In practical terms, nTMS hotspots may be easier to assess but should be interpreted with caution, especially when concerning functional areas of rather low reliability such as the lip or the tongue region. Although nTMS CoGs are more reliable, they may provide misleading information about motor representations since the CoGs do not necessarily coincide with the regions of highest motor responses (e.g., in case of a “banana-shaped” motor representation, the CoG might even lie within a region of almost no response). Thus, CoGs cannot replace hotspots or mapping the entire extent of a region of interest in cases where precise localizing of representation fields is required, for example, in presurgical mappings of patients with tumors affecting the M1 region.

**Spatial reliability — overlap volumes**

The spatial extent of the fMRI motor maps showed only little overlap across all three sessions (32±4%), and slightly better when considering only two sessions (45±4%). For fMRI, other groups reported between-session intersection volumes of on average 36% (Bennett and Miller, 2010) with results ranging from 23% (Havel et al., 2006) to 74% (Yetkin et al., 1996). In the present study, we found no systematic difference in inter-session overlaps between fMRI and nTMS. Even a more detailed analysis of the different body parts did not reveal significant differences between the methods. One factor influencing total overlap volume reported in previous studies (e.g. Yetkin et al.: 1 h; Havel et al.: 6 days) is the inter-session interval, which overall was not confirmed in the present study. However, most of the studies
published before have considered only two instead of three (or more) sessions which may have led to better overlap values. Indeed, our results demonstrated that the mean overlap of two sessions was usually 30–50% higher (fMRI: 45 ± 7% short-term vs. 46 ± 10% long-term; p = 0.51; nTMS: 44 ± 6% short-term vs. 54% ± 6% long-term; p = 0.08) than the respective value of the triple session overlap. Of note, the spatial reliability of overlap volumes depends on the mapping resolution and acceptance thresholds. Hence, greater overlap volumes might have been detected with different neuroimaging or TMS parameters compared to the “standard” settings of the current study. Furthermore, nTMS overlap volumes methodologically differ from fMRI maps since nTMS intersection volumes predominantly reflect variations of the surface distribution while fMRI intersections reflect “true” 3D overlaps. However, in the present study, overlap volumes were usually below 50% irrespective of the mapping technique. Therefore, spatial reliability in terms of the size of a given motor map is low for both fMRI and nTMS.

Spatial reliability – voxel-wise intraclass correlation coefficients

Similar to the results of the overlap volumes, average voxel-wise ICCs revealed only poor stability of TMS/fMRI clusters (ICC: 0.30/0.45) with fMRI showing an overall better between-session stability (Table 2). The ICC data computed for the fMRI hand area mapping (0.44 for all three session, 0.57 for sessions 1 and 2) are comparable to values reported in the literature which all range between 0.37 and 0.82 (Bosnell et al., 2008; Friedman et al., 2008; Gountouna et al., 2010; Kong et al., 2007). However, in the current study, each ICC map also revealed a sub-cluster of voxels with excellent reproducibility (ICC>0.75), especially for the fMRI tongue representation (Fig. 5). This finding is somewhat surprising as we expected tongue movement to be associated with more head motion artifacts (Gracco et al., 2005) which would have increased the variance and decreased the ICC. In contrast, especially nTMS ICCs of the tongue and face areas were poor for both short-term and long-term comparisons, probably owing to the technical/biological problems (direct stimulation effects, discomfort) when mapping this region with TMS. Even for the larger sample of the mixed long-term reliability analysis, ICC values remained poor for face and tongue nTMS mappings. These data suggest that nTMS is not the method of choice when aiming at mapping these regions.

In summary, given the recommendation that reliability values should exceed 0.90 to ensure sufficient validity of functional cortex mapping (Portney and Watkins, 2000), both fMRI and nTMS seem to be less suited for precisely delineating the spatial extent of a functional region.

Conclusion

Test–retest reliability of fMRI and nTMS depends on a number of factors. Despite the fact that we aimed at minimizing confounding effects like inter-session differences in coil position, coregistration errors or fluctuation in vigilance, it appears to be impossible to completely exclude their influences on the results. However, this is exactly what we intended to assess in the current study, i.e., the reliability of the data given a standard TMS and fMRI set-up (with all their imminent sources of variations influencing reliability). When considering peak voxel activity and location, both fMRI and nTMS can be considered highly reliable, especially for the hand representation area. Here, the factor time – at least in the range covered by the present study – had no relevant influence on reliability. Nevertheless, longer inter-session periods (e.g., one year) might lead to reduced retest reliability. In contrast to the rather good reproducibility of hotspot and CoG coordinates, the reliability of mapping the spatial extent of a given functional region was moderate to poor for both nTMS and fMRI. Higher reliability estimates might have been achieved with longer nTMS/fMRI acquisition times enabling better signal averages. However, increasing fMRI scanning time may lead to fatigue and discomfort, thereby increasing signal variance. Likewise, longer nTMS sessions with more MEPs per node will lead to even longer acquisition times than the 1.5–2 h needed in the present study, which might have further decreased subject vigilance and hence signal stability. When directly comparing nTMS and fMRI, the nTMS hotspots were less stable than the local fMRI activation maxima. Moreover, nTMS was clearly inferior in mappings of the face and tongue area due to co-stimulation of facial muscles and nerves, leading to false positive MEPs in several subjects. Likewise, ICC values of long-term intervals >3 weeks were poor for face and tongue mappings, even when considering a larger sample size. Thus, nTMS mapping results acquired for these motor regions should be interpreted with caution. A limitation to the generalizability of our results is the relatively small sample size, especially with respect of the nTMS mappings of the face and the tongue which were often not feasible in all subjects and sessions. However, given the fact that the nTMS reliability values were already relatively poor when considering two sessions only (n = 8/9 subjects), it appears rather unlikely that increasing the sample size would lead to higher reliability values for a three-session interval due to the biologically and technically immanent constraints that nTMS has for mapping the cortical representations of the head.

Overall, this study shows that fMRI and TMS have very good reliability for localizing the core region of motor areas. It remains to be elucidated, whether this finding also applies to patients with brain lesions. In a clinical setting, fMRI might be less useful in brain tumor patients with motor impairments, reduced alertness or perception and potential changes in neurovascular coupling. Thus, nTMS might be superior for such special occasions which should be clarified in future studies.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2012.10.046.

Acknowledgments

We like to thank the MR staff of the Max Planck Institute for Neurological Research for technical support. Moreover, we are grateful to Michaela Wirtz from the Department of Neuroradiology, University of Cologne, and Thomas N. Witz, neurologist from the Department of Neurosurgery, University of Munich. C.G. was supported by a grant of the German Research Foundation (GR 3285/2-1).

References
