Imaging Neuroscience
an open access journal

Research Article

Test-retest reliability of the human connectome: An OPM-MEG study

Lukas Rier[a], Sebastian Michelmann[b], Harrison Ritz[b], Vishal Shah[c], Ryan M. Hill[a,d], James Osborne[c], Cody Doyle[c], Niall Holmes[a,d], Richard Bowtell[a], Matthew J. Brookes[a,d], Kenneth A. Norman[b], Uri Hasson[b], Jonathan D. Cohen[b], Elena Boto[a,d]

[a]Sir Peter Mansfield Imaging Centre, School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, United Kingdom
[b]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, United States
[c]QuSpin Inc., Louisville, CO, United States
[d]Cerca Magnetics Limited, Nottingham, United Kingdom

Corresponding Author: Lukas Rier (lukas.rier@nottingham.ac.uk)

ABSTRACT

Magnetoencephalography with optically pumped magnetometers (OPM-MEG) offers a new way to record electrophysiological brain function, with significant advantages over conventional MEG, including adaptability to head shape/size, free movement during scanning, increased signal amplitude, and no reliance on cryogenics. However, OPM-MEG remains in its infancy, with significant questions to be answered regarding the optimal system design. Here, we present an open-source dataset acquired using a newly constructed OPM-MEG system with a triaxial sensor design, 168 channels, OPM-optimised magnetic shielding, and active background field control. We measure the test-retest reliability of the human connectome, which was computed using amplitude envelope correlation to measure whole-brain (parcellated) functional connectivity, in 10 individuals while they watch a 600 s move clip. Our results show high repeatability between experimental runs at the group level, with a correlation coefficient of 0.81 in the θ, 0.93 in α, and 0.94 in β frequency ranges. At the individual subject level, we found marked differences between individuals, but high within-subject robustness (correlations of 0.56 ± 0.25, 0.72 ± 0.15, and 0.78 ± 0.13 in α, θ, and β respectively). These results compare well to previous findings using conventional MEG and show that OPM-MEG is a viable way to robustly characterise connectivity.

Keywords: optically pumped magnetometers, magnetoencephalography, functional connectivity, reliability, OPM-MEG

1. INTRODUCTION

Magnetoencephalography using optically-pumped magnetometers (OPM-MEG) is an emerging technique to image human brain function (see Brookes et al. (2022) for a review). As with conventional MEG, electrophysiological activity is assessed non-invasively by measuring magnetic fields at the scalp surface generated by neural currents (Cohen, 1968). However, unlike conventional MEG which employs arrays of cryogenically cooled sensors (Cohen, 1972; Hamalainen et al., 1993), OPM-MEG uses small and lightweight detectors—OPMs—which do not require cooling. Cryogenic temperatures place significant restrictions on conventional MEG system design, requiring large and cumbersome instrumentation with sensors fixed rigidly in a one-size-fits-all helmet. OPMs lift these restrictions leading to several advantages, for example, sensors can be positioned closer to the head, increasing signal amplitude and (theoretically) spatial resolution (Nugent et al., 2022); lightweight sensors can be mounted in a wearable helmet, enabling free subject movement during data acquisition; freedom to place sensors anywhere means OPM-MEG can, in principle, adapt to head size, enabling lifespan compliance; and systems are relatively simple to build, site, and operate.

Received: 10 July 2023 Revision: 31 August 2023 Accepted: 1 September 2023 Available Online: 13 September 2023

The capability of OPMs to measure the MEG signal has been shown extensively, for example (Boto et al., 2017; Johnson et al., 2010; Kamada et al., 2015; Sander et al., 2012; Xia et al., 2006), and OPM arrays have been developed which can image brain function accurately and, in some cases, with whole-head coverage (Feys et al., 2022; Hill et al., 2020; Iivanainen et al., 2019; Johnson et al., 2013; Nardelli et al., 2020; Seymour et al., 2021). Improved data quality has been shown in both theory (Boto et al., 2016; Iivanainen et al., 2017) and practice (Boto et al., 2017), including during subject movement (e.g. Boto et al., 2018), though recording during active movement critically depends on background field control (Holmes et al., 2018, 2020; Rea et al., 2021). Applications in children are also beginning to emerge (Hill et al., 2019) with exciting clinical possibilities (Feys et al., 2022). In sum, OPM-MEG systems offer new opportunities which are not possible using conventional neuroimaging. However, OPM-MEG remains nascent technology—there are only a small number of systems worldwide and a few have been tested for robustness. The best ways to design OPMs, sensor arrays, and magnetic shielding are not yet settled, and there is relatively little open-source data available from OPM-MEG systems. In this paper, we aimed to evaluate a recently developed triaxial OPM-MEG instrument (Boto et al., 2022; Rea et al., 2022; Rhodes et al., 2023) via quantitative assessment of test-retest reliability for the measurement of human connectomics. We further intended to generate an open-source dataset to allow other researchers to assess OPM-MEG capabilities.

Our system employs triaxial OPMs which allow independent measurement of the magnetic field along three orthogonal axes (Shah et al., 2020). Despite a slightly higher noise floor compared to conventional (single or dual axis) OPMs, triaxial sensors are an effective means to interrogate the MEG signal (Boto et al., 2022). They also allow increased total signal strength (i.e. each sensor makes three measurements of field) (Brookes et al., 2021; Rea et al., 2022), improved ability to differentiate brain activity from background fields (Brookes et al., 2021; Rea et al., 2022; Tierney et al., 2022), more uniform coverage in infants (Boto et al., 2022), and optimised calibration. In addition to the triaxial array, the system includes magnetic shielding which operates in active feedback configuration (Rea et al., 2021). This means that both low-frequency drifts in the background field and the static (i.e. time-invariant) magnetic field inside a magnetically shielded enclosure (MSE) are suppressed (Holmes et al., 2020), so data are collected in close to "zero" field (Rea et al., 2021).

Over the last two decades, functional connectivity has emerged as an important means to characterise brain health. Data from functional magnetic resonance imaging (fMRI) and MEG have shown that even with the brain "at rest", spatially separate but functionally related regions communicate to form networks. Some networks are associated with sensory processes, others with attention or cognition. These networks are key to healthy brain function and are often perturbed in neurological and psychiatric disorders. MEG offers multiple measures of connectivity (O'Neill, Barratt, et al., 2015) and therefore provides a tool to better understand the neural substrates that underlie communication in the brain (Sadaghiani et al., 2022). In addition, the exquisite time resolution of MEG allows us to look for dynamic changes in network connectivity, at the scale of seconds (O'Neill, Bauer, et al., 2015) and milliseconds (Baker et al., 2014). Consequently, the accurate and reliable measurement of network connectivity plays a critical role for any MEG system. However, connectivity measurement is also a challenge: the distributed nature of networks requires whole-head coverage and since mathematical techniques to characterise connectivity (particularly in the resting state) must be applied to unaveraged data, high-fidelity recordings are paramount.

Functional connectivity has been measured previously using OPM-MEG, during tasks and in the resting state (Boto et al., 2021), with results comparable to a conventional MEG system. However, this was with an early whole-head instrument (50 radial channels) and test-retest robustness was not assessed. Even with conventional MEG, the test-retest reliability of connectivity is challenging, for example, Colclough et al. (2016) showed that while at the group level (~30 subjects) repeatability of connectome estimation is excellent (>95%, based on amplitude envelope correlation), at the individual level reproducibility is closer to 60% (within-subject), and this drops further (<50%) for between-subject comparisons. Liuzzi et al. (2017) showed not only within-subject test-retest correlations of just ~58% using conventional MEG, but also that longer MEG recordings (10 mins relative to 5 mins) and immobilising the head to prevent movement relative to the sensor array significantly improved consistency, to >70%. The extension of such metrics to OPM-MEG would be a significant step forward.

In this paper, we characterise the robustness of connectomics using OPM-MEG. To maximise the chances of high reliability, we used 10-min recordings and, to provide a degree of consistency in brain activity, participants watched a movie clip during the scan. We chose a movie-viewing

paradigm that has been used previously in fMRI, EEG, and electrocorticography (ECoG) (Haufe et al., 2018). This standard task facilitates our objective and provides a new open-source resource with direct equivalence to existing data (Haufe et al., 2018). We quantitatively assess consistency between separate experimental runs and provide a benchmark for the reliability of connectivity measurement using OPM-MEG.

2. MATERIALS AND METHODS

2.1. Subjects and experimental paradigms

Ten participants gave written informed consent to participate in the experiment, which was approved by the University of Nottingham Medical School Research Ethics Committee (Of the 10 subjects, 4 identified as female, 6 as male, all right-handed; the age range of the subjects was 31 $\pm$ 8 (mean $\pm$ standard deviation across subjects.)). Each participant was scanned twice. During both recordings, participants watched the same 600 s clip of the movie "Dog Day Afternoon". The clip selected, which shows the scene of a bank robbery, was identical to that used in previous papers (Haufe et al., 2018; Honey et al., 2012; Lumet, 1975). Subjects remained seated and they were asked to watch the movie; they were free to move though not explicitly encouraged to do so. Subjects continued wearing the sensor helmet between scans (so that a single co-registration of sensor geometry to brain anatomy could be used for both measurements, reducing co-registration error). The gap between the two runs was 1-2 mins. Before the MEG recording, a field-mapping and nulling procedure (Rea et al., 2021) was carried out to control the background magnetic field (see below). MRI scans (acquired using a Phillips Ingenia 3 T MRI system running an MPRAGE sequence, with 1-mm isotropic resolution) were also acquired for all participants.

2.2. The OPM-MEG system

We used an OPM-MEG system averaging 168 channels, constructed from triaxial OPMs, each yielding three independent channels per sensor (Boto et al., 2022; Shah et al., 2020) (QuSpin, Inc. Colorado, USA). Sensors have a noise floor of ~13 fT/sqrt (Hz) and a bandwidth of ~150 Hz. The sensors were spaced evenly around the scalp and mounted in a 3D-printed lightweight helmet (Cerca Magnetics Ltd., Nottingham, UK), affording approximately whole-head coverage. The helmets came in multiple sizes and the best-fitting helmet was chosen for each participant. Outputs of all channels were recorded via a digital data acquisition (DAQ) system (National Instruments, Austin, TX, USA).

Participants were seated on a patient support inside an OPM-optimised magnetically shielded room (MSR) (Cerca Magnetics Ltd., Nottingham, UK). The MSR comprised four layers of mu-metal and one layer of copper, and was equipped with degaussing coils (Altarev et al., 2014) to reduce the magnetisation of the mu-metal layers. The static background field at the centre of this room following degaussing of the inner-most layer is typically ~3 nT. To further control the field, an array of four (QuSpin, first generation) reference OPMs was placed immediately behind the subject to measure background field fluctuations, and a set of biplanar electromagnetic coils were placed on either side of the participant, which enabled the generation of all three uniform fields and five independent linear gradients in a 40 x 40 x 40 cm^3 region enclosing the subject's head. The room also housed a motion tracking system comprising six cameras (OptiTrack Flex 13, NaturalPoint Inc., Corvallis, OR, USA) placed around the MSR, which recorded the movement of infrared-retroreflective markers attached to the bi-planar coils (as a static reference) and the sensor helmet (to monitor head movement).

The OPM sensors, DAQ, storage, and background field compensation were controlled via a single ("acquisition") PC. A second ("stimulus") PC controlled the movie and motion tracking. The visual display was achieved via projection through a waveguide onto a back-projection screen. We used a View Sonic PX748-4K projector positioned outside the MSR, and the screen was placed ~80 cm in front of the subject. The movie was presented at a visual angle of ~13 degrees horizontally and ~9 degrees vertically. Audio was presented through a set of speakers mounted outside the MSR and connected to a waveguide via a plastic tube. A schematic of the full system is shown in Figure 1A; a photograph of a participant wearing the system is shown in Figure 1B.

2.3. Magnetic field control

We used the field-control techniques originally described by Rea et al. (2021). Briefly, after positioning participants in the MSR, the MSR door was closed and the inner mu-metal layer was degaussed. The reference array was used to sample background field fluctuations, and the data were fed back to the (calibrated) bi-planar coils which generated an equal and opposite field. In this way, slow (<3 Hz) changes in the three uniform components of the field and the three gradients varying in z (from the

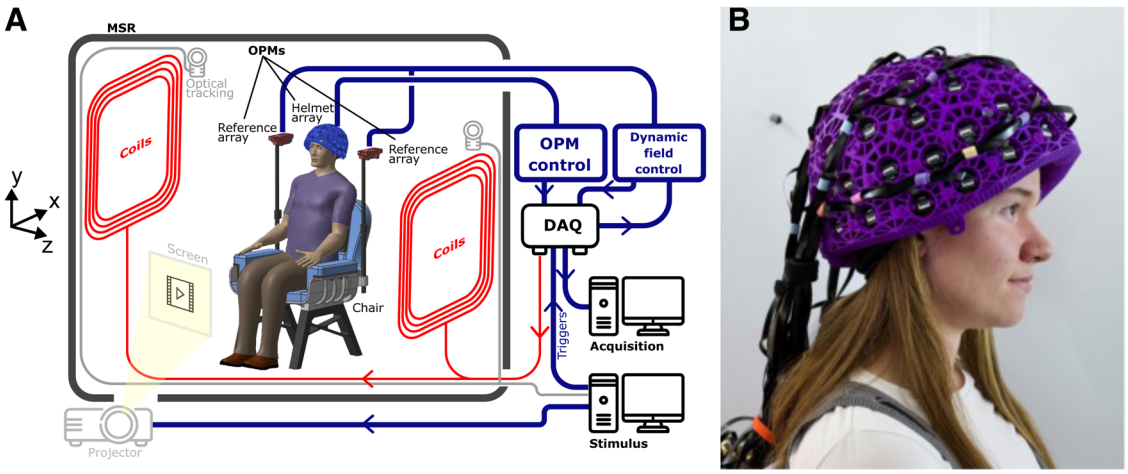


Fig. 1. the OPM-MEG system. (A) Schematic adapted from Rea et al. (2022) showing the OPM-MEG system setup. (B) A photograph of a participant wearing the OPM-MEG system.

participant's right to left) were stabilised (Holmes et al., 2020). This left only a static (i.e. temporally invariant) background field which was measured via a nulling procedure in which participants executed translations and rotations of their heads. The motion of the helmet was tracked for 60 s and 5 OPMs (15 channels) were used to sample the changes in magnetic field induced by the movement. These data were combined, and the background field was modelled using spherical harmonics. The calculated three homogeneous field components and five linear gradients were then compensated using the bi-planar coils. This nulling process was repeated twice (to iteratively improve the estimate) and the modelling was repeated a third time to estimate the 2magnitude of the background field in which the experimental MEG data were captured.

2.4. Data collection and co-registration

A total of 600 s of OPM data were recorded for each participant and each run of the experiment. All OPM channels were sampled at 16-bit resolution with a sampling rate of 1200 Hz. At least once per scanning day, a 90-s measurement with no subject present—termed "empty-room noise data"—was also acquired to verify that the system was working. This meant that in total, seven empty room recordings were also available for analysis (see below) alongside the MEG data.

Immediately following MEG data acquisition, two 3D digitisations of the participants' heads were acquired using an optical imaging system (Einscan H, SHINING 3D, Hangzhou, China)—the first with the helmet on and a second with the helmet removed and a swimming cap used to flatten hair. A 3D surface representing the face and scalp was also extracted from the anatomical MRI. These data were used to enable co-registration of the MEG sensor geometry to brain anatomy. Briefly, the two optical digitisations were segmented, leaving only points around the face, which were then aligned. The second optical digitisation (with the helmet removed) was then aligned to the surface extracted from the MRI. These two steps enabled knowledge of the helmet position relative to the brain. The locations and orientations of the OPMs, relative to the helmet, were known from the 3D printing process and the addition of this information enabled complete co-registration (see also Zetter et al. (2019) and Hill et al. (2020)). This was used subsequently to facilitate forward modelling of the magnetic fields generated by current dipoles in the brain.

2.5. Data analysis

2.5.1. Pre-processing and artefact correction

OPM-MEG data for each experiment (and the corresponding empty noise recordings) were notch-filtered at the mains frequency (50 Hz) using a 2nd-order infinite impulse response filter (Q-factor of 35 at -3 dB), and band-pass filtered (1-150 Hz) using a 4th-order, zero-phase-shift Butterworth filter. The filtered channel-level data and their power spectra were inspected visually for noisy and/or failed channels. We removed a channel if 1) its output variance was close to zero, indicating it was broken; 2) its noise level was obviously high across the sensor bandwidth upon visual inspection (experience showed such

noisy channels were easy to spot). On average, 152 ± 3 clean channels were included in the final analyses. We note that many of the OPMs used in this study were early "handmade" triaxial prototypes, and consequently not as reliable as more recently manufactured sensors. This accounts for the relatively high channel rejection rate compared to what we might hope for in conventional MEG.

Each experimental recording was divided into 5-s epochs which were characterised as "good" or "bad": epochs were inspected visually and trials containing visible motion or muscle artefacts were marked as bad. Additionally, an automatic thresholding procedure was used to remove trials containing large artefacts: specifically, the standard deviation of the 1-150 Hz data within each epoch was calculated independently for each channel. Epochs containing more than one channel with a standard deviation exceeding 3 standard deviations from the mean (calculated over all time) were marked as "bad". On average, 17 ± 5 bad trials (18 ± 4 in run 1, 17 ± 5 in run 2) were removed, resulting in an average of 513 ± 24 s of clean data (mean ± std. deviation across recordings) per run. Independent component analysis (ICA) (FieldTrip implementation—Oostenveld et al., 2011) was used to identify and remove ocular and cardiac artefacts: the data were decomposed into a number of components equal to the channel count and visual inspection of component time courses used to identify the artefacts. Finally, homogeneous field correction (HFC) (Tierney et al., 2021) was applied to attenuate interference from distal sources of magnetic field.

2.6. Source reconstruction

A beamformer (Vrba and Robinson, 2001) was used for source reconstruction. The brain was parcellated into 78 cortical regions, defined by the Automated Anatomical Labelling (AAL) atlas (Gong et al., 2009; Hillebrand et al., 2016; Tzourio-Mazoyer et al., 2002). This was achieved by co-registering the AAL atlas to individual brain space using FLIRT in FSL (Jenkinson and Smith, 2001; Jenkinson et al., 2002). The coordinates of the centre of mass of each AAL region were determined and forward fields for each resulting location were calculated. The forward calculation was implemented using a dipole approximation and a single shell volume conductor model, based on a head shape extracted from the anatomical MRI using FieldTrip (Nolte, 2003). Source reconstruction was repeated using data covariance based on broadband data (1-150 Hz) and six bands of interest (BoIs) encompassing the canonical θ (4-8 Hz), α (8-12 Hz), and β band (13-30 Hz), as well as three ranges within the γ band (γ_1: 30-40 Hz, γ_2: 35-45 Hz, γ_3: 40-48 Hz). Preprocessed data were band-pass filtered to each BoI using a 4th-order, zero-phase-shift Butterworth filter and covariance matrices constructed using data recorded throughout the whole experiment. Covariance matrices were regularised using the Tikhonov method by adding 5% of the maximum singular value of the unregularised matrix to all elements along the leading diagonal. The forward fields and data covariance were used to calculate beamformer weighting parameters, where source orientation was determined as the direction of maximum beamformer projected signal amplitude (Sekihara et al., 2004). Multiplication of the weighting parameters with the data resulted in 7 electrophysiological time series (one for each frequency band) at each of the 78 regions defined by the AAL atlas. This was repeated for every subject and independently for each experimental run.

2.7. Spectral power

To visualise the spectral content across AAL regions, and to examine the consistency of the beamformer projected signals between the two experimental recordings, we performed two analyses. First, we took the broadband (1-150 Hz) beamformer projected data, normalised by its standard deviation, and filtered to each BoI (using a 4th-order, zero-phase-shift Butterworth filter). The variance of the filtered data thus offered an estimate of the relative contribution of each BoI to the signal in a specific region. Applying this to all BoIs and regions allowed us to construct maps showing the spatial signature of the relative contribution of each band to the total signal for each AAL region. Secondly, for each region, we took the broadband beamformer projected data and used Welch's method to estimate the power spectral density (PSD). We also applied the same beamformer weights to project the empty room noise data. This enabled visualisation of not only the consistency of the PSD across recordings but also the relative contribution of empty room noise. We estimated the fractional difference in spectral power between runs as the square root of the sum of squared differences between PSDs, for runs one and two, normalised by the total integral of the overall mean PSD.

2.7.1. Functional connectivity

Functional connectivity between all pairs of AAL regions and for each BoI was calculated using amplitude envelope correlation (AEC) (Brookes et al., 2011; O'Neill, Barratt,

et al., 2015). The narrow-band beamformer projected data were taken for two regions, and pairwise orthogonalisation was applied to reduce the effect of source leakage (Brookes et al., 2012; Hipp et al., 2012). Following orthogonalisation, a Hilbert transform was applied to the data from each region and the analytic signals were calculated. The absolute value of the analytic signals was then used to determine the "Hilbert Envelope" (i.e. the instantaneous amplitude envelope of band-limited oscillations for each region). These envelopes were down-sampled temporally from 1200 to 120 Hz and the Pearson correlation coefficient between the envelopes was used to quantify functional connectivity. This procedure was applied to all ($78^2 - 78 =$) 3003 region pairs within the AAL parcellation, resulting in a whole-brain (parcellated) connectome. The analysis was run independently for each experimental run, participant, and BoI.

To visualise the connectome matrices, they were normalised by dividing each matrix element by the square root of the mean of all squared matrix elements and averaged across subjects (preventing a single subject with high connectivity values from dominating the group average). This produced a group mean connectome for the first and second experimental runs, and each BoI, separately. The matrices were plotted, and in addition, thresholded to keep only the 150 strongest connections which were plotted as lines within a glass brain. We also assessed average global connectivity (i.e. the mean across matrix elements, before normalisation) and the mean paired difference in global connectivity between runs.

We quantified the reliability of the group-average connectomes by calculating the Pearson correlation coefficient (using only matrix elements above the leading diagonal) between the subject averages for the two runs (separately for each BoI). We also assessed the influence of group size: for sample sizes of $N = \{2, 3, \ldots 9\}$, all possible combinations of subjects were drawn, and average connectomes calculated. We then re-measured the between-run correlation. By plotting the mean and standard deviation of these correlations for each N, we were able to estimate the trajectory of consistency with increasing N.

2.7.2. Inter-individual differences

In addition to group analyses, we examined connectivity at the individual level and the sensitivity of our OPM-MEG system to differences between participants. With 10 subjects, each scanned twice, there are 100 independent comparisons between run 1 and run 2 that can be made at the individual level; 10 within-subject comparisons; and 90 between-subject comparisons. For every possible comparison, we measured the Pearson correlation between vectorised matrices (again using only elements above the leading diagonal). We analysed these in two ways. First, we averaged the within- and between-subject correlations, computed the difference in the mean, and tested to see if this difference was significant using a Monte-Carlo test. Specifically, we randomly switched which 10 values were chosen as the within-subject correlations; doing this for 100,000 iterations enabled the construction of an empirical null distribution and allowed us to estimate whether the real difference could have occurred by chance. Second, we performed a "neural fingerprinting" analysis. For every subject, there is one within-subject comparison and nine between-subject comparisons—one might expect that the correlation coefficient for the within-subject comparison should be higher than the other nine values. If it is, that subject can be said to be successfully identified. By repeating this 10 times, we were able to assess how many (out of 10) subjects could be correctly identified based on their MEG connectome data.

All data presented here have been made publicly available (Rier et al., 2022), enabling free access to OPM-MEG data for the neuroimaging community—a core aim of the current study. The ability of our system to capture an accurate reflection of a known magnetic field was also tested—see Appendix 1.

3. RESULTS

Our field modelling showed that—following degaussing of the MSR walls and the application of average coil currents—the magnitude of the uniform magnetic field components inside the MSR was 0.54 ± 0.33 nT, with linear gradients of 1.70 ± 0.75 nT/m. These values dropped to 0.19 ± 0.17 nT and 0.63 ± 0.69 nT/m for the second field mapping. Comparable conditions were achieved previously (Rea et al., 2021).

3.1. Power spectral density

Figure 2A shows the spatial signature of spectral power in different bands of interest during the task. As expected, α oscillations dominate the signal in occipital areas, with high contributions stretching forward to the parietal lobes. β oscillations were highest in sensorimotor regions. Θ oscillations were approximately uniform across the whole head while γ_1 oscillations were most prominent in the frontal areas.

Figure 2B shows example power spectral density plots for six selected AAL regions—left and right superior frontal, postcentral, and calcarine cortex. In all cases, the PSD for run 1 is shown in blue, run 2 in green, and red shows the PSD of the beamformer-projected empty room noise. In agreement with Figure 2A, there are differences between regions—for example, elevated β power is observed in the sensorimotor regions and prominent α peaks exist in the occipital areas. Most importantly, note the high level of consistency between runs: the relative difference was $4 \pm 1\%$ (mean $\pm$ std. deviation) when averaged over all 78 AAL regions; when examining the variation of this difference across brain regions, it was dominated by differences in occipital, parietal, and temporal lobes. The largest difference between runs was elevated α power in run 2, compared to run 1 (Wilcoxon sign rank test, $p = 0.0039$). Differences in power in the other bands did not survive multiple comparisons correction.

For frequencies below ~60 Hz, the projected empty room noise was lower than the signal, implying a good

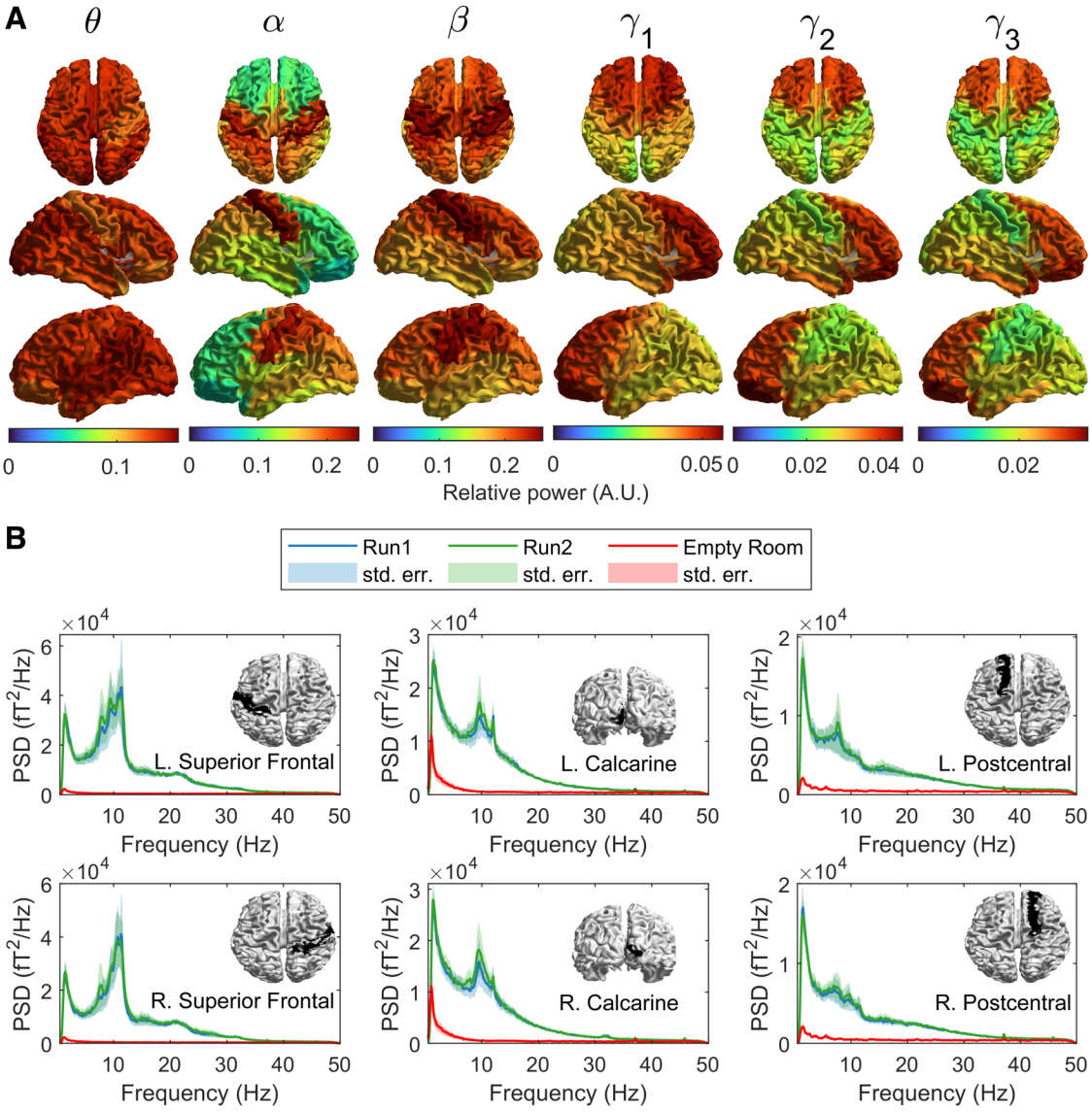


Fig. 2. Spectral Power. (A) Brain plots showing the spatial topographies of relative spectral power averaged across subjects and runs in θ (4-8 Hz), α (8-12 Hz), β (13-10 Hz) and overlapping sub-bands of the γ band (γ_1: 30-40 Hz, γ_2: 35-45 Hz, γ_3: 40-50 Hz). (B) Broad-band power spectra plotted for the regions indicated in the corresponding inset images. Blue and green lines represent the group average spectra for the first and second runs respectively. Shaded areas correspond to the standard error across subjects in each run. Red lines indicate beamformer projected empty room noise.

ratio of signal to sensor noise/interference. On average, the ratio of signal to noise (i.e. the ratio of the green/blue lines to the red line) was 14 ± 8 for θ, 24 ± 18 for α, and 8 ± 4 for β. However, this decreased to 2.7 ± 0.9 for γ_1, 2.0 ± 0.5 for γ_2, and 1.8 ± 0.4 for γ_3, demonstrating how the signal amplitude approaches the sensor noise level with increasing frequency. This is an important point for OPM-MEG sensor design.

3.2. Functional connectivity at the group level

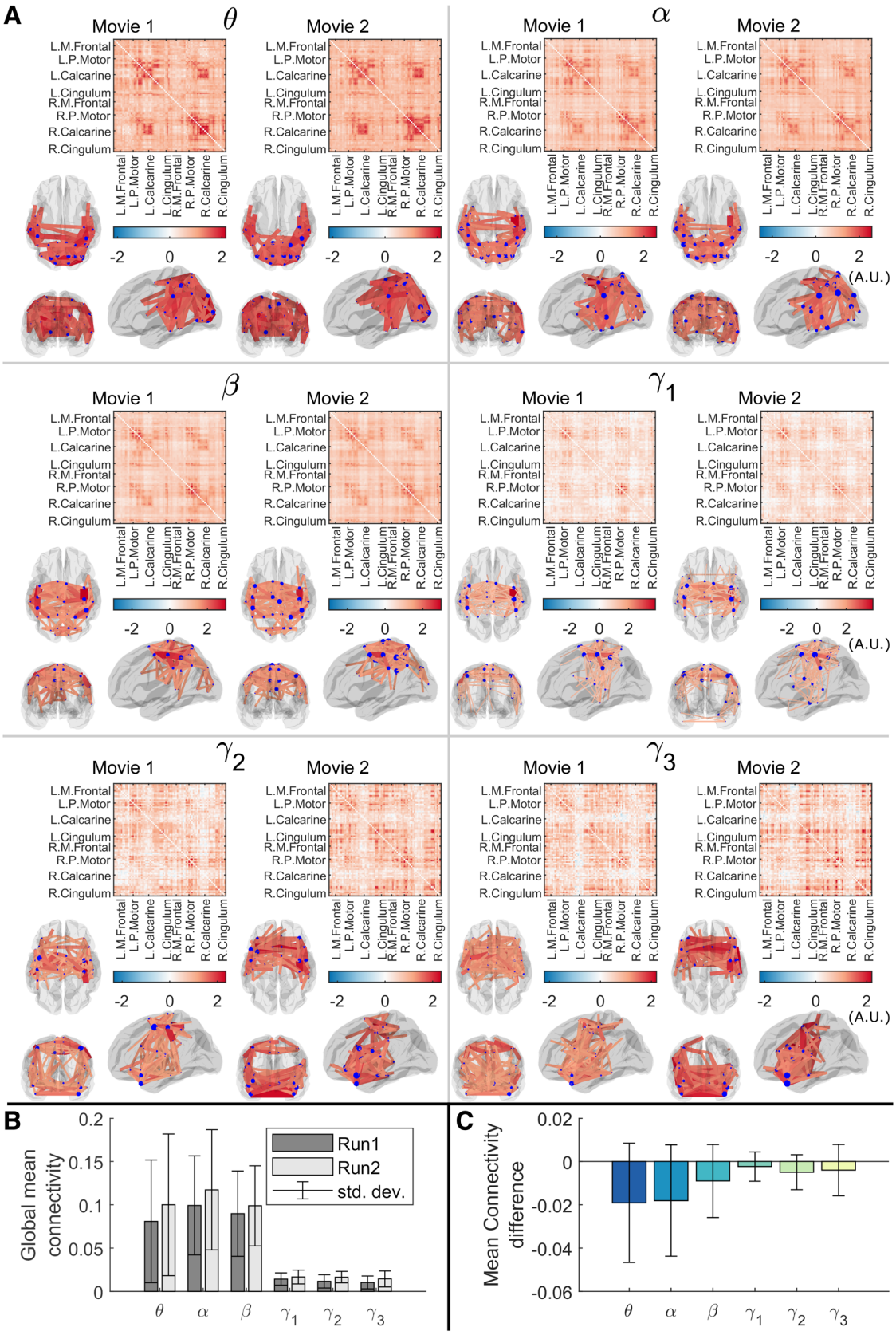
Figure 3A shows group-level connectome results. Connectome matrices are shown alongside glass brain plots in which the lines show the spatial signature of the strongest 150 connections. The blue circles show connectivity strength (i.e. a representation of how connected that brain region is to all other regions). Results for all BoIs are shown. As expected, the spatial signature of connectivity is different in different frequency bands: the α band is dominated by occipital, temporal, and posterior parietal connections; the β band has the highest connectivity strength in sensorimotor regions, with additional frontoparietal and occipital projections. γ_1 highlights a strong sensorimotor network. The θ band has strong posterior connections but with some frontal projections, while the two highest frequency (γ) bands appear to identify frontal and superior parietal connections. These spatial signatures agree with those found using conventional MEG (Hunt et al., 2016). Figure 3B shows the mean global connectivity (averaged over the whole connectome matrix) for the two runs, for each frequency band. Figure 3C shows the difference between runs (i.e. a paired subtraction of global connectivity within each subject, averaged across subjects). In all cases, the bar heights show the mean value and error bars show the standard deviation across participants.

There is a slight trend towards higher global connectivity in the second experimental run compared to the first, though this did not reach significance (a paired Wilcoxon sign rank test on the difference values suggested p-values of 0.05, 0.06, and 0.23 for θ, α, and β bands respectively—no measures survived a multiple comparison correction across bands). Most importantly, in the θ, α, β and low γ bands there is marked similarity in the structure of the connectome matrix across the two separate experimental runs. This is formalised in Figure 4, where panel A shows all matrix elements from run 1 plotted against all matrix elements for run two. Between-run correlation coefficients are shown in panel B as a function of frequency band. Consistency between runs peaks in the β band with a correlation coefficient of 0.935. Correlation is also high for α (0.929) and θ (0.814) but declines with increasing frequency to 0.714, 0.599, and 0.54 for the three γ bands. Figure 4C shows the relationship between sample size (i.e. number of subjects included) and between-run correlation in the group average. The plotted values and error bars represent the mean and standard deviation across all possible combinations. As expected, consistency declines with decreasing group size.

3.3. Individual subject comparisons

Figure 5 shows the individual connectomes for all 10 subjects, for the α band, for both experimental runs. All matrices are distinctly structured and display a marked difference between subjects. However, the consistency across the two runs within each individual is striking. This qualitative observation is formalised in Figure 6A which shows within- and between-subject correlations between connectome matrices. Recall there are 10 possible within-subject comparisons and 90 between-subject comparisons between runs 1 and 2. In Figure 6A, the bars show the mean correlation values while the dots show individual values. The difference between within- and between-subject averages is shown in Figure 6B as a function of frequency band. Within-subject correlation (Fig. 6C) peaked in the β band at 0.78 but was high for θ (0.56) and α (0.72). In agreement with the group result, it drops for the γ bands. The within-/between-subject difference (Fig. 6B) peaked in the α band but according to our Monte-Carlo test was significant in the θ, α, β, and γ_1 bands. In agreement with this, using neural fingerprinting analysis, we were able to correctly identify 7, 10, 8, and

Fig. 3. Group average amplitude envelope correlation and between-run consistency. (A) Group average connectome calculated using amplitude envelope correlation for 2 separate experimental runs, in θ (4-8 Hz), α (8-12 Hz), β (13-10 Hz) and overlapping sub-bands of the γ band (γ_1: 30-40 Hz, γ_2: 35-45 Hz, γ_3: 40-50 Hz). Glass brains show the strongest 150 connections. For visualisation purposes, connectivity values in (A) are scaled by the mean across each matrix. (B) Mean global connectivity (AEC) across subjects for each run and frequency band. (C) Mean connectivity difference between runs across subjects. (B and C) show raw, unscaled correlation values and differences respectively.



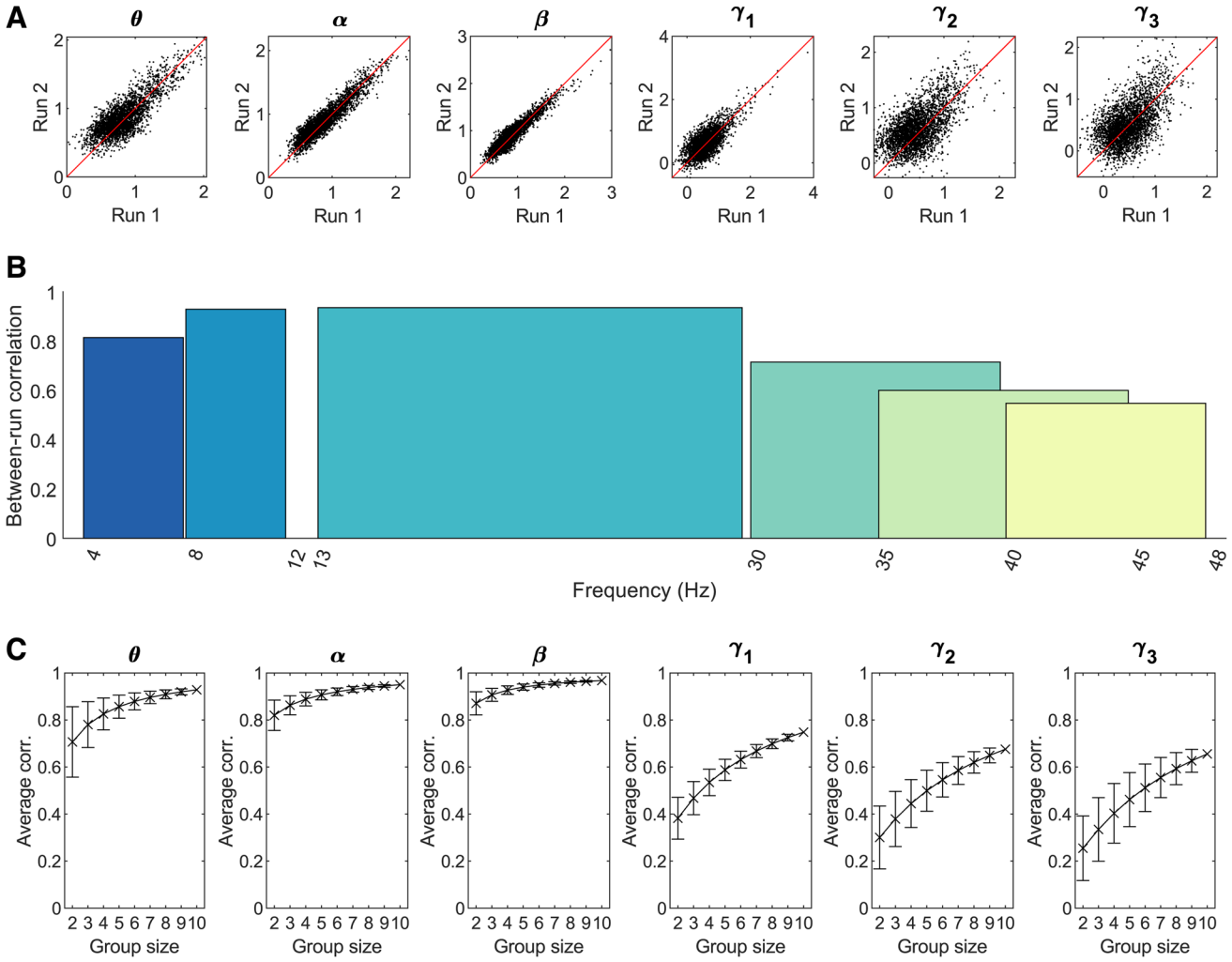


Fig. 4. Between-run reliability of the functional connectome. (A) Scatter plots of group average connectivity values; run 1 plotted against run 2. Black points represent the mean AEC values for each of the 3003 edges in the group average connectomes for both runs. Lines of equality are indicated in red. (B) Bar chart of Pearson correlation coefficients between the group average connectomes for runs 1 and 2. Low-frequency connectomes are highly consistent while the γ sub-bands display more variability between the two runs. (C) The effect of sample size on group average between-run correlation. Crosses represent mean correlation values across possible subsamples for each group size; error bars show the standard deviation across subsamples.

5 individuals in the θ, α, β, and γ_1 bands respectively, by looking at the highest values of correlation across the group.

4. DISCUSSION

The ability to characterise functional connectivity reliably is a critical function of any viable MEG system. However, the measurement of connectivity poses a significant challenge as it relies on high-fidelity (unaveraged) MEG data and whole-brain coverage. Here, we provide a benchmark for the repeatability of both neural oscillatory activity and connectivity across experimental runs using a 168-channel whole-head OPM-MEG device.

At the group level, the correlation between participant-averaged connectivity matrices for runs 1 and 2 was high in the θ (0.814), α (0.929), and β (0.935) bands. However, this fell to 0.714, 0.599, and 0.547 for γ_1, γ_2, and γ_3 respectively. These values compare well to those previously derived for conventional MEG. Colclough et al. (2016), using groups of ~30 individuals, demonstrated a between-group correlation of ~97% in the α band using a similar amplitude envelope correlation metric. While values here are marginally lower, this is likely explained by our groups being smaller

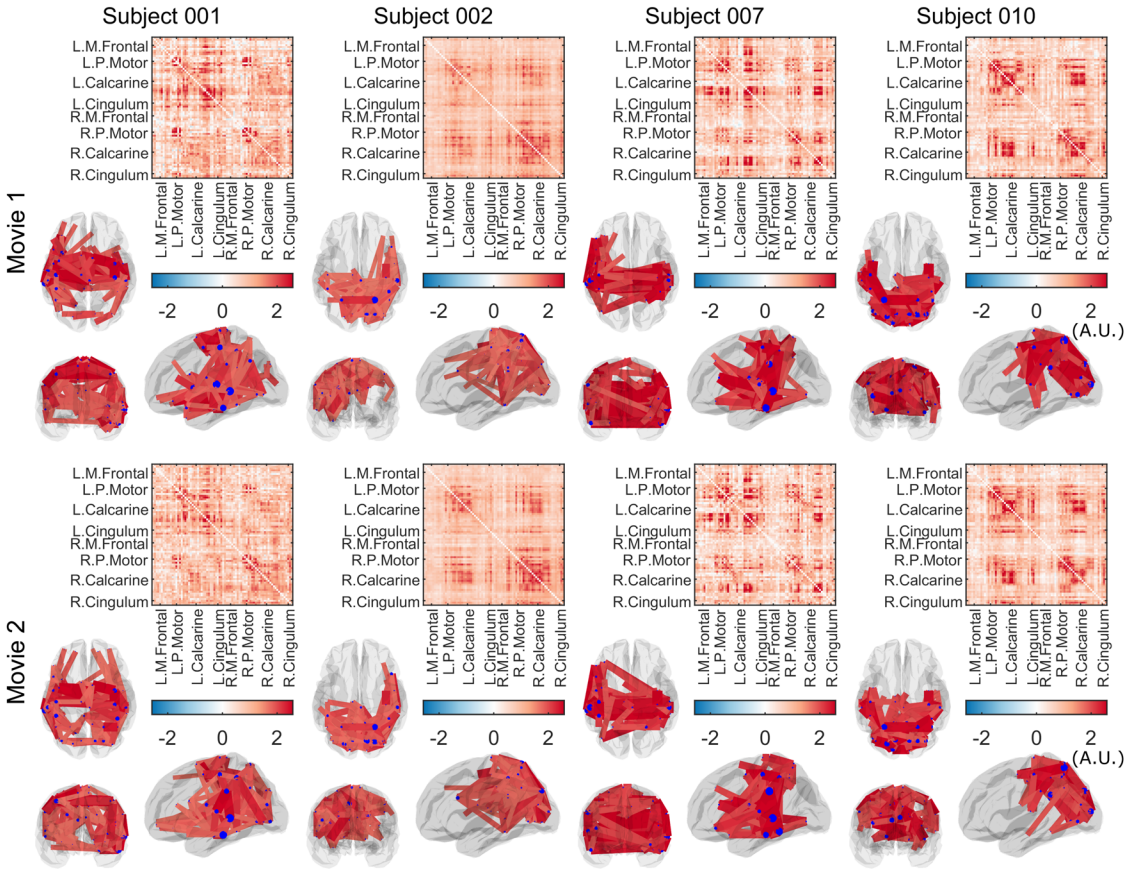


Fig. 5. Individual connectivity matrices: Connectomes and corresponding glass brain plots, for a randomly selected subset of subjects and both experimental runs in the α band (the remaining subjects are shown in Appendix 2, Fig. A2). Note that while variability is high between individuals, results within a single individual are consistent.

(10 people). Figure 4C showed that for bands with relatively lower group-level consistency—such as our γ bands—larger sample sizes can be expected to yield improved consistency. Overall, the high consistency observed in the lower-frequency bands demonstrates that—for group-level measurement—OPM-MEG provides a robust estimate of functional connectivity. The reason for the fall in the γ band can be seen in Figure 2 when comparing the power spectral density from the brain with that from an empty room noise recording. At frequencies above ~30 Hz, the "noise" level is around half of the signal amplitude. Above these frequencies, noise begins to dominate, and measures become unreliable. This agrees with observations in conventional MEG. For example, a previous study of motor network connectivity (Brookes et al., 2011) showed that connectivity between left and right motor cortices in the resting state was measurable up to ~40 Hz; similar observations were found in the frontoparietal and default mode networks (Brookes et al., 2011; Hipp et al., 2012).

At the individual level, correlations were lower. Within-subject correlations were 0.56, 0.72, and 0.78 for the θ, α, and β bands respectively. In comparison, Colclough et al. (2016) observed within-subject consistency of ~58% in the α band. This is somewhat lower than the values observed in our study, though it was estimated using shorter segments of data (5 mins rather than 10 mins). Liuzzi et al. (2017) used amplitude envelope correlation applied to conventional MEG, achieving within-subject consistencies of ~72% in the β band using 560 s of data and with the head clamped into the MEG helmet to eliminate any motion relative to the (SQUID-based) sensors. In line with expectation, the between-subject consistencies were generally much lower, peaking at 0.54 for the β band. Once again, this is in line with expectations from conventional MEG, with Colclough et al. (2016) showing a between-subject correlation of ~45% for the α band. Based on both the group level and individual observations above, the

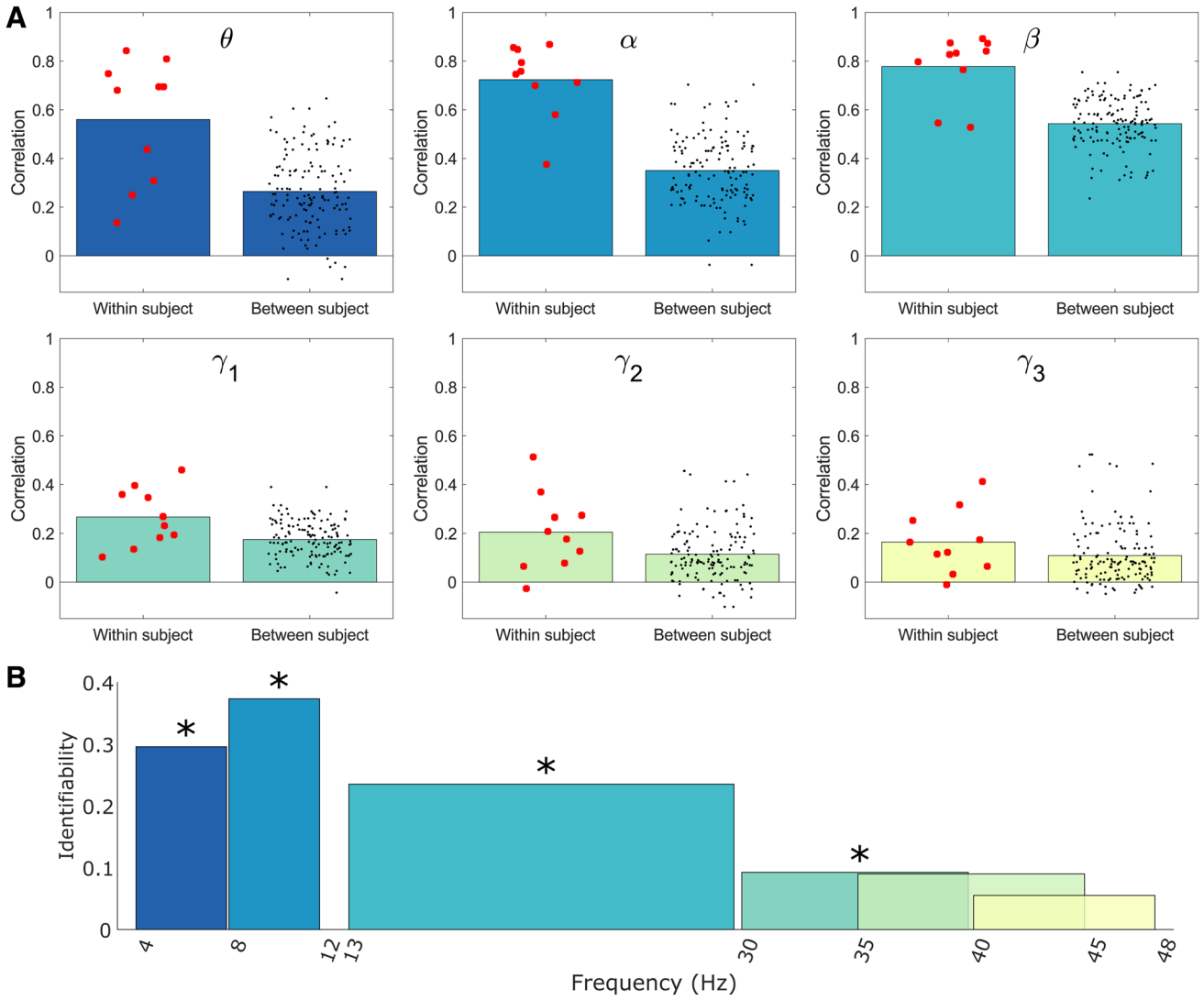


Fig. 6. Individual subject comparison across bands: (A) Within- and between-subject Pearson correlations of the AEC connectomes for each frequency band. (B) Identifiability—the difference between average within- and between-subject correlations indicating the potential for neural fingerprinting across frequency bands. Asterisks indicate statistical significance at the $p < 0.05$ level. p-values were estimated via a permutation test and corrected for multiple comparisons across frequency bands using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995).

repeatability of OPM-MEG compares well with previously published conventional MEG findings.

The drop in correlation values when undertaking between-subject versus within-subject comparisons is the basis for the technique known as neural fingerprinting. Briefly, successful neural fingerprinting requires that a subject can be correctly identified from a group, based on some feature derived from a previous scan. Here, α band connectome matrices enabled successful neural fingerprinting in all 10 subjects, with the β band offering 8 correctly identified individuals and the θ band 7 correctly identified individuals. The γ band was less successful, and this is also reflected in the fact that the within-subject versus between-subject differences were not significant in γ_2 and γ_3. The topic of neural fingerprinting has gained significant traction in recent years (da Silva Castanheira et al., 2021) with the idea that between-subject variance (which is often treated as noise) contains useful and reproducible information. Indeed, it offers the exciting possibility that, by tracking changes in the neural fingerprint, one might enable early detection of disorders (e.g. dementia). The data presented demonstrate that OPM-MEG is a robust platform from which to launch such studies.

At a technical level, there are several limitations of our system which should be addressed. First, the channel count of 168 remains significantly lower than that of conventional MEG systems (which have ~300 channels). In addition, our triaxial sensor measures both the radial and tangentially oriented components of the magnetic field, whereas conventional MEG only measures the radial components. While the use of triaxial sensors has proven to be an excellent means to reduce the effects of non-brain sources of magnetic field (Brookes et al., 2021; Rea et al., 2022; Tierney et al., 2022), the tangential field components are smaller in amplitude and consequently, in terms of absolute signal, OPM-MEG remains disadvantaged compared to cryogenic instrumentation. It is encouraging that, despite the lower channel count, we achieve approximate parity with conventional MEG in terms of repeatability of connectivity measurement. In addition, one significant advantage of the triaxial design is that three-axis measurement enables complete calibration of the sensor and removal of cross-talk artefacts between close-set sensors. This means that, ostensibly, the construction of high-density whole-head OPM systems should be possible in the near future.

One important observation is that, at high frequencies (above ~60 Hz), the signal and empty room noise levels begin to converge. Importantly, this does not mean that OPM-MEG cannot assess high-frequency activity; indeed, several papers (Hill et al., 2019, 2022; Iivanainen et al., 2020) have shown that OPM-MEG can successfully record γ band (>30 Hz) oscillations with similar SNR to that observed in conventional MEG (Hill et al., 2020). However, these previous observations use trial averaging to increase SNR. Our observation suggests that using unaveraged data, gamma responses from the brain (due to the stimulus used here) may be lower amplitude than the inherent noise level of the OPMs. This is also likely the case for cryogenic MEG; however, with OPMs there exist multiple means to enhance SNR beyond what we see in the current study, either by decreasing the inherent sensor noise or by increasing sensor density. This is currently a priority in system development and future OPM-MEG implementations are likely to be able to reliably measure gamma effects, even in unaveraged data.

Finally, there are two aspects of experimental design that should be considered. First, in previous conventional MEG studies, data have typically been recorded in the "pure" resting state (i.e. participants are asked to "sit still and do nothing"). In contrast, here, subjects were asked to watch a movie. This experimental decision was taken to provide consistency between this dataset and those previously collected using the same movie clip with multiple imaging modalities including EEG, fMRI, and ECoG (Haufe et al., 2018). However, the addition of this naturalistic stimulus likely drives brain activity which is synchronised across runs and the extent to which this might help to enhance consistency (beyond what might be expected in resting state) is unknown. This complicates direct comparisons of our measures with those previously presented (e.g. Colclough et al., 2016). This said, a previous study (Lankinen et al., 2018) has shown that, unlike fMRI (where inter-subject correlation of the BOLD response to watching the same movie clip was high) correlations between oscillatory envelopes of band-limited oscillations were relatively low. In addition, correlations were highest in the visual cortex whereas our connectome analysis measures whole brain dynamics. We think it therefore unlikely that the impact of the movie (in contrast to pure resting state) is large. Nevertheless, future studies might look to repeat similar measures using pure resting state. Secondly, the interval between the two separate runs of movie watching was short (around a minute). This was for two reasons: first to be consistent with previous literature on neural fingerprinting (da Silva Castanheira et al., 2021) and second to remove any undesirable effect of co-registration error (which would necessarily be different between runs, if the subject removed the OPM helmet). However, this leaves the question of how stable the neural fingerprint would be if the gap between runs were days or even years long. In a recent (independent) study by our group (Rhodes et al., 2023), we showed that neural fingerprinting was possible despite a gap of (on average) 19 days between runs. While this was using a task-based analysis of θ oscillations (rather than movie watching), it does provide confidence that similar results to those presented here might be possible even if the gap between runs was made larger.

5. CONCLUSION

OPM-MEG offers significant advantages over conventional MEG, and other non-invasive functional imaging modalities including EEG, fNIRS, and fMRI. However, OPM-MEG is also a new technology. Demonstrating both the viability and repeatability of key metrics is a necessary step in the path to adoption. Here, we aimed to test the robustness of whole-brain connectivity across two separate experimental runs of the same movie-watching paradigm. Results

showed that the power spectra of the neural signal, from which connectivity is derived, were consistent across repeats of the experiments, with differences between runs amounting to 4% of the total signal. When assessing connectivity, we demonstrated excellent group-level robustness, with high correlations between connectomes in the θ (0.81), α (0.93), and β (0.94) frequency ranges. At the individual subject level, we found marked differences between individuals, but high within-subject robustness (correlations of 0.56 ± 0.25, 0.72 ± 0.15, and 0.78 ± 0.13 in θ, α, and β respectively). These results compare well to equivalent findings using conventional MEG; they show that OPM-MEG is a viable way to characterise whole-brain connectivity and add significant weight to the argument that OPMs can overtake cryogenic sensors as the fundamental building block of MEG systems.

DATA AND CODE AVAILABILITY

All data used to produce the results presented here will be made available upon acceptance of the manuscript at https://doi.org/10.5281/zenodo.7525341. The MATLAB software used for data analysis will be available at https://github.com/LukasRier/Rier2022_OPM_connectome_test-retest/.

AUTHOR CONTRIBUTIONS

Lukas Rier: Conceptualisation; Data curation; Formal analysis; Investigation; Methodology; Software; Visualisation; and Writing—original draft. Sebastian Michelmann: Methodology; Writing—Review & Editing. Harrison Ritz: Methodology; Writing—Review & Editing. Vishal Shah: Methodology; Resources; and Writing—Review & Editing. Ryan M. Hill: Data curation; Investigation; Methodology; Software; and Writing—Review & Editing. James Osborne: Writing—Review & Editing. Cody Doyle: Writing—Review & Editing. Niall Holmes: Data curation; Investigation; Methodology; Software; and Writing—Review & Editing. Richard Bowtell: Conceptualisation; Funding acquisition; and Writing—Review & Editing. Matthew J. Brookes: Conceptualisation; Funding acquisition; Software; Supervision; Visualisation; and Writing—Original Draft. Kenneth A. Norman: Conceptualisation; Methodology; and Writing—Review & Editing. Uri Hasson: Conceptualisation; Writing—Review & Editing. Jonathan D. Cohen: Conceptualisation; Methodology; and Writing—Review & Editing. Elena Boto: Conceptualisation; Funding acquisition; Data curation; Investigation; Methodology; Supervision; and Writing—Review & Editing.

DECLARATION OF COMPETING INTEREST

V.S. is the founding director of QuSpin, a commercial entity selling OPM magnetometers. J.O. and C.D. are employees of QuSpin. E.B. and M.J.B. are directors of Cerca Magnetics Limited, a spin-out company whose aim is to commercialise aspects of OPM-MEG technology. E.B., M.J.B., R.B., N.H., and R.H. hold founding equity in Cerca Magnetics Limited.

ACKNOWLEDGEMENTS

This work was supported by an Engineering and Physical Sciences Research Council (EPSRC) Healthcare Impact Partnership Grant (EP/V047264/1). We also acknowledge support from the UK Quantum Technology Hub in Sensing and Timing, funded by EPSRC (EP/T001046/1). We thank the Medical Research Council UK for funding OPM sensors (MC_PC_MR/X012263/1). Sensor development was made possible by funding from the National Institutes of Health (R44MH110288). S.M. is funded by Deutsche Forschungsgemeinschaft (DFG), project 437219953. We also express thanks to Cerca Magnetics Ltd. and Metamorphic AM and Added Scientific Ltd. for the useful and productive discussions that led to the design of the lightweight helmet used for OPM-MEG measurements.

REFERENCES

Altarev, I., Babcock, E., Beck, D., Burghoff, M., Chesnevskaya, S., Chupp, T., Degenkolb, S., Fan, I., Fierlinger, P., Frei, A., Gutsmiedl, E., Knappe-Gruneberg, S., Kuchler, F., Lauer, T., Link, P., Lins, T., Marino, M., McAndrew, J., Niessen, B., …, Zechlau, T. (2014). A magnetically shielded room with ultra low residual field and gradient. *Rev Sci Instrum*, *85*, 75106. https://doi.org/10.1063/1.4886146

Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Smith, P. J. P. P., Woolrich, M., Probert Smith, P. J., & Woolrich, M. (2014). Fast transient networks in spontaneous human brain activity. *Elife*, *3*, e01867. https://doi.org/10.7554/eLife.01867

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol*, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Boto, E., Bowtell, R., Krüger, P., Fromhold, T. M., Morris, P. G., Meyer, S. S., Barnes, G. R., & Brookes, M. J. (2016). On the potential of a new generation of magnetometers for MEG: A beamformer simulation study. *PLoS One*, *11*, e0157655. https://doi.org/10.1371/JOURNAL.PONE.0157655

Boto, E., Hill, R. M., Rea, M., Holmes, N., Seedat, Z. A., Leggett, J., Shah, V., Osborne, J., Bowtell, R., & Brookes, M. J. (2021). Measuring functional connectivity with

wearable MEG. *Neuroimage*, *230*, 117815. https://doi.org/10.1016/j.neuroimage.2021.117815
Boto, E., Holmes, N., Leggett, J., Roberts, G., Shah, V., Meyer, S. S., Muñoz, L. D., Mullinger, K. J., Tierney, T. M., Bestmann, S., Barnes, G. R., Bowtell, R., & Brookes, M. J. (2018). Moving magnetoencephalography towards real-world applications with a wearable system. *Nature*, *555*, 657–661. https://doi.org/10.1038/nature26147
Boto, E., Meyer, S. S., Shah, V., Alem, O., Knappe, S., Kruger, P., Fromhold, T. M., Lim, M., Glover, P. M., Morris, P. G., Bowtell, R., Barnes, G. R., Brookes, M. J., Fromhold, M., Lim, M., Glover, P. M., Morris, P. G., Bowtell, R., Barnes, G. R., & Brookes, M. J. (2017). A new generation of magnetoencephalography: Room temperature measurements using optically-pumped magnetometers. *Neuroimage*, *149*, 404–414. https://doi.org/10.1016/j.neuroimage.2017.01.034
Boto, E., Shah, V., Hill, R. M., Rhodes, N., Osborne, J., Doyle, C., Holmes, N., Rea, M., Leggett, J., Bowtell, R., & Brookes, M. J. (2022). Triaxial detection of the neuromagnetic field using optically-pumped magnetometry: Feasibility and application in children. *Neuroimage*, *252*, 119027. https://doi.org/10.1016/j.neuroimage.2022.119027
Brookes, M. J., Boto, E., Rea, M., Shah, V., Osborne, J., Holmes, N., Hill, R. M., Leggett, J., Rhodes, N., & Bowtell, R. (2021). Theoretical advantages of a triaxial optically pumped magnetometer magnetoencephalography system. *Neuroimage*, *236*, 118025. https://doi.org/10.1016/j.neuroimage.2021.118025
Brookes, M. J., Hale, J. R., Zumer, J. M., Stevenson, C. M., Francis, S. T., Barnes, G. R., Owen, J. P., Morris, P. G., & Nagarajan, S. S. (2011). Measuring functional connectivity using MEG: Methodology and comparison with fcMRI. *Neuroimage*, *56*, 1082–1104. https://doi.org/10.1016/j.neuroimage.2011.02.054
Brookes, M. J., Leggett, J., Rea, M., Holmes, N., Boto, E., & Bowtell, R. (2022). Magnetoencephalography with optically pumped magnetometers (OPM-MEG): The next generation of functional neuroimaging. *Trends Neurosci*, *45*, 621–634. https://doi.org/10.1016/j.tins.2022.05.008
Brookes, M. J. J., Woolrich, M. W. W., & Barnes, G. R. R. (2012). Measuring functional connectivity in MEG: A multivariate approach insensitive to linear source leakage. *Neuroimage*, *63*, 910–920. https://doi.org/10.1016/j.neuroimage.2012.03.048
Cohen, D. (1968). Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents. *Science*, *161*, 784–786. https://doi.org/10.1126/SCIENCE.161.3843.784
Cohen, D. (1972). Magnetoencephalography: Detection of the brain's electrical activity with a superconducting magnetometer. *Science*, *175*, 664–666. https://doi.org/10.1126/SCIENCE.175.4022.664
Colclough, G. L. L., Woolrich, M. W. W., Tewarie, P. K. K., Brookes, M. J. J., Quinn, A. J. J., & Smith, S. M. M. (2016). How reliable are MEG resting-state connectivity metrics? *Neuroimage*, *138*, 284–293. https://doi.org/10.1016/j.neuroimage.2016.05.070
da Silva Castanheira, J., Orozco Perez, H. D., Misic, B., & Baillet, S. (2021). Brief segments of neurophysiological activity enable individual differentiation. *Nat Commun*, *12*, 1–11. https://doi.org/10.1038/s41467-021-25895-8
Feys, O., Corvilain, P., Aeby, A., Sculier, C., Christiaens, F., Holmes, N., Brookes, M., Goldman, S., Wens, V., & De Tiège, X. (2022). On-scalp optically pumped magnetometers versus cryogenic magnetoencephalography for diagnostic evaluation of epilepsy in school-aged children. *Radiology*, *304*, 429–434. https://doi.org/10.1148/radiol.212453
Gong, G., Rosa-Neto, P., Carbonell, F., Chen, Z. J., He, Y., & Evans, A. C. (2009). Age- and gender-related differences in the cortical anatomical network. *J Neurosci*, *29*, 15684–15693. https://doi.org/10.1523/JNEUROSCI.2308-09.2009
Hamalainen, M. S., Hari, R., Ilmoniemi, R. J., Knuutila, J., Lounasma, O. V, Lounasmaa, & O. V. (1993). Magnetoencephalography: Theory, instrumentation, and applications to non-invasive studies of the working human brain. *Rev Mod Phys*, *65*, 413–497. https://doi.org/10.1103/RevModPhys.65.413
Haufe, S., DeGuzman, P., Henin, S., Arcaro, M., Honey, C. J., Hasson, U., Parra, & L. C. (2018). Elucidating relations between fMRI, ECoG, and EEG through a common natural stimulus. *Neuroimage*, *179*, 79–91. https://doi.org/10.1016/j.neuroimage.2018.06.016
Hill, R. M., Boto, E., Holmes, N., Hartley, C., Seedat, Z. A., Leggett, J., Roberts, G., Shah, V., Tierney, T. M., Woolrich, M. W., Stagg, C. J., Barnes, G. R., Bowtell, R. R., Slater, R., & Brookes, M. J. (2019). A tool for functional brain imaging with lifespan compliance. *Nat Commun*, *10*, 1–11. https://doi.org/10.1038/s41467-019-12486-x
Hill, R. M., Boto, E., Rea, M., Holmes, N., Leggett, J., Coles, L. A., Papastavrou, M., Everton, S. K., Hunt, B. A. E., Sims, D., Osborne, J., Shah, V., Bowtell, R., & Brookes, M. J. (2020). Multi-channel whole-head OPM-MEG: Helmet design and a comparison with a conventional system. *Neuroimage*, *219*, 116995. https://doi.org/10.1016/j.neuroimage.2020.116995
Hill, R. M., Devasagayam, J., Holmes, N., Boto, E., Shah, V., Osborne, J., Safar, K., Worcester, F., Mariani, C., Dawson, E., Woolger, D., Bowtell, R., Taylor, M. J., & Brookes, M. J. (2022). Using OPM-MEG in contrasting magnetic environments. *Neuroimage*, *253*, 119084. https://doi.org/10.1016/j.neuroimage.2022.119084
Hillebrand, A., Tewarie, P., Van Dellen, E., Yu, M., Carbo, E. W. S., Douw, L., Gouw, A. A., Van Straaten, E. C. W., & Stam, C. J. (2016). Direction of information flow in large-scale resting-state networks is frequency-dependent. *Proc Natl Acad Sci U S A*, *113*, 3867–3872. https://doi.org/10.1073/pnas.1515657113
Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M., & Engel, A. K. (2012). Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nat Neurosci*, *15*, 884–890. https://doi.org/10.1038/nn.3101
Holmes, M., Leggett, J., Boto, E., Roberts, G., Hill, R. M., Tierney, T. M., Shah, V., Barnes, G. R., Brookes, M. J., & Bowtell, R. (2018). A bi-planar coil system for nulling background magnetic fields in scalp mounted magnetoencephalography. *Neuroimage*, *181*, 760–774. https://doi.org/10.1016/j.neuroimage.2018.07.028
Holmes, N., Rea, M., Hill, R. M., Leggett, J., Edwards, L. J., Hobson, P. J., Boto, E., Tierney, T. M., Rier, L., Rivero, G. R., Shah, V., Osborne, J., Fromhold, T. M., Glover, P., Brookes, M. J., & Bowtell, R. (2023). Enabling ambulatory movement in wearable magnetoencephalography with

matrix coil active magnetic shielding. *Neuroimage*, *274*, 120157. https://doi.org/10.1016/J.NEUROIMAGE.2023.120157

Holmes, N., Tierney, T. M., Leggett, J., Boto, E., Mellor, S., Roberts, G., Hill, R. M., Shah, V., Barnes, G. R., Brookes, M. J., & Bowtell, R. (2020). Balanced, bi-planar magnetic field and field gradient coils for field compensation in wearable magnetoencephalography. *Sci Rep*, *9*, 14196. https://doi.org/10.1038/s41598-019-50697-w

Honey, C. J., Thesen, T., Donner, T. H., Silbert, L. J., Carlson, C. E., Devinsky, O., Doyle, W. K., Rubin, N., Heeger, D. J., & Hasson, U. (2012). Slow cortical dynamics and the accumulation of information over long timescales. *Neuron*, *76*, 423–434. https://doi.org/10.1016/J.NEURON.2012.08.011

Hunt, B. A. E., Tewarie, P. K., Mougin, O. E., Geades, N., Jones, D. K., Singh, K. D., Morris, P. G., Gowland, P. A., & Brookes, M. J. (2016). Relationships between cortical myeloarchitecture and electrophysiological networks. *Proc Natl Acad Sci U S A*, *113*, 13510–13515. https://doi.org/10.1073/pnas.1608587113

Iivanainen, J., Stenroos, M., & Parkkonen, L. (2017). Measuring MEG closer to the brain: Performance of on-scalp sensor arrays. *Neuroimage*, *147*, 542–553. https://doi.org/10.1016/J.NEUROIMAGE.2016.12.048

Iivanainen, J., Zetter, R., Grön, M., Hakkarainen, K., & Parkkonen, L. (2019). On-scalp MEG system utilizing an actively shielded array of optically-pumped magnetometers. *Neuroimage*, *194*, 244–258. https://doi.org/10.1016/J.NEUROIMAGE.2019.03.022

Iivanainen, J., Zetter, R., & Parkkonen, L. (2020). Potential of on-scalp MEG: Robust detection of human visual gamma-band responses. *Hum Brain Mapp*, *41*, 150–161. https://doi.org/10.1002/hbm.24795

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med Image Anal*, *5*, 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

Johnson, C., Schwindt, P. D. D., & Weisend, M. (2010). Magnetoencephalography with a two-color pump-probe, fiber-coupled atomic magnetometer. *Appl Phys Lett*, *97*, 243703. https://doi.org/10.1063/1.3522648

Johnson, C. N., Schwindt, P. D. D., & Weisend, M. (2013). Multi-sensor magnetoencephalography with atomic magnetometers. *Phys Med Biol*, *58*, 6065–6077. https://doi.org/10.1088/0031-9155/58/17/6065

Kamada, K., Sato, D., Ito, Y., Natsukawa, H., Okano, K., Mizutani, N., & Kobayashi, T. (2015). Human magnetoencephalogram measurements using newly developed compact module of high-sensitivity atomic magnetometer. *Jpn J Appl Phys*, *54*, 026601. https://doi.org/10.7567/JJAP.54.026601

Lankinen, K., Saari, J., Hlushchuk, Y., Tikka, P., Parkkonen, L., Hari, R., & Koskinen, M. (2018). Consistency and similarity of MEG- and fMRI-signal time courses during movie viewing. *Neuroimage*, *173*, 361–369. https://doi.org/10.1016/J.NEUROIMAGE.2018.02.045

Liuzzi, L., Gascoyne, L. E., Tewarie, P. K., Barratt, E. L., Boto, E., & Brookes, M. J. (2017). Optimising experimental design for MEG resting state functional connectivity measurement. *Neuroimage*, *155*, 565–576. https://doi.org/10.1016/j.neuroimage.2016.11.064

Lumet, S. (Director). (1975). *Dog Day Afternoon* [Film]. Warner Bros.

Nardelli, N. V, Perry, A. r., Krzyzewski, S. P., & Knappe, S. A. (2020). A conformal array of microfabricated optically-pumped first-order gradiometers for magnetoencephalography. *EPJ Quantum Technol*, *7*, 11. https://doi.org/10.1140/epjqt/s40507-020-00086-4

Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoenchephalography forward calculation in realistic volume conductors. *Phys Med Biol*, *48*, 3637–3652. https://doi.org/10.1088/0031-9155/48/22/002

Nugent, A. C., Benitez Andonegui, A., Holroyd, T., & Robinson, S. E. (2022). On-scalp magnetocorticography with optically pumped magnetometers: Simulated performance in resolving simultaneous sources. *Neuroimage Rep*, *2*, 100093. https://doi.org/10.1016/J.YNIRP.2022.100093

O'Neill, G. C., Barratt, E. L., Hunt, B. A. E., Tewarie, P. K., & Brookes, M. J. (2015). Measuring electrophysiological connectivity by power envelope correlation: A technical review on MEG methods. *Phys Med Biol*, *60*, R271–R295. https://doi.org/10.1088/0031-9155/60/21/R271

O'Neill, G. C., Bauer, M., Woolrich, M. W., Morris, P. G., Barnes, G. R., & Brookes, M. J. (2015). Dynamic recruitment of resting state sub-networks. *Neuroimage*, *115*, 85–95. https://doi.org/10.1016/j.neuroimage.2015.04.030

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Oyama, D., Adachi, Y., Yumoto, M., Hashimoto, I., & Uehara, G. (2015). Dry phantom for magnetoencephalography—Configuration, calibration, and contribution. *J Neurosci Methods*, *251*, 24–36. https://doi.org/10.1016/J.JNEUMETH.2015.05.004

Rea, M., Boto, E., Holmes, N., Hill, R., Osborne, J., Rhodes, N., Leggett, J., Rier, L., Bowtell, R., Shah, V., & Brookes, M. J. (2022). A 90-channel triaxial magnetoencephalography system using optically pumped magnetometers. *Ann N Y Acad Sci*, *1517*, 107–124. https://doi.org/10.1111/nyas.14890

Rea, M., Holmes, N., Hill, R. M., Boto, E., Leggett, J., Edwards, L. J., Woolger, D., Dawson, E., Shah, V., Osborne, J., Bowtell, R., & Brookes, M. J. (2021). Precision magnetic field modelling and control for wearable magnetoencephalography. *Neuroimage*, *241*, 118401. https://doi.org/10.1016/j.neuroimage.2021.118401

Rhodes, N., Rea, M., Boto, E., Rier, L., Shah, V., Hill, R. M., Osborne, J., Doyle, C., Holmes, N., Coleman, S. C., Mullinger, K., Bowtell, R., & Brookes, M. J. (2023). Measurement of frontal midline theta oscillations using OPM-MEG. *Neuroimage*, *271*, 120024. https://doi.org/10.1016/J.NEUROIMAGE.2023.120024

Rier, L., Michelmann, S., Ritz, H., Shah, V., Hill, R. M., Osborne, J., Doyle, C., Holmes, N., Bowtell, R., Brookes, M. J., Norman, K. A., Hasson, U., Cohen, J. D., & Boto, E. (2022). Test-retest reliability of the human connectome: An OPM-MEG study (1.1.0) [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.7525341

Vrba, J., & Robinson, S. E. (2001). Signal Processing in Magnetoencephalography. *Methods*, *25*(2), 249–271. https://doi.org/10.1006/meth.2001.1238

Sadaghiani, S., Brookes, M. J., & Baillet, S. (2022). Connectomics of human electrophysiology. *Neuroimage*, *247*, 118788. https://doi.org/10.1016/j.neuroimage.2021.118788

Sander, T. H., Preusser, J., Mhaskar, R., Kitching, J., Trahms, L., & Knappe, S. (2012). Magnetoencephalography with a chip-scale atomic magnetometer. *Biomed Opt Express*, *3*, 981–990. https://doi.org/10.1364/BOE.3.000981

Sekihara, K., Nagarajan, S. S., Poeppel, D., & Marantz, A. (2004). Asymptotic SNR of scalar and vector minimum-variance beanformers for neuromagnetic source reconstruction. *IEEE Trans Biomed Eng*, *51*, 1726–1734. https://doi.org/10.1109/TBME.2004.827926

Seymour, R. A., Alexander, N., Mellor, S., O'Neill, G. C., Tierney, T. M., Barnes, G. R., & Maguire, E. A. (2021). Using OPMs to measure neural activity in standing, mobile participants. *Neuroimage*, *244*, 118604. https://doi.org/10.1016/j.neuroimage.2021.118604

Shah, V., Doyle, C., & Osborne, J. (2020). *U.S. Patent No. 10,775,450*. Washington, DC: U.S. Patent and Trademark Office.

Tierney, T. M., Alexander, N., Mellor, S., Holmes, N., Seymour, R., O'Neill, G. C., Maguire, E. A., & Barnes, G. R. (2021). Modelling optically pumped magnetometer interference in MEG as a spatially homogeneous magnetic field. *Neuroimage*, *244*, 118484. https://doi.org/10.1016/j.neuroimage.2021.118484

Tierney, T. M., Mellor, S., O'Neill, G. C., Timms, R. C., & Barnes, G. R. (2022). Spherical harmonic based noise rejection and neuronal sampling with multi-axis OPMs. *Neuroimage*, *258*, 119338. https://doi.org/10.1016/j.neuroimage.2022.119338

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*, 273–289. https://doi.org/10.1006/nimg.2001.0978

Xia, H., Baranga, A. B. A., Hoffman, D., & Romalis, M. V. (2006). Magnetoencephalography with an atomic magnetometer. *Appl Phys Lett*, *89*, 211104. https://doi.org/10.1063/1.2392722

Zetter, R., Iivanainen, J., & Parkkonen, L. (2019). Optical co-registration of MRI and on-scalp MEG. *Sci Rep*, *9*(1), 5490. https://doi.org/10.1038/s41598-019-41763-4

(*Appendix follows* →)

APPENDIX 1: PHANTOM MEASUREMENT

To confirm that our triaxial OPM-MEG array can accurately measure magnetic fields, and further, that those fields can be used to reliably localise a dipolar current source and reconstruct its time course, we employed a phantom.

We constructed a dry-type dipole phantom as originally described by Oyama et al. (2015) (see also Holmes et al., 2023). Briefly, a triangular electromagnetic coil was constructed (isosceles with 5 mm base and 45 mm height) by winding a single turn of 0.56 mm diameter enamelled copper wire around a 3D-printed guide. The ends of the wire were twisted to avoid stray magnetic fields. The phantom was encased in a Perspex cylinder which was glued to an empty OPM sensor-casing. This allowed for the phantom to be placed in any of the helmet's sensor slots. Once fitted, the dipole was at a fixed (and known) position relative to the sensor array. The base of the triangular wire path was positioned 4 cm radially inward from the chosen slot's position and oriented tangentially with respect to the surface of the helmet. (See Fig. A1A.)

Signal generation

To mimic broad-band brain activity, we generated a signal, 100 s in duration, which was passed through the phantom using a NI-9264 DAC at a sample rate of 4 kHz. To make the signal, white noise was filtered using a polynomial filter (coefficients [-1,0.99]) to generate a characteristic 1/f spectrum. α and β band signals were then added by band-pass filtering white noise using 3rd-order, zero-phase Butterworth filters with passbands of 8-12 and 13-30 Hz respectively. The three separate signals were scaled and summed. The result was scaled such that the maximum amplitude was 1 mV. The signal was also amplified (to a maximum amplitude of 5 V) and sent to an analogue channel to be recorded simultaneously with the MEG data; this provided a "ground-truth". Figure A1C shows a power spectrum of the signal. This process was repeated six times to create six separate signals, which were passed sequentially through the dipole.

Data collection and analysis

Data were collected using the same protocol as that described for the main experiments. Channel-level data and power spectra were visually inspected, and excessively noisy or failed channels were rejected. Data were notch filtered to remove powerline noise and band-pass filtered between 1 and 100 Hz, using a 4th-order zero-phase shift Butterworth filter. Homogeneous field correction was applied.

Each 100 s segment of data was processed independently. A head model, based on the MNI-152 template MRI (1 mm isotropic resolution), was constructed and used to calculate a single-shell forward model (Nolte, 2003). An LCMV beamformer was used to reconstruct a pseudo-Z image showing the source location. Source orientation was determined by estimating the direction of the maximum projected signal amplitude (Sekihara et al., 2004). To improve computational efficiency, only voxels within 3.5 cm of the ground-truth phantom location were included in the source reconstruction. To test the agreement between our simulated signal and the ground truth, for each data segment we measured:

1) The difference between the ground-truth source location and that derived by beamformer reconstruction.
2) The correlation between the reconstructed time course and the original (ground truth) signal sent to the dipole.
3) The correlation between the field vectors measured by the array (when the phantom signal exceeds 60% of its maximum amplitude) and the expected field (i.e. the forward model, calculated assuming a dipole at the ground-truth location/orientation).

Results

In Figure A1B, the left-hand panel shows the mean (over all six data segments) pseudo-Z-statistical image of electrical activity; the spatial discrepancy between the centre of the dipole and the peak in the pseudo-Z-statistical image was 0.7 mm (the peak was at the same voxel for each of the 6 runs). The plots in Figure A1C show a representative segment of the phantom time course (top) and the power spectrum of activity; in both cases, blue shows the beamformer reconstruction and red shows the ground truth. The Pearson correlation between the reconstruction and the original signal was 0.857 $\pm$ 0.003 (mean $\pm$ standard deviation across six segments). Figure A1D shows the modelled (left) and measured (right) fields: in both cases, the vector fields are shown by the arrows; the colour maps show the radial component of the field. These fields represent an average across time points in which the phantom signal exceeded 60% of its maximum

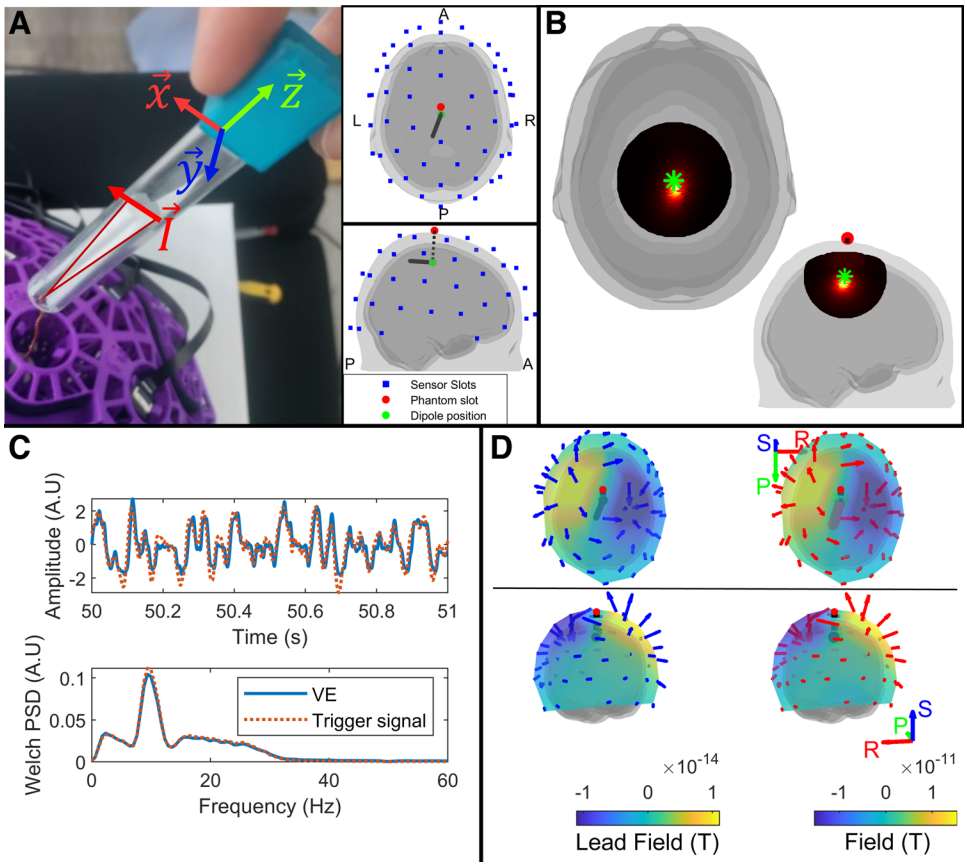


Fig. A1. Results of phantom experiments. (A) The phantom design and a schematic model of the dipole location relative to the OPM sensors. (B) The beamformer reconstructed pseudo-Z-statistical image (average of 6 runs). (C) The graphs on the right show part of the source time course (top) and its power spectrum (bottom). In both cases, the blue trace shows the beamformer reconstruction and the dotted red shows the ground truth. (D) Images show the average measured and modelled magnetic fields from the ground-truth dipole location, with the left (blue arrows) showing the ground truth and the right (red arrows) showing the average measured field from time points where the phantom received >60% of the maximum signal; note the good agreement.

amplitude (to ensure that we were only measuring field correlation when the dipole was switched on). After estimating these fields for each of the six 100s segments, we calculated the correlation between the measured and simulated field. The mean correlation coefficient was 0.99224 ± 0.00004.

In summary, these data show that our array is able to capture an accurate representation of a known magnetic field pattern. This would not be the case (i.e. the measurement would not match the model) if, for instance, there was significant cross-talk between sensors, or background field was adversely affecting sensor gain.

APPENDIX 2:

Fig. A2. Individual connectomes omitted from Figure 5. Connectomes and corresponding glass brain plots, for a randomly selected subset of subjects and both experimental runs in the α band.