Brainprint: identifying individuals from Magnetoencephalography

Shenghao Wu, Aaditya Ramdas, Leila Wehbe

Carnegie Mellon University
{shenghao, aramdas, lwehbe}@cmu.edu

June 18, 2020

Abstract

Neuroimaging tools have been widely adopted to study the anatomical and functional properties of the brain. Magnetoencephalography (MEG), a neuroimaging method prized for its high temporal resolution, records magnetic field changes due to brain activity and has been used to study the cognitive processes underlying various tasks. As the research community increasingly embraces the principles of open science, a growing amount of MEG data has been published online. However, the prevalence of MEG data sharing may pose unforeseen privacy issues. We argue that an individual may be identified from a segment of their MEG recording even if their data has been anonymized. From our standpoint, individual identifiability is closely related to individual variability of brain activity, which is itself a widely studied scientific topic. In this paper, we propose three interpretable spatial, temporal, and frequency MEG featurizations that we term brainprints (brain fingerprints). We show using multiple datasets that these brainprints can accurately identify individuals, and we reveal consistent components of these brainprints that are important for identification. We also investigate how identification accuracy varies with respect to the abundance of data, the level of preprocessing, and the state of the brain. Our findings pinpoint how individual variability expresses itself through MEG, a topic of scientific interest, while raising ethical concerns about the unregulated sharing of brain data, even if anonymized.

1 Introduction

The open science movement [1] is just one example of the increasing awareness of the importance of sharing data and code to promote scientific reproducibility. Specifically, public repositories enable researchers to share their neuroimaging data (fMRI, EEG, MEG, etc) while making sure to censor out individual information [2]. However, data anonymization does not always preserve privacy [3]. Combining different types of information using, for example, record linkage approaches [4] may cause serious privacy violations. This problem is exacerbated when multiple datasets that happen to contain the same individual are available, which is rather common in neuroimaging (e.g. [5]). Hence it is natural to ask if anonymized individuals can be identified from neuroimaging datasets and if so to what degree. Specifically, we ask: do we have a *brainprint*, a brain-activity analog of a fingerprint?

If there is evidence for a brainprint, then researchers may be warned about how easily individual information can be inferred, and it may cause them (and the field) to act with more caution when publishing neuroimaging data online. For instance, it may pave the way for the adoption of more sophisticated data-release mechanisms like differential privacy [6] and homomorphic encryption [7].

Brainprints and privacy Assume there are two multi-subject neuroimaging datasets with overlapping participants: a "source" dataset and a "target" dataset. One possible test of "individual identification" can be cast as follows: can we accurately decide which subject in the source dataset

corresponds to a segment of recording from the target set? This could hypothetically arise naturally in practice: it is very common for university labs (comprising faculty, postdocs, and graduate students) to recruit their own lab members for preliminary studies; these are anonymously released with an associated publication. Assume that one year later, lab member A relocates to city B, and privately volunteers for a study by a public hospital that tracks the effect of a drug (or some intervention) on patients in early stages of early-onset Alzheimer's, while collecting MEG data. If this data is also anonymously released at a future point, brainprints could plausibly be used to detect a common participant, thus identifying that A has Alzheimer's because only one member of the lab moved to city B. This would already be a gross unintended violation of privacy, but one can further imagine that an insurance company uses this to prove that a condition was pre-existing at the time of the first scan (before the individual themselves knew), or use it to decide individual-level pricing.

Brainprints and individual variability In multi-subject, multi-session neuroimaging data, there exists "within-session" variability across subjects in the same session and "cross-session" variability of the same subject cross sessions [8]. In Figure 1, we conceptualized possible indications with respect to different combinations of the two types of variability: low variability in both within-session (subjects are similar) and cross-session (a subject's data is consistent across session) suggests a higher statistical power for detecting average group effects with fixed sample size, thereby facilitating reproducibility [9, 10], whereas high cross-session and low within-session variability (e.g. subject 1's data in session 1 is very different from their data for session 2, but somehow very similar to subject 2's data in session 1) may indicate session-specific artifacts (e.g. the scanner was faulty during the recording of session 1 for all subjects). Individual identifiability is closely related to individual variability as high identifiability reflects high within-session (subjects are different from each other) and low cross-session variability (subjects are similar to themselves). This in turn indicates a *consistent* variability among individuals, which in itself is an important topic of scientific enquiry [8, 11]. Understanding sources of consistent variability can help learn the underpinnings of disease or more generally to map the relationship of brain structure and activity to individual behavioral characteristics.

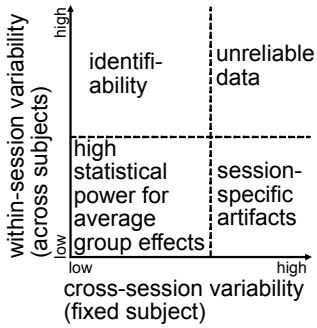

Figure 1: Individual identifiability and variability.

Related work Individual identification has been studied using EEG and fMRI for the purpose of biometric authentication and investigation of the biological difference among individuals. Various features have been proposed, for example, time-series-based statistics and features extracted by a deep neural network [12, 13, 14]. These features may yield high accuracy but are tricky to interpret since they are complicated functions of the data. Another type of feature, functional connectome, has been proven to yield high identification accuracy in fMRI data [15, 16] and used to identify monozygotic twins in MEG [17]. This feature is more interpretable due to its well-understood biological implications. There has also been a line of research looking at inter-subject variability and test-retest reliability using MEG connectivity metrics [8, 18]. However, these studies didn't consider the problem in the context of individual identification.

High-level strategy In this paper, we first ask if individual identification can be easily achieved using a random forest classifier on within-session MEG data without featurization (Section 2). Since there was high identification accuracy, we looked into what contributes to the high accuracy by proposing a set of interpretable features and test their performance (Section 2). Given that the features achieve high classification accuracy within session, we proceeded to look at their performance on multi-session data to rule out environmental session-specific artifacts (Section 3). Since the identification

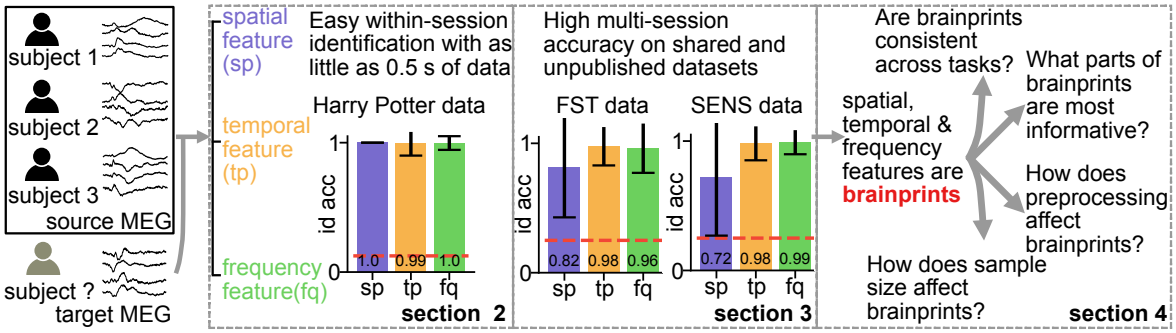

Figure 2: Graphical summary of the paper.

accuracy was preserved, we found it highly plausible that the features are brainprints. We leveraged the interpretability of these features to see if different states of the brain affect brainprints, which specific components are most informative, and how sample size and preprocessing affect brainprints (Section 4). In Section 2-4, we focused on identification alone and finally, we revisit its link to privacy and variability in the discussion. A graphical summary of our results is in Figure 2.

2 Within-session identification with raw and interpretable features is surprisingly easy

We first investigated if a subject can be identified using MEG data of the same session.

Within- vs cross- session We call a pair of *source* and *target* sets "within-session" if, for each individual, both datasets were collected in the same visit to the scanner. For example, two blocks of a resting-state recording of a participant collected on the same day are within-session. If the two datasets are collected on different days for each individual, they are "cross-session". For example, a resting state recording on day 1 and another resting-state recording on day 2 are cross-session. Individuals with within-session data may be easier to identify since the source and target data were collected under almost the same environment.

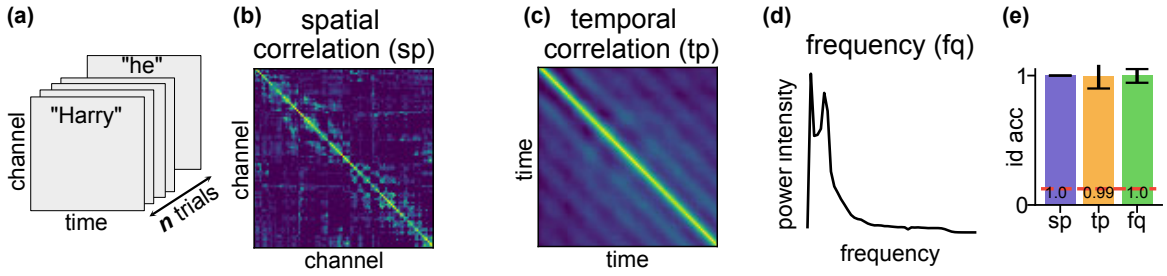

Figure 3: Within-session identification on Harry Potter data. **a.** Shape of the Harry Potter data. **b.** **sp** is a 102×102 correlation matrix. **c.** **tp** is a 100×100 correlation matrix. **d.** **fq** is a vector in $\mathbb{R}^{51}$ where 51 is the number of frequency values. **e.** Identification accuracy with the three features. Red dashed line represents the chance level. Error bars across subjects and identification runs.

Harry Potter data We considered a MEG dataset of *eight* subjects during a reading task, shared with us by the authors of [19]. Subjects were asked to read a chapter of Harry Potter [20] while each word was presented for 0.5 s on a screen. There were 306 sensors at 102 locations where each

location has one magnetometer and two planar gradiometers whose signal was averaged. The sampling frequency of the data was 1000 Hz which was further downsampled to 200 Hz. Details about the preprocessing of all the datasets in this paper can be found in Supplemental Section A. The data was parsed into trials where each trial corresponds to the MEG recording when an individual was reading a word. Specifically, the trials of subject k is $\{X_i^k \in \mathbb{R}^{102\times 100}\}_{i=1}^{I_k}$ where I_k is the number of trials for subject k, 102 represents the number of spatial channels, and 100 represents the number of temporal points in the trial (Figure 3 (a)). Since the recording of each individual was collected in one session, we simply split the data into a target and source dataset for the within-session identification task.

Random forest identification with raw features We first trained a random forest classifier with 256 estimators by first concatenating all the trials of all the subjects along the time dimension, resulting in $\mathbf{X} \in \mathbb{R}^{102\times N}$ where $N = \sum_{k=1}^{8} 100 I_k$ is the total number of time points of all the subjects. The training data is $\{X(:,i) \in \mathbb{R}^{102}\}_{i=1}^{N}$ corresponding to the signal across all channels at one time point, and the training label is $y_i \in \{1, 2, \cdots, 8\}$. Data was z-scored by channel separately on training and testing data. A 10-fold cross-validation yielded a $94.1 \pm 0.18\%$ classification accuracy (mean $\pm$ standard error across folds). The high accuracy suggests that, when the source and target set are from the same session, we can identify individuals with high accuracy *from only 0.005s of MEG!*

Interpretable MEG features The high accuracy in random forest with raw features may not directly explain what specific brain-related information makes the identification easy. MEG enjoys high temporal resolution and adequate spatial resolution, enabling it to capture various characteristics, such as the spatial connectivity, temporal rhythm, and frequency of brain activities [21, 22, 23]. Hence we would like to consider a set of *interpretable* features representing these well-known characteristics of brain activities in MEG. In particular, we propose the following three features (Figure 3 (b)-(d)):

1- Spatial correlation (**sp**): Pearson correlation between channels averaged over time
2- Temporal correlation (**tp**): Pearson correlation between time points averaged over channels
3- Frequency (**fq**): power spectrum averaged over channels

sp is the spatial correlation between different sensors which may be related to individual-specific correlated activities between areas of the brain or the anatomy of the subject. [8, 24]. A high value in the **tp** matrix indicates highly synchronous brain signals between two temporal points, which may be individual-specific when the same stimuli were presented. A relevant study shows that the temporal change of brain activities in auditory steady-state responses are different between individuals [25]. **fq** represents the distribution of the power intensity of signal frequency. Specific frequency bands may be unique to an individual [23].

Identification using 1NN We adopted a 1NN identification similar to [15] to study if the three features are *brainprints* for the within-session identification task. We performed $R = 100$ identification runs. In identification run r, we randomly split the Harry Potter dataset into non-overlapping source and target set, z-scored the source and target by channel separately, and computed the feature $x_{i,r,f}^{\alpha}$ averaged over $n = 300$ randomly sampled trials using data $\alpha \in \{\text{target, source}\}$ for subject i and $F \in \{\mathbf{sp}, \mathbf{tp}, \mathbf{fq}\}$. The features from the target to the source set were matched with a labeling with replacement protocol :

$$\hat{y}(x_{i,r,F}^{\text{target}}) = \underset{j\in\{1,2,\cdots,K\}}{\text{argmax}} \; m(x_{i,r,F}^{\text{target}}, x_{j,r,F}^{\text{source}})$$

where $K = 8$ is the total number of subjects and $m(\cdot,\cdot)$ is the similarity function measuring the similarity between the two features. We used Pearson correlation as our similarity function. The identification accuracy for individual i and feature F is $\frac{1}{R}\sum_{r=1}^{R} \mathbb{1}_{\hat{y}(x_{i,r,F}^{\text{target}})=i}$. The averaged identification accuracy for feature F is $\frac{1}{KR}\sum_{i=1}^{K}\sum_{r=1}^{R} \mathbb{1}_{\hat{y}(x_{i,r,F}^{\text{target}})=i}$.

The identification accuracy is shown in Figure 3 (e): with $n = 300$ trials all three features achieve near-perfect identification accuracy. The high identification accuracy with these features suggests they are *brainprints*, at least for identifying individuals within a session. It turns out that much fewer number of trials is required to attain a similar accuracy (Supplemental Section C). The high accuracy based on merely 0.5 s of data for **sp** and 25 s for **tp** and **fq** is striking since the insufficient data usually leads to inaccurate estimates of these features unless the individual patterns are strong.

3 Cross-session identification confirms the existence of brainprints

The high within-session identification accuracy suggests **sp, tp**, and **fq** are individual-specific within a session. Artifacts such as environmental noise and equipment configurations, however, might be the main contributing factor to the accuracy of the within-session experiment in Section 2. In the current section, we examined the consistency of the three features when the same type of task data was collected from each subject on multiple sessions. This setting tests if the features are preserved over time, i.e. if they are indeed *brainprints* and not mere artifacts.

FST and SEN data We looked at two multi-session datasets:

1- FST data [26], shared online:[1] individuals saw faces with each face appearing on the screen. Each trial lasted 0.5 s. There were 4 subjects and 4 sessions. The sampling frequency was 1000 Hz and was downsampled to 200 Hz. Intervals between consecutive sessions were several days.

2- SEN data (unpublished anonymized citation): individuals read sentences. Each trial lasted 0.5 s. There were 4 subjects and 3 sessions. The sampling frequency was 1000 Hz and was downsampled to 200 Hz. Intervals between consecutive sessions ranged from days to weeks.

The shape of one trial of the two datasets is 102 channels by 100 time points, the same as the Harry Potter data. We used 300 trials to create features for each run of identification.

Cross-session identification with 1NN We used the similar identification approach as defined in Section 2. When the source set and target set were from the same session, we split the dataset into non-overlapping sets as we did in the within-session identification. We didn't split data when the source and target data are from different sessions (Figure 4 (a)) since there is no potential data leakage. We z-scored the data by channel on the source and target separately.

Rank accuracy In addition to the identification accuracy in Section 2 , we used a relaxed version, the *rank accuracy*. The rank accuracy of individual i on one run of identification (suppressing notations of feature F and run r) is defined as $\frac{1}{K}\text{rank}(m(x_i^{\text{target}}, x_i^{\text{source}}))$ where K is the number of subjects, $\text{rank}(m(x_i^{\text{target}}, x_i^{\text{source}}))$ is over $\{m(x_i^{\text{target}}, x_j^{\text{source}}), j = 1, 2, \cdots, K\}$. The rank accuracy equals to 1 if the feature of the same subject has the largest similarity between the source and target sets among all K subjects, and is $\frac{1}{K}$ if the similarity is the smallest. The rank accuracy captures more information in a failure case where an individual is mis-identified.

As shown in Figure 4, both **tp** and **fq** achieved almost perfect average identification and rank accuracy on both FST and SEN data. **sp** achieved lower but still well above-chance accuracy. The high cross-session identification accuracy of **sp, tp**, and **fq** confirms that it is reasonable to call them brainprints for individual identification in MEG. We included several visual comparisons between the brainrpints of different subjects and sessions in Supplemental Section B to show the consistency of the brainprints across sessions.

Figure 4(d) and (g) indicate that the lower identification accuracy for **sp** was due to low accuracy on a small set of individuals. The identification accuracy of them is not consistently low across all session pairs which is likely due to the change of the spatial alignment of the sensors [27].

4 A closer look at brainprints

The high performance and interpretability of the brainprints make it enticing to study the factors and the underlying mechanism for identification.

4.1 Consistent spatial brainprint between resting-state and task sessions

We looked at the performance of these features between two sessions of different types collected on the same day to test their consistency between different brain states. We compared the features between a

[1] `https://figshare.com/articles/FST_raw_data/4233107`

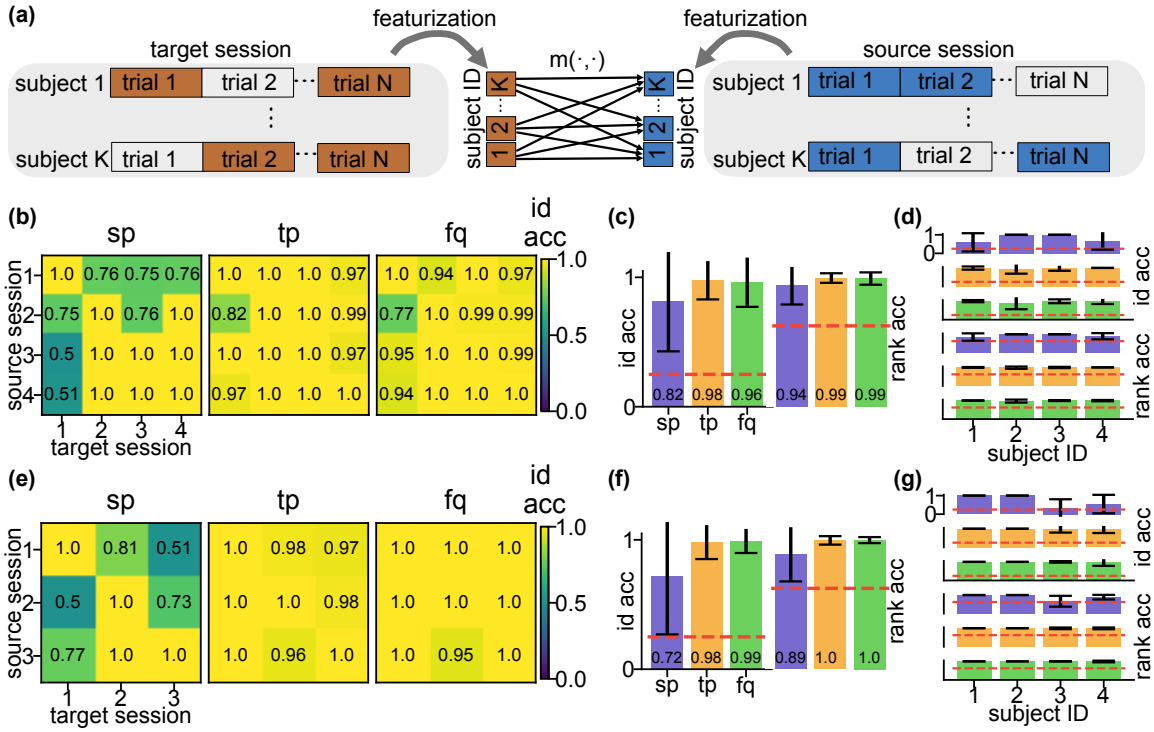

Figure 4: Cross-session identification results on FST and SEN data. **a.** Schema of the cross-session identification task for one identification run **b.** Identification accuracy using the three features on FST data. The red dashed line represents the chance level. The within-session accuracy (diagonal entries) were computed using the same procedure in Section 2 to avoid data leakage. **c.** Average identification accuracy and rank accuracy across source-target pairs on FST data. Within-session accuracy (diagonal elements in **b**) were excluded from the average. Error bars were computed with respect to number of cross sessions × number of subjects × number of featurizations. **d.** Identification and rank accuracy on FST data by individual. **e-g.** Similar to **b-d** but on SEN data.

resting-state session in which subjects rest in the scanner and do not perform a task and a task-MEG session where subjects view images and perform a working memory task.

HCP data We looked at the Human Connectome Project data[2][5] to see if the brainprints are consistent between resting and task states. There were two sessions, one resting-state recording and one working-memory (WM) task recording where the stimuli were images for the participants to remember. Each trial of the WM corresponded to the 2.5 s of the recording after the onset of the stimulus. The two datasets had 77 subjects in common and we only looked at these subjects. There were 146 channels and the signal was downsampled to 200 Hz. The two sessions were collected on the same day with a break of several hours. We used 200 trials for featurization for each run of identification due to fewer number of total trials as compared to the aforementioned datasets.

As shown in Figure 5, **sp** yielded a 77% cross-task identification accuracy, well above the 1.3% random baseline. Therefore, the spatial fingerprint is consistent between different brain states. This result also confirms a similar finding in fMRI [15]. **tp** and **fq** did not perform as well as **sp**, perhaps expectantly, since the temporal rhythm and frequency involved might be different between resting-state and task [28, 29]. The cross-session difference of **tp** and **fq** of the same subject may be larger than

[2] `https://www.humanconnectome.org/study/hcp-young-adult`

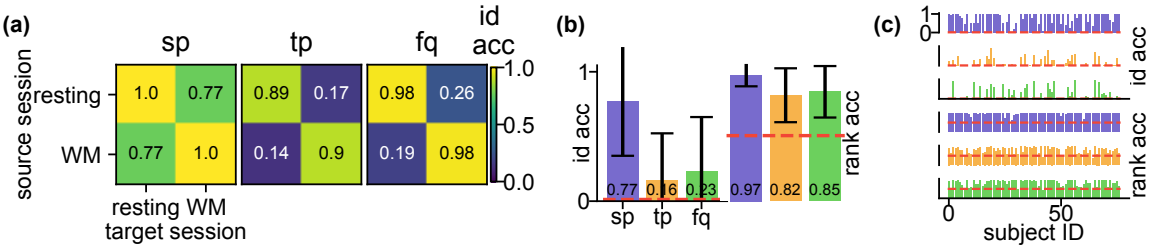

Figure 5: Cross-task identification results between resting and task MEG on HCP data. **a.** Identification accuracy with the three features. Similar to Figure 4 (b). **b.** Average cross-task identification and rank accuracy. Resting-resting and WM-WM accuracy(diagonal elements in **b**) were excluded from the average. **c.** Identification and rank accuracy by individual. Error bars omitted for visualization.

or comparable to the within-session difference between subjects. The rank accuracy of **tp** and **fq** (Figure 5(c)), however, suggests that the majority of subjects still have similar **tp** and **fq** to the same subject in a different session than a random subject. Hence more complicated matching method may be proposed to further boost the performance of these two brainprints.

4.2 Not every part of a brainprint is equally important

We divided the three brainprints into subfeatures:

1- **sp**: The sensors were partitioned into 8 subgroups according to the map in Fig 1 of [30]: Left Frontal (LF), Right Frontal (RF), Left Temporal (LT), Right Temporal (RT), Left Parietal (LP), Right Parietal (RP), Left Occipital (LO), Right Occipital (LO). Each subfeature was the rows and columns of the spatial correlation matrix corresponding to the sensors in one of the eight groups

2- **tp**: The 100 temporal points were divided into 10 consecutive segments containing 10 time points. Each subfeature was the rows and columns of the temporal correlation matrix corresponding to one of the ten segments.

3- **fq**: Each subfeature was the segment of the frequency feature vector corresponding to $[f, f + 10]$ Hz where $f \in \{0, 2, \cdots, 90\}$ Hz.

For both SEN and FST, the correlations between sensors within LO area and between LO and RP yielded high accuracy (Figure 6(a) and Supplemental Section B). LO is involved in visual processing [31] and RP is involved in sensory integration [32], both of which are functions recruited by the experimental task. Due to the nature of the sampled signal and the physical properties of the skull, each MEG sensor samples coarsely from the brain, making it hard to say whether MEG spatial correlation effectively corresponds to functional connectivity, especially for nearby sensors [8]. However, the correlations between faraway groups of sensors, for example, LT and RT, still have good accuracy suggesting it may be due to actual functional correlation between these areas, or perhaps that the difference skull shapes that contribute to the high **sp** accuracy.

For both SEN and FST data, the super-diagonal of the heat map for temporal subfeatures (Figure 6(b) and Supplemental Section B) had high accuracy. The super-diagonal entries correspond to the cross-correlation of the MEG signal between two consecutive segments of 0.05 s. Hence the rhythm of the signal within a short segment of time contributes to identification accuracy, which can also be seen from the banded structure of **tp** (Figure 3(c)). Moreover, the correlations between the first and third 0.05 s segment and between fourth and fifth 0.05 s yield considerably high accuracy. These time periods overlap with the time we expect the brain is processing word and picture stimuli [33].

The power intensity of frequencies between 10 and 30 Hz yielded the highest accuracy on both SEN and FST data. This range roughly corresponds to the alpha and beta frequency band which are related to the resting state and response to attentive cognitive tasks [34].

For the three brainprints, higher accuracy seems to be associated with the components of features with more stimuli-driven activity: the occipital lobe, the time around the stimulus, and frequency bands the with highest power intensities. Indeed, MEG signal is most sensitive to transient, coordinated firings of many neurons that happen after stimulus onset. This commonality indicates the possibility that higher accuracy is related to event-related signals, which in turn suggests that identifiability might be caused by different subjects responding differently to the stimulus. However, this observation could also be explained by a signal-to-noise ratio argument: regions, time-points, or frequencies related to stimulus processing correspond to parts of the underlying brain signal with higher amplitudes (while the ambient noise amplitude is constant). It might be that the increase in signal magnitude make the (spatial, temporal or frequency) activity patterns that are specific to a subject more detectable by increasing their amplitude relative to the ambient noise, even if these patterns are not inherently related to stimulus processing and are just consistent features of a subject's brain activity.

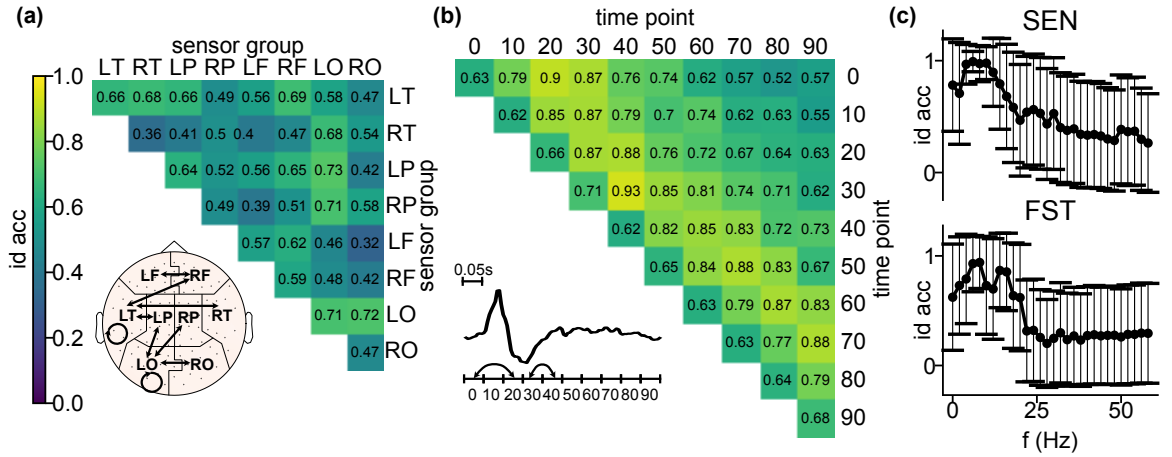

Figure 6: Identification accuracy with components of the brainprints. See Supplemental Section B for **a,b** of FST. **a.** Subfeatures of **sp**.Inset is the sensor group layout and edges correspond to the entries with over 0.6 accuracy. Topomap plotted using the python MNE package [35]. **b.** Subfeatures of **tp**. Inset is an example signal averaged across channels and trials for one subject and edges correspond to the entries of the heatmap with over 0.9 accuracy. **c.** Identification accuracy with respect to certain frequency bands for SEN and FST data. f larger than 60 Hz truncated since the curve will be flat.

4.3 Identification accuracy change with data size and preprocessing

Various factors may affect identification accuracy. The accuracy increase with the amount of data used for computing **sp**, **fp**, and **fq** (Figure 7(a)) as the sampling variance becomes smaller. In general, with 50 s of data, the brainprints perform well on cross-session identification of the same task. **sp** becomes reasonably accurate on the HCP dataset with 100 trials corresponding to 250 s of recording, possibly because the identification is more difficult with more number of subjects.

Preprocessing may also affect identification accuracy. The changes in accuracy were all statistically significant (Chi-square test, $p < 0.001$) when the raw data was preprocessed for all the three features (Figure 7(b)). For both FST and SEN, preprocessing yielded better accuracy for **tp** and **fq**. The artifact removal and temporal filtering in the preprocessing pipeline might have prevented session-specific noise from contaminating individual-specific features, resulting in higher accuracy. There is opposite evidence on the effects of preprocessing on **sp** between FST and SEN data (for FST: 63% VS 82% , for SEN, 89% VS 72%). There was one difference in the preprocessing pipeline for both datasets:

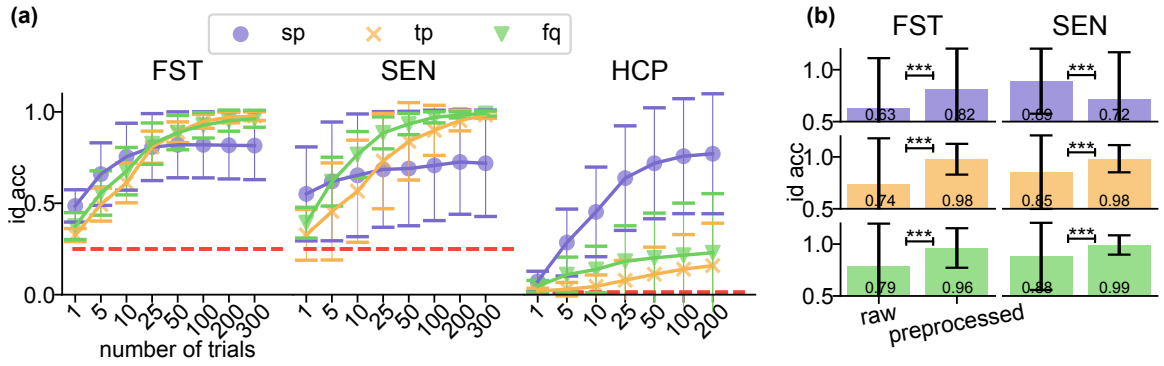

Figure 7: Factors affecting identification accuracy. **a.** Identification accuracy with respect to the number of trials used for featurization. **b.** Comparison of identification accuracy of the three features computed on raw and fully preprocessed data. A Pearson's Chi-square test was performed on the binary identification outcomes across sessions, individuals, and identification runs between the raw and preprocessed data. The same color represents the same feature as in **a**.

FST preprocessing did not include head position correction due to a lack of head position recordings. This alone does not justify our results, as identifiability using **sp** increases after prepossessing when not performing head position correction but decreases when performing it, and it's expected that head position correction would improve identifiability by recentering each subject's data to the same position in each session. Head position correction might be changing the signal in unexpected ways. Future work and analysis of additional datasets are required to investigate this result.

5 Discussion

Overlapping subjects We assumed both the target and source datasets had the same pool of participants in the scope of this paper. If we don't know if one subject from the target set is included in the source set, more complicated classification methods such as conformal prediction [36] may be used to account for the case when no label in the source set can be assigned to the subject.

Revisiting brainprints and privacy The high accuracy of across-session identification is an alarming message to consider before releasing MEG data. The difference in the accuracy between raw and preprocessed data suggests, for example, encrypting the data with session-specific noise may lower identification accuracy. Advances in federated machine learning, through which models can be trained on many datasets in a distributive manner without revealing all datasets to any of the researchers [37], might be a source of solutions to the problem of identifiability when sharing human subject data such as MEG or other neuroimaging data.

Revisiting brainprints and individual variability The existence of brainprints are examples of certain functions of the MEG data with high inter-subject variability preserved across sessions. For example, the high accuracy with **tp** suggests the existence of individual variability in their temporal response to the same stimuli. Understanding brainprints will facilitate the understanding of the underlying anatomical and functional variability between individuals.

Improving brainprints More complicated features can be proposed which combine the spatial, temporal, and frequency information. For example, functional connectivity at different bands has been used to identify twins from other participants [17]. Other feature similarity function $m(\cdot, \cdot)$ may also be used to improve accuracy [38].

References

[1] Thomas Landrain, Morgan Meyer, Ariel Martin Perez, and Remi Sussan. Do-it-yourself biology: challenges and promises for an open science and technology movement. *Systems and synthetic biology*, 7(3):115–126, 2013.

[2] Krzysztof Gorgolewski, Oscar Esteban, Gunnar Schaefer, Brian Wandell, and Russell Poldrack. Openneuro—a free online platform for sharing and analysis of neuroimaging data. *Organization for human brain mapping. Vancouver, Canada*, 1677(2), 2017.

[3] Latanya Sweeney. Simple demographics often identify people uniquely. *Health (San Francisco)*, 671:1–34, 2000.

[4] Halbert L Dunn. Record linkage. *American Journal of Public Health and the Nations Health*, 36(12):1412–1416, 1946.

[5] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.

[6] Jaewoo Lee and Chris Clifton. Differential identifiability. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1041–1049, 2012.

[7] Craig Gentry. Fully homomorphic encryption using ideal lattices. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 169–178, 2009.

[8] Giles L Colclough, Mark W Woolrich, PK Tewarie, Matthew J Brookes, Andrew J Quinn, and Stephen M Smith. How reliable are meg resting-state connectivity metrics? *Neuroimage*, 138:284–293, 2016.

[9] Katherine S Button, John PA Ioannidis, Claire Mokrysz, Brian A Nosek, Jonathan Flint, Emma SJ Robinson, and Marcus R Munafò. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5):365–376, 2013.

[10] Russell A Poldrack, Chris I Baker, Joke Durnez, Krzysztof J Gorgolewski, Paul M Matthews, Marcus R Munafò, Thomas E Nichols, Jean-Baptiste Poline, Edward Vul, and Tal Yarkoni. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature reviews neuroscience*, 18(2):115, 2017.

[11] Vincent Wens, Mathieu Bourguignon, Serge Goldman, Brice Marty, Marc Op De Beeck, Catherine Clumeck, Alison Mary, Philippe Peigneux, Patrick Van Bogaert, Matthew J Brookes, et al. Inter-and intra-subject variability of neuromagnetic resting state networks. *Brain topography*, 27(5):620–634, 2014.

[12] M Poulos, M Rangoussi, N Alexandris, and A Evangelou. Person identification from the eeg using nonlinear signal classification. *Methods of information in Medicine*, 41(01):64–75, 2002.

[13] Lan Ma, James W Minett, Thierry Blu, and William SY Wang. Resting state eeg-based biometrics for individual identification using convolutional neural networks. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2848–2851. IEEE, 2015.

[14] Xiang Zhang, Lina Yao, Salil S Kanhere, Yunhao Liu, Tao Gu, and Kaixuan Chen. Mindid: Person identification from brain waves through attention-based recurrent neural network. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–23, 2018.

[15] Emily S Finn, Xilin Shen, Dustin Scheinost, Monica D Rosenberg, Jessica Huang, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, 18(11):1664, 2015.

[16] Oscar Miranda-Dominguez, Brian D Mills, Samuel D Carpenter, Kathleen A Grant, Christopher D Kroenke, Joel T Nigg, and Damien A Fair. Connectotyping: model based fingerprinting of the functional connectome. *PloS one*, 9(11), 2014.

[17] M Demuru, AA Gouw, A Hillebrand, CJ Stam, BW Van Dijk, P Scheltens, BM Tijms, E Konijnenberg, M Ten Kate, A Den Braber, et al. Functional and effective whole brain connectivity using magnetoencephalography to identify monozygotic twin pairs. *Scientific reports*, 7(1):1–11, 2017.

[18] Pilar Garcés, María Carmen Martín-Buro, and Fernando Maestú. Quantifying the test-retest reliability of magnetoencephalography resting-state functional connectivity. *Brain connectivity*, 6(6):448–460, 2016.

[19] Leila Wehbe, Ashish Vaswani, Kevin Knight, and Tom Mitchell. Aligning context-based statistical models of language with brain activity during reading. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 233–243, 2014.

[20] JK Rowling. *of Book: Harry Potter and the Sorcerer's Stone*. Year: Scholastic Press/1997, 1997.

[21] Matthew J Brookes, Joanne R Hale, Johanna M Zumer, Claire M Stevenson, Susan T Francis, Gareth R Barnes, Julia P Owen, Peter G Morris, and Srikantan S Nagarajan. Measuring functional connectivity using meg: methodology and comparison with fcmri. *Neuroimage*, 56(3):1082–1104, 2011.

[22] Francesco De Pasquale, Stefania Della Penna, Abraham Z Snyder, Christopher Lewis, Dante Mantini, Laura Marzetti, Paolo Belardinelli, Luca Ciancetta, Vittorio Pizzella, Gian Luca Romani, et al. Temporal dynamics of spontaneous meg activity in brain networks. *Proceedings of the National Academy of Sciences*, 107(13):6040–6045, 2010.

[23] Douglas Owen Cheyne and Paul Ferrari. Meg studies of motor cortex gamma oscillations: evidence for a gamma "fingerprint" in the brain? *Frontiers in human neuroscience*, 7:575, 2013.

[24] Sofie S Meyer, James Bonaiuto, Mark Lim, Holly Rossiter, Sheena Waters, David Bradbury, Sven Bestmann, Matthew Brookes, Martina F Callaghan, Nikolaus Weiskopf, et al. Flexible head-casts for high spatial precision meg. *Journal of neuroscience methods*, 276:38–45, 2017.

[25] H-RM Tan, Joachim Gross, and PJ Uhlhaas. Meg—measured auditory steady-state oscillations show high test–retest reliability: A sensor and source-space analysis. *Neuroimage*, 122:417–426, 2015.

[26] Mark D Vida, Adrian Nestor, David C Plaut, and Marlene Behrmann. Spatiotemporal dynamics of similarity-based neural representations of facial identity. *Proceedings of the National Academy of Sciences*, 114(2):388–393, 2017.

[27] Qiong Zhang, Jelmer P Borst, Robert E Kass, and John R Anderson. Inter-subject alignment of meg datasets in a common representational space. *Human brain mapping*, 38(9):4287–4301, 2017.

[28] Sheng Zhang and Chiang-Shan R Li. Task-related, low-frequency task-residual, and resting state activity in the default mode network brain regions. *Frontiers in psychology*, 3:172, 2012.

[29] Omer Grigg and Cheryl L Grady. Task-related effects on the temporal and spatial dynamics of resting-state functional connectivity in the default network. *PloS one*, 5(10), 2010.

[30] Yegang Hu, Chunli Yin, Jicong Zhang, and Yuping Wang. Partial least square aided beamforming algorithm in magnetoencephalography source imaging. *Frontiers in neuroscience*, 12:616, 2018.

[31] Tony Ro, Bruno Breitmeyer, Philip Burton, Neel S Singhal, and David Lane. Feedback contributions to visual awareness in human occipital cortex. *Current biology*, 13(12):1038–1041, 2003.

[32] Oliver Jakobs, Robert Langner, Svenja Caspers, Christian Roski, Edna C Cieslik, Karl Zilles, Angela R Laird, Peter T Fox, and Simon B Eickhoff. Across-study and within-subject functional connectivity of a right temporo-parietal junction subregion involved in stimulus–context integration. *Neuroimage*, 60(4):2389–2398, 2012.

[33] Riitta Salmelin. Clinical neurophysiology of language: the meg approach. *Clinical Neurophysiology*, 118(2):237–254, 2007.

[34] Pravat K Mandal, Anwesha Banerjee, Manjari Tripathi, and Ankita Sharma. A comprehensive review of magnetoencephalography (meg) studies for brain functionality in healthy aging and alzheimer's disease (ad). *Frontiers in Computational Neuroscience*, 12:60, 2018.

[35] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S Hämäläinen. Mne software for processing meg and eeg data. *Neuroimage*, 86:446–460, 2014.

[36] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(Mar):371–421, 2008.

[37] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.

[38] Manasij Venkatesh, Joseph Jaja, and Luiz Pessoa. Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification. *NeuroImage*, 207:116398, 2020.

[39] Eric Jones, Travis Oliphant, and Pearu Peterson. Scipy: Open source scientific tools for python. 2001.

[40] Samu Taulu and Matti Kajola. Presentation of electromagnetic multichannel data: the signal space separation method. *Journal of Applied Physics*, 97(12):124905, 2005.

[41] Samu Taulu and Juha Simola. Spatiotemporal signal space separation method for rejecting nearby interference in meg measurements. *Physics in Medicine & Biology*, 51(7):1759, 2006.

[42] Mikko A Uusitalo and Risto J Ilmoniemi. Signal-space projection method for separating meg or eeg into components. *Medical and Biological Engineering and Computing*, 35(2):135–140, 1997.

Supplementary Material

A. Data preprocessing

Here we list the preprocessing steps applied to the four types of data sets: Harry Potter (HP), SEN, FST, and Human Connectome Project (HCP). A summary is listed in table S1. For all data sets, we used an order 8 Chebyshev type I anti-aliasing filter in Python Scipy package[39] for downsampling. For any within-session identification task, data was z-scored within its corresponding type of dataset (target vs source). Some steps of preprocessing were performed using the python MNE package [35].

1- **HP/SEN**: The 306-channel Elekta Neuromag system was used for the recording. Source-space separation (SSS) along with Maxwell filtering and their temporal extension (tSSS) [40, 41] were used for bad channel correction, head position correction, and electromagnetic artifacts removal. Empty room artifacts were removed. $1 \sim 150$ Hz bandpass filter and 60 & 120 Hz notch filter were used to remove line noise. Heartbeats and eyeblinks artifacts were removed with signal-space projection (SSP) [42]. The data was downsampled to 200 Hz and z-scored by channel within each subject and session.

2- **FST** (preprocessing pipeline was included in the source code): The 306-channel Elekta Neuromag system was used for the recording. Source-space separation (SSS) along with Maxwell filtering and their temporal extension (tSSS) were used for bad channel correction and electromagnetic artifacts removal. Empty room artifacts were removed. We didn't perform head position correction since there was no head position data. $1 \sim 150$ Hz Bandpass filter and 60 & 120 Hz Notch filter were used to remove line noise. Heartbeats and eyeblinks artifacts were also removed with SSP. The data was downsampled to 200 Hz and z-scored by channel within each subject and session.

3-**HCP**: Both resting and WM data sets were already preprocessed and downloaded from the HCP database[3]. The details of the preprocessing pipeline can be found at `https://www.humanconnectome.org/storage/app/media/documentation/s1200/HCP_S1200_Release_Reference_Manual.pdf`. MAGNES 3600 (4D Neuroimaging, San Diego, CA) system was used for the recording. For WM data, we looked at the TIM partition which corresponds to $-1.5 \sim 2.5$ s relative to the onset of the image. For both resting and WM data, the sampling frequency of the preprocessed data is 508.63 Hz, and 2 s of data were selected from each trial. This corresponds to the whole 1018 time points in the resting data and $[763 : 1780]$-th time point for the WM data (corresponding to $0 \sim 2$ s relative to the onset of the image). The 2 s data was then downsampled to 101.73 Hz. Data was z-scored by channel within each subject and each data type (resting and WM). We looked at the 146 channels which were marked "good" among all the 77 overlapping subjects between resting and WM.

Table S1: Summary of the preprocessing stpes for HP, SEN, FST, and HCP data

Steps	**HP/SEN**	**FST**	**HCP**
bad data	corrected	corrected	removed
head position	corrected	not corrected	not corrected[4]
electromagnetic artifacts	removed using SSS	removed using SSS	removed with bad data
empty room artifacts	removed	removed	removed[5]
band filtering	$1 \sim 150$ Hz	$1 \sim 150$ Hz	$1.3 \sim 150$ Hz
notch filtering	60 & 120 Hz	60 & 120 Hz	$59-61$&$119-121$ Hz
ECG (heartbeat) artifacts	removed with SSP	removed with SSP	removed with ICA
EOG (eyeblink) artifacts	removed with SSP	removed with SSP	removed with ICA
downsampling	200 Hz	200 Hz	101.73 Hz
z-scoring	by channel within subject, session	same	same
shape of a trial $[channels, timepoints]$	$[102, 100]$	$[102, 100]$	$[146, 204]$

[3] `https://www.humanconnectome.org/study/hcp-young-adult`

[4] No continuous recording of head position was available in HCP data

[5] page 68 of `https://www.humanconnectome.org/storage/app/media/documentation/s1200/HCP_S1200_Release_Reference_Manual.pdf`

B. Details of the brainprints

B1. Sensor layout for FST, SEN, and HP data

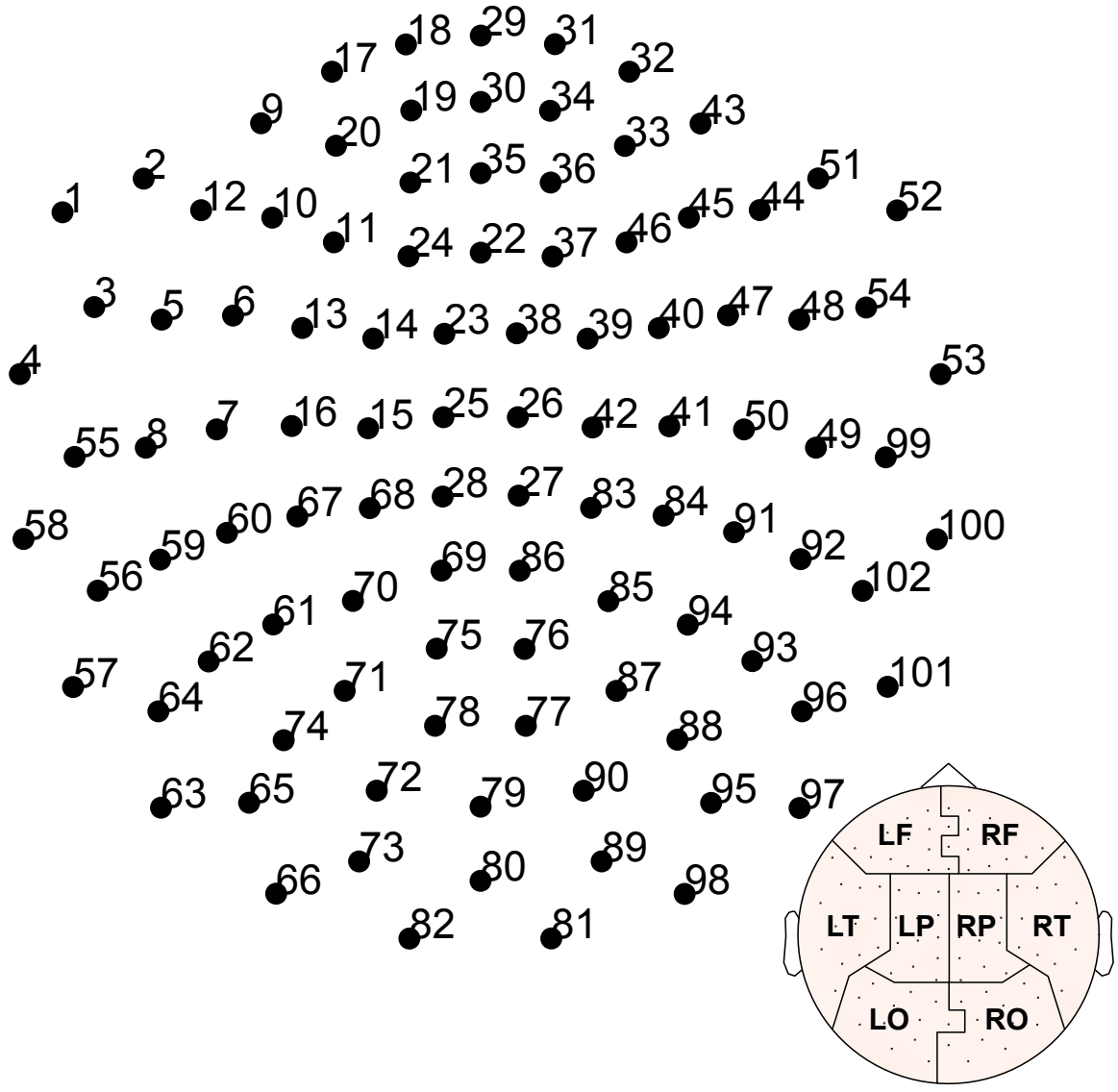

Figure S1: Layout of the sensors for FST, HP, and SEN data (306-channel Elekta Neuromag system). Channel numbers are consistent with the channel index in Fig S2. Inset is the partitioning of the sensors same as Figure 6 (a) of the main text.

B2. Example brainprints of FST data

Note: we have emphasized the importance of preserving individual privacy throughout the paper. Since the FST data set is published online and our way of computing brainprints (as either discussed in the main text or the source code) will eventually be publicly available, showing individual brainprints will not reveal new information about the subjects. Hence we decided to include the following examples of brainprints to show more intuition behind the high identification accuracy of the three brainprints.

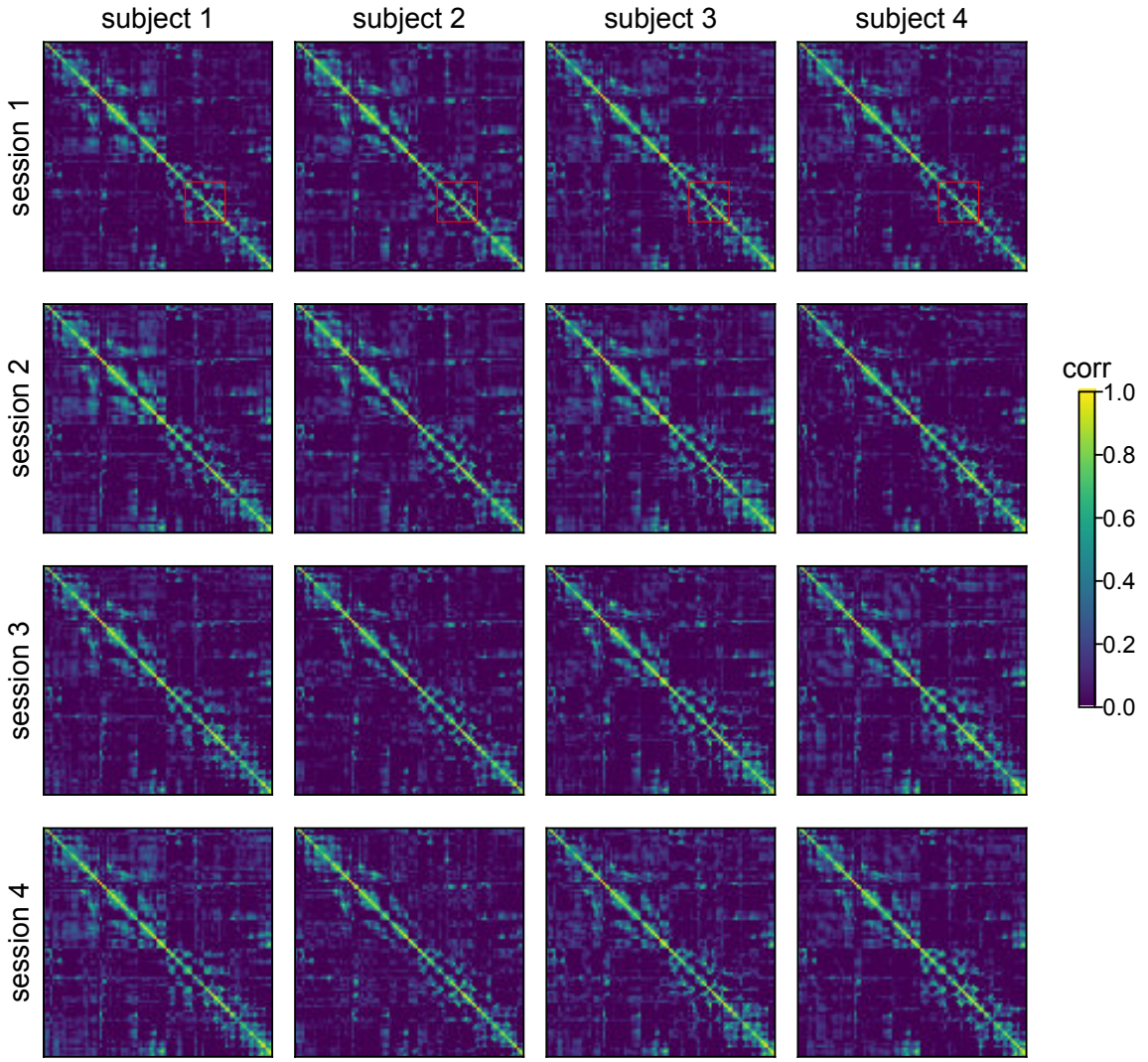

Figure S2: Example **sp** (spatial connectivity) of FST data. Each heatmap represents a 102×102 spatial correlation matrix. The general patterns of the correlation matrices are similar to each other. Some subsets of the heatmap, for example, the bottom-right corner, the top-left corner, and the red rectangle areas are more consistent within a subject and different between subjects. This suggests that only the interactions among a subset of sensors are individual-specific. The red rectangle areas, in particular, roughly correspond to the correlations within the left occipital (LO) lobe which yields the highest identification accuracy on both FST and SEN data (see Figure 6(a) and Figure S5). More complicated comparison algorithms may be proposed to focus on these specific subsets to improve the identification accuracy.

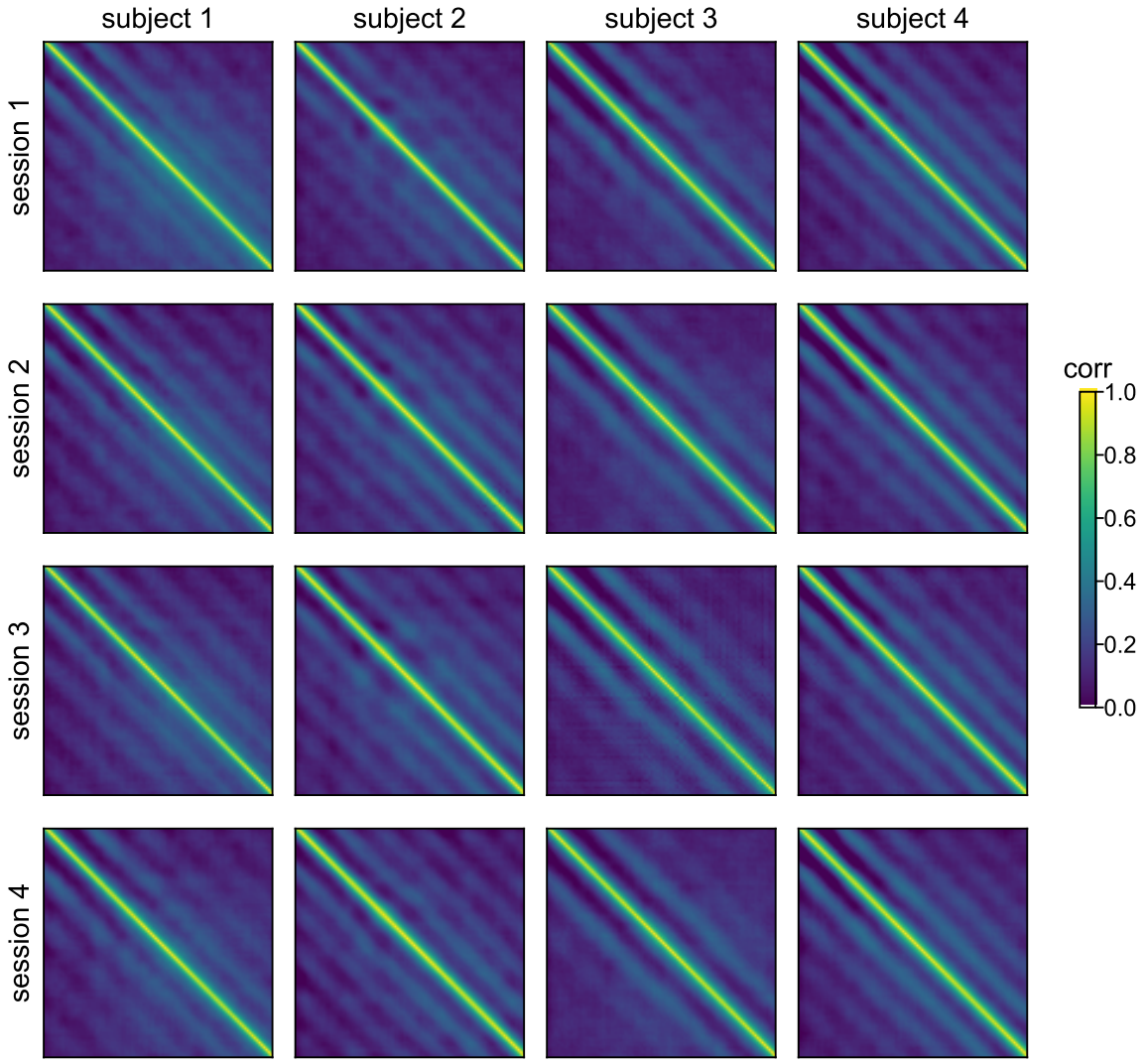

Figure S3: Example **tp** (temporal connectivity) of FST data. Each heatmap represents a 100×100 temporal correlation matrix. The banded structure of the matrices are preserved for the same subject across sessions, and are different between subjects in terms of the number of bands and the relative locations of the bands. The banded structure indicates that there are stronger correlations of the signal with itself at certain lags. In other words, looking at the auto-correlation of the signal or even cross-correlation between different channels may reveal interesting results about the temporal dynamics of the brain activities. The individual-specific band structures also confirm the findings in Figure 6 (b) that correlations of the signal with itself at certain lags are best able to identify individuals.

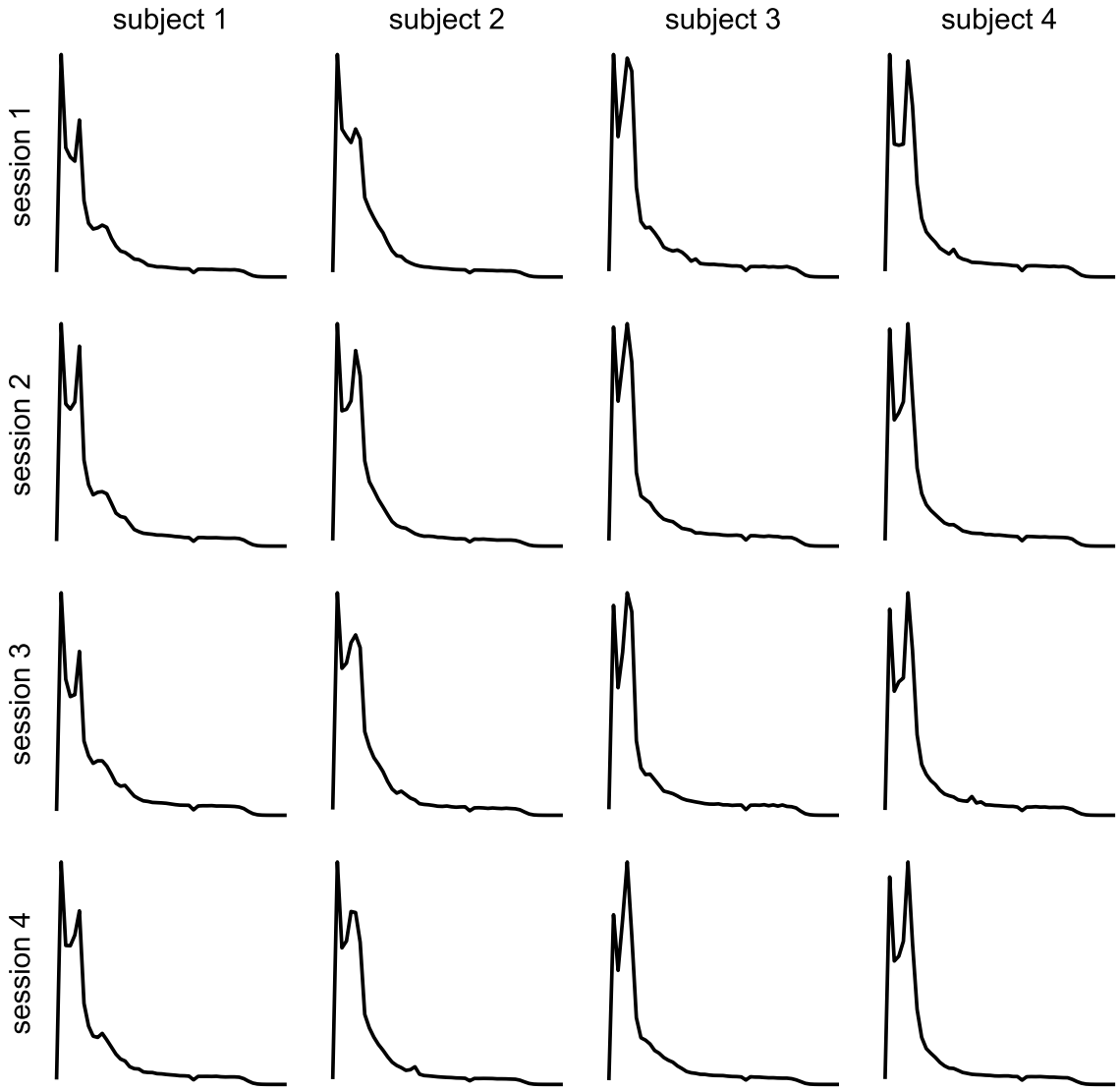

Figure S4: Example **fq** (frequency) of FST data. Each plot represents spectrum (averaged across channels) vs. frequencies (Hz), where the range of frequencies is $[0, 100]$ with a 2 Hz increment. For all subjects, there are two peaks in the power spectrum. The two peaks correspond to around 5 and 10 Hz. The relative height of the two peaks as well as the shape of the curve near the two peaks are consistently unique to an individual across sessions and different across individuals. There are also small peaks near 20 Hz for some subjects. These frequencies with higher amplitudes seem to align with the results shown in Figure 6 (c) where the frequency band near 10 Hz yields the highest identification accuracy. Hence the components of **fq** associated with more stimuli-driven activity or larger signal-to-noise ratio seem to yield better results.

B3. Identification accuracy with components of brainprints for FST data

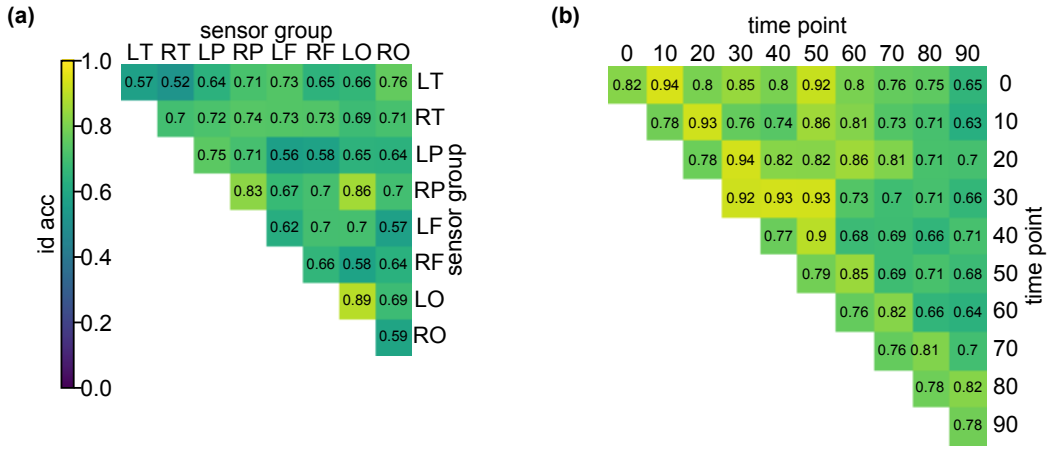

Figure S5: Identification accuracy with sub features for **a: sp** and **b: tp** in FST data (similar to Figure 6 in the main text). For both FST and SEN, the within-LO and LO-RP correlations yield high identification accuracy. Similarly, for both FST and SEN, the super-diagonal and the correlations between the fourth and fifth 0.05 s yield high accuracy. The consistency of the results on the two data sets suggest that our conclusions in Section 4.2 are not due to experiment-specific artifacts.

C. Identification accuracy vs. sample size for Harry Potter data

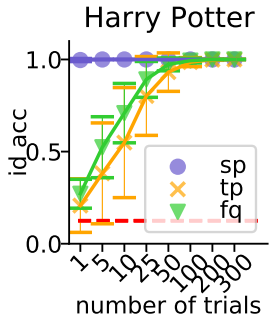

Figure S6: Identification accuracy of **sp, tp**, and **fq** on the Harry Potter data. Each trial is 0.5s in length. The trends for **tp** and **fq** are similar to that of the cross-session data (SEN and FST). **sp** requires as few as one trial to achieve a perfect accuracy. This indicates strong spatial patterns in the HP data which are specific to each individual. This is expected since HP does not have more than one session, and the identification accuracy for **sp** may be lower if there are multiple sessions in HP data, similar to what we have observed on FST and SEN data.