Supplementary Data

Common Symbolization Approach for Pairs of Magnetoencephalography Time Series

Initially, this procedure entails transcribing pairs of time series into two distinct symbolic sequences that share a common codebook (set of symbols). The size and content of the codebook are data dependent. The unsupervised Neural Gas (NG) algorithm maintained a relatively low computational load. Given signals ${}^{\mathrm{A}}\mathrm{x_t}$ and ${}^{\mathrm{B}}\mathrm{x_t}$ from a pair of channels A and B, time-delay vectors were first reconstructed from each time series. These vectors take the form $\mathbf{x_t} = \{x_t, x_{(t+\tau)}, \cdots, x_{(t+(m-1)\tau)}\}$ where m is the embedding dimension, τ denotes the time lag, and t = 1,2,..,T runs over the time points. Then, the two individual sequences of time-delay vectors are collectively gathered in data matrices:

$${}^{A}\mathbf{X}_{[\mathrm{T}\times\mathrm{m}]}=[{}^{\mathbf{A}}\mathbf{x}_1|{}^{A}\mathbf{x}_2|\ldots|{}^{\mathbf{A}}\mathbf{x}_T] \;\&\; {}^{B}\mathbf{X}_{[\mathrm{T}\times\mathrm{m}]}=[{}^{B}\mathbf{x}_1|{}^{B}\mathbf{x}_2|\ldots|{}^{B}\mathbf{x}_T]$$

Next, the two trajectories were brought to a common reconstructed space by forming the overall data matrix:

$$^{AB}\mathbf{X}_{[2\mathrm{T}\times\mathrm{m}]} = [{}^{\mathbf{A}}\mathbf{X}|{}^{B}\mathbf{X}]. \tag{1}$$

The partition of all tabulated m-dimensional vectors into groups of homogenous patterns is the most direct way to summarize the temporal variations in the dynamics of each original time series and describe them with a common vocabulary.

In our approach, a codebook of k code vectors was designed by applying the NG algorithm to the data matrix ${}^{\mathrm{AB}}\mathbf{X}$, which is of size [$\sim$2T$\times$m]. The NG algorithm is an artificial neural network model, which converges efficiently to a small number of k<<T of codebook vectors $\{\mathbf{M_i}\}_{i=1:k}$ using a stochastic gradient descent procedure with a softmax adaptation rule that minimizes the average distortion error (Martinetz et al., 1993).

In the encoding stage, each of the 2T vectors was assigned to the nearest code vector. By replacing the original vectors with the assigned code vectors, we could rebuild the two vectorial time series with a measurable error. If we denote the reconstructed (i.e., decoded) version of the vectorial time series as $^{AB}\mathbf{X}^{\mathbf{rec}}(t)$, we can estimate the fidelity of the overall encoding procedure with the following index, which is the total distortion error divided by the total dispersion of the original vectors:

$$n_{Distortion} = \frac{\sum_{t=1}^{2T} ||{}^{AB}\mathbf{x}(t) - {}^{AB}\mathbf{x}^{\mathbf{rec}}(t)||^2}{\sum_{t=1}^{2T} ||{}^{AB}\mathbf{x}(t) - \bar{x}||^2}, \quad \mathbf{x} = \frac{1}{2T}\sum_{t=1}^{2T} {}^{AB}\mathbf{x}(t) \tag{2}$$

The smaller the $n_{Distortion}$, the better the encoding. This index gets smaller at higher k values, reaching a plateau for a relative small value of k. In the present study, we considered encoding to be acceptable if it was produced with the smallest k that satisfied the condition that $n_{Distortion}$ should be less than 5%. Hence, we repeatedly applied the NG algorithm with increasing k and measured the reconstruction quality. In this way, we defined the optimal $\mathrm{k_{0,}}$ which in turn defined the codebook to use in the subsequent symbolization scheme. At the vector-quantization stage, each vector of ${}^{\mathbf{A}}\mathbf{X}$ and ${}^{\mathbf{B}}\mathbf{X}$ is assigned (according to the nearest prototype rule) to the most similar among the derived code vectors $\{\mathbf{M_i}\}_{j=1:ko}$. This step completes the mapping of original time series to two symbolic time series ${}^{\mathrm{A}}\mathrm{s_t}$ and ${}^{\mathrm{B}}\mathrm{s_t}$, $t = 1, 2, \ldots T$, which in mathematical notation reads as follows:

$$\begin{aligned}
&[{}^{A}x_t, {}^{B}x_t] \in R^2\\
&{}^{\mathbf{A}}\mathbf{X_t} \xrightarrow{VQ} \mathrm{M_{j1}} \in \{M_i\}_{i=1}^{K_o}, M_i \in R^m\\
&\qquad {}^{B}\mathbf{X_t} \xrightarrow{VQ} \mathrm{M_{j_2}} \in \{M_i\}_{i=1}^{K_o}, M_i \in R^m\\
&{}^{A}x_t \rightarrow {}^{A}S_t = \mathrm{j_1(t)}, \quad {}^{B}x_t \rightarrow {}^{B}S_t = \mathrm{j_2(t)}, \quad \mathrm{j_1, j_2} \in \{1, 2, .., k_o\}.
\end{aligned} \tag{3}$$

In the derived symbolic time series, the temporal dynamics of a pair of sensors is encoded as transitions among adaptively defined (i.e., data-dependent) symbols. We adopted the Ragwitz criterion for optimizing the embedding dimension m and the embedding delay τ (Ragwitz and Kantz, 2002). Optimality of the embedding refers to a minimal prediction error for future samples of the time series. The Ragwitz criterion predicts subsequent values in a time series based on estimates of the probability densities of future values of the nearest neighbors after embedding. The adopted method is based on the minimization of mean squared prediction error (Lindner et al., 2011; Ragwitz and Kantz, 2002).

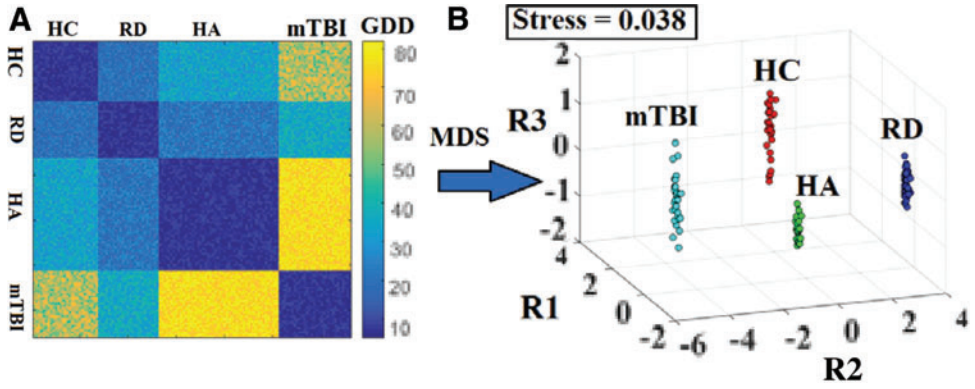
Dominant Coupling Modes Based on the Phase Lag Index

A complementary index of interdependence between time series, in the form of the phase lag index (PLI) (Stam et al., 2007), which is considered to be less susceptive to volume conduction, was also computed. PLI reflects coupling strength by assessing consistent, nonzero phase lags between two time series as reflected in the asymmetry of the distribution of instantaneous phase differences between two signals (i.e., phase differences that deviate from 0 mod π). If no phase coupling exists between two time series, then this distribution is expected to be flat; therefore, any deviation from this flat distribution indicates phase synchronization. An index of the asymmetry of the phase difference distribution can be obtained from a time series of phase differences $\Delta\Phi(tk)$, $k = 1\ldots N$ according to the following formula:

$$\mathrm{PLI} = |<\mathrm{sign}[\Delta\Phi(tk)]>| \tag{4}$$

PLI values range between 0 and 1. A PLI of zero indicates either no coupling or coupling with a phase difference centered around 0 mod π. A PLI of 1 indicates perfect phase locking at a value of $\Delta\Phi$ different from 0 mod π.

Subject-specific functional connectivity networks were then computed using identical procedures as in the case of symbolic mutual information, followed by topological



SUPPLEMENTARY FIG. S1. Group differences revealed through OMST based on PLI. **(A)** Dissimilarity matrix of subject-specific FCGs ($N = 132$) based on the GDD metric. **(B)** MDS was applied to the dissimilarity matrix of GDD values, which were rescaled and projected to a common three-dimensional space. Stress indicates the% loss of information due to the dimensionality reduction process via the MDS algorithm. FCGs, functional connectivity graphs; GDD, graph diffusion distance; HA, healthy adults; HC, typically developing (healthy) children; MDS, multidimensional scaling; mTBI, mild traumatic brain injury patients; OMST, orthogonal minimal spanning tree; PLI, phase lag index; RD, reading-disabled children.

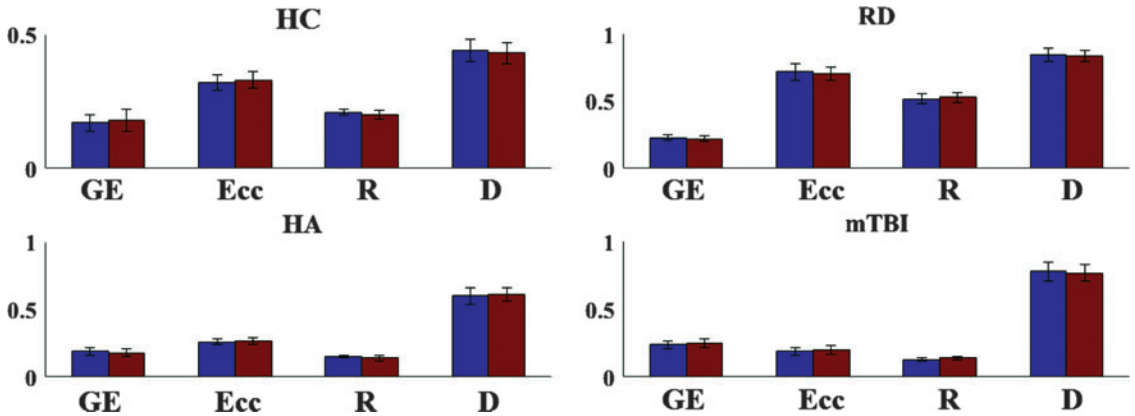
filtering using orthogonal minimal spanning tree (OMST). Then, we estimated the graph diffusion distance matrix for every pair of individual functional brain networks and, in the context of multidimensional scaling, we projected the derived pairwise distances in a common three-dimensional space. As shown in Supplementary Figure S1, the spatial separation of the four groups of participants is unequivocal.

Reliability of OMST Network Metrics

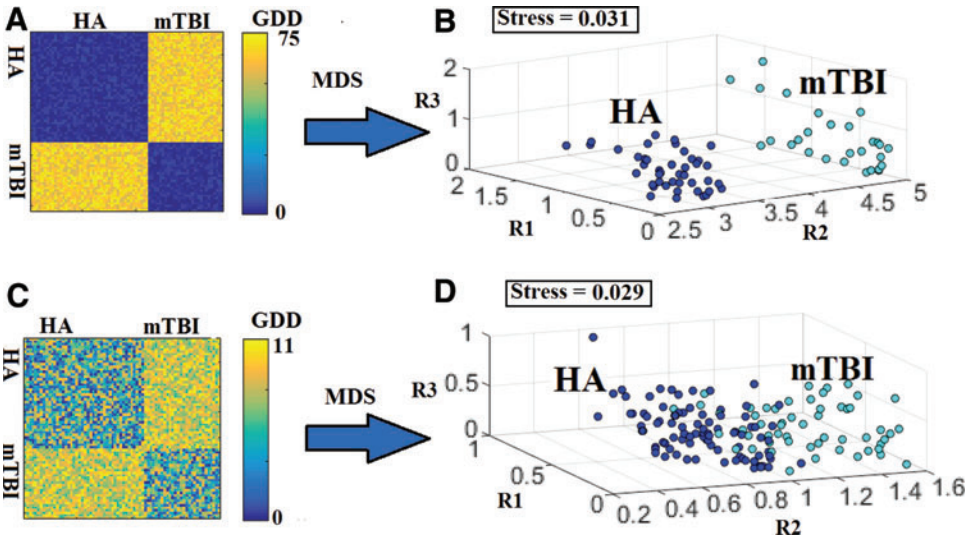
To ensure that gender-related effects did not contribute to the age- and/or diagnostic group-differences, we computed network metrics for gender-matched subgroups. As shown in Supplementary Figure S2, average network metrics were very similar between split-half subgroups ($p > 0.15$ in all cases). Men/women ratios and average age (±1SD) were as follows: 15/15 (32.3 ± 7.12 years) for the first and 14/16 (33.8 ± 8.37 years) for the second healthy adult subgroup; 9/6 (32.6 ± 9.67 years) and 8/7 (32.1 ± 10.12 years) for the two mild traumatic brain injury (mTBI) subgroups; 6/8 (10.32 ± 2.5 years) and 6/7 (10.69 ± 2.8 years) for the two typically developing child subgroups; and finally, 6/7 (11.35 ± 2.68 years) and 5/7 (10.88 ± 2.3 years) for the reading-disabled child subgroups.

OMST-Based Differences Between Age- and Gender-Matched Healthy Adults and mTBI Patients

Although the two child groups were matched on gender ratio, there was a higher percentage of women among healthy adults (62%) compared to the mTBI group (43%). To ensure that gender did not contribute to group discrimination, the analysis was repeated on a subgroup of healthy adults consisting of 22 men and 22 women and a subgroup of adults with a history of mTBI consisting of 14 men and 14 women. As shown in Supplementary Figure S3, the



SUPPLEMENTARY FIG. S2. Split-half, group-averaged network metrics of the topologically filtered PLI-based FCGs. Blue and purple bars indicate two randomly formed, gender-matched subgroups. D, diameter; Ecc, eccentricity; GE, global efficiency; HA, healthy adults; HC, typically developing (healthy) children; mTBI, mild traumatic brain injury patients; R, radius; RD, reading-disabled children.



SUPPLEMENTARY FIG. S3. Topological filtering of GDD values using orthogonal (upper panel) and conventional Minimal Spanning Tree (lower panel) for gender-matched healthy adult (HA) vs. mTBI subgroups. **(A, C)** Dissimilarity matrices of subject-specific FCGs ($N = 72$) based on the GDD metric using OMST and MST, respectively. Corresponding MDS results are displayed **(B, D)**. Stress indicates the% loss of information due to the dimensionality reduction process via the MDS algorithm. R1, R2, and R3 indicate the three prespecified dimensions used to plot participant cases through MDS.

capacity of OMST to derive functional connectivity graphs that clearly differentiate healthy adults from adults with a history of mTBI was maintained (as well as the superiority of this method over conventional minimal spanning trees).

Classification Results Based on Relative Power

Relative power spectrum was calculated based on the power spectral density (PSD) for each artifact-free magnetoencephalography epoch, as the Fourier transform of the autocorrelation function. PSD was then normalized to range from 0 to 1. We followed a feature selection strategy using Multi-Cluster/Class Feature Selection (MCFS) for both pairwise and multi-group classification analysis (Cai et al., 2010).

Leave-one-out cross-validation (LOOCV) scheme was employed separately for k-nearest neighbor and support vector machine with RBF kernel classifier. At every round of LOOCV, the MCFS was applied to the N-1 subjects and the prediction of the model was validated on the N subject.

Finally, we selected the features with the maximum representation among the N rounds of the LOOCV. We searched over a range of features from 1 up to 100 to get the maximum accuracy.

Pairwise and multigroup classification results are presented in Supplementary Table S1.

Supplementary Table S1. Average (±SD) Classification Performance Over Five Runs of the k-Nearest Neighbor and Linear (Support Vector Machine) Classifiers Using Relative Power Features

	k-NN classifier			*Linear SVM*			*No. of features per frequency band* (δ-θ-α_1-α_2-β_1-β_2- β_3-$\gamma 1$-$\gamma 2$)
	Accuracy (%)	*Sensitivity*	*Specificity*	*Accuracy (%)*	*Sensitivity*	*Specificity*	
Typical vs. RD children	67.3 (8.7)	65.4 (6.1)	65.9 (9.3)	70.2 (10.5)	67.8 (9.2)	67.0 (9.5)	4-3-1-2-3-0-1-0-0
Typical children vs. healthy adults	62.3 (9.1)	62.2 (6.9)	62.1 (6.2)	62.3 (6.8)	60.1 (7.5)	59.1 (7.8)	9-3-1-2-10-8-0-7-9
Healthy adults vs. mTBI	68.1 (9.1)	65.9 (5.3)	64.8 (10.1)	71.6 (9.7)	67.9 (9.8)	66.9 (10.8)	7-4-0-0-0-0-0-0-0
Multigroup classification	53.2 (6.1)	52.7 (6.5)	53.0 (5.8)	57.8 (6.3)	58.1 (7.2)	58.8 (6.8)	8-3-3-2-7-7-0-8-10

k-NN, k-nearest neighbor; NI, nonreading impaired children; RD, reading-disabled children; SVM, support vector machine.

References

Cai D, Zhang C, He X. 2010. Unsupervised feature selection for multi-cluster data. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, DC, USA, 2010. New York, NY: ACM Press, pp. 333–342.

Lindner M, Vicente R, Priesemann V, et al. 2011. TRENTOOL: a Matlab open source toolbox to analyze information flow in time series data with transfer entropy. BMC Neurosci 12:119.

Martinetz TM, Berkovich SG, Schulten KJ. 1993. Neural-gas network for vector quantization and its application to time-series prediction. IEEE Trans Neural Netw Learn Syst 4:558–569.

Ragwitz M, Kantz H. 2002. Markov models from data by simple nonlinear time series predictors in delay embedding spaces. Phys Rev E 65:056201.

Stam CJ, Nolte G, Daffertshofer A. 2007. Phase lag index: assessment of functional connectivity from multichannel EEG and MEG with diminished bias from common sources. Hum Brain Mapp 28:1178–1193.