Supplementary Materials

1. Dataset I: simulated quasi-stationary data

We first generate 28 source time-series using order-3 random-coefficient autoregressive models driven by super-Gaussian noise (simulated 250-Hz sampling rate, 135k samples). The nine total minutes of simulated source data are partitioned into three 3-min segments, each with a different set of 16 active sources. For each of the 2^3=8 possible combinations of active/inactive sources over three segments, e.g. active-inactive-active, etc., we simulate 4 sources that follow such a pattern, excluding the inactive-inactive-inactive pattern in which no sources are active. This results in 16 active sources in each segment with 8 active sources changing between segments. Each source is assigned to the center of a randomly selected cortical region in the MNI brain atlas as an equivalent current dipole with its dipole moment oriented normal to the cortical surface. Finally, the simulated source-level activities are projected through a zero-noise three-layer boundary element method (BEM) forward model (MNI Colin27), yielding simulated 16-channel EEG data (selected from a 64-channel BioSemi montage). More details are included in Hsu et al. (2015).

2. Decomposition errors of ICA models

2.1 Model errors

For a known ground truth (N-by-N) mixing matrix **A**, the model error $\mathrm{E}^{(\mathrm{h})}$ of the learned unmixing matrix $\mathbf{W}_{\mathrm{h}}$ and sphering matrix **S** can be defined (Dougals, 2001):

$$\mathrm{E}^{(\mathrm{h})} = \frac{1}{N-1}\left[N - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{\max\limits_{1\le j\le m}\left|\boldsymbol{C}_{ij}^{(h)}\right|^2}{\sum_{j=1}^{N}\left|\boldsymbol{C}_{ij}^{(h)}\right|^2} + \frac{\max\limits_{1\le j\le m}\left|\boldsymbol{C}_{ji}^{(h)}\right|^2}{\sum_{j=1}^{N}\left|\boldsymbol{C}_{ji}^{(h)}\right|^2}\right)\right]$$

where $\boldsymbol{C}^{(h)} = \boldsymbol{W}_h \cdot \boldsymbol{S} \cdot \boldsymbol{A}$. This measure is a normalized total cross-talk error and accounts for scale and permutation ambiguities. In the case of perfect reconstruction, $\mathrm{E}^{(\mathrm{h})}$ equals zero. For three quasi-stationary segments and the three ground-truth mixing matrices $\mathbf{A}_{\mathrm{k}}$, we report the sum of logarithm of the model error for each segment k, i.e., total log error $= \sum_{k=1}^{3} \log E_k$, where $E_k = \min\limits_{h=1\ldots H} E_k^{(h)}$.

2.2 Signal-to-interference ratios (SIR)

SIR estimates the log-scaled normalized mean-squared errors of the decomposed time series of the i-th component, $\widehat{s_\iota}(t)$, compared with ground-truth source activities, $s(t)$, defined as

$$SIR_i = \min_{j=1\ldots N} 10\log_{10}\left(\frac{\sum_{t=1}^{T} s_j(t)^2}{\sum_{t=1}^{T}\left(\hat{s}_i(t)^2 - s_j(t)^2\right)}\right)$$

(Salazar et al., 2010 b) where T is the number of data samples. The final SIR is averaged across all components of the best-matched model (i.e., the model with the smallest model error) in each quasi-stationary segment, and then averaged across the three segments.

2.3 KL divergence

The KL divergence quantifies the difference between the estimated PDF $P_i(x)$ of the i-th source and the ground-truth source PDF $Q(x)$, defined as:

$$\text{KL-Div}_i = \min_{j=1\ldots N} \frac{1}{2}\left[D_{KL}(P_i||Q_j) + D_{KL}(Q_j||P_i)\right]$$

where $D_{KL}(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$.

$Q(x)$ is approximated by histograms because the ground truth source activity is generated by an autoregressive process instead of a parametric model. The final KL-Div is averaged across all components of the best-matching model for each quasi-stationary segment and then averaged across the three segments.

3. A note on the cross-validation approach

Note that the current cross-validation approach was applied to the model probabilities of AMICA decompositions of the "combined training and testing data." It would be possible to fit a single AMICA model to the training data and to then apply the pre-trained model to the test data to obtain out-of-sample predictions. However, the purpose of the cross-validation analysis here was to test the consistency and interpretability of AMICA model probabilities across time. To be clear, this analysis is not intended as a new practical method for sleep-stage prediction across recordings, but as an exploratory test of the utility of AMICA model probabilities for separating functionally distinct periods within a continuous data set. Given this goal, applying AMICA simultaneously to both the training and test data segments was necessary to produce model probabilities that allowed us to ascertain the consistency of model switching across the decomposition. Other studies might use a full cross-validation scheme to assess the feasibility of applying the results of AMICA decomposition for out-of-sample prediction and/or for selecting hyperparameters including the most suitable number of AMICA models.

4. Identifiability of multi-model AMICA

Given that multi-model AMICA decomposition models three layers of mixing with a large number of parameters (Fig. 1 in the main text), the issue of identifiability – whether varying sets of model parameters across the three layers may equally well account for the decomposed data – is further discussed below.

The concern for the identifiability of the AMICA three-tier model should apply only to the first two tiers (models and sources) as the third (source distributions) is entirely defined by the first two. Once the models and their IC sources are fixed, the source distributions are fit to the histograms of the derived IC activations. The mixture of extended Gaussian models used in AMICA serves only to simplify the overall model by approximating each IC histogram with a few (3 in this study) extended Gaussian functions, greatly reducing the required number of parameters while sharpening the ICA learning. To disable the source distribution modeling would only degrade model performance (albeit with lower computational cost) as seen in (Delorme et al., 2012), a test in which performance of single-model AMICA compared favorably to that of Infomax ICA and other BSS algorithms. Similarly, Extended Infomax ICA outperformed the original Infomax ICA by somewhat increasing the flexibility of its source-density assumptions (Lee et al., 1999). AMICA further individualizes the learning rule for each source using more flexible and adaptive source density models.

As to the identifiability in the first two stages (models and sources), Bell and Sejnowski (1995) had already shown that single Infomax ICA model decomposition is identifiable so long as (1) the number of sources matches the number of channels, and (2) there is at most one Gaussian component (as more than one would not be separable). Therefore, the question can be reduced to whether introducing additional models maintains identifiability provided that the within-model identifiability conditions are met. Answering this question is likely non-trivial and could be addressed in future work. Currently, identifiability at this level does not seem to be an issue as shown by the clear and consistent interpretability of model probabilities in these two experiments.

References

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. Neural computation, 7(6), 1129-1159.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., Makeig, S., 2012. Independent EEG sources are dipolar. PloS one 7 (2), e30135.

Douglas, S. C., 2001. Blind signal separation and blind deconvolution. Handbook of neural network signal processing 6, 1-28.

Hsu, S.-H., Pion-Tonachini, L., Jung, T.-P., Cauwenberghs, G., 2015. Tracking non-stationary EEG sources using adaptive online recursive independent component analysis. In: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE. IEEE, pp. 4106-4109.

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural computation, 11(2), 417-441.

Salazar, A., Vergara, L., Serrano, A., Igual, J., jan 2010b. A general procedure for learning mixtures of independent component analyzers. Pattern Recognition 43 (1), 69{85.