# Learning Recurrent Waveforms within EEGs

Austin J. Brockmeier, *Member, IEEE*, and Jose C. Principe, *Fellow, IEEE*

***Abstract*—*Goal:* We demonstrate an algorithm to automatically learn the time-limited waveforms associated with phasic events that repeatedly appear throughout an electroencephalogram. *Methods:* To learn the phasic event waveforms we propose a multi-scale modeling process that is based on existing shift-invariant dictionary learning algorithms. For each channel, waveforms at different temporal scales are learned based on the assumption that only a few waveforms occur in any window of the time-series, but the same waveforms reoccur throughout the signal. Once the waveforms are learned the timing and amplitude of the phasic event occurrences are estimated using matching pursuit. To summarize the waveforms learned across multiple channels and subjects, we analyze their frequency content, their similarity to Gabor-Morlet wavelets, and perform shift-invariant k-means to cluster the waveforms. A prototype waveform from each cluster is then tested for differential spatial patterns between different motor imagery conditions. *Results:* On multiple human EEG datasets, the learned waveforms capture key characteristics of signals they were trained to represent, with a consistency in waveform morphology and frequency content across multiple training sections and initializations. On multichannel datasets, the spatial amplitude patterns of the waveforms are also consistent and can be used to distinguish different modalities of motor imagery. *Conclusion:* We explored a methodology that can be used for modeling the recurrent waveforms in EEG traces. *Significance:* The methodology automatically identifies the most frequent phasic event waveforms in EEG, which could then be used as features for automatic evaluation and comparison of EEG during sleep, pathology, or mentally engaging tasks.**

***Index Terms*—Biomedical signal processing, clustering, dictionary learning, EEG, sparse coding.**

## I. Introduction

AMPLIFIED voltage recordings across the human scalp reveal a diversity of spatiotemporal oscillations and patterns [1], [2], [3], [4], [5] commonly referred to as brain waves. The electroencephalogram (EEG) records a mixture of the electrical activity of the brain along with potentials arising from eye and face muscles and movements. The interesting portion of the signal is the superposition of the potentials generated from the electrochemical activity in the neocortex [5], [6]. Research has demonstrated that signal characteristics, such as certain frequencies, spatial locations, and phasic event waveforms are indicative of an individual's cognitive state, the presence or presentation of stimuli, or neural pathology. Many distinct signatures appear as brief, time-limited, oscillations [7] or phasic events [8] such as alpha waves [1], [9], and sleep spindles. Some signatures are ubiquitous and evident to experts from the raw EEG traces, while other patterns such as evoked potentials [1], [2], [10] are time-locked to stimulus presentation. A commonality among these signatures is their non-stationarity: even narrow-band rhythms are time-varying in frequency and amplitude [11], and many waveforms are phasic events that occur transiently.

Due to their non-stationary nature and the superposition of transient and brief oscillatory events, EEGs require careful consideration for analytic examination. Signal processing tools are only appropriate when the underlying assumptions that they are based on are met [12], [13]. Time-frequency analysis, such as multitaper analysis [14] or wavelet decompositions [15], has been the standard tool for exploratory analysis of neural potential signals and is well suited for locally stationary brain rhythms [16], but these representations are not able to separate signals emanating from different sources.

Neural activity from distinct spatial sources occurs simultaneously in the brain. Each source may have distinct time-frequency patterns or waveforms, but in the EEG recordings these patterns are all mixed. By simile, brainwaves are like an audio recording from a busy social event—say a cocktail party—where conversations are occurring simultaneously. Can the individual sources be identified from these seemingly chaotic signals? The cocktail party analogy strikes a chord with the blind source separation community, which has proposed the use of spatial independent component analysis (ICA) and associated techniques to disentangle distinct sources in multichannel EEG [17], [18], [19], [20], [21]. The utility of these techniques is based on the topological organization of the cortex. As the neural circuitry for distinct modalities are spatially isolated they can be associated with unique source signals that are then 'mixed' together in the scalp recordings. The blind source separation problem is then to 'demix' the signals using spatial filters that isolate the signal components from each of the underlying sources.

In the case of a single recording channel, spatial filtering is impossible, but it is possible to decompose the signal into components with distinct time-frequency characteristics. Ideally, each component would be strictly composed of waveforms with similar morphology: a component dedicated to ripples, another to alpha waves, a third to EEG spikes, and so on.[1] Linear filtering is inadequate for morphological separation as it can only segregate signals by frequency. Nonlinear filtering by matching pursuit [24] has been shown to successively

Accepted for publication Nov. 3, 2015. This work was supported by the Defense Advanced Research Projects Agency under Contract N66001-10-C-2008.

A. J. Brockmeier was with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, 32611, USA and is now with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, L69 3GJ, UK (correspondence e-mail: ajbrockmeier@gmail.com).

J. C. Principe is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, 32611, USA.



[1]Morphological component analysis has been applied to denoising and separation tasks for images [22], [23], which in essence are single-channel two-dimensional signals.

separate components with different underlying morphology in EEG [25], [26], [27], [28], [29], [30], and has been adapted to multichannel recordings [27], [31].

Matching pursuit is based on the assumptions that any portion of the time-series signal can be approximated using relatively few elements from a 'dictionary' of component waveforms. For instance, this dictionary may be a set of Gabor-Morlet wavelets (sinusoids with Gaussian envelopes) [32] and Dirac delta functions appearing at different translations. Each instance of waveform, its amplitude, and timing are referred to as an atom, and together they form an atomic decomposition of the signal [33]. Using the correct basis allows a meaningful representation of the signal using relatively few atoms.

Instead of leaving the dictionary as a design choice, which must be predefined, the waveforms can be learned directly from the data based on the higher-order statistics of the signal [34], using techniques known as dictionary learning or sparse coding [35], [36], [37], [38], [39]. An alternative to dictionary learning is to apply ICA directly to time-embedded vectors from a single channel [40], [41], [42]: a technique that has been shown to be successful in learning the constituent waveforms on EEG and other biomedical signals. In either sparse coding or single-channel ICA, waveforms can be learned from patches (windows) of signals, but these dictionaries will contain copies of similar waveforms at many different shifts. Shift-invariant dictionary learning, also known as convolutional sparse coding, proposes to learn only a few explicit waveforms and represent the rest of the dictionary by translating these waveforms [43], [44], [45], [46], [47], [48], [49], [50], [51]. These approaches avoid estimating redundant copies of the same waveform at different shifts. However, some of these approaches neglect the phase information and only learn the magnitude spectrum of the waveforms. For EEG it is important to maintain the phase information to investigate phasic events such as evoked potentials. Although there is yet no natural way to incorporate shift-invariance into single channel ICA, it will preserves waveform shape, with a sign ambiguity, and the redundancy issues of single-channel ICA can be alleviated by using a post-hoc waveform selection. A greedy approach for waveform subset selection was previously proposed [52] and evaluated on synthetic data.

Shift-invariant dictionary learning has been fruitfully applied to EEG for the detection of evoked potentials [53]. We emphasize that the estimated dictionary (set of waveforms) are data-dependent descriptors of the signal and that learning a small dictionary guarantees that any approximation reuses the same set of waveforms. This can be contrasted with a decomposition using a predefined dictionary, which is not constrained on the number of unique waveforms to use. A comparison of a predefined to a learned dictionary on a segment of EEG is shown in Fig. 1. Learning a small dictionary also allows different signals (or the conditions that generated them) to be compared by contrasting the sets of learned waveforms. For instance, the waveform shape could be compared across different conditions such as ages or neuropathy.

In the rest of this study, we focus on adapting existing shift-invariant dictionary learning algorithms to the problem of estimating multiscale dictionaries of EEG waveforms.

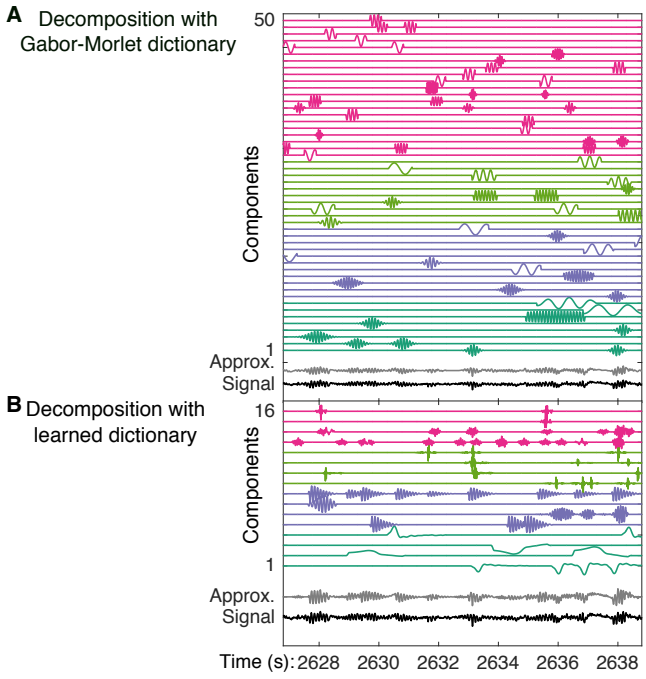



Fig. 1. Comparison of two different sparse decompositions of a novel EEG segment using the same number of atoms. (A) Predefined Gabor-Morlet wavelet dictionary, which explains 82% of the variance of the test portion of the signal. (B) Dictionary learned on a disjoint section of the same recording using single-channel ICA and waveform subset selection, which explains 72% of the variance, but reuses the same 16 waveforms repeatedly throughout the decomposition, whereas the predefined dictionary uses different subsets.

Specifically, we combine shift-invariant dictionary learning algorithms with a multistage modeling approach. The multistage modeling approach avoids the difficulties involved in estimating many waveforms of different lengths at once. The results in the main body use the matching-pursuit with SVD update [54], [47], [55], but the supplementary material (available at http://ieeexplore.ieee.org) contains results using single-channel ICA with a post-hoc waveform selection algorithm [52] and a comparison of these two algorithms on synthetic data. Another synthetic example is used to illustrate the difference between decompositions using learned dictionaries versus using a pre-defined Gabor-Morlet dictionary.

Multiple publicly available human EEG datasets are used to demonstrate and empirically verify the consistency of the estimation. We try to highlight how the learned waveforms can be used to characterize and distinguish EEGs under different conditions or from different subjects. We consider only single channel models, and run the waveform estimation on all the channels independently and in parallel. Then in post-hoc analysis we determine if the waveforms are consistently estimated across multiple subjects and spatial locations, and organize them into clusters using shift-invariant k-means [56]. Additionally, we examine the spatial amplitude patterns associated with the cluster prototypes using a model-based approach for investigating global spatial patterns across the scalp [57]. The spatial patterns are then used to classify segments of EEG during motor imagery on a single-trial basis [58]. The MATLAB code for the described methodology is available online at http://cnel.ufl.edu/~ajbrockmeier/eeg/.

## II. Model and Estimation Method

In this section we introduce the sparsely excited multiple-input single-output (MISO) system, discuss generative models, introduce the least-squares framework with non-negativity constraints on the sparse sources, mention the single-channel ICA with waveform subset selection, and finally discuss a multistage subset deflation approach to learn waveforms across multiple scales.

### A. Multiple-Input Single-Output Model

We assume the signal of interest is formed by a linear MISO system where each sparse source excites a distinct waveform to form a component [59]. The signal is a uniform mixture of these components.

Let $x(t)$ be a combination of $P$ component signals $\{y_p(t)\}_p, p \in \{1,\dots,P\}$ observed in the presence of noise $e(t)$. Overall, this is a multiple-input single-output (MISO) linear system with sparse inputs. Each component, $y_p(t)$, has a unique waveform $a_p(t)$ and sparse source $s_p(t)$ consisting of a weighted train of delta functions:

$$x(t) = e(t) + \hat{x}(t) = e(t) + \sum_{p=1}^{P} y_p(t) \tag{1}$$

$$y_p(t) = \int_{-\infty}^{\infty} s_p(t-u)a_p(u)du \tag{2}$$

$$s_p(t) = \sum_i \alpha_{p,i}\delta(t-\tau_{p,i}) \quad p = 1,\dots,P. \tag{3}$$

The summation of the components is a noise-free signal $\hat{x}(t)$.

The atomic representation of $\hat{x}(t)$ consists of a set of source indices, amplitudes, and timings $\{(p_i,\alpha_i,\tau_i)\}_i$. Using this set the model signal can be rewritten as:

$$\hat{x}(t) = \sum_i \int_{-\infty}^{\infty} \alpha_i\delta(t-\tau_i-u)a_{p_i}(u)du. \tag{4}$$

Similarly, each component signals can be described by the impulse response of the filter $a_p(t)$ and the set of excitation times and amplitudes $\{(\alpha_j,\tau_j)\}_{j\in\mathcal{I}_p}$ where $\mathcal{I}_p = \{i : p_i = p\}$.

### B. Model Estimation

Stochastically, the sparse source signals activation times can be described by a point process; a realization of a point process is a train of Dirac delta functions. In the model above, a marked point process is required that also describes the amplitude of the impulses [60]. A marked point process is fully described by a joint distribution over both the timing and amplitude of the impulses. With a distribution over the noise, a complete generative model can be posed, but solving it is intractable [60] and approximations are necessary.

For simplicity, we do not utilize a full generative model. Instead, we assume a maximum likelihood approach where the atomic representation $\{(p_i,\alpha_i,\tau_i)\}_{i=1}^{L}$ and the sparse inputs $\{s_p(t)\}, p \in \{1,\dots,P\}$ are fixed and let the user select $L$ and $P$ and the duration of each waveforms model parameters. We assume uncorrelated white noise and optimize the parameters to minimize the mean squared error.

In the case of correlated noise, the noise covariance can be additionally estimated and used to whiten the signal [60]. This is important if the waveforms of interest are sufficiently different from the colored noises—such as action potentials [61], [62] or evoked potentials [63]. However, in the case of action potentials (spike trains) and evoked potentials the estimation can start with an initial template for each waveform. We assume no prior information on the shape or frequency, and assume the model will account for all signal correlation.

### C. Source Estimation

Even though each component is a linear convolution of the source and waveform, linear filtering is inadequate to separate the original components from the combination if the waveforms' frequency responses overlap. There are two regimes in which it is possible to resolve the inputs to a MISO system from a single output: spectrally disjoint waveforms (corresponding to sparsity in the frequency domain) or sufficiently sparse input (corresponding to temporally disjoint input). In the former case, linear band-pass filtering is sufficient. In the latter case, recovering the sparse source as a train of Dirac deltas requires an overcomplete basis of the signal: shifted versions of the underlying waveforms appearing at each time-shift. With an overcomplete basis, linear analysis is not meaningful [64], and sparsity constraints on the sources are necessary to recover them. The resulting problems can be relaxed to convex optimization problems or solved using iterative algorithms [65]. Matching-pursuit (MP) [24] provides a greedy approximation to solve the sparsity constrained problem.

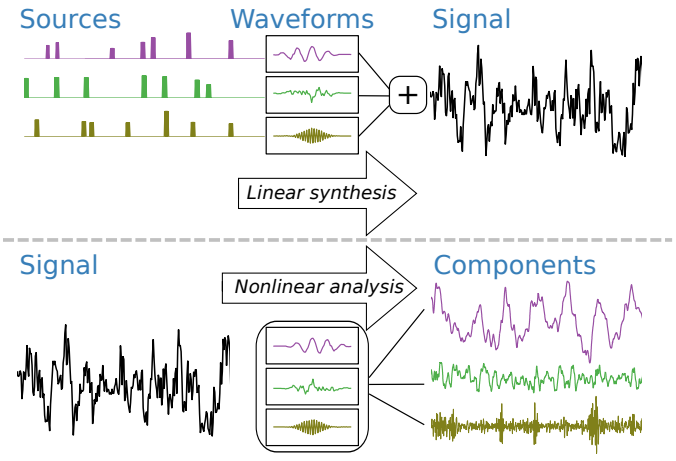



Fig. 2. A depiction of the assumed model and the signal flow. The observed signal is assumed to be linearly synthesized by convolving the sparse source with time-limited waveforms; the resulting components are added together. Nonlinear analysis separates the signal back into its constituent components

### D. Least Squares Estimation

Assuming a signal plus white noise model and that the number of waveforms $P$ and source excitations $L$ are known, the blind system identification problem can be posed as a

least-squares optimization over $\mathcal{A} = \{a_p(t)\}_{p=1}^{P}$ and $\mathcal{S} = \{(p_i, \alpha_i, \tau_i)\}_{i=1}^{L}$:

$$\min_{\mathcal{A},\mathcal{S}} J(\mathcal{A},\mathcal{S}) = \left\| x(t) - \sum_{i=1}^{L} \alpha_i \int_{-\infty}^{\infty} \delta(t-\tau_i) a_{p_i}(u) du \right\|_2^2. \tag{5}$$

Jointly solving for both $\mathcal{A}$ and $\mathcal{S}$ is difficult because the source estimates are intrinsically linked to the waveforms. It is necessary to perform an alternating optimization.

Assuming $\mathcal{A}$ is fixed, a greedy optimization of $\mathcal{S}$ can be made using matching pursuit. At each iteration of matching pursuit, the atom (consisting of the timing, amplitude, and waveform index) that explains the most energy remaining in the residual of the signal is selected. The residual signal is updated by removing this single-atom reconstruction. This updated residual is used as the input to the next iteration.

Given the atomic decomposition $\mathcal{S} = \{(p_i, \alpha_i, \tau_i)\}_{i=1}^{L}$, either the sources or the individual components can be computed via (3) or (2), respectively. Then the sources are fixed, and the set of waveforms $\mathcal{A}$ is updated via least squares.

The alternating optimization between time-series matching pursuit and least squares updates has been proposed as a general tool for shift-invariant dictionary learning [54], [47], [55]. It is the time-series extension of the popular dictionary learning algorithm K-SVD [66]. For conciseness, we refer to it as MP-SVD.

*1) MP-SVD:* For practical digital implementation, we consider the case when the time series (1) is discretely sampled, and the convolution operators are replaced by finite summations. In this framework we only allow integer shifts, but this approximation can be avoided using continuous basis pursuit [50]. Let $\mathbf{a}$ denote an $M$-length waveform and $T_\tau(\mathbf{a})$ denote the translation of the waveform to begin at time $\tau$. Correspondingly, let $W_\tau(\mathbf{x})$ denote the windowing function that extracts an $M$-length window from signal $\mathbf{x}$ starting at time $\tau$. (Here we have used a Tukey window [67], which is also known as a tapered cosine window, with parameter of $0.1$.) Using these notations the objective function can be written in terms of vectors as

$$\min_{\{\mathbf{a}_p\}_{p=1}^{P}, \{(p_i,\alpha_i,\tau_i)\}_{i=1}^{L}} \left\| \mathbf{x} - \sum_{i=1}^{L} \alpha_i T_{\tau_i}(\mathbf{a}_{p_i}) \right\|_2^2. \tag{6}$$

To update the waveforms, we first assume we have an estimate of the components using the current waveforms. Let $\mathbf{x}^{(p)}$ denote the signal consisting only of the estimate of the $p$th component and the error signal

$$\mathbf{x}^{(p)} = \mathbf{e} + \mathbf{y}_p = \mathbf{x} - \sum_{q \in \{1,\ldots,P\} \setminus p} \mathbf{y}_q \tag{7}$$

where $\mathbf{y}_p = \sum_{j \in \mathcal{I}_p} \alpha_j T_{\tau_j}(\mathbf{a}_p)$ and $\mathcal{I}_p = \{i : p_i = p\}$. Only the patches when the waveform is active are needed to update the waveform. These patches are collected into a matrix:

$$X_p = \left[ W_{\tau_j}(\mathbf{x}^{(p)}) \right]_{j \in \mathcal{I}_p}. \tag{8}$$

Treating the amplitudes as a nuisance parameter (the previous estimates are ignored) and *assuming that none of the patches overlap* the following optimization problems for the optimal waveform are equivalent

$$\arg\min_{\|\mathbf{a}\|=1} \min_{\mathbf{v}} \left\| \mathbf{x}^{(p)} - \sum_{i:p_i=p} \mathbf{v}_i T_{\tau_i}(\mathbf{a}) \right\|_2^2 \tag{9}$$

$$\arg\min_{\|\mathbf{a}\|=1} \min_{\mathbf{v}} \left\| X_p - \mathbf{a}\mathbf{v}^\mathsf{T} \right\|_F^2 = \max_{\|\mathbf{a}\|=1} \mathbf{a}^\mathsf{T} X_p X_p^\mathsf{T} \mathbf{a}. \tag{10}$$

The updated waveform is selected as the eigenvector of the matrix $X_p X_p^\mathsf{T}$ corresponding to the largest eigenvalue. This eigenvector is also the primary singular vector of the columns of the $X_p$ corresponding to the best rank-1 approximation [68]. Assuming these were the correct timings, this update minimizes the reconstruction cost for these patches.

*2) Non-negative amplitude constraint:* In the case of non-oscillatory waveforms in EEG, it is worthwhile to preserve the polarity of the waveforms. This can be done by allowing only non-negative amplitudes during the matching pursuit and the waveform update:

$$\arg\min_{\|\mathbf{a}\|=1} \min_{\mathbf{v}\geq 0} \left\| X_p - \mathbf{a}\mathbf{v}^\mathsf{T} \right\|_F^2. \tag{11}$$

Unlike the unconstrained case, there is no analytic solution to this problem [69]. However, let $X_P = \sigma_1 \mathbf{a}\mathbf{v}^\mathsf{T}$ be the rank-1 SVD, if $\mathbf{v}$ is strictly positive then it is the solution to (11). If not, then a local minima can be found by alternating between

$$\mathbf{v} \leftarrow \max(0, (1-\lambda)\mathbf{v} + X_p^\mathsf{T}\mathbf{a}) \tag{12}$$

$$\mathbf{a} \leftarrow \frac{X_p \mathbf{v}}{\sqrt{\mathbf{v}^\mathsf{T} X_p^\mathsf{T} X_p \mathbf{v}}} \tag{13}$$

where $\max$ enforces the elements to be non-negative and $\lambda$ is the step size of a proximal gradient update [70]. The non-negative amplitude constraint with step size of $\lambda = 1$ is used in the rest of this study.

*3) Waveform Initialization:* The non-linear least-squares cost function (5) may have many local optima. From different initializations, it is unlikely to estimate the same waveforms. It is possible to initialize the waveforms from a predefined set of wavelets, and then 'optimize' them further. However, to avoid biasing the waveforms to any particular shape, we initialize the waveforms as random vectors with entries independently drawn from the normal distribution. Ideally, the optimized waveforms across multiple initializations should be similar.

### *E. Single-Channel ICA with Greedy Subset Selection*

Single-channel ICA is an alternative approach to least squares estimation that avoids the explicit estimation of the sources during learning. Instead, the sparse sources' statistical properties are used in the filter estimation [71]. Single-channel ICA uses windows of the time series as the input vectors to independent component analysis (ICA) [34]. Previous studies [41], [42] have demonstrated that the fixed point algorithm FastICA [72] can efficiently estimate the waveforms in a multiple-input-single-output model. Essentially, FastICA estimates the waveforms using a non-linear feedforward network without explicit estimation of the sources [52]. The main drawback of single-channel ICA is that many of the waveforms

are redundant—they are the same waveform appearing at different shifts—and others are artifacts caused by the multi-unit ICA estimation constraints.

To solve these problems, we use a greedy subset selection algorithm to choose a non-redundant subset of the estimated waveforms with the goal of minimizing the reconstruction error [52]. The first step in the greedy algorithm is to approximate the training signal using matching pursuit with each waveform individually. Herein, we enforce non-negativity constraints during matching pursuit, and since ICA is invariant to polarity, we use both the waveform and its negation as candidate waveforms. The resulting approximations are treated as the basis vectors for approximating the same signal using orthogonal matching pursuit (OMP) [73]. At each iteration, OMP includes the basis vector which minimizes the reconstruction error. The algorithm is terminated when the number of included basis vectors equals the number of desired waveforms. The waveforms corresponding to the included basis vectors form the selected subset. This post-hoc subset selection algorithm can be applied to find a signal-specific subset of a larger shift-invariant dictionary. In particular, we use it to transform a dictionary of Gabor-Morlet wavelets into a much smaller, data-dependent model (examples are included in the supplementary material). When a training portion of the signal is used to select the subset, this approach can be seen as an compromise between using pre-defined dictionaries and fully adaptive dictionaries.

### F. Multistage Waveform Estimation

In EEGs, waveforms may have widely different time-scales. This motivates learning a multiscale waveform dictionary, which can be difficult to achieve in a single optimization. To address this, we propose a multistage approach to simplify the estimation of multiple waveforms at different scales. It consists of a subset deflation approach that greedily estimates a set of waveforms at each stage. After convergence, multiple passes of matching pursuit are run to remove the contribution of the waveforms before the resulting residual is passed as the input signal to the next stage. We use a coarse-to-fine approach where longer waveforms are estimated before shorter waveforms. While this approach is *ad hoc*, it is well suited for EEG where low-frequency/long duration signals explain more of the variance in the signal.

Before estimation, the user must select the parameters of the sparsely excited multiple-input-single-input model. Most importantly the user must select the number of waveforms and their length. These choices will depend on the time-scales of interest and the application. The other important choice is the assumed rate of the sources, which determines the number of atoms, that is, the number of waveform occurrences used in approximation/decomposition. Only the total number of occurrences for each scale needs to be set as the particular number of occurrences of each waveform is based on how often it is used in the matching pursuit-based decomposition. Since the modeling is based on sparse sources, the total number of occurrences should be kept relatively low.

Multiple models with different waveform lengths, numbers of waveforms, and approximations with different number of atoms can be trained and compared using model selection criterion. To compare models, we assume the background activity is white noise with constant variance. We show an example of using the Bayesian information criterion (BIC) to select the number of waveforms in the supplementary material.

## III. Waveform Meta-Analysis

When different shift-invariant dictionaries are learned across multiple sections, channels, and subjects there is a need to summarize the characteristics of the large number of resulting waveforms. To do this, we fit the waveforms using parametric models and perform shift-invariant vector quantization on the waveforms to group them into clusters. For multichannel datasets, we use a prototypical waveform to represent each cluster and analyze the waveform's ability to differentiate between known conditions using its spatial amplitude patterns.

### A. Gabor Fit of Waveforms

We see how well each waveform is modeled by real-valued Gabor-Morlet wavelets of varying frequency, bandwidth, and phase: $g_{f,\phi,\sigma}(t) = \exp(-t^2/(2\sigma^2))\cos(2\pi f t + \phi)$, where $f, \phi$ are the frequency and phase, and $\sigma$ is the standard deviation of the temporal envelope. In lieu of an uniform sampling over the space of these parameters [31] we use a logarithmic scaling over frequency and temporal envelope, and three phases $[0, \pi/4, \pi/2]$. Any time shift is accounted for by performing the matching using maximum absolute cross-correlation

$$c(g,h) = \max_t \frac{|\int_{-\infty}^{\infty} g(\tau)h(\tau - t)d\tau|}{\|g\|\|h\|}. \tag{14}$$

The discrete time version of the above can be computed efficiently by using the discrete Fourier transform.

To characterize the frequency content of waveforms that match Gabor-Morlet wavelets, we compute the peak frequency and 3dB bandwidth from the power spectral density (PSD) of each waveform. To get a robust estimate of the PSD, we use multitaper analysis [14], [16]. Specifically, the waveforms are zero-padded to a specified length, which controls the frequency resolution, and projected onto a small number of discrete prolate spherical sequences. The magnitude of the Fourier transform is computed for each projection and a uniform average is taken to provide the multitaper estimate. Using the resulting peak frequency avoids the bias caused by the grid of center frequencies used in constructing the set of wavelets.

### B. Average Cross-correlation between Waveform Sets

For a measure of similarity between two sets of waveforms we propose to use the average cross-correlation. Specifically, if $\mathcal{G}$ is a set of $|\mathcal{G}|$ waveforms ($|\mathcal{G}|$ indicating the number of elements in $\mathcal{G}$) and $\mathcal{H}$ is a set of $|\mathcal{H}|$ waveforms, we compute the average cross-correlation $\bar{c}(\mathcal{G}, \mathcal{H})$ as

$$\max\left\{\frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}}\max_{h\in\mathcal{H}} c(g,h), \frac{1}{|\mathcal{H}|}\sum_{h\in\mathcal{H}}\max_{g\in\mathcal{G}} c(g,h)\right\} \tag{15}$$

where $c(g,h)$ is the maximum correlation for waveforms $g(t)$ and $h(t)$, which is normalized so that the zero-lag auto-correlation is 1.

### C. Shift-invariant Vector Quantization

Clustering is useful to organize a set of waveforms into groups without a predefined criterion such as frequency or bandwidth. Standard clustering algorithms developed for data represented as vectors are inappropriate for shift-invariant waveforms since they are ignorant of the shifts necessary to align the waveforms. One option is to represent each waveform by its power spectral density. An obvious drawback of using the power spectral density to represent the waveform is that it discards the phase spectrum of the waveform. Alternatively, one can use clustering algorithms that directly use the maximum cross-correlation between normalized waveforms, such as the one-dimensional case of circular invariant k-means [56]. In this algorithm, each cluster mean is associated with a new shift-invariant waveform, formed as the average of all the aligned waveforms in the cluster.

As with standard k-means [74] the user must select the number of centers. As shift-invariant k-means optimizes the minimal mean-squared error of each waveform to its cluster center, it is straightforward to use this error to guide model selection. Various criteria can be used for this, in particular Calinski-Harabasz's criterion [75], which uses the inter/intra-cluster variance ratio (pseudo F-score) and was previously used for circular-invariant k-means [56]. However, the set of waveforms may not consist of separate clusters, but instead, the set may exhibit smooth variations across the continuous space of waveforms.[2] Nonetheless, shift-invariant k-means can be used as vector quantization rather than cluster identification, where increasing the number of clusters gives a finer-grained view of the space. For illustrative purposes, we use a manageable set of 9 clusters.

### D. Spatial Amplitude Patterns

In multichannel records, we identify the spatial amplitude patterns associated with the occurrence of each waveform [58]. We use the timing of a waveform's occurrences based solely on the activation times on the channel it originated from. Given the set of timings, we assume the spatiotemporal pattern is time-locked, and we record the vector of cross-correlation with the waveform on the other channels at each time point. These timings can be taken from the atomic decomposition of the single channel obtained from matching pursuit. Alternatively, they may be obtained by greedily choosing the most significant (in terms of amplitude) instances of cross-correlation while avoiding times that are within a fixed window of previous estimates. This is a faster approximation since each waveform amplitude patterns are estimated independently.

Let $\mathbf{x}_m, \quad m \in \{1, \ldots, M\}$ denote a channel of a $M$-channel EEG recording, and let $\mathbf{x}_\star$ denote the channel from which waveform $\mathbf{a}$ (with duration $N$) was estimated. This is the 'anchor' channel and each temporal alignment is based solely on it. Let $\mathcal{T}_k$ denote the set of $k$ timings. The spatial amplitude vectors are found via

$$\tau^k = \mathrm{argmax}_{\tau:|\tau-t|>N, \forall t \in \mathcal{T}_k} \langle \mathbf{x}_\star, T_\tau \mathbf{a} \rangle \tag{16}$$

$$v_m^k = \langle \mathbf{x}_m, T_{\tau^\star} \mathbf{a} \rangle \quad m = 1, \ldots, M \tag{17}$$

Let $V = [\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^k]$ be a matrix where each column is the spatial amplitude vectors of a different occurrence of the waveform. The primary spatial pattern of the waveform corresponds to eigenvector of $VV^\mathsf{T}$ with the largest eigenvalue.

The rationale for using a channel-anchored approach is two-fold: firstly, the timings are maximized for the channel from which the waveform is estimated—if a waveform is completely localized, then using the rest of the channels would bias the timing estimation; and secondly, the non-negativity of the amplitude can be preserved on the original channel while allowing the polarity to flip on other channels. The latter point is especially important depending on the type of channel referencing used. For instance if channels correspond to electrode differences then many waveforms may have different polarities. The channel-anchored approach is an alternative to using multichannel matching pursuits [27] that find the timing using an equal contribution of all the channels.

## IV. Case Study 1: Specificity and Consistency

The first group of datasets we use is publicly available from the Department of Epileptology at the University Hospital of Bonn. The details of the recording are available in the original publication [76]. This set contains 5 EEG datasets with varying characteristics including 'intracortical EEG', recorded by depth electrodes targeting the hippocampal formations. Each of the datasets contain one-hundred 23.6s segments recorded at 173.61Hz. Segments are not necessarily from the same channel nor the same subject; they were cut out of continuous multichannel recordings to avoid artifacts.

The datasets denoted A and B are from 5 healthy volunteers recorded using the standard 10-20 EEG montage. In dataset A the subjects were awake and relaxed with eyes open, and in dataset B the subjects have their eyes closed. Datasets C, D, and E are presurgical recordings used for diagnosis from 5 subjects who had resections of one of the hippocampal formations for the control of seizures. Dataset D was recorded from the epileptogenic zone, and dataset C was recorded from the contralateral hippocampus. Both of these datasets are seizure free, whereas dataset E contains segments from all implanted electrodes during sessions exhibiting ictal spikes.

The multistage modeling consisted of 4 stages with 4 waveforms each. In each stage, the waveform had discrete lengths of 200, 100, 80, and 40, corresponding to approximately 1.15s, 576ms, 460ms, and 230ms respectively. The multistage estimation process used 4 passes of non-overlapping matching pursuit to remove the model approximation before the next stage of waveform estimation. For multitaper analysis, we used 4 tapers and a sequence length of 2000, resulting in a frequency resolution of less than 0.1Hz.

The waveforms estimated using the first 25 sections of each dataset are shown in Fig. 3. The specificity of the waveform shape and frequency content in the different datasets is evident:

[2]To better visualize the full space, one option is to implement a shift-invariant versions of Kohonen's self-organizing map.

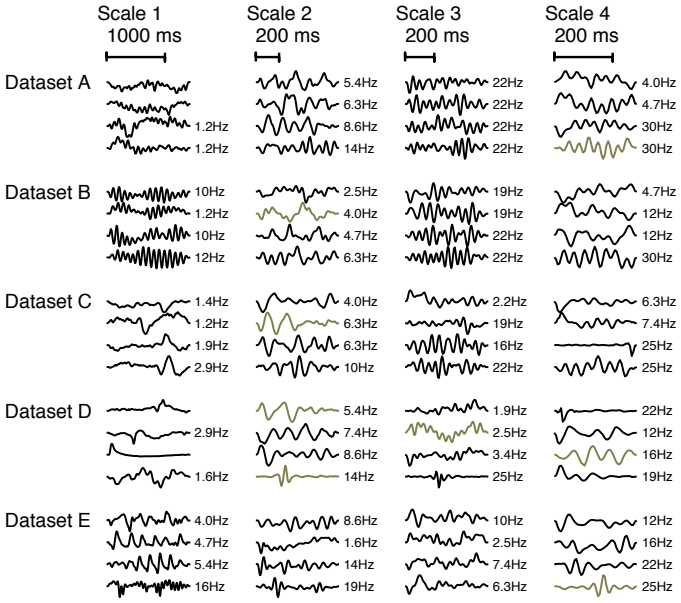



Fig. 3. Case Study 1: Waveforms estimated across the 5 single-channel, ongoing EEG datasets at 4 different scales. Waveforms are scaled to maximum absolute amplitude, those with thicker and lighter coloring matched Gabor-Morlet wavelets with cross-correlation above 0.8, those with peak frequency noted matched with cross-correlation above 0.5.

on dataset B, where the subjects' eyes were closed, all four waveforms at the longest scale correspond to alpha waves; sharp positive waves were estimated from dataset D; and on dataset E oscillatory waveforms in the theta and beta range were estimated. Similar results were obtained when single-channel ICA was used to estimate the waveforms and are included in supplementary material along with example segments of each dataset and signal approximations based on both models. The supplementary material also contains results obtained using subset selection on a dictionary of Gabor-Morlet wavelets.

The specificity of the waveforms to the datasets by their frequency content was assessed for waveforms that matched Gabor-Morlet wavelets. The best matches between the estimated waverforms and Gabor-Morlet wavelets were computed across 4 disjoint sets of 25 contiguous sections and 4 Monte Carlo runs. A scatter plot of the waveforms peak frequency and Q-values are shown for each dataset in Fig. 4. The distribution of parameters appears consistent with the similarity of datasets' recording location/conditions: dataset A has many high-Q waveforms above 20Hz, in dataset B these are absent and replaced by waveforms near 10Hz (alpha waves), datasets C and D have similar distributions, and dataset E has many low-Q waveforms with frequency above 20Hz corresponding. To further analyze the specificity of the waveforms for the different conditions, we computed the average cross-correlation—maximized across shifts—between sets of waveforms across the sections and Monte Carlo runs. The averages across all four scales are shown in Table I.

To further assess the consistency of the waveform estimation, especially the variation across multiple Monte Carlo initializations, we experimented with the number of sections used in training the shift-invariant waveforms. For each dataset

TABLE I
CROSS-CORRELATION BETWEEN WAVEFORMS OF DIFFERENT SUBJECTS

| Subset that match Gabor-Morlet wavelets | | | | | |
|---|---|---|---|---|---|
| | A | B | C | D | E |
| A | 1.00 | 0.72 | 0.66 | 0.73 | 0.72 |
| B | 0.72 | 1.00 | 0.43 | 0.51 | 0.44 |
| C | 0.66 | 0.43 | 1.00 | 0.64 | 0.48 |
| D | 0.73 | 0.51 | 0.64 | 1.00 | 0.59 |
| E | 0.72 | 0.44 | 0.48 | 0.59 | 1.00 |
| All waveforms | | | | | |
| | A | B | C | D | E |
| A | 1.00 | 0.66 | 0.67 | 0.68 | 0.53 |
| B | 0.66 | 1.00 | 0.60 | 0.60 | 0.53 |
| C | 0.67 | 0.60 | 1.00 | 0.75 | 0.57 |
| D | 0.68 | 0.60 | 0.75 | 1.00 | 0.60 |
| E | 0.53 | 0.53 | 0.57 | 0.60 | 1.00 |

we compared the average cross-correlation between waveforms estimated across Monte Carlo initialization to that of waveforms estimated on different sections; the results are shown in Fig. 5. For all datasets and number of sections, the cross-correlation between different initializations is higher than between different subjects as determined by a one-tailed sign test (p-value of 0.031, effect size of 0.83).

We also calculated the average cross-correlation between the waveforms estimated on different datasets. In this case the assessment was done only on the first scale. A one-tailed Wilcoxon rank sum test was used to determine that the correlation among different initialization is higher than between different sections (p-value of 0.00033 and effect size of 0.88 ) and between different subjects (p-value of 0.00397 and effect size of 0.84) for all number of sections used for training. Average, standard deviation error bars, and the cross-correlation values are plotted against the number of training sections in Fig. 6.

This analysis was done using the particular model described above. The model complexity can be justified by using BIC. Specifically, we fix the number of stages and waveform lengths, and find the number of waveforms to minimize BIC; 3 or 4 waveforms per scale was optimal on all subjects. The plot of BIC versus number of waveforms, and the waveforms themselves, are included in supplementary material. We found the estimated waveform shapes to be consistent when the number of waveforms or atoms was varied, both for the MP-SVD and single-channel ICA dictionary learning algorithms and when the greedy subset selection algorithm is applied to a Gabor-Morlet wavelet dictionary. The waveforms and a comparison of the approximations using different number of atoms are also included in the supplementary material.

## V. CASE STUDY 2: MULITCHANNEL, MULTIPLE SUBJECTS

For the second group of datasets, we used the BCI competition III dataset IV(a) [77], provided by Fraunhofer FIRST, Intelligent Data Analysis Group (Klaus-Robert Müller, Benjamin Blankertz), and Campus Benjamin Franklin of the Charité–University Medicine Berlin, Department of Neurology, Neurophysics Group (Gabriel Curio). Healthy human subjects performed visually cued segments of left hand, right hand, and right foot motor imagery while seated. Each cue lasted

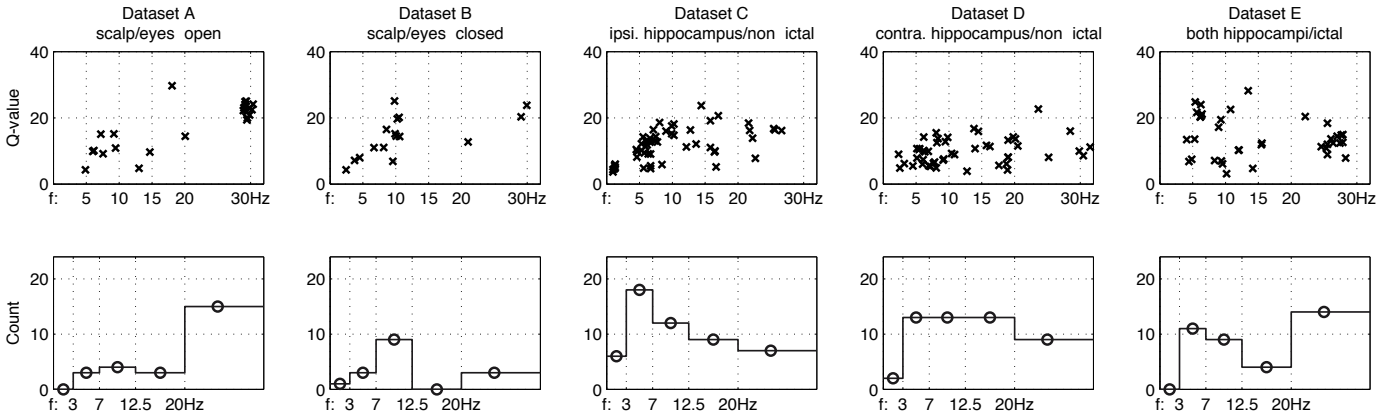



Fig. 4. Case Study 1: Distribution of peak frequency and Q-value (peak frequency divided by bandwidth) for the subset of waveforms that matched Gabor-Morlet wavelets (cross-correlation $\geq 0.8$). Points correspond to waveforms estimated at 4 scales from 4 disjoint sets of 25 contiguous sections across 4 Monte Carlo runs (maximum number of points in a plot is 256). Bottom row shows a histogram of peak frequencies for waveforms.

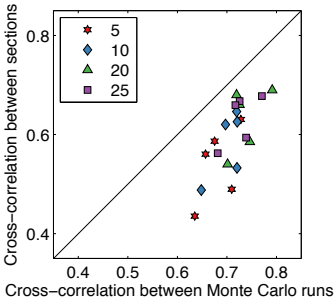



Fig. 5. Case Study 1: Scatter plot of cross-correlation between Monte Carlo initializations versus between sections. The correlation between runs is always higher than the correlation between different sections.

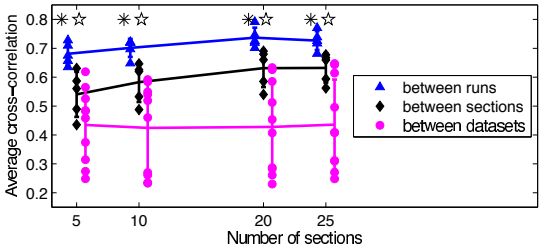



Fig. 6. Case Study 1: Comparison of cross-correlation between waveforms in the first scale estimated on different Monte Carlo initialization, disjoint sets of sections, and between the 5 datasets. Stars and asterisks indicate that the correlation between different initialization runs is significantly higher than correlation between different sections and between different subjects.

3.5s and periods of rest in between had pseudorandom lengths between 1.75s and 2.25s. For each subject, one continuous record was provided per subject along with the timings and labels of right hand and foot cues, with 140 trials of each. The provided recordings were downsampled to 100Hz.

We filtered each channel using a high-pass (0.5Hz) and low-pass (35Hz) first-order Butterworth filter. Noisy channels were removed from analysis on two subjects: channel 2 on 'aa' and 118 on 'ay'. Each record was broken up into four equal length sections. The first section was used in the estimation of waveforms, the remaining sections were kept for testing the decomposition and statistical analysis of the spatial extent between the different classes of motor imagery. The multistage modeling parameters were set similarly to the single-channel case: 4 stages with 4 waveforms each with waveform lengths of 200, 100, 80, and 40, which correspond to 2s, 1s, 0.8s, and 0.4s, respectively.

The results that follow are based on using the MP-SVD dictionary learning algorithm. A complementary version of each result using single-channel ICA is included in the supplementary material. In both cases, after estimation of the waveforms at a particular scale, 4 passes of matching pursuit were ran and the model approximation removed. An example of the decomposition using different numbers of approximation passes is also shown in the supplementary material.

We assessed the oscillatory characteristics of the estimated waveforms by matching them to Gabor-Morlet wavelets and calculating their PSD using multitaper analysis with 4 tapers and a sequence length of 2000. The peak frequency and 3dB bandwidth were recorded from those that matched a wavelet with a cross-correlation above 0.8. A Q-value for each matching waveform was computed as the peak frequency divided by the bandwidth. Fig. 7 shows a scatter plot of the waveforms' peak frequency and Q-values across all channels and subjects. The Q-values are especially high near 10Hz corresponding to alpha waves.

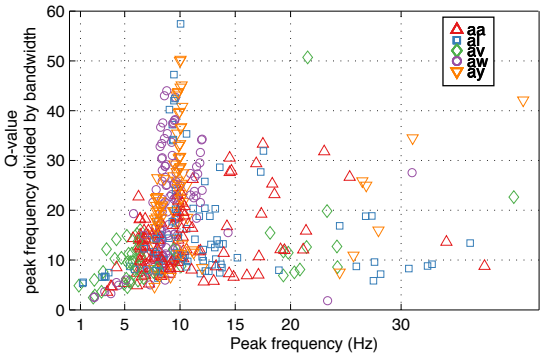



Fig. 7. Case study 2: Scatter plot of peak frequency and Q-value, defined as the peak frequency divided by bandwidth, of waveforms that matched Gabor-Morlet wavelets.

We used cluster analysis to collectively characterize the waveforms estimated across the multiple channels and subjects. For each scale, we ran shift-invariant k-means (with $k = 9$ clusters) on the set of waveforms estimated across different channels and subjects. The cluster centroids are shown in Fig. 8. All of the waveforms in each cluster were also compared to Gabor-Morlet wavelets. Most clusters had very few waveforms matching, but notable exceptions are clusters 1.5, 1.6, and 1.7 (that is the 5th, 6th, and 7th largest clusters in the 1st scale) that seem to correspond to alpha waves. In comparison, many more of the waveforms estimated by single-channel ICA matched Gabor-Morlet wavelets (results shown in the supplementary material).

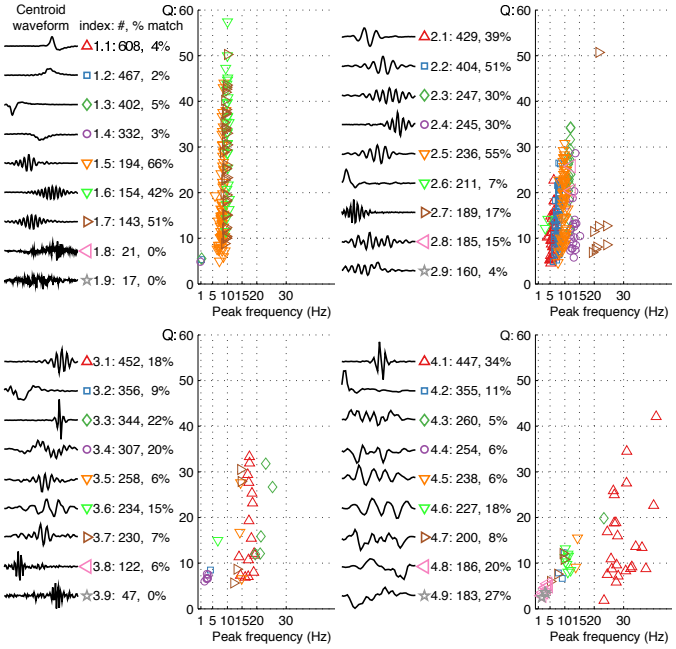



Fig. 8. Case study 2: Cluster centroid waveforms and scatter plot of peak frequency and Q-value for waveforms that matched Gabor-Morlet wavelets. A unique cluster index, the number of waveforms the cluster contains, and the percentage of waveforms that match Gabor-Morlet wavelets is listed to the right of each cluster centroid waveform.

For each cluster of waveforms, we compute the average power spectral density, and choose a prototypical waveform to represent the cluster. The first step in choosing the prototype was to remove waveforms estimated on channels that were flagged as noisy in any of the other subjects; this ensured that the same channel for all subjects could be used for the channel-anchored approach described in Section III-D. The maximum cross-correlation (14) between all pairs of remaining waveforms in the cluster was computed. Cross-correlations less than 0.5 were set to 0, and the waveform that had the largest average cross-correlations to the other waveforms was selected as the cluster prototype.

Descriptors of the clusters in terms of the distribution of channels the waveforms originated from, the average spectral density, and subject distribution are shown in Fig. 9 for waveforms with a duration of 1s. The cluster descriptors for the other scales are shown in the supplementary material. The prototype waveforms for the clusters exhibit a variety of morphologies. Waveforms for clusters 2.1-2.5 correspond to bandpass filters. Cluster 2.1 is the largest cluster at this scale, with a frequency range of 4-8Hz (theta) and distribution across the scalp. Cluster 2.3 has a peak frequency of 11Hz and waveforms from most subjects were estimated from electrodes over the sensorimotor cortex. Cluster 2.5 has a peak frequency of 9.65Hz and originated predominantly from the frontal cortex. Cluster 2.8 has a peak frequency in 12.35Hz and originates near the motor cortex for most subjects.

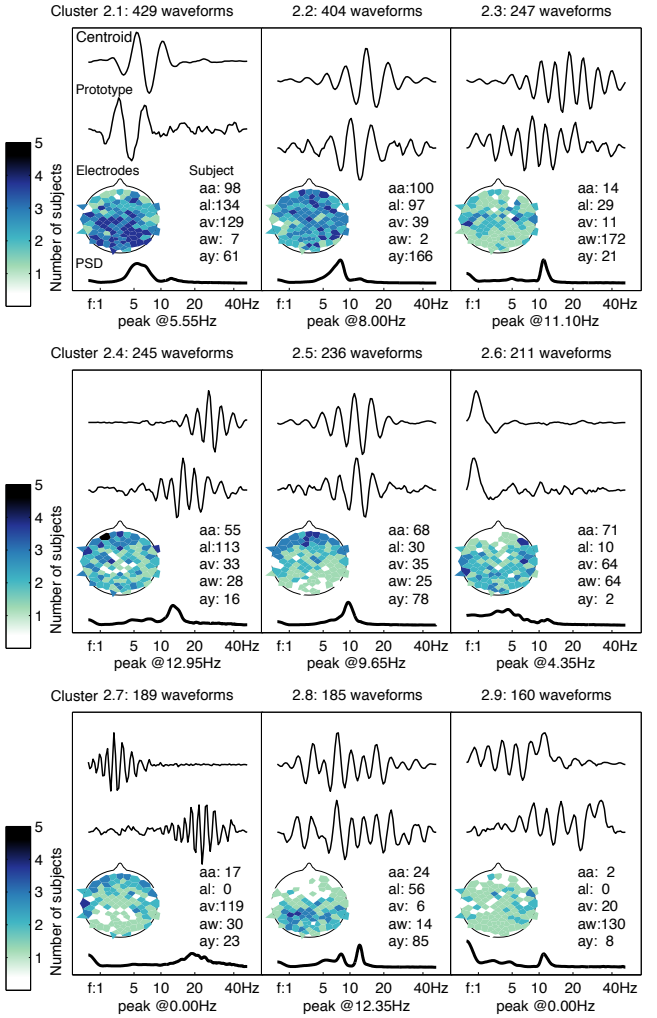



Fig. 9. Case study 2: Cluster descriptors for waveforms with a 1 second duration. Each subplot shows the cluster centroid, prototypical waveform, electrode distribution shown on an unfolded scalp map where the color intensity indicates the number of subjects with waveforms originating from that electrode, and the power spectral density over all waveforms in the cluster.

Finally, we assessed the discrimination between the spatial amplitude patterns of the estimated waveforms during different motor imagery modalities. Fisher's linear discriminant analysis (LDA) was applied to the training data for each waveform and subject. This was done by collecting the spatial amplitude vectors per class, computing the class conditional means, removing the common mean, and computing a common covariance. We found the covariance matrices to be ill-conditioned (LDA requires a matrix inversion), so we added a scaled identity matrix for regularization and robustness [78], choosing a relatively high regularization of 0.75 relative to

the dimension-normalized trace of the covariance matrix. The linear discriminant weights were computed from amplitudes in the same section as the waveform estimation (35 trials per class). The spatial patterns in the remaining sections (total of 105 trials per class) were used for testing. The feature value is the inner product between the linear discriminant and the amplitude patterns for each occurrence during the two types of motor imagery (the mean was removed and amplitude patterns occurring in the same trial were averaged). Classification was done based on the sign of the output. A two-tailed Wilcoxon rank sum test was used to determine if the medians of the resulting values differed between the classes for each cluster and subject. The statistical significance cut-off was set at 0.05 and a Bonferroni correction was applied to the resulting p-values to accommodate for multiple testing of the 36 clusters: the original p-value needed to be less than 0.00139 to be deemed significant.

The single-trial classification accuracy for significant subject-waveform pairs are shown along with the discriminant vector (spatial weights) in Fig. 10. Subject 'al' had the most discriminating patterns and is often reported to have the highest single-trial classification rate in other cross-validation experiments. The highest classification rate occurred for waveforms with peak frequency of 11Hz to 13Hz (prototypes for clusters 2.4, 2.5, and 2.8), within the range of mu rhythms associated with motor imagery. The differential spatial amplitude patterns for the prototype of cluster 2.3 (peak frequency of 11Hz, also within the mu range) was deemed significant on 4 of the 5 waveforms. This analysis highlights the possibility of using shift-invariant dictionary learning as an automatic feature engineering tool for classifying segments of EEGs.

## VI. Relation to Other Work

Learning the recurrent waveforms in EEG is a macroscopic version of the modeling used for spike sorting; however, we are not aided by any prior knowledge on the shape of the waveforms, which makes the problem even more difficult than spike sorting. Recent improvements in spike sorting methodologies are based on allowing overlaps [61] and continuous shifts of waveforms via approximations of translation operators [60]. It is possible to improve the matching pursuit approximations used herein by allowing continuous shifts.

The data-dependent decomposition, which shift-invariant dictionary learning provides, resembles the empirical mode decomposition (EMD). EMD is a model-free time-series analysis technique [79] that has been applied to EEG [80]. The iterative process of EMD is also similar to the multiscale approach we have used. The benefits of our approach versus EMD are the estimate of the sources, in terms of the atomic decomposition, and the learned waveforms, which can be used on novel segments of the signal.

Learning waveforms and then analyzing their spatial extent is complementary to spatial ICA, which finds spatially distinct sources and then analyzes their time-frequency content or uses a bank of band-passes filters to first separate the signals before computing spatial ICA [19]. Working in the time domain enable time-series decompositions to separate morphologically distinct waveforms that may overlap in frequency before assessing their spatial patterns.

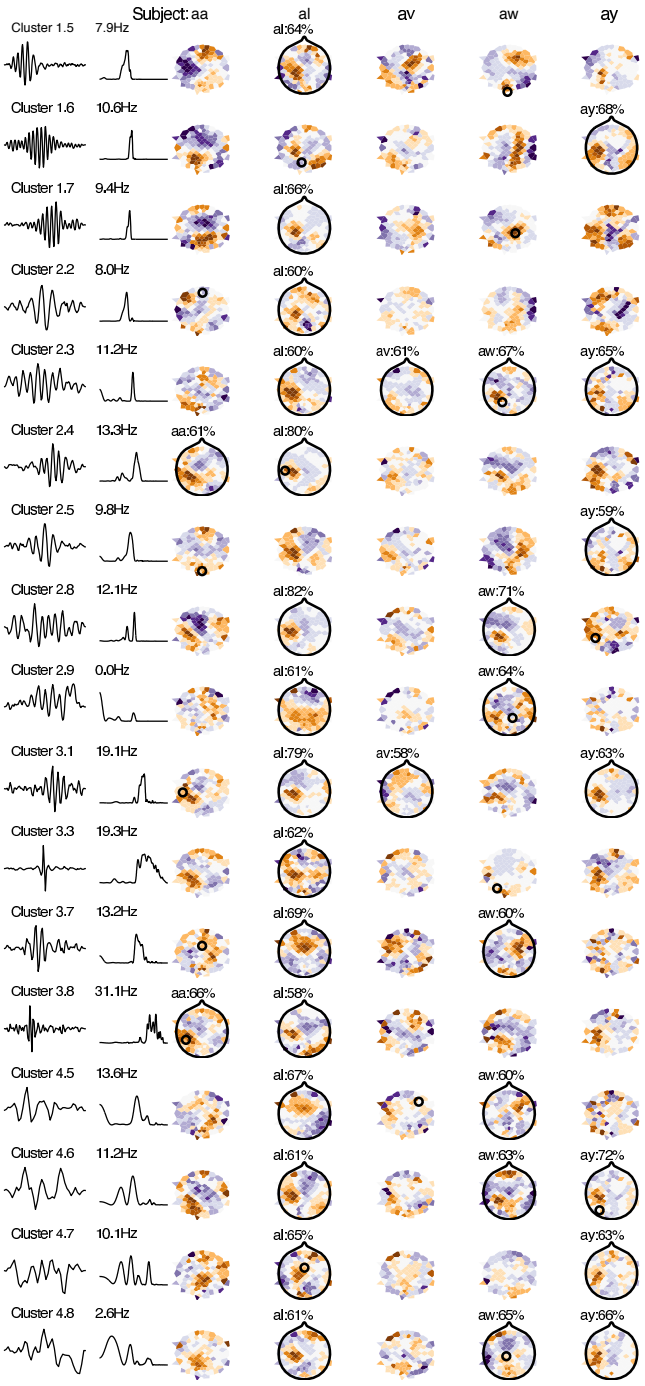



Fig. 10. Case study 2: Prototype waveforms, power spectral density, and spatial weight pattern corresponding to linear discriminate analysis between the spatial amplitudes during two classes of motor imagery. In the color version available online, purple and orange correspond to different signs of the weights. The amplitude patterns where based on the originating electrode of the prototype waveform, which is circled on the originating subject. For a significant waveform/subject pair, the percent accuracy over the testing set is listed. The training/testing split was 70/210.

## VII. Conclusion

With time-frequency analysis it is difficult to separate components in EEG corresponding to waveforms with unique mor-

phology. Previous studies have shown atomic decompositions obtained by matching pursuit using dictionaries of waveforms to be useful for this purpose [25], [27], [28], [29], [30]. However, these methods do not learn any model of the signal, and atomic decompositions on disjoint sections are allowed to use completely different waveforms.

We model each component of an EEG channel as a convolution of a waveform with a sparse source, where the sparsity is based on the assumption that in any given window only a few of the sources are active. To learn this model we apply algorithms that learn data-dependent shift-invariant dictionaries [54], [47], [55], [53]. The modeling constrains the same waveforms to reoccur throughout the recording, with the shape of each waveform adapted to the characteristics of the signal. Since only a small number of waveforms are learned for each channel, the waveforms serve as data-dependent features and are useful for comparing channels, subjects, or conditions.

The results demonstrate that the estimation is consistently able to learn waveforms that are specific to the morphology of the signal. Waveforms estimated during different conditions were distinguished both by their shape and frequency content. For the multichannel datasets, the waveforms estimated over different portions of the scalp differed in morphology and their frequency content showed recognizable localizations. We highlighted how the spatial extent of the waveforms could be used to distinguish between different types of motor imagery.

Automatically learning recurrent waveforms directly from single-channel signals is a general, unsupervised modeling approach. Coupled with appropriate meta-analysis such as clustering and spatial analysis, this method allows a researcher to gain a better understanding of the phasic events and oscillations inherent in EEG signals. For instance, this methodology can be applied to analyze EEG segments with a large presence of phasic events such as EEGs during sleep or deep brain recordings from the hippocampus.

## ACKNOWLEDGMENT

The authors would like to thank Sara Burke, Aysegul Gunduz, Andrew Maurer, and Gavin Philips for their insights.

## REFERENCES

[1] E. D. Adrian and B. H. Matthews, "The Berger rhythm: Potential changes from the occipital lobes in man," *Brain*, vol. 57, no. 4, pp. 355–385, 1934.

[2] L. Ciganek, "The EEG response (evoked potential) to light stimulus in man," *Electroencephalogr. Clin. Neurophysiol.*, vol. 13, no. 2, pp. 165–172, 1961.

[3] G. Buzsáki and A. Draguhn, "Neuronal oscillations in cortical networks," *Science*, vol. 304, no. 5679, pp. 1926–1929, 2004.

[4] T. Sejnowski and O. Paulsen, "Network oscillations: Emerging computational principles," *J. Neurosci.*, vol. 26, no. 6, pp. 1673–1676, 2006.

[5] G. Buzsáki *et al.*, "The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes," *Nat. Rev. Neurosci.*, vol. 13, no. 6, pp. 407–420, 2012.

[6] P. Nunez and R. Srinivasan, *Electric fields of the brain: The neurophysics of EEG*. Oxford University Press, USA, 2006.

[7] X. Wang, "Neurophysiological and computational principles of cortical rhythms in cognition," *Physiol. Rev.*, vol. 90, no. 3, pp. 1195–1268, 2010.

[8] W. J. Freeman, *Mass action in the nervous system*. Academic Press New York, 1975.

[9] H. Berger, "Über das elektrenkephalogramm des menschen," *Eur. Arch. Psychiatry Clin. Neurosci.*, vol. 87, no. 1, pp. 527–570, 1929.

[10] L. A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, 1988.

[11] W. J. Freeman, "Origin, structure, and role of background EEG activity. Part 1. Analytic amplitude," *Clin. Neurophysiol.*, vol. 115, no. 9, pp. 2077–2088, Sep. 2004.

[12] J. Gross, "Analytical methods and experimental approaches for electrophysiological studies of brain oscillations," *J. Neurosci. Methods*, vol. 228, pp. 57–66, 2014.

[13] J. C. Principe and A. J. Brockmeier, "Representing and decomposing neural potential signals," *Curr. Opin. Neurobiol.*, vol. 31, pp. 13–17, 2015, SI: Brain rhythms and dynamic coordination.

[14] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[15] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proc. IEEE*, vol. 84, no. 4, pp. 626–638, 1996.

[16] B. Babadi and E. Brown, "A review of multitaper spectral analysis," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1555–1564, May 2014.

[17] S. Makeig *et al.*, "Independent component analysis of electroencephalographic data," in *Adv. Neural Inf. Process. Syst.*, D. Touretzky *et al.*, Eds. MIT Press, 1996, no. 8, pp. 145–151.

[18] ——, "Dynamic brain sources of visual evoked responses," *Science*, vol. 295, no. 5555, p. 690, 2002.

[19] J. Anemüller *et al.*, "Complex independent component analysis of frequency-domain electroencephalographic data," *Neural Networks*, vol. 16, no. 9, pp. 1311 – 1323, 2003.

[20] M. Zibulevsky and Y. Y. Zeevi, "Extraction of a source from multichannel data using sparse decomposition," *Neurocomputing*, vol. 49, no. 1, pp. 163–173, 2002.

[21] A. Delorme *et al.*, "Independent EEG sources are dipolar," *PLoS ONE*, vol. 7, no. 2, p. e30135, 02 2012.

[22] M. Elad *et al.*, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Appl. Comput. Harmon. Anal.*, vol. 19, no. 3, pp. 340– 358, 2005.

[23] J. Bobin *et al.*, "Morphological component analysis: An adaptive thresholding strategy," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2675–2681, 2007.

[24] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[25] P. Durka and K. Blinowska, "Analysis of EEG transients by means of matching pursuit," *Ann. Biomed. Eng.*, vol. 23, no. 5, pp. 608–611, 1995.

[26] P. J. Durka *et al.*, "Stochastic time-frequency dictionaries for matching pursuit," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 507–510, 2001.

[27] P. Durka *et al.*, "Multichannel matching pursuit and EEG inverse solutions," *J. Neurosci. Methods*, vol. 148, no. 1, pp. 49–59, Oct. 2005.

[28] K. Blinowska, "Methods for localization of time-frequency specific activity and estimation of information transfer in brain," *Int. J. Bioelectromagn.*, vol. 10, no. 1, pp. 2–16, 2008.

[29] C. G. Bénar *et al.*, "Pitfalls of high-pass filtering for detecting epileptic oscillations: a technical note on "false" ripples," *Clin. Neurophysiol.*, vol. 121, no. 3, pp. 301–310, 2010.

[30] N. Jmail *et al.*, "A comparison of methods for separation of transient and oscillatory signals in EEG," *J. Neurosci. Methods*, vol. 199, no. 2, pp. 273–289, 2011.

[31] R. Kuś *et al.*, "Multivariate matching pursuit in optimal Gabor dictionaries: Theory and software with interface for EEG/MEG via Svarog," *Biomed. Eng. Online*, vol. 12, no. 1, p. 94, 2013.

[32] D. Gabor, "Theory of communication. Part 1: The analysis of information," *J. Inst. Electrical Eng.-Part III: Radio and Commun. Eng.*, vol. 93, no. 26, pp. 429–441, 1946.

[33] S. S. Chen *et al.*, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[34] A. Bell and T. Sejnowski, "Learning the higher-order structure of a natural sound," *Network: Comp. Neural*, vol. 7, no. 2, pp. 261–266, 1996.

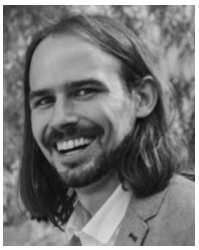
[35] B. Olshausen *et al.*, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[36] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vision Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.

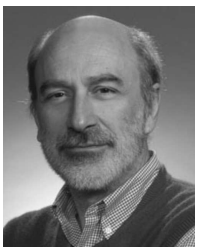
[37] M. Lewicki, "Efficient coding of natural sounds," *Nat. Neurosci.*, vol. 5, no. 4, pp. 356–363, 2002.

[38] J. Mairal *et al.*, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, 2010.

[39] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, March 2011.
[40] C. James and D. Lowe, "Extracting multisource brain activity from a single electromagnetic channel," *Artif. Intell. Med.*, vol. 28, no. 1, pp. 89 – 104, 2003.
[41] M. Davies and C. James, "Source separation using single channel ICA," *Signal Process.*, vol. 87, no. 8, pp. 1819–1832, 2007.
[42] F. Lucena *et al.*, "Statistical coding and decoding of heartbeat intervals," *PLoS ONE*, vol. 6, no. 6, p. e20227, 06 2011.
[43] T. Virtanen, "Separation of sound sources by convolutive sparse coding," in *ISCA Tutorial Res. Work. Stat. Percept. Audio Process.*, 2004.
[44] T. Blumensath and M. Davies, "Sparse and shift-invariant representations of music," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 1, pp. 50–57, 2006.
[45] E. Smith and M. Lewicki, "Efficient auditory coding," *Nature*, vol. 439, no. 7079, pp. 978–982, 2006.
[46] R. Grosse *et al.*, "Shift-invariant sparse coding for audio classification," in *Proc. 23rd Conf. Uncert. Artif. Intell.*, 2007.
[47] B. Mailhé *et al.*, "Shift-invariant dictionary learning for sparse representations: Extending K-SVD," in *16th Euro. Signal Process. Conf.*, 2008.
[48] ——, "Dictionary learning for the sparse modelling of atrial fibrillation in ECG signals," in *IEEE Int. Conf. Acoust. Speech Signal Process.*, 2009, pp. 465–468.
[49] D. C. Balcan and M. S. Lewicki, "Point coding: Sparse image representation with adaptive shiftable-kernel dictionaries," in *Signal Proc. Adaptive Sparse Structured Representations*, 2009.
[50] C. Ekanadham *et al.*, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4735–4744, 2011.
[51] Q. Barthelemy *et al.*, "Shift & 2D rotation invariant sparse coding for multivariate signals," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1597–1611, April 2012.
[52] A. J. Brockmeier and J. C. Principe, "Explicit versus implicit source estimation for blind multiple input single output system identification," in *IEEE Int. Conf. Acoust. Speech Signal Process.*, 2015, pp. 2140–2144.
[53] Q. Barthélemy *et al.*, "Multivariate temporal dictionary learning for EEG," *J. Neurosci. Methods*, vol. 215, no. 1, pp. 19–28, 2013.
[54] M. Aharon, "Overcomplete dictionaries for sparse representation of signals," Ph.D. dissertation, Technion-Israel Institute of Technology, Faculty of Computer Science, 2006.
[55] J. Thiagarajan *et al.*, "Shift-invariant sparse representation of images using learned dictionaries," in *IEEE Work. Mach. Learn. Signal Process.*, Oct 2008, pp. 145–150.
[56] D. Charalampidis, "A modified k-means algorithm for circular invariant clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1856–1865, 2005.
[57] Y. Ruiz *et al.*, "A method to study global spatial patterns related to sensory perception in scalp EEG." *J. Neurosci. Methods*, vol. 191, no. 1, pp. 110–118, Aug. 2010.
[58] A. J. Brockmeier *et al.*, "Locating spatial patterns of waveforms during sensory perception in scalp EEG," in *Annu. Int. Conf. IEEE Eng. Med. Bio. Soc.*, Sept. 2012, pp. 2531–2534.
[59] M. Lewicki and T. Sejnowski, "Coding time-varying signals using sparse, shift-invariant representations," *Adv. Neural Inf. Process. Syst.*, pp. 730–736, 1999.
[60] C. Ekanadham *et al.*, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4735–4744, 2011.
[61] J. W. Pillow *et al.*, "A model-based spike sorting algorithm for removing correlation artifacts in multi-neuron recordings," *PLoS ONE*, vol. 8, no. 5, p. e62123, 2013.
[62] C. Ekanadham *et al.*, "A unified framework and method for automatic neural spike identification," *J. Neurosci. Methods*, vol. 222, pp. 47–55, 2014.
[63] J. C. de Munck *et al.*, "A maximum-likelihood estimator for trial-to-trial variations in noisy MEG/EEG data sets," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 12, pp. 2123–2128, 2004.
[64] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Comput.*, vol. 13, no. 4, pp. 863–882, 2001.
[65] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, October 2004.
[66] M. Aharon *et al.*, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.
[67] F. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978.
[68] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.
[69] N. Gillis and A. Kumar, "Exact and heuristic algorithms for semi-nonnegative matrix factorization," *SIAM J. Matrix Anal. Appl.*, vol. 36, no. 4, pp. 1404–1424, 2015.
[70] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.
[71] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 312–321, 1990.
[72] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, 1999.
[73] Y. Pati *et al.*, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Conf. Rec. 26th Asilomar Conf. Signals Syst. Comp.*, 1993, pp. 40–44.
[74] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, 1982.
[75] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Stat.-Theory*, vol. 3, no. 1, pp. 1–27, 1974.
[76] R. G. Andrzejak *et al.*, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, p. 061907, 2001.
[77] B. Blankertz *et al.*, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, 2006.
[78] J. H. Friedman, "Regularized discriminant analysis," *J. Am. Stat. Assoc.*, vol. 84, no. 405, pp. 165–175, 1989.
[79] N. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London Ser. A*, vol. 454, no. 1971, pp. 903–995, 1998.
[80] B. Mijović and *et al.*, "Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 9, pp. 2188–2196, sept. 2010.



**Austin J. Brockmeier** (S'05-M'15) received the B.S. degree in computer engineering in 2009 from the University of Nebraska–Lincoln while attending the Peter Kiewit Institute in Omaha, NE, USA. He earned the Ph.D. degree in electrical engineering in 2014 from the University of Florida in Gainesville, FL, USA. He is currently a Research Associate in the Department of Electrical Engineering and Electronics at the University of Liverpool, UK. His research interests are signal and information processing, machine learning, and diverse applications of exploratory data analysis including biomedicine and public health.



**José C. Príncipe** (M'83-SM'90-F'00) received the Ph.D. degree in electrical engineering from the University of Florida, Gainesville, FL, USA, in 1979. He was a Professor at the University of Aveiro, Portugal, from 1980 to 1987. He is a BellSouth, Distinguished Professor of Electrical and Biomedical Engineering at the University of Florida where he is Founding Director of the University of Florida Computational NeuroEngineering Laboratory. His research interests are centered on advanced signal processing and machine learning, brain-machine interfaces, and the modeling and applications of cognitive systems. He is a Fellow of the IEEE, ABME, and AIBME. He is the past Editor-in-Chief of the IEEE Transactions on Biomedical Engineering, past Chair of the Technical Committee on Neural Networks of the IEEE Signal Processing Society, Past-President of the International Neural Network Society, and a recipient of the IEEE EMBS Career Award and the IEEE Neural Network Pioneer Award.

# Learning Recurrent Waveforms within EEGs
## *Supplementary Material*

Austin J. Brockmeier and Jose C. Principe

### S1. Model Selection

Multiple models with different waveform lengths, numbers of waveforms, and approximations with different numbers of atoms can be trained and compared using model selection criteria. To compare models, we assume the background activity is white noise with constant variance. For instance, we show how the Bayesian information criterion (BIC) can be used for selecting the number of waveforms. Specifically, in the case of Gaussian white noise, the model with the minimal Bayesian information criterion is chosen:

$$BIC = \log RSS - \frac{C}{N} \log N \tag{1}$$

where $C$ is degrees of freedom, $N$ is the length of the signal, and $RSS$ is the sum of the squared values of the residual, that is the portion of the signal that is not explained by the model. For the multiple-input-single-output model with $L$ atoms $C = L + \sum_{p=1}^{P} (M_p - 1)$, where $M_p$ is the length of the $p$th waveform out of $P$. BIC evaluated across different numbers of waveforms for each of the subjects is shown in Fig. S1.

However, we note that BIC is not consistent for choosing the number of atoms across different signal lengths. That is, if the signal is twice as long we should expect that twice the number atoms should be allowed. However, because of the $\log N$ term, the amount of variance each atom should explain increases as the signal length increases. As an alternative one could use the Akaike information criterion (AIC)

$$AIC = \log RSS - \frac{2C}{N}, \tag{2}$$

which allows an increase in the number of atoms proportional to the relative increase in signal length. In practice we note that the AIC is very lenient allowing many more atoms than appear necessary. As a harsher penalty we propose a simple and intuitive penalty that can be used to choose the number of atoms and thus as a stopping criterion for a matching pursuit decomposition. Specifically, assuming a white noise error model, an atom must explain more variance than expected from a Kronecker delta atom, which could perfectly fit the largest time-point in the noise. Based on this intuition, we keep an atom if explains more than three times the square root of the remaining signal power. This stopping criterion, which refer to as 3-std, naturally adapts to the amount of noise in the signal and to how well the atoms fit the signal. If the candidate atoms have a poor fit it will stop early. We use this stopping criterion on the synthetic experiments detailed in Section S4.

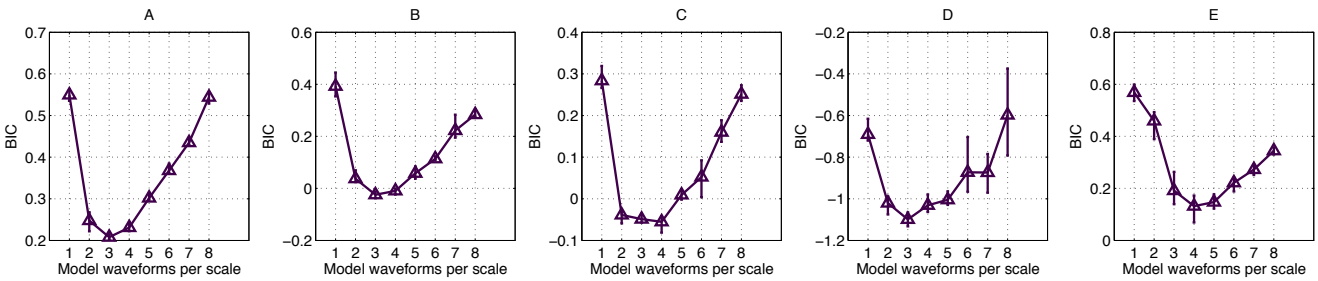



Fig. S1. Case Study 1: Bayesian information criterion (BIC) evaluated across the number of waveforms per scale for each of the 5 single-channel, ongoing EEG datasets, when the number of atoms, scales, and waveform length is fixed.

## S2. Parameter Exploration on Case Study 1

In this section we include some additional results that compare the MP-SVD dictionary learning, single-channel ICA (SC-ICA) with subset selection, and Gabor-Morlet dictionary subset selection across different number of waveforms. The case study used datasets publicly available from the Department of Epileptology at the University Hospital of Bonn. The details of the recording are available can be found in Andrzejak *et al., Physical Review E,* 2001.

Fig. S2 shows a side-by-side comparison of the 4 waveforms learned/selected at each scale across the 5 single-channel EEG datasets. Fig. S3, Fig. S4, and Fig. S5 show the waveforms learned when different numbers of waveforms per scale were allowed.

Fig. S6 show approximations of testing segments of each dataset using the 4 waveforms shown in Fig. S2 with enough atoms per scale such that each scale can cover 10% of the signal with non-overlapping waveforms. Fig. S7 shows the results with enough atoms to provide 100% coverage per scale. The approximation error (in terms of proportion of variance explained) for the MP-SVD and single-channel ICA across different numbers of waveforms and atoms is reported in Table S1. As the number of atoms increases the proportion of variance explained by MP-SVD reaches 95% using only two waveforms per scale. Even with eight waveforms SC-ICA is not able reach 90% and has much higher variance. Table S2 also reports the results using MP-SVD along with using the subset of Gabor waveforms. Using 5 Gabor waveforms per scale is able to achieve 95% variance explained. The Gabor subset performs worse than the waveforms learned by MP-SVD but much better than those learned by SC-ICA and selected using the same subset selection algorithm. This means that the set of waveforms found by SC-ICA is suboptimal in terms of mean-squared error; however, we note that they still appear to be tuned to the different characteristics of the datasets as can be seen in Fig. S4.


Fig. S2. Case Study 1: Waveforms estimated across the 5 single-channel, ongoing EEG datasets at 4 different scales. (Top left) Waveforms learned by MP-SVD. (Top right) Waveforms learned by single-channel ICA with subset selection. (Bottom) Waveforms selected from a Gabor-Morlet wavelet dictionary using subset selection. Waveforms are scaled to maximum absolute amplitude, those with lighter coloring matched Gabor wavelets with cross-correlation $\geq$ 0.8.

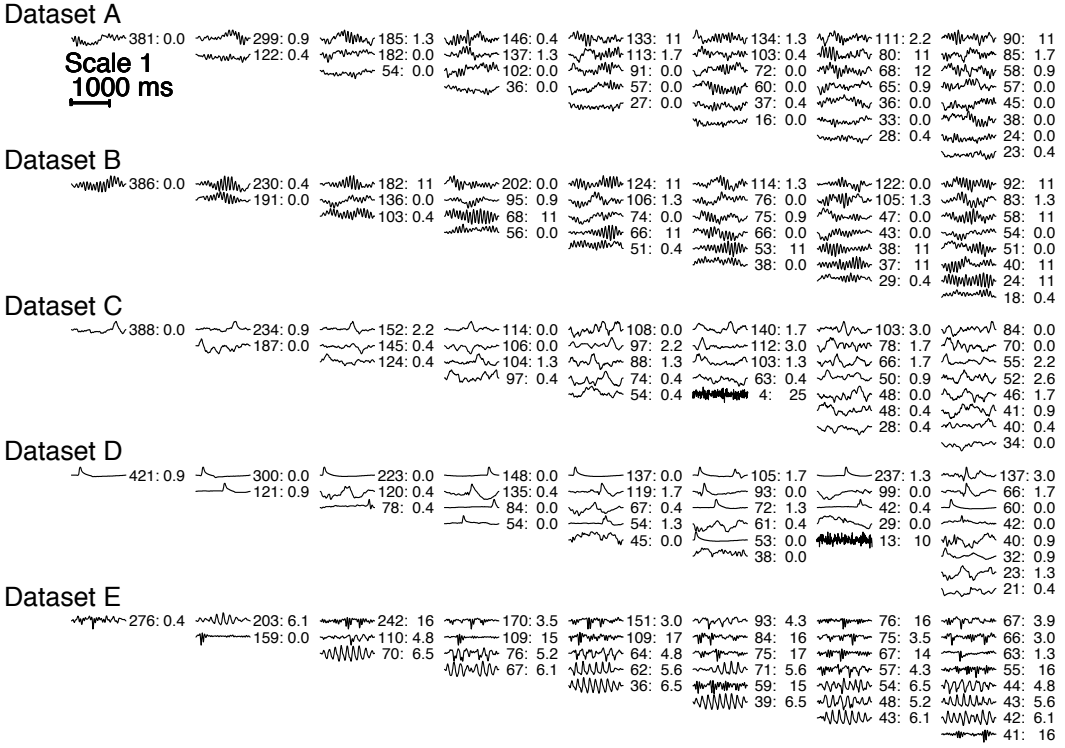



Fig. S3. Case Study 1: Waveform sets estimated using the alternating matching pursuit with SVD update (MP-SVD) on the 5 single-channel, ongoing EEG datasets. Each column corresponds to models with different number of waveforms. The two number listed to the right of each waveform indicate the number of times the waveform was used in the approximation of the training signal, and the peak frequency of the waveform. Many waveforms appear consistently as the number of total waveforms is increased.

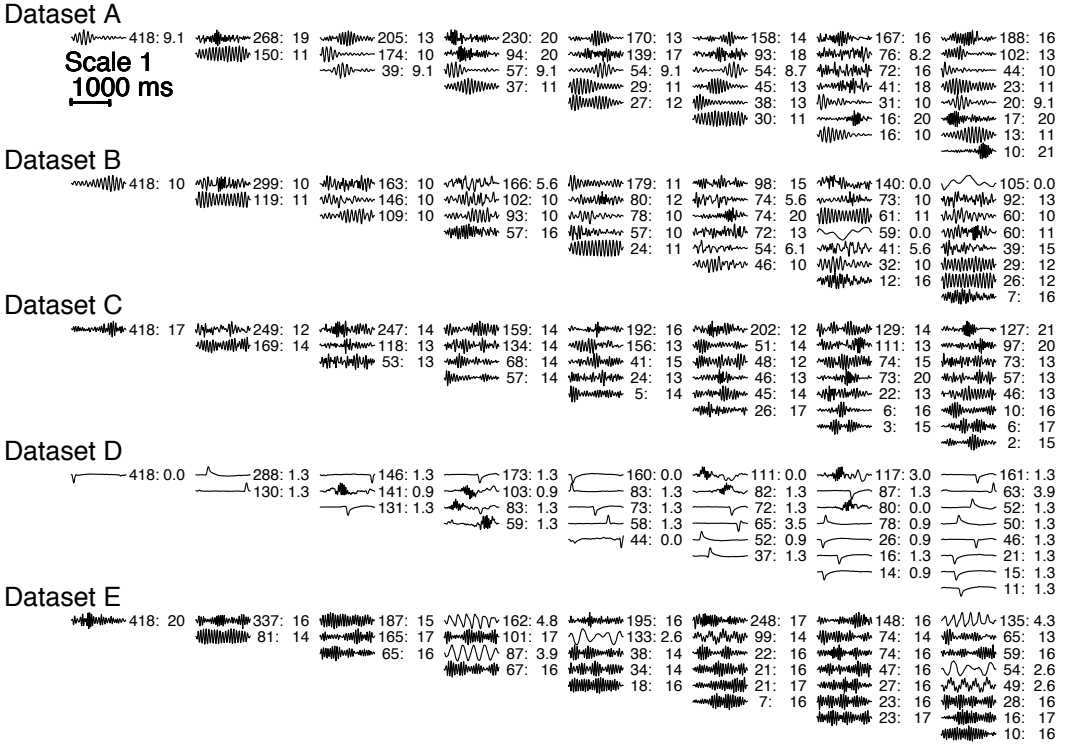



Fig. S4. Case Study 1: Waveform sets estimated using the single-channel ICA with subset selection on the 5 single-channel, ongoing EEG datasets. Each column corresponds to models with different number of waveforms. The two number listed to the right of each waveform indicate the number of times the waveform was used in the approximation of the training signal, and the peak frequency of the waveform. Many waveforms appear consistently as the number of total waveforms is increased.

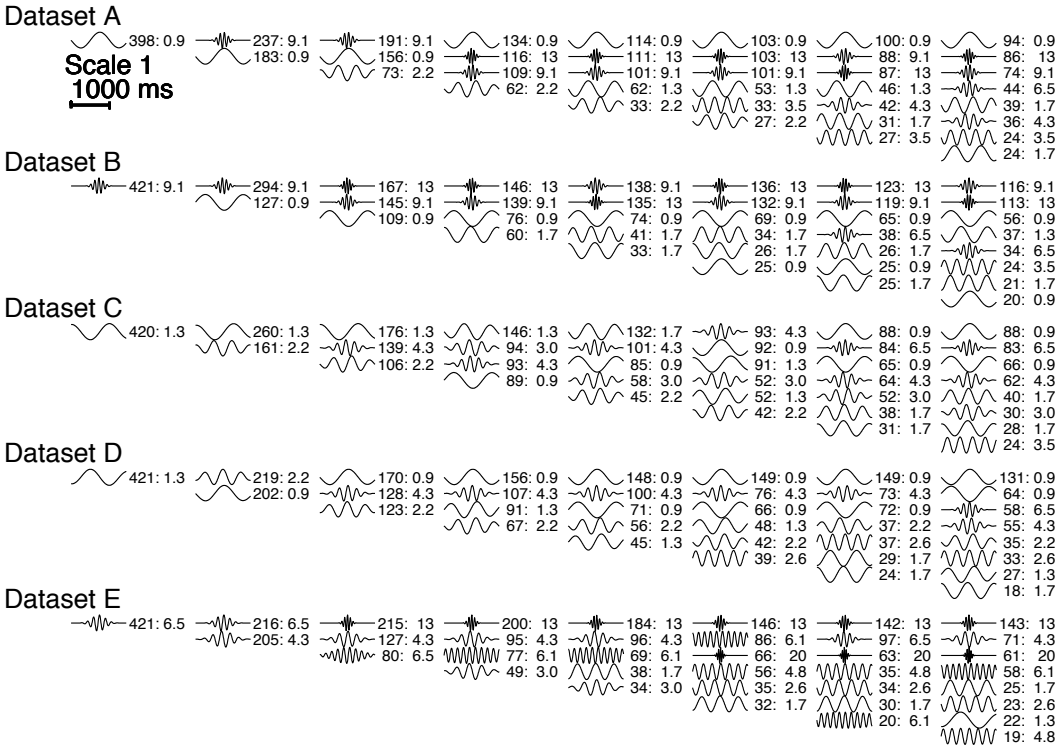



Fig. S5. Case Study 1: Gabor-Morlet wavelet sets selected using greedy dictionary subset selection on the 5 single-channel, ongoing EEG datasets. Each column corresponds to models with different number of waveforms. The two number listed to the right of each waveform indicate the number of times the waveform was used in the approximation of the training signal, and the peak frequency of the waveform. Many waveforms appear consistently as the number of total waveforms is increased.

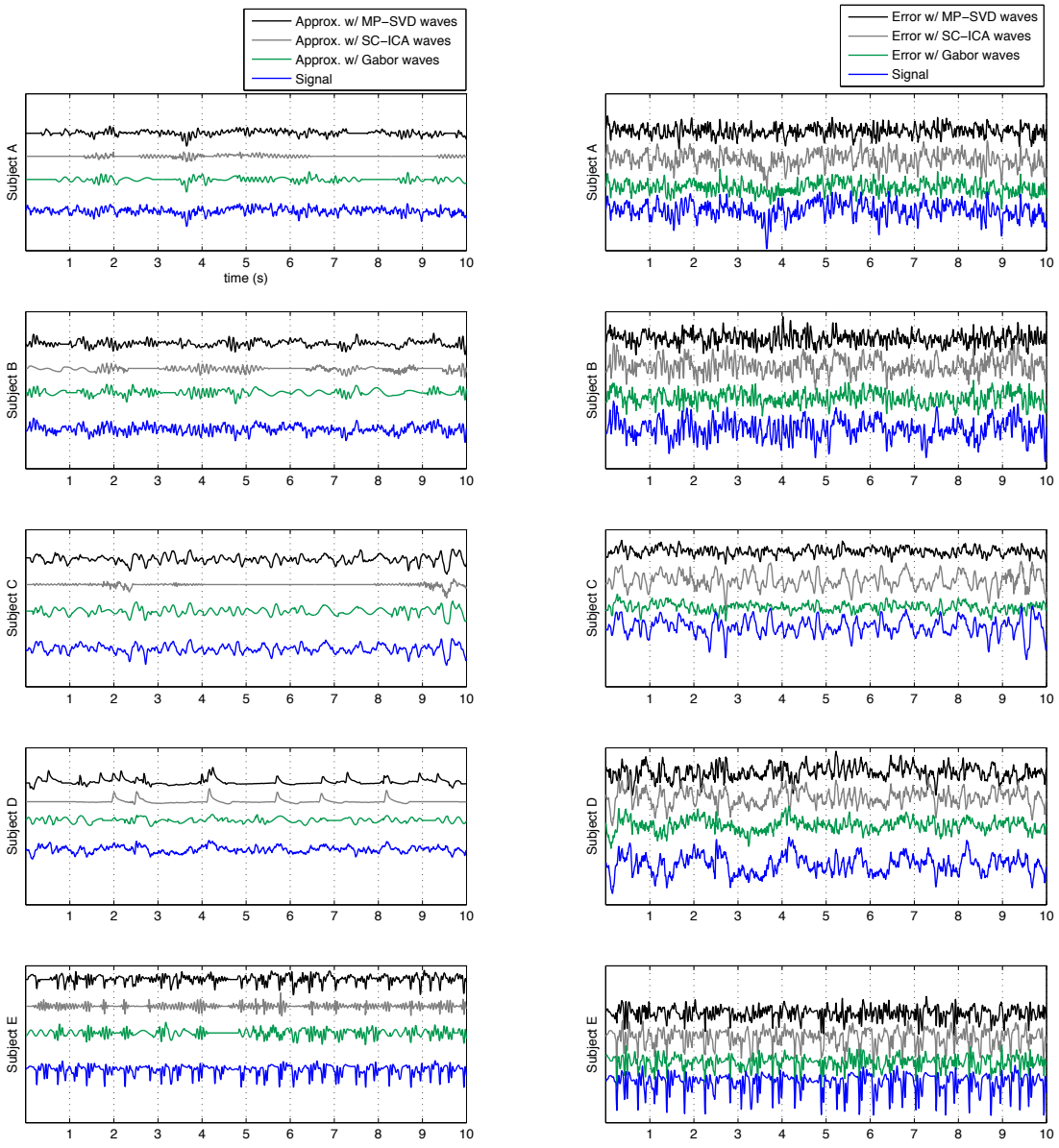



Fig. S6. Case Study 1: Example approximations and residual (error) for a segment of each of the 5 single-channel, ongoing EEG datasets using waveforms learned via the alternating matching pursuit with SVD update (MP-SVD), single-channel ICA with subset selection (SC-ICA), and subset selected Gabor-Morlet wavelets (Gabor waves). The number of atoms is set such that the approximation waveforms cover 10% of the original signal per scale. At this level of sparsity the approximations are truly sparse, but many waveforms remain in the error.

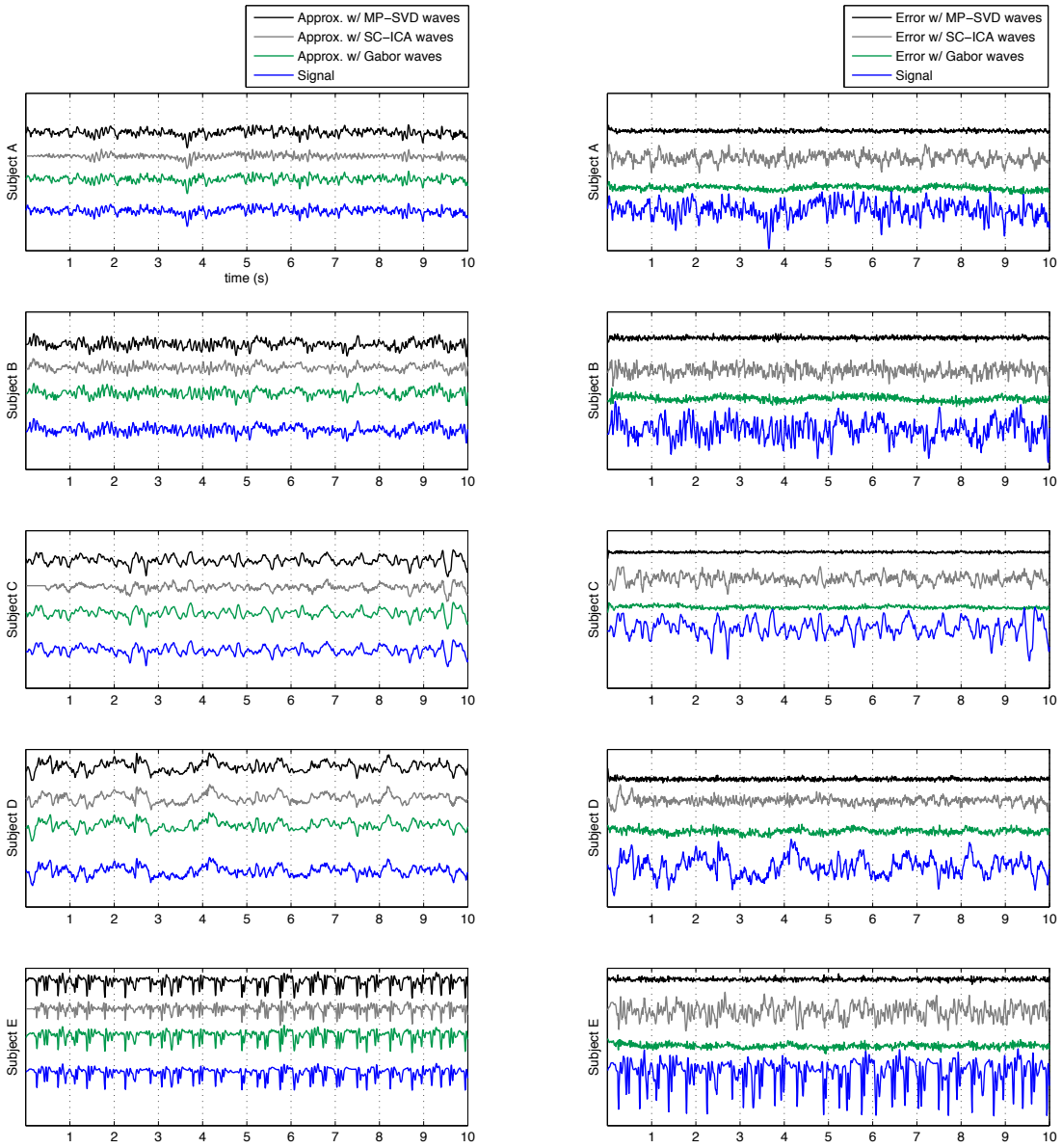



Fig. S7. Case Study 1: Example approximations and approximation errors for a segment of each of the 5 single-channel, ongoing EEG datasets using waveforms learned via the alternating matching pursuit with SVD update (MP-SVD), single-channel ICA with subset selection (SC-ICA), and subset selected Gabor-Morlet wavelets (Gabor waves). The number of atoms is set such that the approximation waveforms cover 100% of the original signal per scale. At this level of sparsity the approximations explain most of the signal. The approximation error using MP-SVD waveforms is minimal, the Gabor waveform approximation exhibits some low-frequency trends, but the error from single-channel ICA has much higher variance.

TABLE S1

CASE STUDY 1: COMPARISON OF THE NORMALIZED APPROXIMATION ERROR FOR THE TESTING PORTION OF THE 5 SINGLE-CHANNEL, ONGOING EEG DATASETS USING WAVEFORMS LEARNED VIA THE ALTERNATING MATCHING PURSUIT WITH SVD UPDATE (MP-SVD) VERSUS SINGLE-CHANNEL ICA WITH SUBSET SELECTION (SC-ICA). THE COLUMNS CORRESPOND TO THE FREQUENCY OF THE ATOMS ACROSS ALL SCALES, AND THE ROWS CORRESPOND TO THE NUMBER OF WAVEFORMS IN EACH SCALE OF THE MODEL. ENTRIES INDICATE THE AVERAGE AND STANDARD DEVIATION ACROSS THE 5 DATASETS.

| | 2.69Hz | 6.80Hz | 13.63Hz | 21.03Hz | 27.37Hz | 41.44Hz | 53.29Hz | 111.14Hz |
|---|---|---|---|---|---|---|---|---|
| Proportion of variance explained with MP-SVD waveforms | | | | | | | | |
| 1 | 0.43±0.18 | 0.58±0.13 | 0.69±0.11 | 0.73±0.10 | 0.76±0.09 | 0.81±0.07 | 0.85±0.05 | 0.84±0.06 |
| 2 | 0.48±0.16 | 0.64±0.13 | 0.77±0.09 | 0.83±0.07 | 0.85±0.07 | 0.89±0.06 | 0.92±0.04 | 0.95±0.02 |
| 3 | 0.50±0.16 | 0.67±0.13 | 0.79±0.08 | 0.85±0.06 | 0.88±0.05 | 0.91±0.04 | 0.94±0.02 | 0.96±0.01 |
| 4 | 0.50±0.13 | 0.69±0.12 | 0.81±0.07 | 0.87±0.05 | 0.90±0.04 | 0.93±0.02 | 0.95±0.02 | 0.97±0.01 |
| 5 | 0.51±0.17 | 0.70±0.11 | 0.82±0.07 | 0.88±0.04 | 0.91±0.03 | 0.94±0.02 | 0.95±0.02 | 0.97±0.01 |
| 6 | 0.50±0.17 | 0.69±0.10 | 0.81±0.03 | 0.89±0.04 | 0.91±0.03 | 0.94±0.02 | 0.96±0.02 | 0.98±0.01 |
| 7 | 0.49±0.17 | 0.70±0.11 | 0.83±0.06 | 0.89±0.04 | 0.92±0.03 | 0.95±0.02 | 0.96±0.01 | 0.98±0.01 |
| 8 | 0.50±0.16 | 0.71±0.11 | 0.83±0.05 | 0.89±0.04 | 0.92±0.03 | 0.95±0.02 | 0.96±0.01 | 0.98±0.01 |
| Proportion of variance explained with SC-ICA waveforms | | | | | | | | |
| 1 | 0.20±0.20 | 0.25±0.21 | 0.34±0.22 | 0.39±0.26 | 0.39±0.24 | 0.41±0.14 | 0.39±0.11 | 0.54±0.18 |
| 2 | 0.23±0.22 | 0.29±0.20 | 0.34±0.19 | 0.48±0.17 | 0.46±0.21 | 0.51±0.21 | 0.62±0.16 | 0.64±0.14 |
| 3 | 0.23±0.21 | 0.32±0.23 | 0.39±0.23 | 0.46±0.22 | 0.53±0.21 | 0.58±0.21 | 0.65±0.14 | 0.78±0.13 |
| 4 | 0.25±0.22 | 0.28±0.12 | 0.43±0.22 | 0.50±0.19 | 0.53±0.14 | 0.63±0.17 | 0.70±0.15 | 0.83±0.08 |
| 5 | 0.25±0.21 | 0.33±0.19 | 0.44±0.21 | 0.49±0.20 | 0.58±0.17 | 0.64±0.17 | 0.73±0.12 | 0.85±0.08 |
| 6 | 0.26±0.21 | 0.37±0.21 | 0.46±0.19 | 0.46±0.20 | 0.58±0.19 | 0.67±0.17 | 0.74±0.13 | 0.86±0.10 |
| 7 | 0.26±0.21 | 0.35±0.21 | 0.51±0.18 | 0.59±0.18 | 0.62±0.18 | 0.70±0.16 | 0.75±0.13 | 0.84±0.13 |
| 8 | 0.28±0.22 | 0.37±0.23 | 0.47±0.19 | 0.59±0.18 | 0.61±0.17 | 0.69±0.17 | 0.74±0.17 | 0.88±0.08 |
| Difference of squared error: SC-ICA minus MP-SVD | | | | | | | | |
| 1 | 0.23±0.06 | 0.33±0.10 | 0.34±0.12 | 0.33±0.18 | 0.37±0.16 | 0.40±0.07 | 0.46±0.13 | 0.30±0.17 |
| 2 | 0.25±0.08 | 0.35±0.09 | 0.43±0.11 | 0.34±0.12 | 0.40±0.16 | 0.38±0.17 | 0.30±0.13 | 0.30±0.13 |
| 3 | 0.26±0.09 | 0.34±0.13 | 0.40±0.16 | 0.39±0.17 | 0.35±0.17 | 0.34±0.17 | 0.29±0.12 | 0.19±0.13 |
| 4 | 0.25±0.11 | 0.40±0.07 | 0.37±0.16 | 0.37±0.15 | 0.37±0.12 | 0.30±0.15 | 0.25±0.14 | 0.14±0.08 |
| 5 | 0.25±0.08 | 0.37±0.11 | 0.37±0.16 | 0.38±0.16 | 0.33±0.14 | 0.29±0.15 | 0.22±0.10 | 0.12±0.08 |
| 6 | 0.24±0.09 | 0.32±0.12 | 0.35±0.18 | 0.43±0.17 | 0.33±0.16 | 0.27±0.15 | 0.22±0.12 | 0.12±0.10 |
| 7 | 0.23±0.10 | 0.35±0.13 | 0.33±0.13 | 0.30±0.15 | 0.30±0.16 | 0.25±0.15 | 0.21±0.12 | 0.13±0.13 |
| 8 | 0.22±0.11 | 0.33±0.14 | 0.37±0.14 | 0.30±0.15 | 0.31±0.15 | 0.26±0.15 | 0.22±0.17 | 0.10±0.07 |
| Squared error ratio (log 2): a value of 1 indicates indicates SC-ICA has twice the squared error of MP-SVD | | | | | | | | |
| 1 | 0.51±0.17 | 0.84±0.13 | 1.09±0.17 | 1.12±0.33 | 1.32±0.17 | 1.70±0.33 | 2.03±0.68 | 1.54±0.68 |
| 2 | 0.59±0.14 | 1.02±0.18 | 1.56±0.25 | 1.62±0.29 | 1.94±0.39 | 2.18±0.31 | 2.21±0.28 | 2.76±0.41 |
| 3 | 0.63±0.20 | 1.06±0.27 | 1.55±0.21 | 1.90±0.39 | 2.00±0.30 | 2.24±0.25 | 2.54±0.42 | 2.50±0.64 |
| 4 | 0.57±0.16 | 1.30±0.56 | 1.56±0.29 | 1.96±0.36 | 2.30±0.33 | 2.39±0.37 | 2.37±0.59 | 2.47±0.45 |
| 5 | 0.64±0.19 | 1.20±0.31 | 1.62±0.30 | 1.98±0.24 | 2.22±0.30 | 2.44±0.43 | 2.49±0.36 | 2.31±0.68 |
| 6 | 0.61±0.22 | 1.04±0.15 | 1.41±0.54 | 2.28±0.39 | 2.19±0.30 | 2.45±0.51 | 2.53±0.51 | 2.26±1.03 |
| 7 | 0.56±0.21 | 1.16±0.22 | 1.60±0.29 | 1.84±0.37 | 2.15±0.44 | 2.41±0.48 | 2.45±0.62 | 2.43±1.13 |
| 8 | 0.53±0.24 | 1.10±0.25 | 1.65±0.22 | 1.91±0.36 | 2.31±0.37 | 2.44±0.54 | 2.43±0.92 | 2.34±0.80 |
| Win count (n=5) for MP-SVD: p-value for 5 is 0.0312: 5 wins is significant at a level of 0.05. | | | | | | | | |
| 1 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 2 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 7 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 8 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

TABLE S2

CASE STUDY 1: COMPARISON OF THE NORMALIZED APPROXIMATION ERROR FOR THE TESTING PORTION OF THE 5 SINGLE-CHANNEL, ONGOING EEG DATASETS USING WAVEFORMS LEARNED VIA THE ALTERNATING MATCHING PURSUIT WITH SVD UPDATE (MP-SVD) VERSUS A SUBSET (WITH THE SAME SIZE) SELECTED FROM A GABOR-MORLET WAVELET DICTIONARY (GABOR). THE COLUMNS CORRESPOND TO THE FREQUENCY OF THE ATOMS ACROSS ALL SCALES, AND THE ROWS CORRESPOND TO THE NUMBER OF WAVEFORMS IN EACH SCALE OF THE MODEL. ENTRIES INDICATE THE AVERAGE AND STANDARD DEVIATION ACROSS THE 5 DATASETS.

| | 2.71Hz | 6.80Hz | 13.59Hz | 20.74Hz | 27.50Hz | 40.74Hz | 54.39Hz | 107.69Hz |
|---|---|---|---|---|---|---|---|---|
| Proportion of variance explained with MP-SVD waveforms | | | | | | | | |
| 1 | 0.42±0.16 | 0.57±0.13 | 0.68±0.10 | 0.72±0.09 | 0.75±0.09 | 0.80±0.07 | 0.83±0.05 | 0.81±0.07 |
| 2 | 0.48±0.17 | 0.64±0.13 | 0.76±0.08 | 0.82±0.07 | 0.85±0.07 | 0.89±0.06 | 0.91±0.04 | 0.94±0.02 |
| 3 | 0.48±0.15 | 0.65±0.13 | 0.79±0.08 | 0.85±0.06 | 0.88±0.05 | 0.91±0.04 | 0.93±0.04 | 0.96±0.01 |
| 4 | 0.46±0.15 | 0.68±0.12 | 0.80±0.07 | 0.86±0.04 | 0.90±0.04 | 0.93±0.02 | 0.94±0.02 | 0.97±0.01 |
| 5 | 0.47±0.14 | 0.69±0.10 | 0.81±0.07 | 0.87±0.04 | 0.91±0.03 | 0.94±0.02 | 0.95±0.02 | 0.97±0.01 |
| 6 | 0.48±0.14 | 0.69±0.11 | 0.80±0.03 | 0.88±0.04 | 0.91±0.03 | 0.94±0.02 | 0.96±0.02 | 0.98±0.01 |
| 7 | 0.48±0.16 | 0.69±0.11 | 0.82±0.05 | 0.89±0.04 | 0.92±0.03 | 0.95±0.02 | 0.96±0.01 | 0.98±0.01 |
| 8 | 0.47±0.14 | 0.69±0.10 | 0.82±0.03 | 0.89±0.04 | 0.92±0.02 | 0.95±0.02 | 0.96±0.01 | 0.98±0.01 |
| Proportion of variance explained with selected Gabor waveforms | | | | | | | | |
| 1 | 0.31±0.11 | 0.43±0.11 | 0.55±0.05 | 0.59±0.06 | 0.62±0.06 | 0.67±0.05 | 0.70±0.05 | 0.73±0.05 |
| 2 | 0.35±0.11 | 0.51±0.10 | 0.62±0.08 | 0.67±0.06 | 0.72±0.06 | 0.77±0.04 | 0.81±0.03 | 0.87±0.02 |
| 3 | 0.36±0.10 | 0.54±0.11 | 0.64±0.07 | 0.73±0.06 | 0.76±0.06 | 0.82±0.04 | 0.86±0.02 | 0.92±0.02 |
| 4 | 0.38±0.10 | 0.55±0.10 | 0.68±0.10 | 0.75±0.07 | 0.79±0.05 | 0.85±0.04 | 0.89±0.03 | 0.93±0.02 |
| 5 | 0.38±0.10 | 0.56±0.10 | 0.68±0.09 | 0.76±0.08 | 0.80±0.06 | 0.88±0.04 | 0.91±0.02 | 0.95±0.01 |
| 6 | 0.39±0.10 | 0.57±0.10 | 0.71±0.08 | 0.79±0.07 | 0.83±0.06 | 0.88±0.05 | 0.91±0.03 | 0.96±0.01 |
| 7 | 0.39±0.09 | 0.57±0.10 | 0.71±0.08 | 0.79±0.07 | 0.84±0.06 | 0.89±0.05 | 0.92±0.03 | 0.96±0.02 |
| 8 | 0.40±0.09 | 0.58±0.09 | 0.72±0.08 | 0.79±0.06 | 0.84±0.05 | 0.90±0.04 | 0.92±0.03 | 0.96±0.01 |
| Difference of squared error: Gabor minus MP-SVD | | | | | | | | |
| 1 | 0.11±0.06 | 0.14±0.03 | 0.13±0.07 | 0.13±0.07 | 0.13±0.07 | 0.13±0.05 | 0.13±0.04 | 0.08±0.04 |
| 2 | 0.13±0.07 | 0.13±0.05 | 0.14±0.03 | 0.15±0.06 | 0.13±0.09 | 0.12±0.06 | 0.10±0.06 | 0.07±0.01 |
| 3 | 0.11±0.07 | 0.11±0.04 | 0.14±0.04 | 0.12±0.04 | 0.12±0.05 | 0.09±0.05 | 0.07±0.03 | 0.04±0.01 |
| 4 | 0.09±0.06 | 0.13±0.04 | 0.12±0.04 | 0.12±0.04 | 0.11±0.03 | 0.08±0.02 | 0.06±0.02 | 0.04±0.01 |
| 5 | 0.08±0.07 | 0.13±0.04 | 0.12±0.05 | 0.11±0.05 | 0.10±0.03 | 0.06±0.02 | 0.05±0.01 | 0.02±0.01 |
| 6 | 0.09±0.07 | 0.12±0.04 | 0.10±0.06 | 0.09±0.04 | 0.08±0.03 | 0.06±0.03 | 0.04±0.02 | 0.02±0.01 |
| 7 | 0.09±0.08 | 0.12±0.04 | 0.11±0.04 | 0.10±0.04 | 0.08±0.03 | 0.06±0.03 | 0.04±0.02 | 0.02±0.01 |
| 8 | 0.07±0.07 | 0.11±0.03 | 0.10±0.06 | 0.10±0.03 | 0.08±0.03 | 0.05±0.02 | 0.04±0.02 | 0.01±0.01 |
| Error ratio (log 2): a value of 1 indicates Gabor has twice the squared error of MP-SVD | | | | | | | | |
| 1 | 0.28±0.26 | 0.44±0.21 | 0.56±0.45 | 0.63±0.46 | 0.71±0.54 | 0.80±0.54 | 0.86±0.35 | 0.61±0.47 |
| 2 | 0.38±0.37 | 0.51±0.36 | 0.73±0.33 | 0.95±0.54 | 1.02±0.77 | 1.22±0.78 | 1.29±0.86 | 1.21±0.40 |
| 3 | 0.33±0.30 | 0.47±0.35 | 0.84±0.43 | 0.98±0.55 | 1.13±0.57 | 1.12±0.74 | 1.15±0.65 | 1.08±0.30 |
| 4 | 0.27±0.27 | 0.55±0.37 | 0.69±0.17 | 0.94±0.34 | 1.13±0.45 | 1.19±0.36 | 1.03±0.31 | 1.17±0.28 |
| 5 | 0.25±0.26 | 0.57±0.28 | 0.77±0.31 | 0.89±0.24 | 1.10±0.25 | 1.04±0.15 | 1.07±0.35 | 0.91±0.10 |
| 6 | 0.28±0.28 | 0.53±0.35 | 0.54±0.29 | 0.88±0.30 | 0.96±0.17 | 1.06±0.19 | 1.00±0.12 | 0.78±0.15 |
| 7 | 0.29±0.34 | 0.53±0.33 | 0.71±0.20 | 0.97±0.28 | 0.98±0.24 | 1.05±0.19 | 0.95±0.20 | 0.72±0.27 |
| 8 | 0.22±0.26 | 0.47±0.23 | 0.61±0.25 | 0.92±0.23 | 0.94±0.16 | 0.96±0.21 | 0.92±0.26 | 0.73±0.19 |
| Win count (n=5) for MP-SVD: p-value for 5 is 0.0312, 4 wins is insignificant at a level of 0.05. | | | | | | | | |
| 1 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 2 | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 5 |
| 3 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 5 |
| 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 7 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 8 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

## S3. Comparison of Results for Case Study 2

In this section, we include additional results for waveforms estimated using MP-SVD along with a set of results for waveforms estimated using single-channel ICA. The case study uses the motor imagery dataset: BCI competition III dataset IV(a), Blankertz *et al., IEEE Trans. Neural Syst. Rehabil. Eng.,* 2006, provided by Fraunhofer FIRST, Intelligent Data Analysis Group (Klaus-Robert Müller, Benjamin Blankertz), and Campus Benjamin Franklin of the Charité–University Medicine Berlin, Department of Neurology, Neurophysics Group (Gabriel Curio).

Fig. S8 shows a comparison of the center frequency and Q-value of waveforms estimated by MP-SVD and single-channel ICA. Fig. S9 and Fig. S10 show cluster descriptors (prototype, scatter plot of center frequency and Q-value) for the clusters found by applying shift-invariant k-means to the waveforms learned across all subjects and channels. Fig. S11 and Fig. S12 show the topographic distribution of originating channels of the waveforms assigned to each cluster and the spectrum. Fig. S13 and Fig. S14 show the spatial weight pattern corresponding to a significant linear discriminate analysis between the spatial amplitudes during two classes of motor imagery.

Finally, Fig. S15 and Fig. S16 show the components of the decompositions, and the total approximation, using the waveforms on a specific channel and subject using different numbers of waveforms.

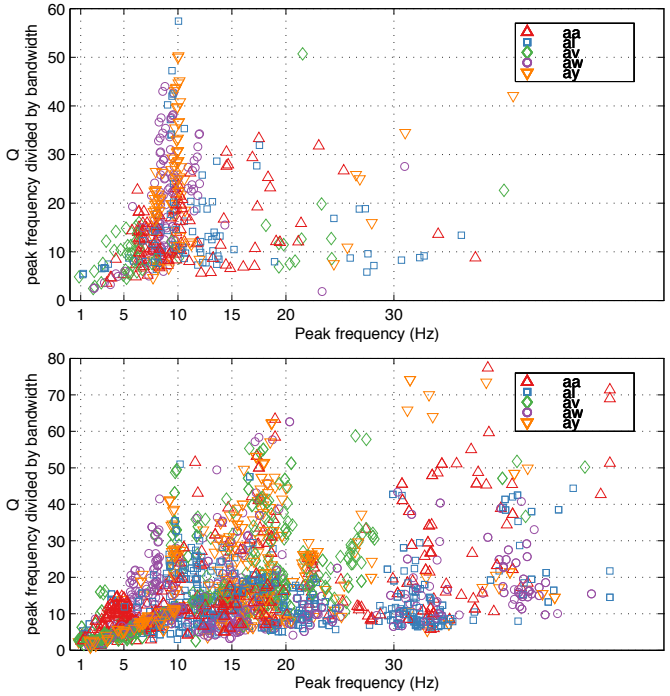



Fig. S8. Case study 2: Scatter plot of peak frequency and Q-value (peak frequency divided by bandwidth) of waveforms that matched Gabor-Morlet wavelets. (Top) Waveforms estimated using matching pursuit with SVD update. (Bottom) Waveforms estimated using single-channel ICA.

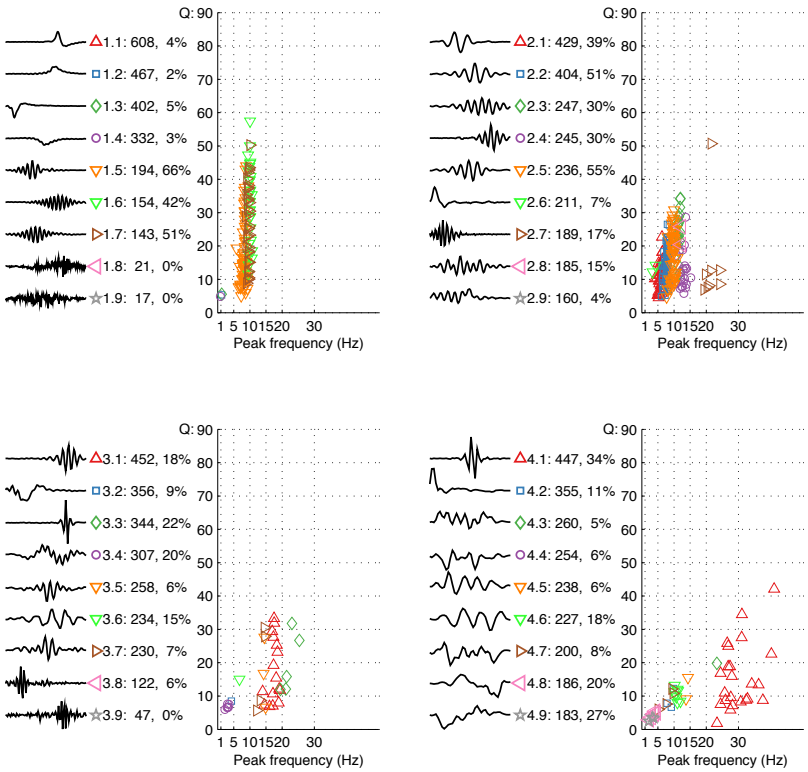



Fig. S9. Case study 2 (MP-SVD): Cluster centroids and scatter plot of peak frequency and Q-value for those that matched Gabor-Morlet wavelets. The legend on the left side of each plot lists unique cluster index, the number of waveforms, and the percentage that match Gabor-Morlet wavelets. Most of the matching waveforms correspond to alpha waves estimated in the first scale.

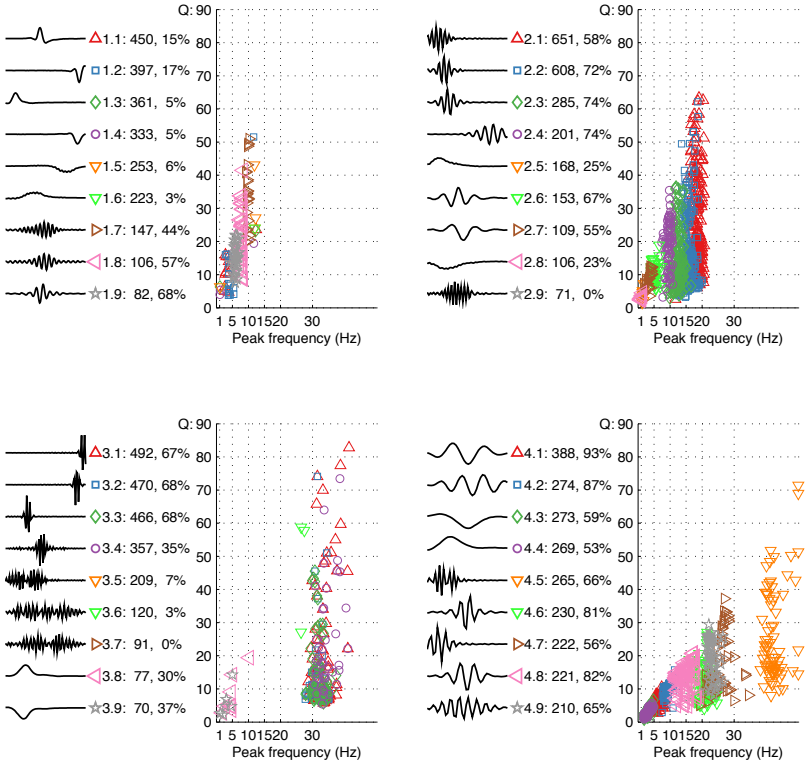



Fig. S10. Case study 2 (SC-ICA): Cluster centroids and scatter plot of peak frequency and Q-value for those that matched Gabor-Morlet wavelets. The legend on the left side of each plot lists unique cluster index, the number of waveforms, and the percentage that match Gabor-Morlet wavelets. Contrary to MP-SVD, many of waveforms match Gabor wavelets across all scales.

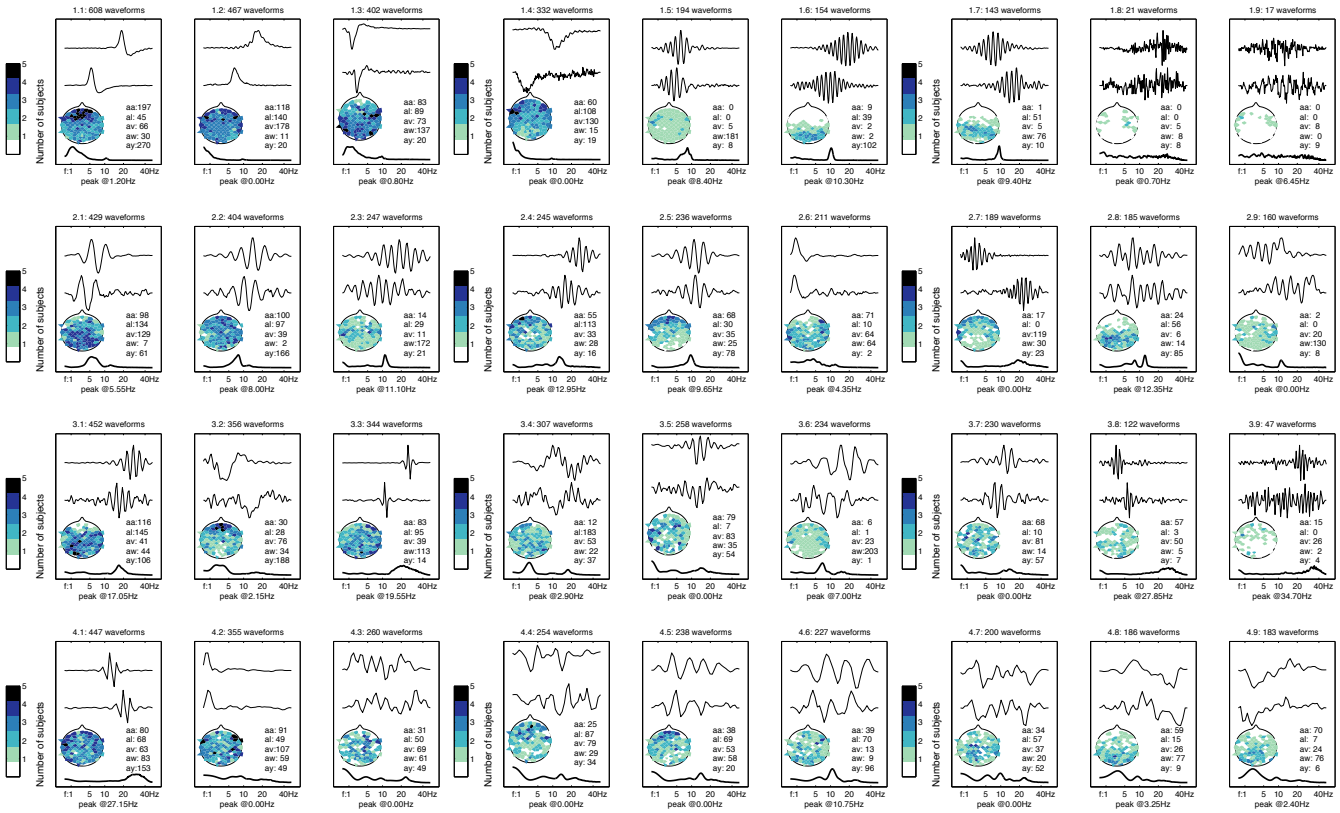



Fig. S11. Case study 2 (MP-SVD): Cluster descriptors for the waveforms estimated using matching pursuit with SVD-update across all scales. Each subplot shows (from top to bottom) the cluster mean, prototypical waveform, electrode distribution shown on an unfolded scalp map where the color intensity indicates the number of subjects with waveforms originating from that electrode, and the power spectral density over all waveforms in the cluster.

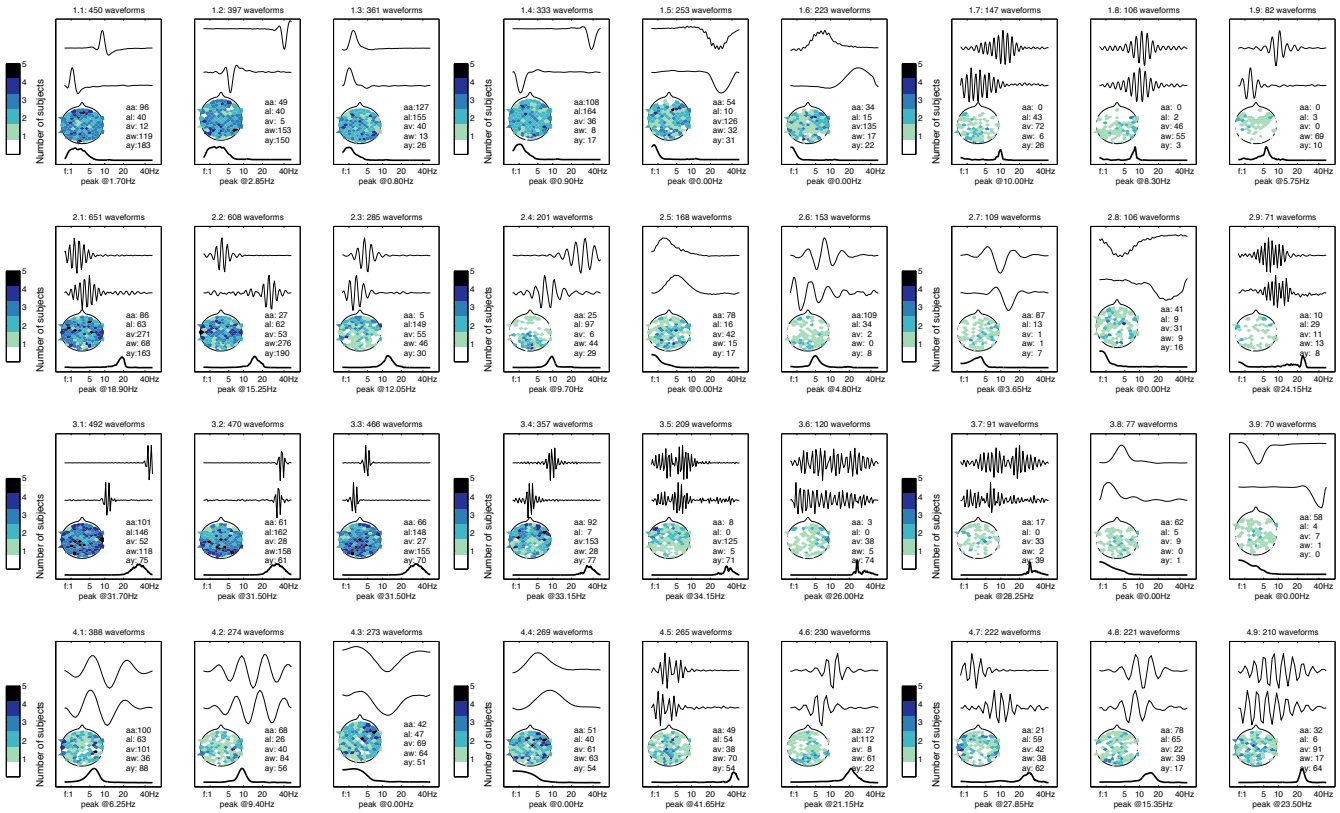



Fig. S12. Case study 2 (SC-ICA): Cluster descriptors for the waveforms estimated using single-channel ICA across all scales. Each subplot shows (from top to bottom) the cluster mean, prototypical waveform, electrode distribution shown on an unfolded scalp map where the color intensity indicates the number of subjects with waveforms originating from that electrode, and the power spectral density over all waveforms in the cluster.

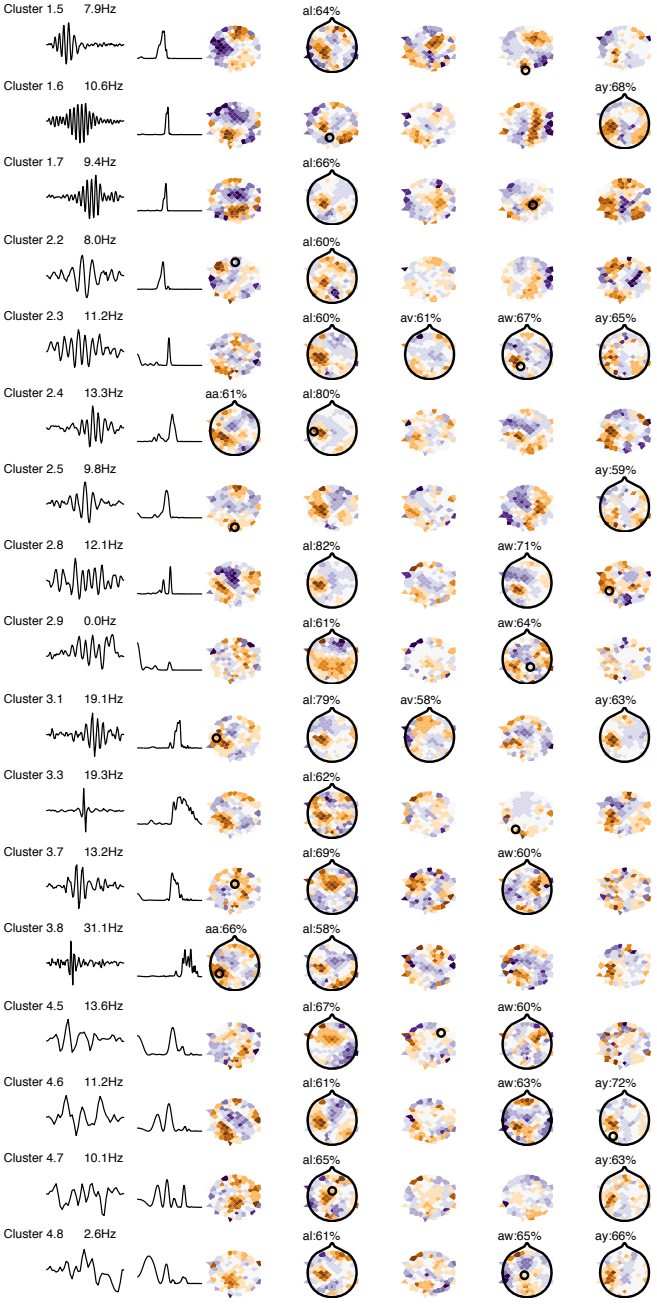



Fig. S13. Case study 2 (MP-SVD): Prototype waveforms and spatial weight pattern corresponding to a significant linear discriminate analysis between the spatial amplitudes during two classes of motor imagery. Waveforms estimated using matching pursuit with SVD update. The originating subject and electrode is circled. The color intensity is in arbitrary units. For a significant waveform/subject pair, the percent accuracy over the testing set is listed. The training/testing split was 70/210.

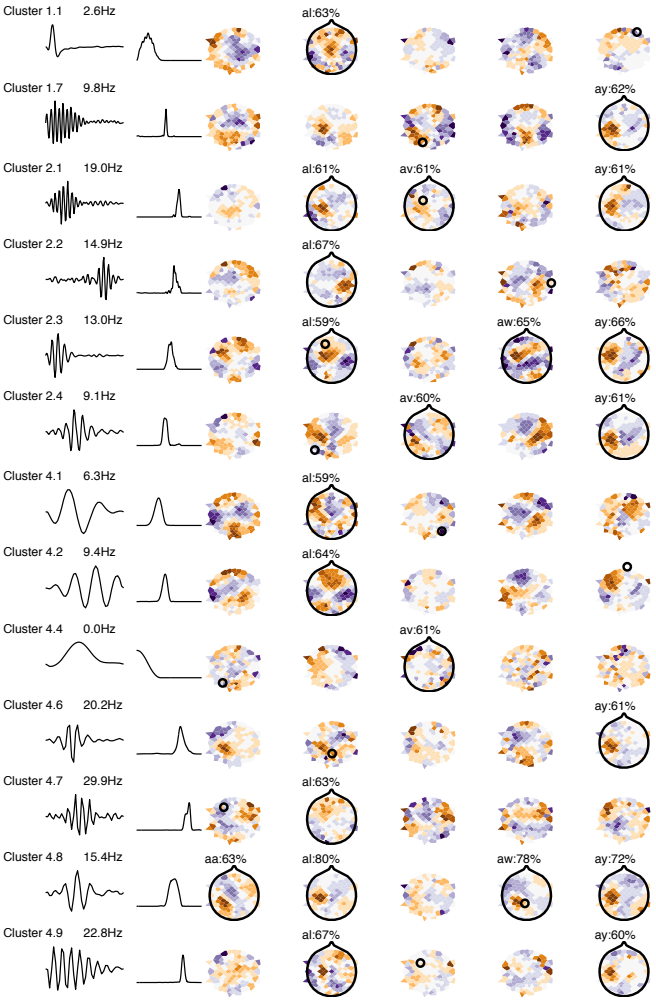



Fig. S14. Case study 2 (SC-ICA): Prototype waveforms and spatial weight pattern corresponding to a significant linear discriminate analysis between the spatial amplitudes during two classes of motor imagery. Waveforms estimated using single-channel ICA. The originating subject and electrode is circled. The color intensity is in arbitrary units. For a significant waveform/subject pair, the percent accuracy over the testing set is listed. The training/testing split was 70/210.

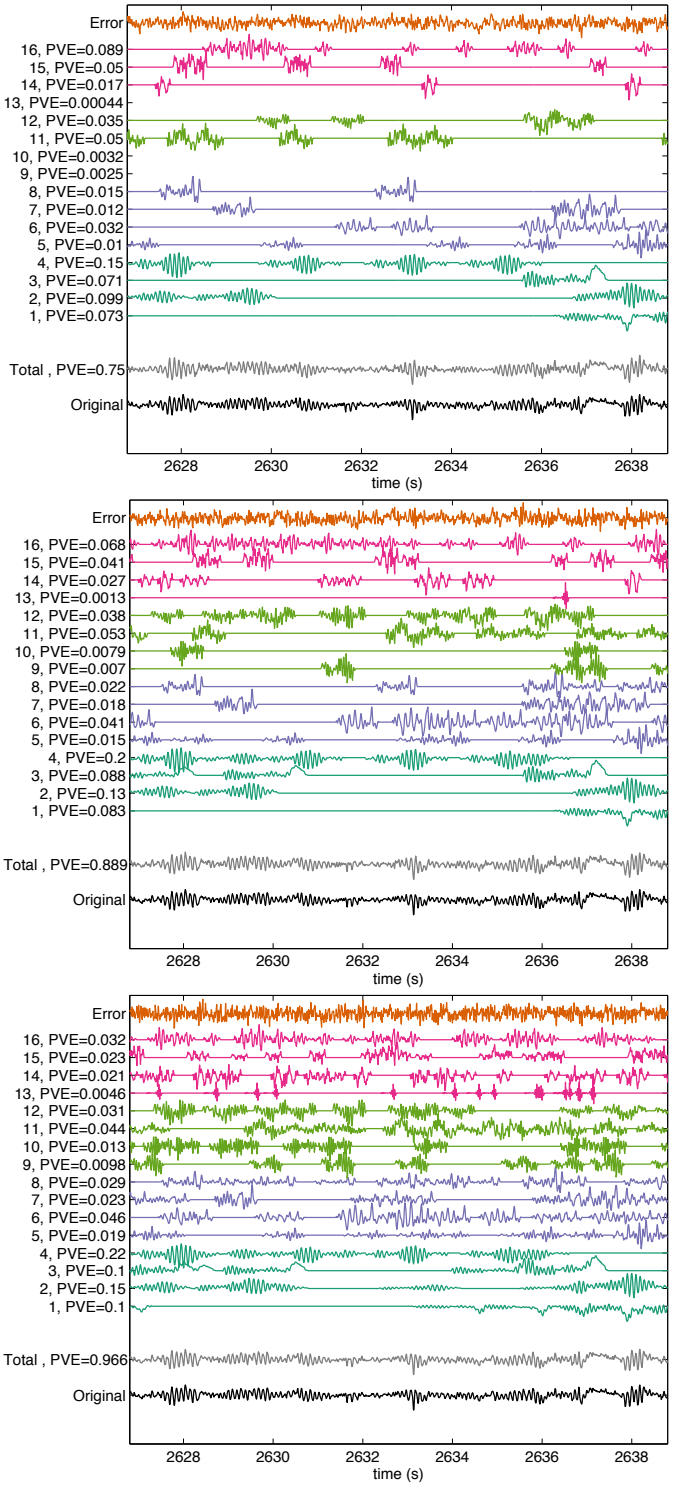



Fig. S15. Case study 2 (MP-SVD): Example decomposition using waveforms estimated with matching pursuit with SVD update and different number of atoms. Proportion of variance explained (PVE) is reported for each component corresponding to a single waveform and for the total approximation. Waveforms and decomposition come from subject 'al' channel O1 in the last quarter of record.

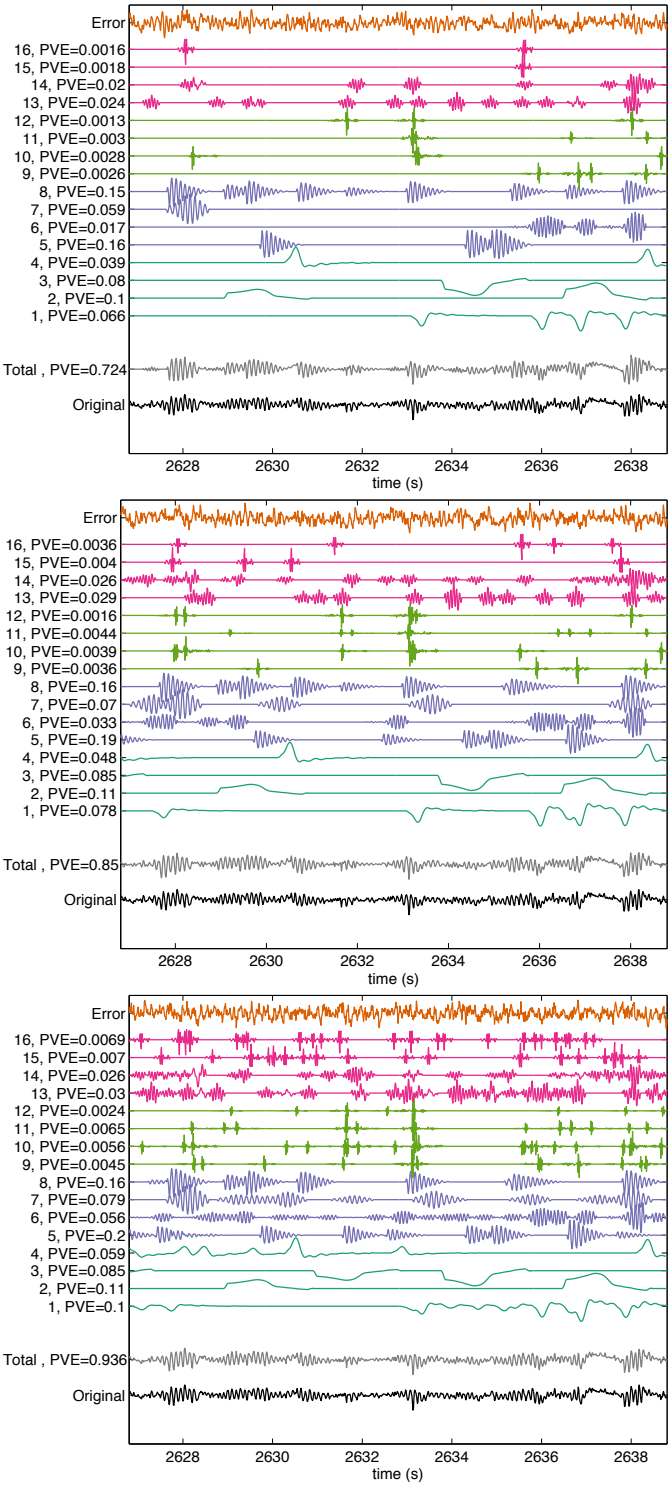



Fig. S16. Case study 2 (SC-ICA): Example decomposition using waveforms estimated with single-channel ICA and different number of atoms (number of atoms per plot matches corresponding plot in Fig. S15). Proportion of variance explained (PVE) is reported for each component corresponding to a single waveform and for the total approximation. Waveforms and decomposition come from subject 'al' channel O1 in the last quarter of record.

## S4. Synthetic Experiments

Two sets of synthetic data experiments were conducted to illustrate the methodology. The first set is used to compare shift-invariant learning with single-channel ICA with and without using a 4-stage modeling procedure. We compared the waveform estimation quality when the true waveforms are a non-redundant (cross-correlation between any waveforms is less than 0.8) set of 16 Gabor-Morlet wavelets or a set of 16 of the Daubechies 4 (db4) wavelet packets. The length of the waveforms was 100 time points. The majority of this simulation matches a previous synthetic simulation we performed Brockmeier and Principe, *IEEE Int. Conf. Acoust. Speech Signal Process.,* 2015.

For each run, an output signal with 10,000 samples was created by feeding sparse source signals through the MISO system. The source signals are independent, marked point processes with a homogeneous Poisson point process for the timing and a Gaussian distribution with mean and standard-deviation of $(1, \frac{1}{3})$ for amplitudes. In the experiments, the rate of the Poisson process controls the sparsity of the sources. The overall rate was varied between 5% and 100%; these rates correspond to average rates of 0.3% and 6.25% for the individual waveforms. White noise with two different standard-deviations $(\frac{1}{2}, 1)$ was used to test the estimation in both a low noise and a moderate noise situation.

The algorithms were run with the correct number of waveforms, but they were not given any information on the sparsity. Matching pursuit assumes a fixed cardinality of the sources. For the single-channel ICA, FastICA with 40-unit symmetric estimation and the $\tanh(\cdot)$ activation function was used. In practice, most of the waveforms in the 40-unit estimation are meaningless so the orthogonal matching pursuit-based subset selection is used to select the predefined number of waveforms.

For each run, the waveform estimation performance is quantified as the cross-correlation between the true waveform and the best matching estimated waveform. This quantity is averaged across the 16 waveforms and the results are collected across 5 Monte Carlo generations of source and noise activity. The mean and standard deviation across the Monte Carlo runs are displayed as error bars in Fig. S17 and Fig. S18.

For each run, the waveform estimation performance is also quantified in terms of recall, that is, the fraction of true waveforms that have cross correlation $\geq$0.8 to an estimated waveform. This quantity is averaged across the 16 waveforms and the results are collected across 5 Monte Carlo generations of source and noise activity. The mean and standard deviation across the Monte Carlo runs are displayed as error bars in Fig. S19 and Fig. S20.

The second set of experiments is used to illustrate the decomposition ability of the methodology, specifically, its ability to estimate the true components of a known signal. The underlying waveforms are two distinct Gabor-Morlet wavelets and two triangular waves, one is negative and one is positive; all of the waveforms are 100 time-points long. The sources are constant amplitude trains of delta functions. The Bernoulli probability that any source is active at a given time point ranges from 0.5% to 4%. For example, the expected number of waveforms active at any point is 1 for a 1% source. The overall signal has zero mean, and is meant to be an extremely rough representative of EEG waveforms, namely time-limited rhythms and evoked potentials. The signal is observed in the presence of Gaussian white noise, for which we vary the standard deviation.

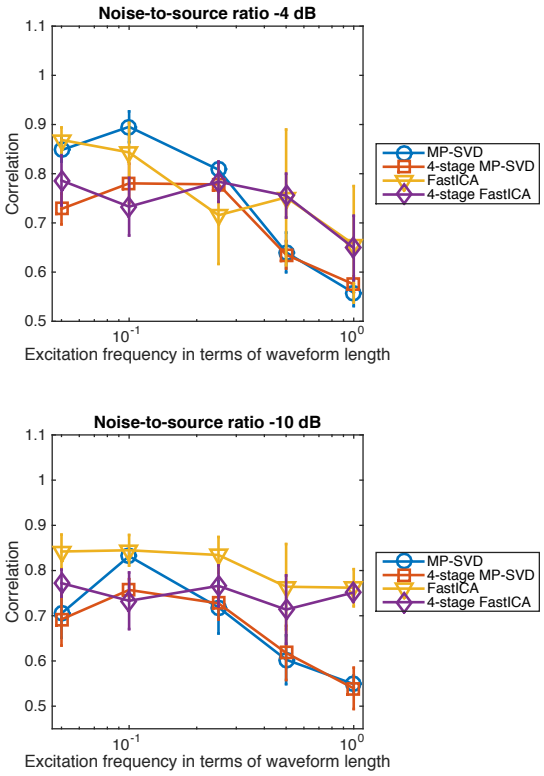



Fig. S17. Synthetic experiment with Gabor-Morlet wavelets: Average cross-correlation of the estimated waveforms to their best matched true waveforms when the estimation is done using matching pursuit with SVD update (MP-SVD) versus single-channel ICA using the FastICA algorithm, and with the case that all 16 waveforms are estimated at once versus a 4-stage estimation.

After learning, we compare the ability of different modeling and decomposition techniques to extract the noise-free signal and its components. For the learned models we use the 3-std stopping heuristic described in Section S1. The Gabor dictionary is restricted to use the same number of atoms as the Gabor subset dictionary used. In addition, we report the proportion of variance explained of the estimated components using the true components; a low value indicates a true component is split between the estimated components, as would happen when many Gabor wavelets are used to approximate a non-Gabor waveform. The results are reported in Table S3. In terms of approximation of the signal the full Gabor dictionary is consistently the best. In terms of estimating the true components, MP-SVD's median performance is the best for the lowest rate case, and the subset selection algorithm performs the best on average. Besides the lowest rate case, the maximum performance of SC-ICA is the highest, indicating that if multiple replications were used SC-ICA may be a more competitive. Interestingly, MP-SVD has the highest rate of proportion of variance explained for the estimated components, which indicates that each of the

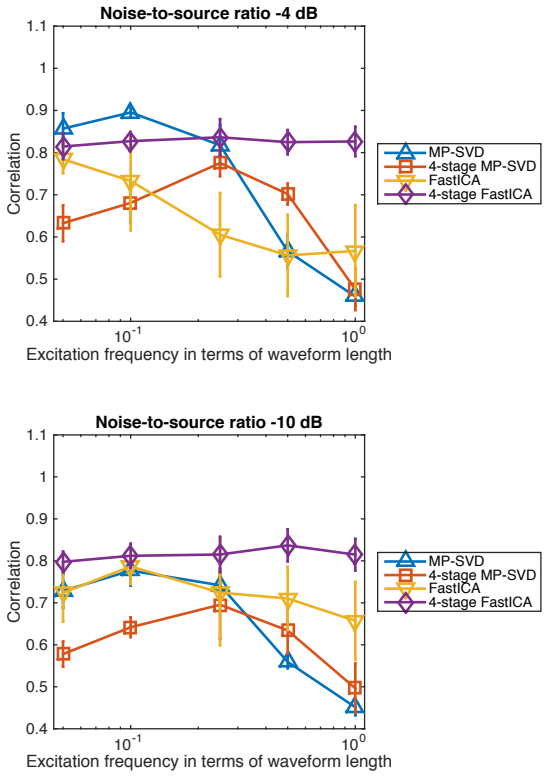



Fig. S18. Synthetic experiment with Daubechies-4 (db4) wavelet packets: Average cross-correlation of the estimated waveforms to their best matched true waveforms when the estimation is done using matching pursuit with SVD update (MP-SVD) versus single-channel ICA using the FastICA algorithm, and with the case that all 16 waveforms are estimated at once versus a 4-stage estimation.

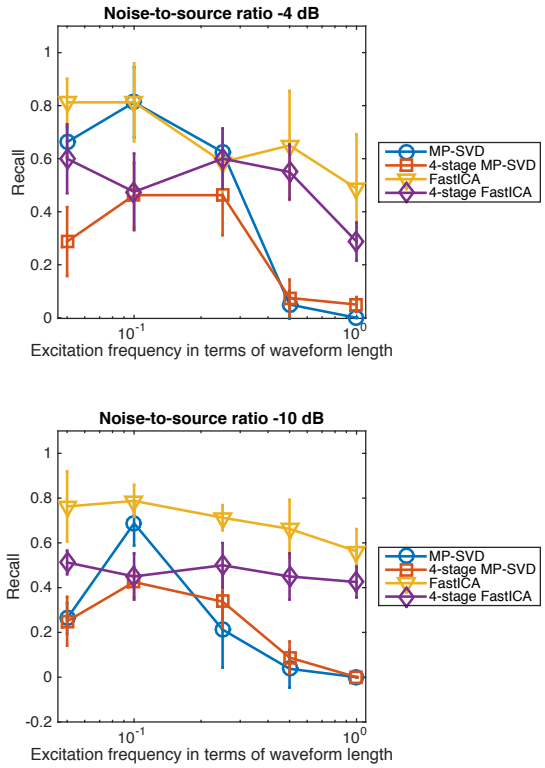



Fig. S19. Synthetic experiment with Gabor-Morlet wavelets: Average recall (proportion of true waveforms that have cross correlation $\geq$0.8 to an estimated waveform).

estimated components is often one of the true components or a mixture of them. Overall, the results indicate that sparsity has a large effect on MP-SVD and SC-ICA. The subset selection algorithm is able to perform very well in terms of estimating the true components. Its performance is similar or better than the dictionary learning algorithms for higher rate sources and is much better than the full dictionary on this simple signal. These results show how the proposed methodology can find the underlying components of a signal. Examples of the decompositions for the various methods are shown in Fig. S21.

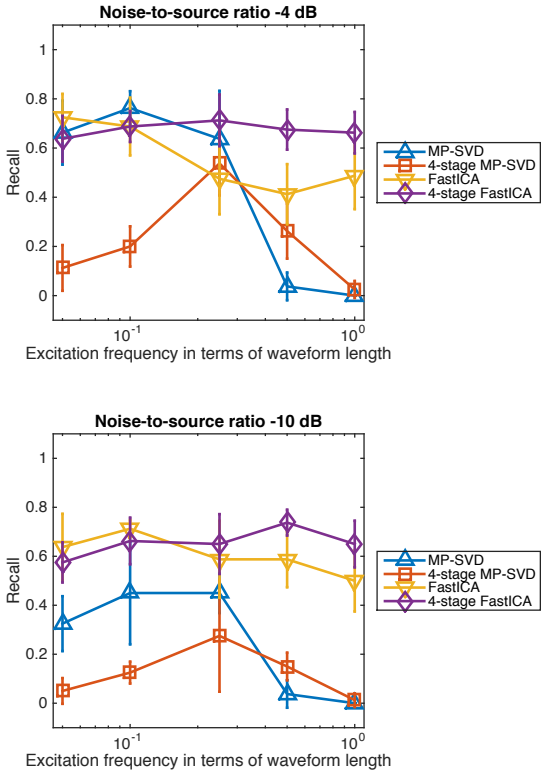



Fig. S20. Synthetic experiment with Daubechies-4 (db4) wavelet packets: Average recall (proportion of true waveforms that have cross correlation $\geq$0.8 to an estimated waveform).

TABLE S3

SYNTHETIC SIGNAL DECOMPOSITION PERFORMANCE IN TERMS OF PERCENTAGE (%) OF THE VARIANCE EXPLAINED: NOISE-FREE SIGNAL BY THE APPROXIMATION, TRUE COMPONENTS BY THE ESTIMATED COMPONENTS, AND THE ESTIMATED COMPONENTS BY THE TRUE COMPONENTS. THE MAXIMUM, MEDIAN, AND STANDARD DEVIATION ACROSS 20 MONTE CARLO RUNS IS REPORTED FOR DIFFERENT SOURCE RATES AND NOISE LEVELS.

| | Signal | | | True comp. | | | Estimated comp. | | | Signal | | | True comp. | | | Estimated comp. | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | max | med. | std | max | med. | std | max | med. | std | max | med. | std | max | med. | std | max | med. | std |
| | Source rate of 0.5%, noise std of 0.1 | | | | | | | | | Source rate of 1%, noise std of 0.1 | | | | | | | | |
| MP-SVD | 99 | 93 | 6 | 96 | 69 | 14 | 96 | 69 | 15 | 92 | 91 | 2 | 68 | 63 | 12 | 68 | 59 | 10 |
| SC-ICA | 96 | 83 | 18 | 87 | 46 | 18 | 85 | 49 | 22 | 95 | 76 | 12 | 71 | 28 | 19 | 76 | 30 | 23 |
| Gabor subset | 94 | 90 | 2 | 84 | 65 | 6 | 81 | 64 | 6 | 90 | 87 | 2 | 66 | 63 | 1 | 65 | 56 | 5 |
| Gabor | 98 | 98 | 1 | 55 | 52 | 3 | 5 | 3 | 1 | 98 | 93 | 2 | 54 | 49 | 2 | 4 | 1 | 1 |
| | Source rate of 2%, noise std of 0.1 | | | | | | | | | Source rate of 4%, noise std of 0.1 | | | | | | | | |
| MP-SVD | 89 | 85 | 19 | 55 | 36 | 16 | 59 | 25 | 15 | 85 | 82 | 19 | 35 | 25 | 7 | 48 | 7 | 14 |
| SC-ICA | 95 | 78 | 9 | 71 | 19 | 18 | 71 | 3 | 20 | 91 | 82 | 6 | 27 | 20 | 6 | 20 | 3 | 6 |
| Gabor subset | 88 | 85 | 2 | 63 | 61 | 2 | 53 | 45 | 4 | 86 | 83 | 1 | 59 | 58 | 1 | 42 | 35 | 3 |
| Gabor | 96 | 92 | 1 | 49 | 45 | 3 | 1 | 1 | 0 | 95 | 91 | 1 | 40 | 35 | 2 | 1 | 0 | 0 |
| | Source rate of 0.5%, noise std of 0.2 | | | | | | | | | Source rate of 1%, noise std of 0.2 | | | | | | | | |
| MP-SVD | 96 | 91 | 3 | 94 | 69 | 10 | 94 | 67 | 10 | 90 | 88 | 19 | 67 | 61 | 18 | 67 | 56 | 14 |
| SC-ICA | 91 | 70 | 26 | 68 | 52 | 17 | 63 | 46 | 17 | 93 | 72 | 19 | 80 | 49 | 14 | 78 | 38 | 21 |
| Gabor subset | 91 | 89 | 2 | 82 | 66 | 6 | 79 | 64 | 6 | 88 | 86 | 2 | 65 | 63 | 1 | 63 | 57 | 5 |
| Gabor | 95 | 94 | 1 | 59 | 48 | 4 | 5 | 3 | 1 | 95 | 90 | 2 | 51 | 47 | 2 | 5 | 1 | 1 |
| | Source rate of 2%, noise std of 0.2 | | | | | | | | | Source rate of 4%, noise std of 0.2 | | | | | | | | |
| MP-SVD | 86 | 83 | 2 | 54 | 27 | 10 | 60 | 20 | 16 | 84 | 82 | 11 | 37 | 27 | 7 | 45 | 1 | 14 |
| SC-ICA | 92 | 68 | 16 | 78 | 33 | 16 | 79 | 17 | 20 | 86 | 75 | 7 | 31 | 17 | 7 | 19 | 0 | 4 |
| Gabor subset | 86 | 84 | 1 | 62 | 60 | 1 | 53 | 44 | 4 | 84 | 82 | 2 | 60 | 56 | 1 | 40 | 33 | 2 |
| Gabor | 95 | 90 | 2 | 44 | 40 | 2 | 1 | 1 | 0 | 94 | 90 | 1 | 39 | 33 | 2 | 1 | 0 | 0 |

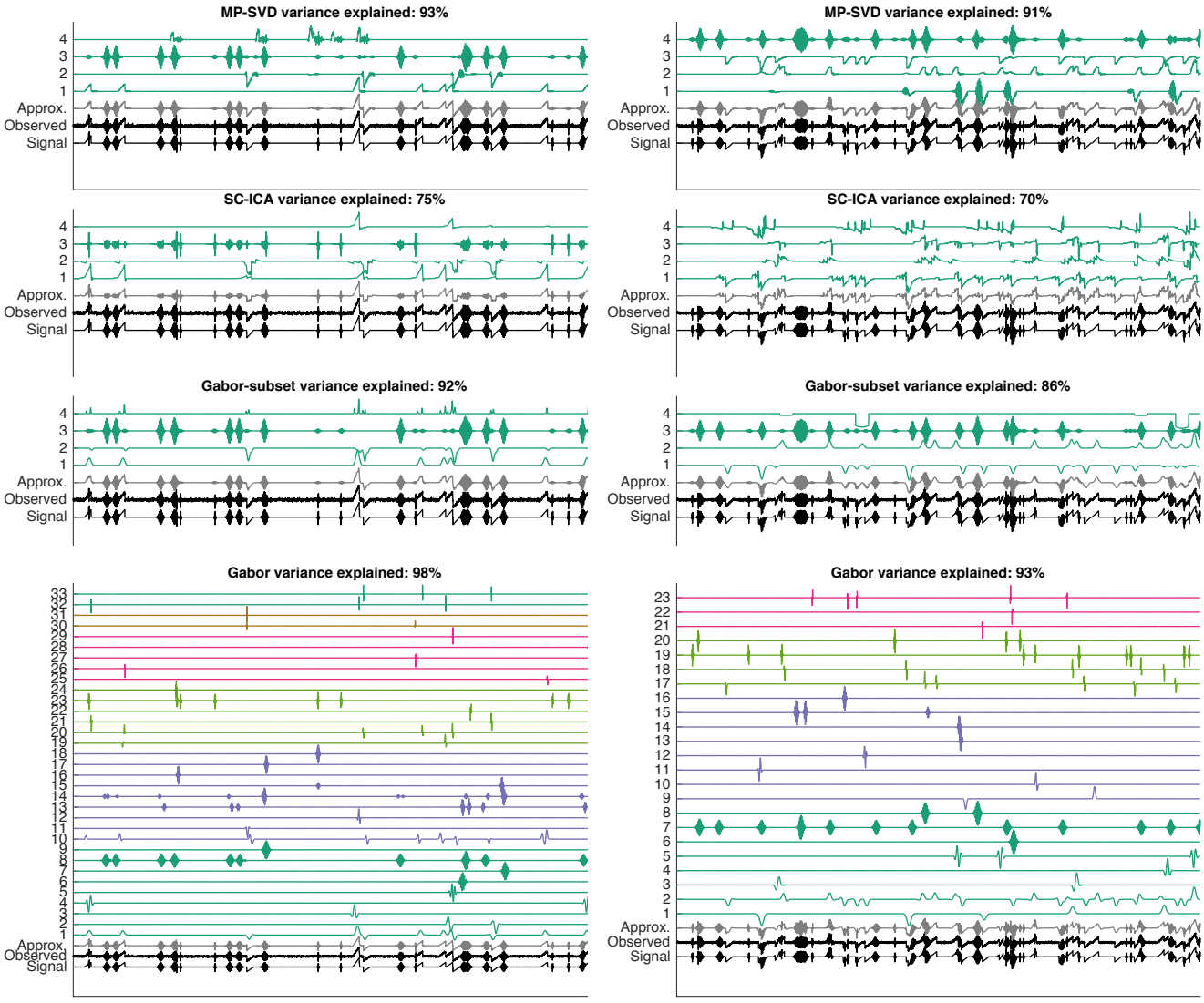



Fig. S21. Example decompositions of a synthetic signal using different shift-invariant dictionaries. Panels from top to bottom correspond to using waveforms learned by MP-SVD, learned by single-channel ICA with subset selection, selected from a Gabor-Morlet dictionaries, and from a full Gabor dictionary. (Left) Signals with an overall source rate with Bernoulli probability of 0.5% (Right) Source rate of 1%. The additive white Gaussian noise has standard deviation of 0.2, compared to the unit-amplitude of the waveforms. Number of atoms selected by the 3-std stopping heuristic.