Computational and Structural Biotechnology Journal 27 (2025) 4943–4951

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

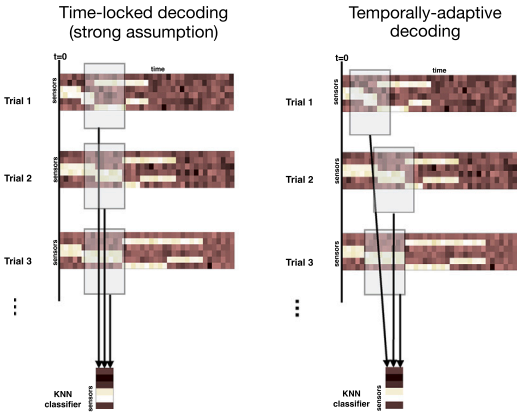
Research article

ADA: A decoding algorithm for temporally-variable brain responses

Pablo Oyarzo[a,b], Radoslaw M. Cichy[a], Diego Vidaurre[b,c,d,*]

[a] Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany
[b] Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[c] Centre de Recerca Matematica, Barcelona, Spain
[d] Department of Psychiatry, Oxford University, Oxford, UK

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Brain decoding
Cognitive neuroscience
Temporal variability
MEG
Machine learning

ABSTRACT

Decoding mental contents from brain activity is a long-standing goal in theoretical neuroscience and neural engineering. While current methods perform well in tasks with externally timed events, such as perception or motor execution, decoding covert cognitive processes like imagery or memory recall remains challenging due to uncertainty in the timing of underlying neural dynamics. In these settings, neurophysiological responses are not reliably linked to observable behaviour and likely vary in latency across trials. This complicates the use of time-locked analysis techniques, which perform decoding time point by time point across trials, thus assuming consistent signal timing. This problem corresponds to an understudied class of supervised learning where input features may be effectively mislabelled and need to be aligned across cases. To address this, we present the Adaptive Decoding Algorithm (ADA), a nonparametric method based on a two-level prediction. First, we estimate, for each trial, the temporal window most likely to reflect task-relevant signals; second, we decode the test trials based on the selection of informative windows. Using controlled simulations as well as a model of memory recall based on real perception data, we show that ADA outperforms alternative methods that assume fixed temporal structure. These results provide evidence that explicitly accounting for trial-specific timing can substantially improve decoding performance when the timing of relevant neural activity is unknown.

* Corresponding author at: Department of Clinical Medicine, Aarhus University, Aarhus, Denmark.
Email address: dvidaurre@cfin.au.dk (D. Vidaurre).

https://doi.org/10.1016/j.csbj.2025.10.044
Received 7 June 2025; Received in revised form 21 October 2025; Accepted 21 October 2025
Available online 3 November 2025

1. Introduction

Decoding internal cognitive states from patterns of brain activity is a central challenge in neuroscience. In its standard formulation, this task is treated as a supervised learning problem: given repeated recordings of brain activity during stimulus presentation or behavioural execution, a model is trained to predict the corresponding condition or label [1,2]. Use cases span basic neuroscience, where decoding is used to investigate the timing and localization of cognitive processes [3] as well as clinical applications [4] and brain-computer interfaces [5,6].

Our capacity to perform decoding of brain signals depends at least on two factors. One is the *spatial specificity* of task-relevant brain activity and how well it can be accessed by the chosen recording modality. For instance, electrophysiological data such as electroencephalography (EEG) and magnetoencephalography (MEG) provide good sensitivity in motor and occipital cortices, facilitating decoding in movement and visual tasks [7], but the signal-to-noise ratio is lower in other areas. The other, which is the focus of this paper, is *temporal specificity* - i.e. whether the timing of the relevant neural process is known to the experimentalist, or whether it can be reasonably assumed to be tightly locked to an observed behaviour or external cue. This would be the case in a standard visual experiment where subjects are passively shown different types of images (one per trial), such that brain responses are time-locked to stimulus presentation. This allows trials to be aligned for decoding. Unfortunately, this assumption breaks down in covert processes such as imagery, planning, or memory recall, where neural responses may occur at different latencies across trials [8,9]. Such variability may reflect differences in internal decision timing, fluctuations in attention, or other volitional factors. This poses a challenge for standard decoding approaches, which require a prespecified alignment of the trials, and rely on time-locking or averaging to improve signal-to-noise ratio [10,11]. While some previous studies have attempted to accommodate temporal variability by using strategies such as time-generalization matrices [12], dynamic state modelling [13], or spatiotemporal pattern analysis [14], these methods often rely on fixed heuristics, post-hoc adjustments, or are not designed as general-purpose decoding tools. A relevant effort in this direction is Temporally Unconstrained Decoding Analysis (TUDA) [13], which embeds decoding within a Hidden Markov Model (HMM) to capture trial-wise variability in the mapping between neural activity and task variables. Each state constitutes a distinct decoder whose dynamics trace shifting correspondence between activity and labels. Unlike ADA, TUDA is interpretive rather than predictive: it identifies when decoders are active but cannot straightforwardly generalize to unseen trials, as state estimation requires known labels. The need remains for approaches that directly model trial-specific variability in a principled and predictive framework.

To address this issue, we propose a method, Adaptive Decoding Algorithm or ADA, that does not assume temporal alignment across trials, but instead estimates when task-relevant activity is likely to occur within each trial and uses this information to guide the decoding. In what follows, we formalize the problem and the algorithm, and evaluate its performance in both simulated data and a model of memory recall based on real perception data.

Contributions. Our contributions are threefold. First, we formalize the decoding problem under temporal uncertainty, highlighting how trial-specific variability in response latency leads to feature misalignment. Second, we introduce ADA, which identifies informative time windows during training and uses them to guide test-time decoding without assuming temporal alignment. Third, we evaluate ADA in both controlled simulations and a biologically motivated memory recall scenario, demonstrating that accounting for timing variability improves decoding performance in settings where standard methods fail.

2. Material and methods

2.1. Problem formulation: decoding under temporal uncertainty

Let $\mathbf{X} \in \mathbb{R}^{T\times p\times N}$ denote the recorded brain activity from a single subject across N trials with T time samples and p channels per trial; $\mathbf{Y} \in \{-1,+1\}^N$ represents the corresponding stimulus or behavioural labels; and $\mathcal{Z}$ denotes the latent neural process that generates task-relevant activity in $\mathbf{X}$ and depends on $\mathbf{Y}$. Thus, each trial $\mathbf{x}_n \in \mathbb{R}^{T\times p}$ is a multichannel time series, and each label $y_n \in \{-1,+1\}$. Although we focus on binary classification for clarity, the approach can straightforwardly be extended to multiclass and regression tasks.

The key difficulty lies in the temporal variability of $\mathcal{Z}$: while it may reliably encode information about $\mathbf{Y}$, its timing can vary across trials. The observed signals $\mathbf{X}$ reflect a mixture of this process, other ongoing oscillatory brain activity at multiple frequencies [15], and measurement noise. Critically, while we assume that $\mathbf{X}$ contains information about $\mathcal{Z}$, it does not necessarily do so during the entire recording or at the same time for all trials.

To formalize this, we introduce an unobserved binary matrix $\mathbf{I} \in \{0,1\}^{T\times N}$, where $I_{tn} = 1$ if time point t in trial n reflects activity from $\mathcal{Z}$, and 0 otherwise. For instance, visual information typically reaches cortex after approximately 75 ms post-stimulus [16], so early time points are likely uninformative. However, in higher-order cognitive processes brain responses are much less stereotyped, and the assumption of consistent timing across columns of $\mathbf{I}$ is less justified. Estimating $\mathbf{I}$, therefore, allows us to identify when, within each trial, task-relevant processing occurs, offering a window into the temporal dynamics of internal cognitive states.

In summary, we aim not only to predict $\mathcal{Y}$ for new trials, but also to estimate $\boldsymbol{I}$ so that we can have a read-out of when $\mathcal{Z}$ develops on a trial-by-trial basis. For a summary of the notation used in the manuscript see Table 1.

2.2. ADA: algorithmic strategy for temporal selection and classification

The objective of ADA is to estimate $\boldsymbol{I}$ for a dataset $\boldsymbol{X}$ —that is, when $\mathcal{Z}$ is elicited in each trial – and to use this information to predict the label of out-of-sample trials. ADA is based on the K-nearest neighbours (KNN) classifier [17], which in its standard form is defined as

$$p(\mathrm{y} = c \mid \mathbf{x}, \mathbf{X}, \mathbf{y}, K) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{I}(y_{\nu(k)} = c), \tag{1}$$

where $\mathbf{x}$ is an out-of-sample trial and y its label, $\mathbf{X}$ and $\mathbf{y}$ are the training data and labels, K is the number of neighbours, $\nu(k)$ indexes the k-th nearest neighbour to $\mathbf{x}$, and $\mathbb{I}(\cdot)$ is the indicator function that returns one if its argument is true, and zero otherwise. Classification is performed by assigning to $\mathbf{x}$ the class c through a majority voting scheme from its K nearest neighbours in the training set.

More generally, given a labelled set of trials and a distance metric, the KNN rule predicts the label of an unlabelled trial by finding its K closest labelled trials and taking a majority vote over their labels. Here we used a weighted variant, which gives more influence to closer neighbours by replacing the vote with a similarity-weighted sum.

To account for trial-specific temporal variability, we can make the KNN rule depend on the variable $\boldsymbol{I} \in \{0,1\}^{T\times N}$, which specifies when $\mathcal{Z}$ is active within each trial, without assuming consistent timing across the dataset. In practice, each trial is segmented into W overlapping temporal windows of length L. Instead of estimating the full matrix $\boldsymbol{I}$, we define a coarse-grained $N \times W$ binary matrix $\boldsymbol{H}$, where each row contains a single one indicating the most informative window for that trial, and zeros elsewhere. This provides a lower-dimensional approximation of $\boldsymbol{I}$, under the assumption that a contiguous window of length L suffices to capture the relevant activity of $\mathcal{Z}$.

To estimate $\boldsymbol{H}$, we unfold the data $\boldsymbol{X}$ into a matrix $\mathbf{D} \in \mathbb{R}^{WN\times Lp}$, where each row $\mathbf{d}_j$ corresponds to a window, and the columns represent

Table 1
Notation used in the manuscript: symbols, domains, shapes, and brief descriptions. Scalars show shape "—"; arrays list explicit shapes.

Symbol	Domain	Dimensions	Description
Observed data			
N	$\mathbb{N}$	—	Number of trials.
T	$\mathbb{N}$	—	Number of time samples per trial.
p	$\mathbb{N}$	—	Number of channels.
t, c, n	index sets	—	Indices: $t \in \{1,\dots,T\}$, $c \in \{1,\dots,p\}$, $n \in \{1,\dots,N\}$.
$\mathbf{X}$	$\mathbb{R}$	$T \times p \times N$	Recorded brain activity.
$\mathbf{Y}$	$\{-1,+1\}$	N	Vector of trial labels.
$\mathcal{Z}$			Latent neural process generating task-relevant activity.
$\mathbf{x}_n$	$\mathbb{R}$	$T \times p$	Multichannel time series for trial n.
y_n	$\{-1,+1\}$	—	Label for trial n.
$\mathbf{I}$	$\{0,1\}$	$T \times N$	Binary mask for $\mathcal{Z}$-related activity over time and trials.
Algorithm			
L	$\mathbb{N}$	—	Window length (in samples).
W	$\mathbb{N}$	—	Number of windows per trial.
K	$\mathbb{N}$	—	Number of neighbours in kNN.
$\mathbf{D}$	$\mathbb{R}$	$WN \times Lp$	Data structured by windows.
$\mathbf{H}$	$\mathbb{R}$	$T \times N$	binary matrix with informative windows per trial.
$\mathbf{d}_j$	$\mathbb{R}$	Lp	Feature vector for window j (row of $\mathbf{D}$).
$\mathbf{r}$	$\{-1,+1\}$	WN	Trial label replicated across its W windows.
a	$\mathbb{R}$	WN	vector of window-level accuracy.
$\hat{\boldsymbol{\beta}}$	$\mathbb{R}$	Lp	coefficients predicting a from $\mathbf{D}$.
α	$\mathbb{R}_{>0}$	—	Ridge regularization coefficient.
κ	$\mathbb{N}$	—	Number of informative windows selected for inference.
Experiments			
$\boldsymbol{x}^{(c)}$	$\mathbb{R}$	T	Class-specific evoked component.
$\boldsymbol{x}^{(nc)}$	$\mathbb{R}$	T	Non-class-specific background component.
ρ	$[0,1]$	—	mixing coefficient between $\boldsymbol{x}^{(c)}$ and $\boldsymbol{x}^{(nc)}$.
p_0	$\mathbb{N}_{<=p}$	—	Number of relevant channels.
s	$\mathbb{N}_{<T}$	—	Trial-specific latency index for evoked response.
σ	$\mathbb{N}$	—	Temporal dispersion parameter for sampling s.

the concatenation of signal values across p channels and L time points. Correspondingly, we define a label vector $\mathbf{r} \in \{-1,1\}^{WN}$ by expanding $\mathbf{y}$, so that each window inherits the label of its parent trial.

Given $\boldsymbol{D}$ and $\boldsymbol{r}$, we apply a window-level weighted KNN classifier in a leave-one-trial-out fashion, modifying the standard KNN rule in Eq. (1) by introducing similarity-based weights:

$$\hat{r}_j = f\left(\mathbf{d}_j, \mathbf{D}^{(-j)}, \mathbf{r}^{(-j)}, K\right) = \sum_{k=1}^{K} \frac{r_{\nu(k)} w_k}{K}, \tag{2}$$

where $\mathbf{D}^{(-j)}$ and $\mathbf{r}^{(-j)}$ refer to the training data after excluding the j-th window and all other windows from the same trial. The index vector $\boldsymbol{\nu}$ contains the position of the K nearest neighbours to $\boldsymbol{d}_j$ within $\boldsymbol{D}^{(-j)}$, and weights $w_1,\dots,w_k$ measure the contribution of each nearest neighbour to the prediction. These weights are computed as

$$w_k = \max\left(S_C(\mathbf{d}_j, \mathbf{d}_{\nu(k)}),\, 0\right), \tag{3}$$

where $S_C(\cdot,\cdot)$ denotes the cosine similarity between vectors. This ensures that only neighbours with non-negative similarity influence the prediction.

Applying Eq. (2) across the training set yields a prediction $\hat{r}_j \in [-1,1]$ for each training window, which we compare to the ground truth r_j to compute a window-level accuracy

$$a_j = \hat{r}_j \cdot r_j, \quad \forall j = 1,\dots,WN, \tag{4}$$

where positive values indicate correct classifications, and negative values indicate incorrect ones. The resulting accuracy vector $\mathbf{a}$ captures the discriminative evidence provided by each window.

To estimate $\boldsymbol{H}$ we select, for each trial, the window j with the highest a_j. We denote this matrix as $\hat{\boldsymbol{H}}$.

To estimate the timing of the cognitive process on a trial-by-trial basis, we fit a ridge regression model to predict the amount of information that each window contains about y:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{j=1}^{WN} (a_j - \mathbf{d}_j \boldsymbol{\beta})^2 + \alpha \|\boldsymbol{\beta}\|_2^2, \tag{5}$$

where α is a regularization coefficient set to a small positive value. This corresponds to the first-level prediction.

Given these elements, at test time, an unseen trial $\mathbf{x}$ is segmented into W windows $\mathbf{d}_1,\dots,\mathbf{d}_W$, each of which is scored by applying $\hat{\boldsymbol{\beta}}$. The window with maximum predicted accuracy, $\mathbf{d}_{\max}$, is then used for classification as

$$\hat{y} = \operatorname{sign} f\left(\mathbf{d}_{\max}, \mathbf{D}_{\hat{\boldsymbol{H}}}, \mathbf{r}_{\hat{\boldsymbol{H}}}, K\right), \tag{6}$$

which is equivalent to Eq. (2) but uses only the training windows indexed by $\hat{\boldsymbol{H}}$. This is the second-level prediction.

This procedure can be generalized to allow multiple windows per trial to participate in the prediction. Specifically, the matrix $\boldsymbol{H}$ may contain up to κ ones per row. Similarly, the κ most informative windows are selected at test time. Thus, if $\kappa > 1$, the final prediction integrates their contributions as

$$\hat{y} = \operatorname{sign} \sum_{l=1}^{\kappa} f\left(\mathbf{d}_l, \mathbf{D}_{\hat{\boldsymbol{H}}}, \mathbf{r}_{\hat{\boldsymbol{H}}}, K\right), \tag{7}$$

where $\mathbf{d}_1,\dots,\mathbf{d}_\kappa$ are the top-κ windows by predicted accuracy.

In summary, ADA operates via non-parametric prediction and trial-wise window selection. Its key hyperparameters are: K (number of neighbours), L (window length), and κ (number of windows per trial for prediction). A schematic of the approach is depicted in Fig. 1

2.3. Justification of the approach

We opted for a non-parametric, similarity-based approach to label prediction, rather than standard discriminative methods based on linear

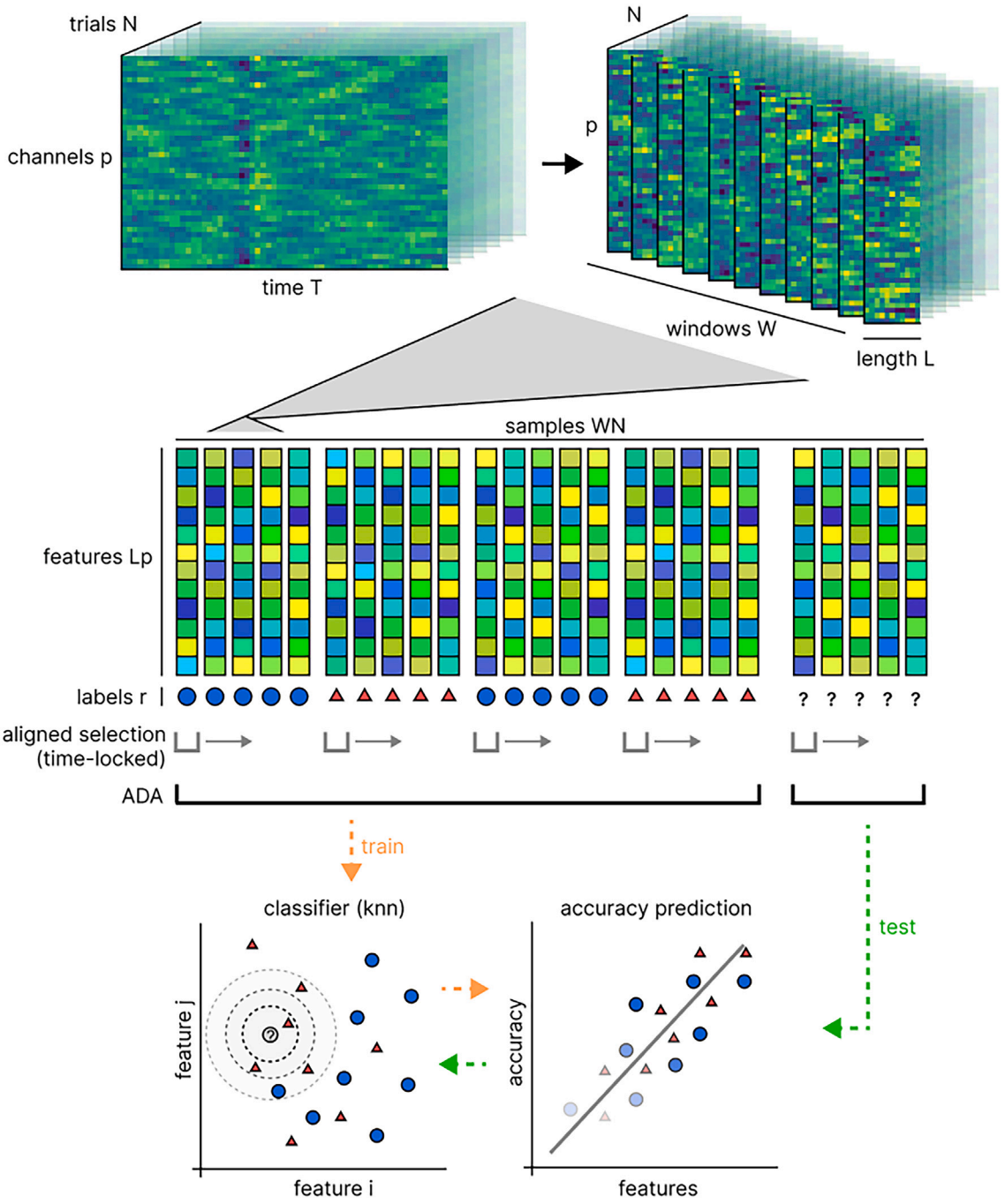


Fig. 1. Overview of ADA. Neural recordings $\mathbf{X} \in \mathbb{R}^{T \times p \times N}$ are segmented into W overlapping temporal windows of length L, forming a feature matrix $\mathbf{D} \in \mathbb{R}^{WN \times Lp}$ with replicated labels. Trials are divided into train and test partitions. Unlike to time-locked approaches that decode each window independently, ADA integrates information across all windows for inference. During training, a KNN decoder computes window-level accuracies $a \in \mathbb{R}^{WN}$, quantifying the informativeness of each temporal segment and summarized in matrix $\hat{\mathbf{H}} \in \mathbb{R}^{T \times N}$. A ridge regression then learns a mapping $\hat{\beta} \in \mathbb{R}^{Lp}$ from window features to informativeness. At test time, ADA uses $\hat{\beta}$ to score windows for each test trial, selects the top-κ informative windows, and decisions are aggregated with the learned KNN across the selected windows for the final label prediction.

regression, linear discriminant analysis (LDA) or support vector machine (SVM), which are more common in neural decoding [1,2]. There are two main reasons for this choice. First, nonparametric methods like KNN make minimal assumptions about the data distribution, which enhances robustness in settings with limited trials and high trial-to-trial variability. Second, discriminative models trained across all time windows are prone to learning systematic but irrelevant differences across the trials, such as global shifts in amplitude or frequency content, that can separate conditions without capturing the actual emergence of the relevant process $\mathcal{Z}$. In such cases, classification performance may reflect temporal position rather than the presence of informative neural activity. In contrast, our similarity-based strategy focuses on detecting recurrence of class-specific patterns across trials, regardless of their absolute position in time. This makes it better suited for detecting transient, temporally variable processes.

We further employ a weighted KNN rule to account for the fact that many windows do not carry any discriminative signal -i.e. $\mathcal{Z}$ is not active during those segments. By down-weighting dissimilar neighbours, the

method reduces the influence of noisy or irrelevant windows, improving robustness even when the estimation of $\boldsymbol{H}$ is imperfect.

To estimate the amount of information that each window contains about the label (and therefore the temporal activation of the cognitive process), we use standard ridge regression (Equation (5)). This choice reflects a trade-off between simplicity and robustness: while windows associated with high accuracy often share recognizable patterns, windows with low or no predictability vary more widely. Regularised regression provides a stable estimation framework to prevent overfitting to heterogeneous inputs.

Finally, we use cosine similarity instead of Pearson correlation to preserve low-frequency information. Pearson correlation subtracts the mean from each window, which suppresses slow components that may carry relevant signals. In contrast, cosine similarity retains the full spectral content by operating on the raw signal. Additionally, we set negative similarities to zero, as such values may arise from windows in an anti-phase relationship, which need not imply functional dissimilarity nor convey information about the stimulus.

It is important to note that ADA is a modular framework. The choices made here are only one possible configuration, and the algorithm can readily be adapted to use other base classifiers. For example, if a neighbour-based classifier is used, the similarity metric can be replaced by any distance measure or learned embedding that defines the pattern space. Other classifiers can also be implemented provided they yield a per-window scalar that can serve as input to the scoring model. We use linear ridge regression, but any regressor, including variants with alternative penalties or fractional formulation, can be used instead.

2.4. Baseline method

In order to assess our capacity to model between-trial variability and leverage this information to improve predictions, we need to benchmark ADA against a method that does not model between-trial temporal variability but is otherwise equivalent regarding any other aspect. To make this comparison as fair as possible, we chose a baseline where we assume that $\boldsymbol{H}$ does not vary across trials, i.e., that $\boldsymbol{h}_{j_1} = \boldsymbol{h}_{j_2}$ for all trials j_1, j_2. In this case, the method is reduced to a simpler sliding-window approach. Following the notation above, this would correspond to each row of $\boldsymbol{H}$ containing zeros everywhere except at the position corresponding to the current window. We implemented this comparison using KNN, SVM, and LDA decoders, each in its time-locked variant (tl-KNN, tl-SVM, tl-LDA) performing one prediction per testing trial and window, and then integrating the predictions across windows by averaging all the votes across windows. Note that this averaging approach would not noticeably penalise having windows where there is no effect, since, asymptotically, in the absence of any signal the votes for these windows would average out to zero across trials. The hyperparameters of the time-locked models included the window length L for all decoders, the number of neighbours K for tl-KNN, the regularization parameter $C = 5$ for tl-SVM, and the shrinkage coefficient λ for tl-LDA, estimated using the Ledoit-Wolf method [18].

3. Results

We benchmarked ADA against a method that does not model between-trial temporal variability, as described in Section 2.4. We conducted both controlled simulations and an analysis of real MEG data. The simulations allowed us to systematically explore the effect of having different degrees of temporal variability in the data as well as ADA's sensitivity to hyperparameter choices. The real-data experiment provided a testbed with realistic signal complexity, enabling us to probe whether ADA maintains its advantage under more natural conditions.

3.1. Simulation study

To evaluate the performance of ADA under controlled conditions, we generated synthetic data using Genephys [19], a generative model of electrophysiological signals that enables precise control over oscillatory dynamics, evoked responses, background noise, and temporal variability. Genephys allows for different types of effects. For simplicity, we focused on the additive oscillatory responses, which adds a stimulus-specific oscillatory wave on top of the ongoing dynamics. We defined two signal components and then combined them according to the parameters defined in our simulation. Specifically, each dataset comprised two additive sources: (i) a task-independent background signal $\mathbf{x}^{(\mathrm{nc})}$, modelled as sinusoidal oscillator whose amplitude and frequency follow first-order autoregressive processes (AR1); and (ii) a class-specific evoked response $\mathbf{x}^{(\mathrm{c})}$, consisting of a 5 Hz sinusoidal carrier modulated by a Gaussian-shaped response function centered at trial latency s. A half-cycle phase shift distinguished the two classes. Thus, the observed signal was defined as:

$$\mathbf{x} = \rho\,\mathbf{x}^{(\mathrm{c})} + (1-\rho)\,\mathbf{x}^{(\mathrm{nc})}, \tag{8}$$

where $\rho \in [0, 1]$ controls the relative strength of the effect. Fig. 2b shows example single-trial and averaged signals across different ρ values.

This scheme was implemented for a dataset containing p channels, of which p_0 were task-relevant. Relevant channels included both $\mathbf{x}^{(\mathrm{nc})}$ and $\mathbf{x}^{(\mathrm{c})}$, while the remaining $p - p_0$ channels were set to $\rho = 0$.

To implement temporal variability, $\mathbf{x}^{(\mathrm{c})}$ was anchored to a latent trial-specific index s, indicating the time point at which the discriminative effect peaked. For each trial, s was drawn from a categorical distribution parameterised by a discrete Gaussian profile $(\eta_1, \ldots, \eta_T)$, where T is the number of time points. This distribution was generated by binning a Gaussian curve (mean $T/2$, standard deviation σ) and normalising it to sum to one. Three illustrative distributions and their corresponding sampled values of s are shown in Fig. 2a for $\sigma = 1$, 5, and 20. A channel-specific, trial-shared temporal jitter was then added to model the spatial dispersion of the effect. We did not project to sensor space nor add correlated noise, so baseline cross-channel correlations were absent. For each iteration of the simulation, trials were assigned to two equally-sized stimulus classes. Training and testing sets were generated independently using the same parameters.

Given this simulation setting, we systematically varied the following simulation parameters (while leaving the rest by default [1]):

- The temporal dispersion σ, controlling the trial-to-trial variability of the effect;
- The signal mixing coefficient ρ, modulating class-related discriminability;
- The number of informative channels p_0.

For each parameter combination, 100 independent datasets were generated, each comprising $N = 200$ training trials and $N_{\mathrm{test}} = 200$ test trials, with $p = 40$ channels and $T = 100$ time points (1 s at 100 Hz).

3.1.1. Sensitivity to signal properties

Here, we evaluated the influence of the data-generating parameters on classification accuracy, using $K = 20$, $L = 30$, and $\kappa = 4$.

First, we examined the impact of temporal variability by varying σ between 1 and 20, with $\rho = 0.5$ and $p_0 = 20$ fixed. As shown in Fig. 3a, the accuracy of baseline models degraded markedly with increasing σ while ADA maintains a stable performance, with a gap that increases proportionally to the dispersion parameter. While all ADA formulations are robust to variability, their overall accuracy diverges, with KNN yielding the highest and SVM the lowest performance. This is the most important comparison, since it speaks to the main limitation of the standard approaches that ADA is aiming to solve.

Next, we tested the effect by varying ρ from 0.3 to 0.7, with $\sigma = 10$ and $p_0 = 20$ (Fig. 3b). Accuracy improves for all methods as ρ increases, but ADA-KNN consistently outperforms every other model. The remaining ADA variants and the time-locked baselines perform comparably

[1] https://genephys-doc.readthedocs.io/

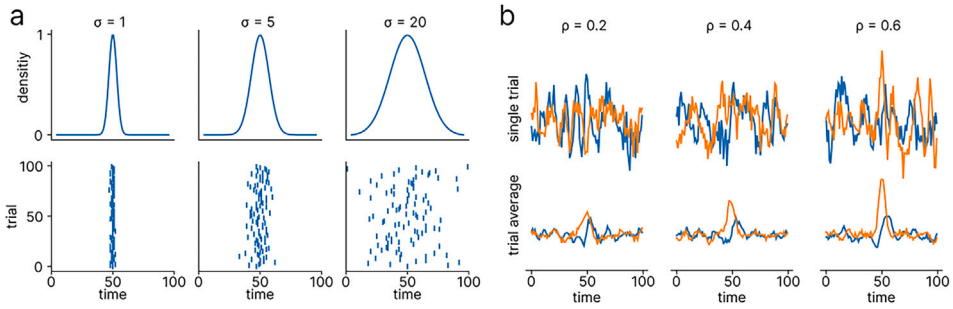


Fig. 2. Simulated data generation. a) Trial-wise variability is introduced by sampling latencies indices s from a categorical distribution $\boldsymbol{\eta}$, determined by the dispersion parameter σ. Top: three examples of $\boldsymbol{\eta}$; bottom: corresponding samples of s across trials of a dataset. b) Single-trial (top) and averaged (bottom) signals for a representative relevant channel at different levels of the signal mixing factor ρ. Blue and orange lines represent different conditions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

across the entire range. This comparison reveals that ADA-KNN is reasonably robust to noisy scenarios.

Finally, we assessed the effect of sparsity by varying p_0 between 10 and 30, with $\rho = 0.5$ and $\sigma = 10$ (Fig. 3c). As expected, all methods performed worse for greater levels of sparsity, but ADA-KNN continued to show superior accuracy. Similarly to the previous test, this comparison was intended to show the robustness of ADA with respect to varying effect sizes.

In addition, we assessed sensitivity to data volume and computational load. Performance remained stable across total trial counts and test-set sizes, except for ADA-SVM, which degraded rapidly with smaller N (Figure S1). During training, ADA-KNN required more memory to store pairwise similarities, while ADA-SVM required more compute time due to leave-one-out fitting. At testing, all ADA variants showed comparable memory use, with ADA-KNN being slightly slower on the present data, potentially limiting online deployment, yet optimizable through more efficient implementation. ADA-LDA was generally the most efficient variant (Figure S2).

Overall, these experiments demonstrate that ADA in its non-parametric formulation is able to model between-trial variability while being robust to decreasing effect sizes and larger amounts of noise. Parametric variants also exhibit robustness to temporal variability, reinforcing the general reliability of the ADA framework.

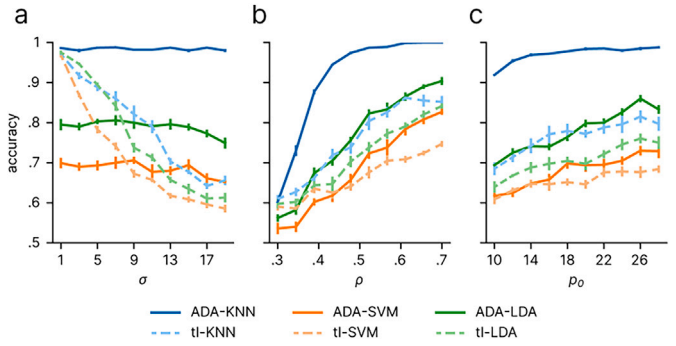


Fig. 3. Performance as a function of the data properties. Decoding accuracy of ADA using different decoders: KNN (blue), SVM (orange), and LDA (green) and their time-locked baselines (dashed lines) as a function of a) temporal dispersion σ, b) signal mixing coefficient ρ, and c) sparsity parameter p_0. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1.2. *Sensitivity to algorithm parameters*

Here, we tested the sensitivity of ADA's performance to variations in the algorithm's hyperparameters. We first tested the hyperparameters that are common to all models, the window length L, then examined the effect of κ, which applies only to ADA, and finally K, which is specific to the KNN variant.

We fixed $p_0 = 20$, $\rho = 0.5$, and $\sigma = (10, 20)$, corresponding to scenarios of moderate and high temporal variability correspondingly. Note that these results are not meant to generalize across all possible datasets, as the optimal configuration may depend on the specific characteristics of the signal.

Regarding L, all models show improved accuracy as the window length increases, with ADA-KNN performing consistently better across the tested range (Fig. 4a). However, the optimal value of L is likely data-dependent, as it interacts with the temporal extent and spectral content of the underlying neural signal.

For κ, ADA-KNN benefits from using multiple predictive windows per trial (Fig. 4b), which helps compensate for potential errors in estimating $\boldsymbol{H}$. In contrast, the parametric variants show little benefit, likely because their decision functions are trained on data containing all noisy samples.

Finally, for K, Fig. 4c shows that ADA-KNN maintains a high performance across a wide range of values, with only a slight drop at the lowest end. By contrast, tl-KNN performs better with lower values of K, likely because the method already integrates across multiple windows and thus requires fewer neighbours per window.

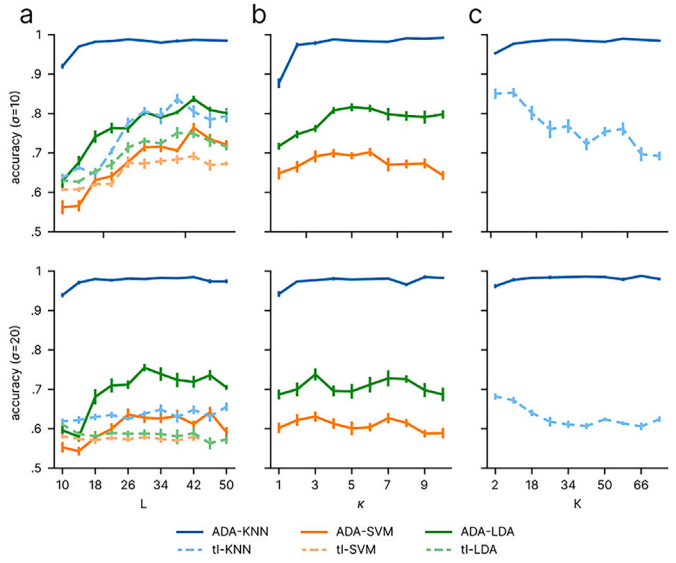


Fig. 4. Performance as a function of the algorithm hyperparameters under moderate and high temporal variability. Decoding accuracy of ADA using different decoders: KNN (blue), SVM (orange), and LDA (green) and their time-locked baselines (dashed lines) as a function of a) window length L, b) κ number of predictive windows per trial (ADA only), and c) number of neighbours K (KNN only). Upper panels correspond to moderate temporal variability ($\sigma = 10$) and lower panels to high temporal variability ($\sigma = 20$). Default values: $K = 20$, $L = 30$, $\kappa = 4$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

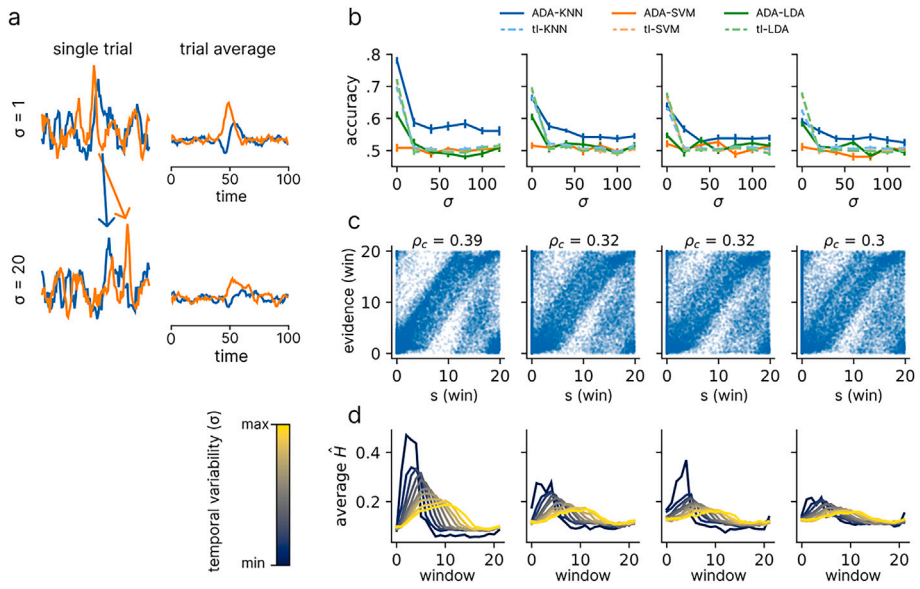


Fig. 5. Decoding on a simulated visual memory recall experiment from real perception data. a) Visual memory recall is emulated by circularly shifting MEG signals on each trial. b) Performance of ADA using different decoders: KNN (blue), SVM (orange), and LDA (green) and their time-locked baselines (dashed lines) as a function of between-trial temporal variability σ, with 95 % confidence intervals. c) Relationship between the ground-truth temporal shift s and ADA-KNN evidence center $|\hat{H}\hat{r}|$. Each point represents one simulated trial. d) For each σ, the cross-trial average of $\boldsymbol{H}$ $(W \times 1)$ reflects the relevance of each window across time as estimated by ADA. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Overall, these experiments demonstrate that all models exhibit robustness to moderate changes in hyperparameter settings. This property is particularly advantageous in practical applications where exhaustive parameter tuning may be limited by data availability or computational constraints.

3.2. Visual memory recall

Visual memory recall, or imagery, refers to the volitional reactivation of perceptual representations [20,21]. In experimental settings, participants are typically presented with a set of stimuli and later cued to imagine one of them [22,23]. Compared to visual perception, the onset of the relevant brain activity during recall is not externally locked and varies across trials, posing a challenge for standard decoding methods.

Provided that decoding in imagery experiments is hampered by other confounds unrelated to temporal variability (imprecise trial labelling, trials in which the subject did not engage with the task, etc.), we constructed a benchmark using real MEG data from a visual perception task [22] so that we could isolate the effects of temporal variability while maintaining realistic signal properties. The data consist of trials where participants viewed either a face or a house, with recordings sampled at 100 Hz, and source-reconstructed to $p = 74$ cortical regions spanning the cortex [24]. We used data from four participants, each with approximately 200 trials of 1s duration. To simulate imagery-like timing variability while retaining real signal properties, we circularly shifted each trial forward by a latency s_n drawn from $\mathcal{U}(0, \sigma)$, with σ controlling the magnitude of temporal misalignment. This procedure maintains the trial content but randomises its timing. Fig. 5a shows a representative example.

We evaluated the ADA variants and their respective time-locked baselines across $\sigma \in \{0, \ldots, 150\}$, running 100 random train/test splits per subject and per choice of σ, using hyperparameters $K = 20$, $L = 30$, and $\kappa = 5$. This configuration reflects a stable setting identified with the simulation study and conforms to standard bias–variance considerations ($K \approx \sqrt{N}$). The window length (L) corresponds to the temporal scale of early visual responses (300 ms), avoiding excessive temporal smoothing, while a small κ preserves interpretability in window selection and reduces overfitting risk.

Fig. 5b shows decoding accuracy for each σ. For low variability, time-locked baselines achieve similar or higher performance, likely due to ADA's flexibility exceeding the requirements of this setting. However, as temporal variability increases, the performance of the baseline models and parametric ADA drops steadily, while ADA-KNN remains stable. This pattern is consistent across all subjects.

We assessed the temporal precision of ADA-KNN by correlating the model's estimated information center with the ground-truth temporal shift. For each trial, the circular position of the information was computed as the angular mean of the selected window positions $\hat{H}$, weighted by the absolute continuous prediction $\hat{r}$. The resulting angle was compared to the ground-truth shift angle using circular–circular correlation [25]. Statistical significance was assessed for each subject using $10,000$ permutations of the true shift, and Bonferroni corrected across subjects; all corrected $p < .001$. These results, shown in 5c demonstrate that ADA-KNN consistently tracks the imposed temporal displacement.

As a more explicit illustration of the captured temporal variability, Fig. 5d shows the cross-trial average of $\boldsymbol{H}$ for each value of σ in ADA-KNN. For low σ, most weight concentrates near the trial onset, consistent with the timing of the original visual evoked response. For higher σ, the weight distribution broadens, indicating that ADA-KNN adapts its window selection to the increased temporal dispersion.

These results demonstrate that ADA is capable of maintaining decoding performance under conditions of substantial between-trial temporal variability in real MEG signals, and that it can recover interpretable estimates of when predictive information is likely to occur.

4. Discussion

In this paper, we address the problem of predicting a behavioural variable from brain activity on a trial-by-trial basis, in settings where the neurophysiological underpinnings of the behavioural variable are covert; that is, when we cannot know exactly when they unfold. Our proposed method, ADA, addresses this challenge by separating the estimation process into two stages: first, identifying when the signal is predictive of the label; then, conditional on this selection, predicting the label itself. This two-stage formulation allows the model to adaptively

focus on informative temporal segments within each trial, avoiding the assumption of fixed temporal alignment across trials.

Interpreting these results benefits from comparison with earlier approaches to trial-wise temporal variability in neural data. It is well established that the timing of brain responses, even at the level of individual neurons, can vary considerable from trial to trial [26–28]. In the case of spike train data, dynamic time warping (DTW) has been proposed to account for this variability [29]. While standard DTW tends to overfit in many neural datasets due to noise, parametric simplifications show considerably better results [30]. However, the problem addressed in this work differs fundamentally. Rather than aligning trials based on overall signal similarity, our objective is to identify transient segments of the signal that are selectively informative about experimental conditions, and may explain only a small portion of the total variance. In machine learning terms, this is a supervised learning problem, whereas DTW and its variants are generally unsupervised. Some supervised extensions exist; for instance, in spike train modelling DTW can be combined with a generative model conditional on $\mathbf{y}$, which can be used for prediction after model inversion [31]. However, these approaches rely on strong distributional assumptions (e.g., Poisson firing) that do not generalize easily to continuous, oscillatory signals such as M/EEG. Other probabilistic models designed to capture between-trial variability, such as Gaussian process factor analysis [32], also fall under unsupervised learning and do not explicitly model stimulus- or task-dependent representations. Finally, switching state-space models or nonparametric HMMs, which are also unsupervised, can in theory detect temporal variability across trials [33,34], yet these methods aim at characterising the most dominant sources of variability, irrespective of what are the most informative features for classification.

In contrast, by combining similarity-based prediction with data-driven window selection, ADA targets condition-discriminative segments without making strong generative assumptions or relying on fixed timing. The explicit separation between temporal localization and classification also enhances interpretability, allowing researchers to identify when neural signals carry relevant information on a trial-by-trial basis. To our knowledge, it is the first method to offer a unified supervised framework for jointly identifying informative time segments and generalizing across trials, applicable to continuous signals without requiring modality-specific assumptions.

From a machine learning perspective, ADA addresses a supervised learning scenario in which the identity of the covariates is not well defined, i.e. where their definition depends on latent, trial-specific timing. While the problem of label noise in the dependent variable has been extensively studied [35,36], less attention has been paid to uncertainty or misalignment in the independent variables. Here we exemplify temporal variability with imagery as a self-initiated cognitive processes. However, even in strictly time-locked paradigms, neural dynamics can vary in latency across trials, and informative patterns may emerge gradually as perceptual evidence accumulates or categorical decisions are formed [3,37,38]. A related form of variability arises in dynamic stimulation paradigms –such as language or movie perception–, where the relevant features of the sensory input unfold over time, and recognition can occur at different moments within the continuous stream [39,40]. ADA represents a step toward filling this gap, providing a general framework for learning from temporally misaligned features.

Our implementation uses a leave-one-out KNN decoder. However, ADA does not rely on this specific choice, and its logic can be extended to any model capable of providing window-level accuracy estimates. KNN offers this without training weights or hidden regularization, demonstrating that adaptive inference is possible even with a simple, non-parametric classifier operating on raw signals. This simplicity entails two main limitations. First, the computational cost scales quadratically with the number of training trials and windows, since all pairwise distances must be computed. Although this remains cheaper than training complex parametric models such as SVMs, deployment requires retaining training data rather than a compact parameter set. Second, the KNN is sensitive to high-dimensional and noisy feature spaces: when distances become uniformly large, discrimination power degrades ("curse of dimensionality"). These effects could be mitigated through dimensionality reduction, feature selection, or whitening. Future versions may incorporate more powerful decoders. Convolutional neural networks, for instance, could learn discriminative spatiotemporal representations directly from the signal, reducing dimensionality while retaining local structure. Likewise, recurrent or attention-based architectures could model temporal dependencies across multiple events, extending the current single-event formulation. Such extensions would enhance scalability and expressivity of the framework while preserving its core adaptive inference principle.

In summary, ADA provides a principled framework for decoding neural activity when the timing of informative signals varies across trials. By comparing ADA against an alternative that assumes no temporal variability (i.e., a fixed-window strategy with integrated predictions), we found that classification accuracy can itself be predicted from the signal in a trial-by-trial basis and in a temporally resolved manner, and that this information can be leveraged to improve decoding under temporal uncertainty.

CRediT authorship contribution statement

Pablo Oyarzo: Conceptualization, Data curation, Investigation, Resources, Validation, Writing – original draft, Formal analysis, Methodology, Software, Visualization, Writing – review & editing. **Radoslaw M. Cichy:** Supervision, Resources, Writing – review & editing. **Diego Vidaurre:** Formal analysis, Investigation, Project administration, Software, Validation, Writing – original draft, Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Visualization, Writing – review & editing.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used chatGPT in order to edit grammar. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

PO was supported by the Scholarship Program of the National Agency for Research and Development (2019-72200281). R.M.C. is supported by German Research Council (DFG) grants (CI 241/1–1, CI 241/1–3, CI 241/1–7 and INST 272/297–1), the European Research Council (ERC) starting grant (ERC-StG-2018-803370) and the ERC Consolidator grant (ERC-CoG-2024101123101). DV is supported by a Novo Nordisk Foundation Emerging Investigator Fellowship (NNF19OC-0054895) and an ERC Starting Grant (ERC-StG-2019–850404).

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.csbj.2025.10.044.

Code and data availability

All code use in this article, including the scripts to reproduce the figures, is available at https://github.com/oyarzou/csbj25. The simulated data can be generated using the code provided in https://github.com/vidaurre/genephys. The real MEG data are openly available from the original publication [22] at https://data.ru.nl/collections/di/dcc/DSC_2017.00072_245

References

[1] Haynes J, Rees G. Decoding mental states from brain activity in humans. Nat Rev Neurosci 2006;7:523–34.
[2] Grootswagers T, Wardle SG, And TAC. Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. J Cogn Neurosci 2017;29:677–97.
[3] Cichy RM, Pantazis D, Oliva A. Resolving human object recognition in space and time. Nat Neurosci 2014;17:455–62.
[4] Brodersen KH, et al. Decoding the perception of pain from fMRI using multivariate pattern analysis. NeuroImage 2012;63:1162–70.
[5] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. Clin Neurophysiol 2002;6: 767–91.
[6] Van Erp J, Lotte F, Tangermann M. Brain-computer interfaces: beyond medical applications. Computer 2012;45:26–34.
[7] Hari R, Puce A. MEG-EEG Primer. Oxford: Oxford University Press; 2017.
[8] Stokes MG. Activity-silent working memory in prefrontal cortex: a dynamic coding framework. Trends Cogn Sci 2015;19:394–405.
[9] Dijkstra N, Mostert P, Lange FPD, Bosch S, van Gerven MA. Differential temporal dynamics during visual imagery and perception. Elife 2018;7:e33904.
[10] Grootswagers T, Robinson AK, Carlson TA. The representational dynamics of visual objects in rapid serial visual processing streams. NeuroImage 2019;188:668–79.
[11] Stokes M, Spaak E. The importance of single-trial analyses in cognitive neuroscience. Trends Cogn Sci 2016;20:483–6.
[12] King J-R, Dehaene S. Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn Sci 2014;18:203–10.
[13] Vidaurre D, Myers NE, Stokes M, Nobre AC, Woolrich MW. Temporally unconstrained decoding reveals consistent but time-varying stages of stimulus processing. Cereb Cortex 2019;29:863–74.
[14] Higgins C, et al. Spatiotemporally resolved multivariate pattern analysis for M/EEG. Hum Brain Mapp 2022;43:3062–85.
[15] Buzsaki G, Draguhn A. Neuronal oscillations in cortical networks. Science 2004;304:1926–9.
[16] Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature 1996;381:520–2.
[17] Fix E, Hodges JL. Discriminatory analysis. Nonparametric discrimination: consistency properties. Int Stat Rev 1989;57:238–47.
[18] Ledoit O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. J Multivar Anal 2004;88:365–411.
[19] Vidaurre D. A generative model of electrophysiological brain responses to stimulation. Elife 2024;12:RP87729.
[20] Cichy RM, Heinzle J, Haynes J-D. Imagery and perception share cortical representations of content and location. Cereb Cortex 2012;22:372–80.
[21] Tong F. Imagery and visual working memory: one and the same? Trends Cogn Sci 2013;17:489–90.
[22] Dijkstra N, Ambrogioni L, Vidaurre D, van Gerven M. Neural dynamics of perceptual inference and its reversal during imagery. eLife 2020;9:e53588.
[23] Xie S, Kaiser D, Cichy RM. Visual imagery and perception share neural representations in the alpha frequency band. Curr Biol 2020;30:2621–7.
[24] Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 2010;53:1–15.
[25] Jammalamadaka SR, Sengupta A. Topics in circular statistics. World Scientific; 2001.
[26] Churchland MM, Shenoy KV. Temporal complexity and heterogeneity of single-Neuron activity in premotor and motor cortex. Nat Neurosci 2007;97:4235–57.
[27] De Ruyter van Steveninck RR, Lewen GD, Strong SP, Koberle R, Bialek W. Reproducibility and variability in neural Spike trains. Science 1997;275:1805–8.
[28] Faisal AA, Selen LPJ, Wolpert DM. Noise in the nervous system. Nat Rev Neurosci 2008;9:292–303.
[29] Berndt D, Clifford J. Using dynamic time warping to find patterns in time series, In: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining; 1994. p. 359–70.
[30] Williams AH, et al. Discovering precise temporal patterns in large-scale neural recordings through robust and interpretable time warping. Neuron 2020;105:246–59.
[31] Lawlor PN, Perich MG, Miller LE, Kording KP. Linear-nonlinear-time-warp-poisson models of neural activity. J Comput Neurosci 2018;45:173–91.
[32] Byron MY, et al. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. Adv Neural Inf Process Syst 2009:1881–8.
[33] Fox E, Sudderth EB, Jordan MI, Willsky AS. Bayesian nonparametric inference of switching dynamic linear models. IEEE Trans Signal Process 2011;59:1569–85.
[34] Vidaurre D, et al. Spectrally resolved fast transient brain states in electrophysiological data. Neuroimage 2016;126:81–95.
[35] Muhlenbach F, Lallich S, Zighed DA. Identifying and handling mislabelled instances. J Intell Inf Syst 2004;22:89–109.
[36] Barnett V, Lewis T. Outliers in statistical data. Norwich: Wiley Series in Probability and Mathematical Statistics; 1984.
[37] Kar K, Kubilius J, Schmidt K, Issa EB, DiCarlo JJ. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. Nat Neurosci 2019;22:974–83.
[38] Kar K, DiCarlo JJ. Fast recurrent processing via ventrolateral prefrontal cortex is needed by the primate ventral stream for robust core visual object recognition. Neuron 2021;109:164–76.
[39] Fedorenko E, et al. Neural correlate of the construction of sentence meaning, In: Proceedings of the National Academy of Sciences, vol. 113. 2016, E6256–E6262.
[40] Lahner B, et al. Modeling short visual events through the bold moments video fMRI dataset and metadata. Nat Commun 2024;15:6241.