Trial-by-trial detection of cognitive events in neural time-series

Gabriel Weindel[1,2], Leendert van Maanen[¶1], and Jelmer P. Borst[¶2]
[1]Utrecht University, The Netherlands
[2]University of Groningen, The Netherlands

Measuring the time-course of neural events that make up cognitive processing is crucial to understand the relation between brain and behavior. To this aim, we formulated a method to discover a trial-wise sequence of events in multivariate neural signals such as electro- or magneto-encephalograpic (E/MEG) recordings. This sequence of events is assumed to be represented by multivariate patterns in neural time-series, with inter-event durations following probability distributions. By estimating event-specific multivariate patterns, and between-event duration distributions, the method allows to recover the by-trial onsets of brain responses. We demonstrate the properties and robustness of this hidden multivariate pattern (HMP) method through simulations, including robustness to low signal-to-noise ratio, as typically observed in EEG recordings. The applicability of HMP is illustrated using previously published data from a speed-accuracy trade-off task. We show how HMP provides, for any experiment or condition, an estimate of the number of events, the sensors contributing to each event (e.g. EEG scalp topography), and the durations between each event. Traditional exploration of tasks' cognitive architectures can thus be enhanced by HMP estimates.

Keywords: Electrophysiology; Multivariate patterns; Single-trial analysis; Time-series analysis; Temporal dynamics; Mental chronometry

1 Introduction

The speed of information processing in the brain has connected psychology and physiology since Von Helmholtz assumed the existence of processing steps in the reaction time (RT) to measure neural transmission velocity in humans (von Helmholtz, 1850). Since then, researchers have investigated the nature and speed of the different information processing operations – or cognitive processes. Classical mental chronometry (Posner, 1978) tackled this problem by relying on experimental manipulations. As one of the earliest examples, the subtractive method developed by Franciscus Donders (1868) takes as an indirect measure of a cognitive process's duration the difference in mean RT from two tasks that are assumed to differ in the insertion of that particular process. For example, the duration of a response selection process could be inferred from the RT difference between a task with response selection, and the same task without response selection.

Yet, condition-based differences impose strong assumptions about the true sequence of cognitive processes and knowledge about the way subjects perform different tasks. Nevertheless, the RT as a composite measure of the time needed for different cognitive processes to complete remains a shared assumption between the main theories of mental chronometry (Luce, 1986). This assumption is further fueled by the repeated observation that RT distributions do not obey any of the known and commonly found statistical distributions (Christie & Luce, 1956; Noorani & Carpenter, 2016). Therefore, considerable effort has been devoted to decomposing RT distributions into different underlying distributions, for example by a Fourier transform approach (Green, 1971; Smith, 1990), mixture modelling (Christie & Luce, 1956; Van Maanen et al., 2014), cognitive (e.g., Anderson et al., 2018; Tenison et al., 2016; Van Maanen et al., 2009), mathematical models (e.g. Anders et al., 2016; Brown & Heathcote, 2005; Ratcliff, 1978; Smith, 1995), and combinations (e.g., van der Velde et al., 2022). While all these applications have extended our understanding of several aspects of human cognition through mental chronometry, the reliance on RT alone limits the ability to decide between competing hypotheses (Forstmann et al., 2011).

To extract the nature and unfolding of cognitive processes, neural recordings with high temporal resolution such as electro-encephalography (EEG), magneto-encephalography (MEG) or intra-cranial recordings would stand as strong candidates. This has led Meyer et al. (1988) to call for a modern mental chronometry approach where physiological

Gabriel Weindel Leendert van Maanen Jelmer P. Borst
Correspondence concerning this article should be addressed to Gabriel Weindel, Department of Experimental psychology, Utrecht University, Heidelberglaan 1, Utrecht 3584 CS, The Netherlands. E-mail: g.m.weindel@uu.nl
¶LvM and JPB are Joint Senior Authors

recordings inform our understanding of cognitive processes involved in the RT and vice-versa (see for example Kelly & O'Connell, 2013; J. Miller et al., 1999; Purcell et al., 2010). However, measuring the onset of cognitive events in physiological signals is complicated by their low signal-to-noise ratio, as relevant activities are usually several orders of magnitude weaker than the background noise of these measures.

Most research efforts have therefore relied on signals averaged over many trials (*e.g.,* peri-stimulus time histograms, event related potentials, etc.). Unfortunately, averaging continuous time series across trials is well known to hide and distort single-trial effects (e.g. Burle et al., 2008; Callaway et al., 1984; Luck, 2005). These distortions arise because of the time variation in the by-trial generators of these average curves, making inferences about the timing of the generators complex (Luck, 2005). These distortions are exacerbated when comparing condition-averaged signals: if the process or component time jitter distribution differs between conditions – as could be expected from behavioral (Noorani & Carpenter, 2016; Ratcliff, 1978; Weindel et al., 2021) and electrophysiological studies (O'connell et al., 2012; Smulders et al., 1994) – then both the peak amplitude and latency of the averaged signal will be different across conditions (Mouraux & Iannetti, 2008).

Additionally, when researchers are interested in several components to describe a task, the temporal overlap of components close to each other also distorts the average of the components (Luck, 2005). Lastly, if one assumes that components in an experimental task are sequential to one another (*e.g.,* if the time of the response-related potential is causally affected by the time of the visual-related potentials), then the time jitter in the later component will contain jitter from the previous components. This has prompted researchers to use multiple events to which to time-lock the signal, for example by using either stimulus onset, response onset or other events such as electro-myographical onset (see Burle et al., 2004, for a comparison of the three methods). However, most psychological processes involved in a task will not have a relevant external index to time-lock physiological analysis on.

For true modern mental chronometry (Meyer et al., 1988) it is therefore necessary to have single-trial descriptions of neural time series. Several single-trial time measurements of physiological events have been suggested. Nunez et al. (2019), for example, used EEG to measure a single-trial visual encoding time and related it to experimental manipulations. Ruchkin and Glaser (1978, as described in Smulders et al., 1994) suggested a pattern search for an EEG component – the P3, associated with stimulus evaluation – to obtain by-trial measurements of the timing of the component. Furthermore, electro-myography was used to capture motor execution time and deepen our understanding of the link between the decision and the motor system (Botwinick & Thompson, 1966; Burle et al., 2002; Weindel et al., 2021). All these applications show that the addition of external recordings is a valuable addition for in-depth exploration of mental processing. Nevertheless, these applications only isolated one processing component from the others, while almost any cognitive task consists of multiple processing steps.

In an effort to address these limitations, Anderson et al. (2016) proposed a method that combined hidden semi-Markov models with a multivariate pattern analysis (HsMM-MVPA). For EEG, this method relied on the two main theories of event-related potential generation: the phasic burst (Makeig, 2002) and the synchronized oscillations theories (Basar, 1980, as cited by Anderson et al., 2016). Both theories predict that a significant cognitive event is expected to be visible as a transient change in an EEG time-series. The HsMM-MVPA method assumes that these transient changes are multivariate patterns repeating across all trials in a sequential order. In addition to the sequential appearance of the events, the method assumes that their onsets follow specific (skewed) probability distributions. By fitting the method to EEG or MEG data, and therefore estimating these patterns and their distributions, one obtains a time location in each RT for each of these putative sequential events. Hence, contrary to ERP averaging, or pattern search based on properties from the averaged ERP (Smulders et al., 1994), the HsMM-MVPA method uses expected time, sequentiality of the patterns and channel contribution to uncover single-trial EEG or MEG events across the whole RT time-series. By explicitly assuming a linking function (Schall, 2019) between the detected EEG/MEG events and cognitive processes, researchers can then interpret the resulting HsMM-MVPA solution as the sequence of cognitive operations that happened on each trial.

This method has been successfully applied to a number of tasks, ranging from simple perceptual (Van Maanen et al., 2021) or lexical decision tasks (Berberyan et al., 2021) to more complex tasks, such as visual working memory (Zhang et al., 2018), associative recognition (Anderson et al., 2016; Portoles et al., 2022), and mathematical problem solving (Anderson et al., 2018; Groeneweg et al., 2021). Each of those studies provided new insights on the nature and latencies of cognitive events of the analyzed tasks. For example, in the original paper, Anderson et al. (2016) linked HsMM-MVPA decomposed RT to a cognitive model of the task. Van Maanen et al. (2021) showed how HSMM-MVPA allowed to separate a decision stage from the other stages and could therefore relate the single-trial decision stage times to an evidence accumulation model. Berberyan et al. (2021) showed that word frequency and lexicality effects are driven by a single cognitive processing stage. Zhang et al. (2018) used HsMM-MVPA to combine EEG and electro-corticography and showed how they could dissociate item familiarity and decision stages in the RT.

Notwithstanding those early successes, HsMM-MVPA has yet to be applied to a broader range of research, which

could be achieved by formulating a generalized version of the model by Anderson et al. (2016). The current assumptions, formulation, and implementation of the HsMM-MVPA method may preclude a broader use. On the assumption side, the HsMM-MVPA method used a very specific definition of the pattern and the by-trial distribution while this might not represent all possible use cases of the method. Moreover, the HsMM-MVPA method is usually fitted with a specified number of events, while the optimal number of events is determined by comparing the different solutions using LOOCV (Anderson et al., 2016; Borst & Anderson, 2024). This two-step procedure implies that an a priori range for the number of events can be estimated.

On the formulation side, HsMM-MVPA was presented as a hidden semi-Markov model. Although the method can indeed be seen as a (linear) state transition system with transition probabilities given by a state duration distribution, the usual assumptions and limitations of hidden semi-Markov models do not apply. That is, HsMM-MVPA does not specify an emission distribution per state, and only allows transitions to the next state (Yu, 2010). The mismatch between HsMM-MVPA and the typical specification of hidden Markov models may have confused researchers familiar with such methods. Finally, on the implementation side, the application of the method requires substantial technical expertise and programming skills, therefore reducing the adoption of the method by the broader research community in cognitive sciences.

In the present study we show a general model based on HsMM-MVPA by redefining the method to a broader range of event patterns and onset assumptions. We also developed a new way of applying the method to directly infer the number of events in the time-series without a priori expectations. Moreover we formulate the model without relying on hidden Markov models to ease the understanding and potential use of the model. To distinguish between the previous iteration of the method and separate its formulation from hidden semi-Markov models, we name this generalized version the *hidden multivariate pattern* method (HMP). At the same time, we make the central idea of Anderson and colleagues more accessible to the wider research community by providing an accompanying Python package.

In the remainder of this paper, we will showcase the use of the HMP method both in simulated and real EEG data, after providing a formal description of the method. All analyses that we report throughout the paper are performed with the accompanying Python package (see the online repository, https://github.com/gweindel/hmp, including a set of tutorials), illustrating the kind of results that are readily accessible to the interested researcher.

2 The hidden multivariate pattern method

The hidden multivariate pattern method or HMP describes a multivariate time-series (e.g., EEG or MEG channels as in Figure 1a), observed from time 0 (e.g., stimulus presentation) to an end time T (e.g., the RT), as a linear succession of a given number of events (Figure 1b). These recurrent sequential experimentally-induced brain events are each assumed to be represented by a pattern (Figure 1c) produced by a specific mixture of channels (Figure 1b and d, see Section 2.1.1).

The multivariate patterns are sequential to each other and their onsets follow a probability distribution of which the parameters are estimated from the data (Figure 1e, see Section 2.1.2). The sequentiality of the events implies that the expected onset time of the nth event depends on the expected onset time of all previous events. This sequentiality implies a fairly complex problem that, as in HsMM-MVPA, we resolve using dynamical programming (see Section 2.2.1).

In theory, the method can be applied to any defined time interval and any type of measurement, however, the current presentation (and Python implementation) focuses on the time interval of the RT from a subject using EEG or MEG data. The application of the method results in an estimation of the probabilities for each of the detected events at each sample for each trial in the data used for the estimation (Figure 1h, Section 2.2.4). These probabilities can be used for follow-up analyses, which we will illustrate in Sections 2.5 and 3.2.

2.1 HMP Concepts

2.1.1 Pattern analysis

An HMP model assumes that events are represented by patterns in the data, such as the half-sine, or 'bump', proposed by Anderson and colleagues (Figure 1c). Once a pattern $\mathbf{H}$ is expressed in the same sampling frequency as the recorded time-series, this template is cross-correlated to the signal of each channel c to reflect the similary of that channel at each time point t with the pattern (an example definition of $\mathbf{H}$ can be found in section 2.4). Each sample s is transformed to be the sum of the number of successive samples in the expected pattern width and their product with their corresponding time points of the pattern:

$$h_{tc} = \sum_{i=0}^{v-1} s_{t+i,c}\mathbf{H}_i \tag{1}$$

After this transformation, the sample h_{tc} reflects the cross-correlation of the template with the signal at time t for a given channel c (Figure 1f). This results in a measure of similarity between the template $\mathbf{H}$ and the sensor data s_{tc}.

In the case of several sensors that record brain activity, an HMP event is defined as a linear combination of the C channels. The magnitude of the contribution, ω, is defined as

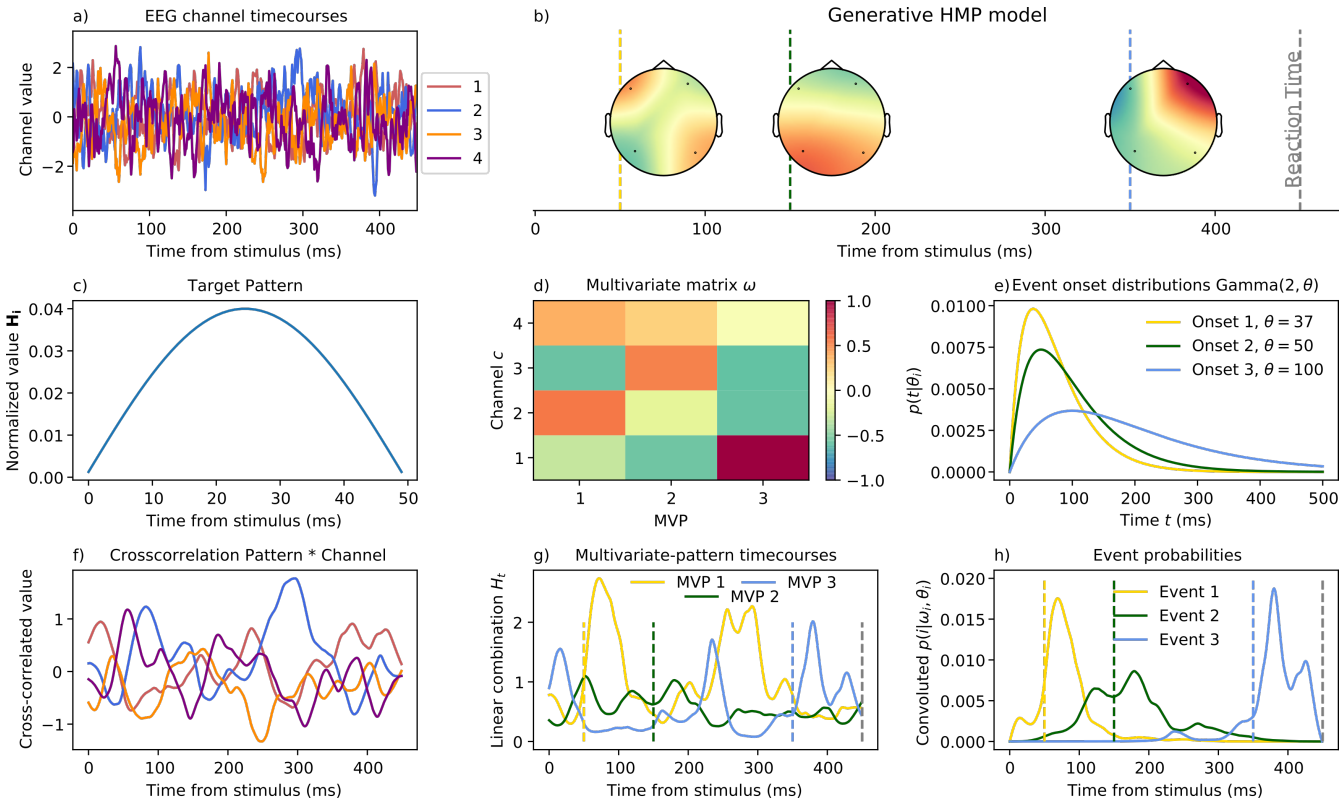

Figure 1

Example of the estimation for a three-event HMP model on simulated data for four EEG channels. Panel a) shows the simulated time-series, locked to stimulus onset time, for the four EEG channels on a trial with an RT of 450 ms. These time-series contain the three multivariate patterns depicted in b) whose onset are at times 50, 150 and 350. The pattern is a 10Hz half-sine represented in panel c), therefore it has a duration of 50 ms and peak amplitude of the patterns in panel b) are at times 75, 175 and 375. The contribution of each electrode to each of the three events are color-coded in panel b) and d). Panel e) shows the probability distributions of the onset of each event relative to the previous event (stimulus in the case of the first event). In order to estimate an HMP model, the channel data are first cross-correlated to the expected pattern resulting in the time-series displayed in f). These cross-correlated signals associated with the mixing matrix displayed in d) give the three multivariate patterns (MVP) time-series for each of the three expected events where higher amplitude indicates higher likelihood of the presence of the event. By imposing the distributions in e) and the assumption of sequentiality, we can estimate the probability of the onset of each event knowing its multi-variate representation and expected onset time, and conditioned to the probability of onsets of the preceding events. In HMP the parameters displayed in d) and e) are estimated from the data and the resulting probabilities in h) are used as proxy for single-trial event position.

a matrix of size $C \times I$ with one value for each sensor c and event i (Figure 1d). The congruence between the channels and the expected pattern i at time t is written as H_{ti} (Figure 1g):

$$H_{ti} = \exp\left(\sum_{k=1}^{C} h_{tc}(\omega_{ci} - \frac{\omega_{ci}^2}{2}\right) \tag{2}$$

The exponential ensures that we consider the absolute magnitudes of the correlation here (similar to Appendix 1 from Anderson et al., 2016). When $I > 1$, ω is a vector of size c for each event i.

2.1.2 *Event onset time*

In HMP, the onset time of event i depends on the onset time of the event $i-1$, expressing their expected dependency. This dependency will be expressed by any distribution on positive real numbers for which the mean onset time can be derived from one parameter (*e.g.,* a gamma distribution with a shape fixed to 2 as in Figure 1e). For the following, we write $p(t|\theta)$ as the probability density function of the onset time t given by a parameter θ, $g(t;\theta)$, divided by the sum of

its elements:

$$p(t|\theta) = \frac{g(t;\theta)}{\sum_{j=0}^{T} g(t_j;\theta)} \tag{3}$$

In the case of only one event ($I = 1$), the overall duration T of a trial (i.e., the RT) is partitioned into two times with durations t and $T - t$. Therefore, only one distribution of event onset needs to be estimated. The probability of the event onset E happening at time t, $p(E = t)$, is given by the likelihood of the expected event expressed by cross-correlation H_t across all sensors, modified by the probability that this event occurs at time $t - 1$ and the time of the response occurs after time $T - t$ given θ (e.g. the yellow curve in Figure 1h):

$$p(E = t|\omega, \theta) = H_t p(t - 1|\theta) p(T - t|T - \theta) \tag{4}$$

In the case of $I > 1$ one θ is estimated for each event. In this case however, one needs to rely on dynamical programming.

2.2 Estimation of an *I* event HMP

The number of free parameters for an I event HMP is $I \times C + I + 1$ with $I + 1$ onset time parameters θ and $I \times C$ magnitudes ω of the channel contributions. When $I > 1$, to compute the probabilities of onset time for event i when $i > 1$, one needs to integrate all possible time locations for the previous $i - 1$ events. Given the complexity of such computation the method relies on dynamic programming using the Baum-Welsh algorithm (Visser, 2011).

2.2.1 Inference

The estimation relies on an adapted version of the forward-backward algorithm usually used in hidden semi Markov models (see Yu, 2010, for a formal description of the algorithm). The forward variable α at a given time t tracks the probability of an event given past sequences from $1 : t$. The backward variable β accounts for the probabilities of coming observations given the presence of an event. Multiplied together, α and β give the likelihood of a an event i being located at time point t.

For an I event model, assuming that the parameters θ_i and ω_i are known, the forward variable α at time t for a given event i is:

$$\alpha_t(i) = \begin{cases} H_{t,i} p(t|\theta_i) & \text{for } i = 1 \\ (\alpha_{1:t}(i-1) * p(t|\theta_i))\, H_{t,i} & \text{for } 1 < i < I \end{cases} \tag{5}$$

In Equation 5, $(\alpha_{1:t}(i - 1) * p(t|\theta_i))$ is the convolution between the value of the forward variable for the previous event and the expected duration before the next event. This convolution allows us to account for the different time points at which the previous event could have happened and captures the core assumption of event sequentiality in an HMP model. This value is then multiplied with the exponentiated sum of the cross-correlation magnitudes (H_i) for the HMP signaling the next event.

The value of the backward variable β is:

$$\beta_t(i) = \begin{cases} p(t|\theta_i) & \text{for } i = I \\ (\beta_{T:t}(i+1) * p(t|\theta_i)) H_{t,i} & \text{for } 0 < i < I \end{cases} \tag{6}$$

The likelihood of a particular event E time location for time point t is then:

$$L[E_i = t|\theta_i, \omega_i] = \alpha_t(i)\beta_t(i) \tag{7}$$

Because of the cross-correlation of the signal to the pattern, one needs to ensure that the samples within a pattern cannot be associated with a new event. For this reason, in the estimation procedure of an HMP model, the samples within half the width of the pattern are censored in Equation 7. This imposes the constraint that the onset time of an event following another has a minimum duration of half of the expected pattern width.

From Equation 7, we can derive the probability of an event i at time t after event $i - 1$ by normalizing over all likelihoods across samples:

$$p(i, t) = \frac{L[E_i = t|\theta_i, \omega_i]}{\sum_{t=1}^{T} L[E_i = t|\theta_i, \omega_i]} \tag{8}$$

Figure 1 represents an example of the estimation steps (bottom row) on simulated EEG data for a three HMP model (top row), given the parametrization described in Section 2.4 (middle row).

2.2.2 Expectation maximization

In most cases, parameters θ_i and ω_i are unknown and need to be estimated from the data. For brevity, we write λ as being the set of the $I + 1$ scale parameters θ and $I \times C$ magnitudes ω. We estimate the joint vector of parameters $\lambda = (\theta, \omega)$ that best describes the data through expectation maximization (EM, Rabiner, 1989). We first make a proposal on the values of λ and compute the log-likelihood of the events summed over the sequence $1 : T$

$$LL[E_{1:T}|\lambda] = \log(\sum_{t}^{T} L[E_t|\lambda_1]) \tag{9}$$

The EM algorithm takes a first proposal to compute the event probabilities of Equation 8 and uses these to compute better parameters λ^*. In the case of the duration between two events, θ^* is computed from the probability-weighted average duration as given by the most likely position of each event, from which is subtracted the previous event position:

$$\theta_i^* = \begin{cases} f(\frac{\sum_1^T p(i,t)t}{T}) & \text{for } i = 1 \\ f(\frac{\sum_1^T p(i,t)t}{T} - \frac{\sum_1^T p(i-1,t)t}{T}) & \text{for } 1 < i \leq I \end{cases} \tag{10}$$

The function f to go from durations between events to scale parameter θ_i^* depends on the chosen distribution (see example for a gamma with shape 2 in Section 2.4).

The magnitudes of the channel contribution to the events are computed based on the average channels values at the event location times scaled by the probability of the event:

$$\omega_{ci}^* = \sum_{1}^{T} h_{ct} p(i,t) \tag{11}$$

Equation 9 is then computed on the basis of the new set of parameters λ^*. In our application, the EM algorithm stops as soon as the relative difference in log-likelihood between λ^* and λ increases less than a given tolerance ϵ:

$$\epsilon > \frac{(LL_{n+1} - LL_n)}{|LL_n|} \tag{12}$$

2.2.3 *Repeated measurement*

In the case of multiple trials, we assume that each trial is an independent new realization of the same event sequence. T is then the maximum duration of all fitted trials. The missing values in the sensor signals for trials shorter than T are imputed with a gain of 0 in Equation 1. EM Equations 10 and 11 are adapted to take the mean across all trials, and Equation 9 is then summed across all trials.

2.2.4 *From event probabilities to by-trial event times*

When fitting an HMP to a time-series, researchers will usually be mainly interested in the probabilities computed from Equation 8. These posterior distributions capture for each event the certainty about the time location of a event based on the estimated model. From there several strategies are possible to attribute a single time to each event at each trial. One can either use the weighted average as in Equation 10 or any other method (mean, max) to turn the event probabilities into single time points to use in a follow-up analysis. In the present manuscript we take the maximal probability of each event as the optimal estimate of the true time points at which the onset of the event is located. Once the most likely time point of each event at each trial has been selected, we compute the values of the channel at the peak of the expected pattern. Using this method allows to average trials and obtain time based topologies within the RT as plotted in the top row of Figure 1 and as illustrated in several following sections. These times can also be used as starting point for additional analyses, see for an example on simulated data Figure 2 and real data in Figure 6.

2.3 Finding the number of events

How many events are present in a task is often part of the scientific question, we derived two ways of estimating how many events are needed to describe the recorded time-series.

2.3.1 *Cumulative fit*

One first strategy to estimate the optimal number of events needed to account for the data, is to exploit the sensitivity of the EM algorithm to local maxima (Wu, 1983). That is, given an initial proposal of the time location parameter θ^*, the EM algorithm finds the nearest local maximum. In our application, this entails that a one-event HMP model with a location parameter initialized at the first sample ($\theta_1^* = 1$) will converge (within a certain tolerance ϵ) to the log-likelihood local maximum of the first event. In this procedure, the magnitudes are typically initialized at 0, although this is not strictly necessary. The first local maximum, being the first event, is thus parametrized by $\lambda_1 = (\theta_1, \omega_1)$.

After identifying the first event this way, we fit a two-event model in which the first event is initialized with the optimal parameters of the one-event HMP model (θ_1, ω_1), and the second event again with magnitudes of zero. The time sample following the most likely event time, as defined by Equation 8, is then used as proposal location parameter for event 2. This way, the EM algorithm will find the second local maximum. Next, a three-event HMP model is fit, initialized in the same way, and so on. This procedure continues until time sample T is used as proposal time location.

Once we cumulatively fit HMP to the data, a model with the n events obtained this way is fitted to the data, using the parameters associated with those events as initial proposals. This model fine tunes the location and magnitude of the optimal number of events that can be found in the data. Simulations (see Section 3.1) and application to real data (see Section 2.4) show that this sequential estimation procedure performs well in recovering the exact number, location, and magnitudes of underlying events. If necessary, this can be followed by an LOOCV procedure, for example to test which events generalize over subjects (Anderson et al., 2016; Van Maanen et al., 2021).

2.3.2 *Alternative fitting strategies*

The cumulative event fit method described above relies on the detection of local maxima where the EM algorithm converges. While this method allows to estimate an HMP model without specifying a number of events a priori, it does however make the assumption that the local maxima can systematically be found by cumulatively fitting HMP models. Alternatively, researchers can fit an HMP model with a pre-specified number of events to explore different event hypotheses. In this case however, one needs to ensure that the parameter space has been properly explored, for example by running several starting points for the parameters of HMP (as done with HsMM-MVPA by Portoles et al., 2022). Once an HMP with a given number of events has been estimated with enough certainty with respect to starting points, researchers can test for the generalization of each event by iteratively removing an event and testing the generalizability

of the resulting $n-1$ HMP with the LOOCV method (see Borst & Anderson, 2024, for a detailed explanation).

2.4 Example HMP parameterization

The general formulation described above ought to be parameterized according to a hypothesis about the pattern of the underlying events. As an example we show how the above derived equations relate to the HsMM-MVPA method described by Anderson et al. (2016). In HsMM-MVPA, the HMP was modelled as a 10Hz half-sine, leading to a "bump"-like pattern of 50 ms. To match the half sine-wave frequency f and the sampling frequency of the signal (f_s), we construct a vector $\mathbf{V}$ of length $v = \frac{f_s}{2f}$ with values $0, 1, \ldots, v-1$. Values within $\mathbf{V}$ are adjusted to the signal sampling frequency (adjusted to ms) by multiplying by $\frac{1000}{f_s}$ and adding $\frac{1000}{2f_s}$ to estimate the sine-wave value half-way between samples. The sine wave associated with f is then sampled accordingly:

$$\mathbf{B} = \sin(2\pi f \frac{\mathbf{V}}{1000})$$

The sum of its squared elements then normalizes the resulting vector:

$$\mathbf{H_i} = \frac{\mathbf{B_i}}{\sum_{j=0}^{v-1} \mathbf{B}_j^2}$$

$\mathbf{H}$ is then the normalized template of a half-sine wave with duration given by f (Figure 1c). In HsMM-MVPA the probability distribution is a gamma with a shape fixed to 2 and a scale estimated by the parameters. To resolve Equation 10 in the estimation procedure, θ is calculated from the average event time onset of a starting point or a previous iteration of the EM algorithm as $\frac{\bar{E}_i}{2}$, where $\bar{E}_i$ is the average time location for event i relative to the onset of the previous onset. Conversely, the average relative time location $\bar{E}_i$ in the case of a gamma with a shape of 2 is given by $2\theta_i$.

2.5 Simulations

To test properties of HMP with different parametrization, we simulate data with HMP's core assumptions of 1) a sequence of events with 2) a multivariate-pattern representation at the channel level and 3) a stochastic time between each of these events. All the simulations rely on the MNE Python package for EEG analysis (version 1.6.1; Gramfort et al., 2014). The simulated data were generated from MNE's sample participant channel positions, variance-covariance matrix and source-space.

2.5.1 Defining simulation parameters

2.5.1.1 Pattern simulation. In the simulations, we choose to implement HMP events as the activation of a source in the simulated participant's source space that yields a topology similar to the target pattern at the channel level. An HMP event is then simulated by the activation of a source following a pattern (half-sine) with a given width/frequency added to ongoing background oscillations. The activation is chosen with a certain amplitude expressed in log base 10 Ampère.

2.5.1.2 Between event duration simulation. These thereby defined events are then inserted at different times in each simulated trial following a chosen probability distribution for the inter-events durations. For each trial, the last simulated event defines the response event and therefore the simulated RT. The number of HMP events in each trial is the number of sources that were simulated in the RT time-series.

In each simulations we assume that all trials share the same number (and order) of events, a constant pattern and distribution for each event.

2.5.2 Demo

For the purpose of illustrating the simulation method we simulate one dataset with 200 trials. In this dataset, four events were generated (three hidden patterns with a width of 50ms and one response production event) with single-trial between event durations drawn from a gamma distribution with a shape of 2 and a scale of 50 (therefore an average between event-duration of 100ms and an average RT of 400ms). Figure 2a shows the average onset times and scalp topologies resulting from the simulated dataset.

The source amplitude at the peak of the half-sine was $10^{-6}A$ leading to single-trial electrode activity illustrated for three trials in Figure 2b. Because these activities are simulated we can display the underlying source activity projected to electrode space based on the participants' forward model without (solid lines) and with noise (transparent lines). The vertical lines in Figure 2b show the times at which the events were inserted for each trial.

Figure 2c shows the average stimulus centered ERP for three selected electrodes. As can be seen, the ERP displays the classical properties with an overall curve much more spread out and of lower amplitude than the single-trial events displayed in Figure 2b. As outlined in the introduction, this results from the between trial time jitter of the events (illustrated here by the small vertical lines at the top of the plot).

By fitting an HMP it is possible to recover that by-trial event time-jitter and for example to center the ERP to the detected single-trial onset times of each event. This procedure is illustrated in both the right panel of Figure 2c and in Figure 2d. In Figure 2c the signal between each event onset times, based on the times estimated by HMP, has been resampled to have the same duration as the average between-event duration. While this involves a strong data distortion, it does allow to represent the HMP centered ERP in the same timeframe as the classical stimulus-centered ERP. Figure 2d represent the average electrode activity around each event after

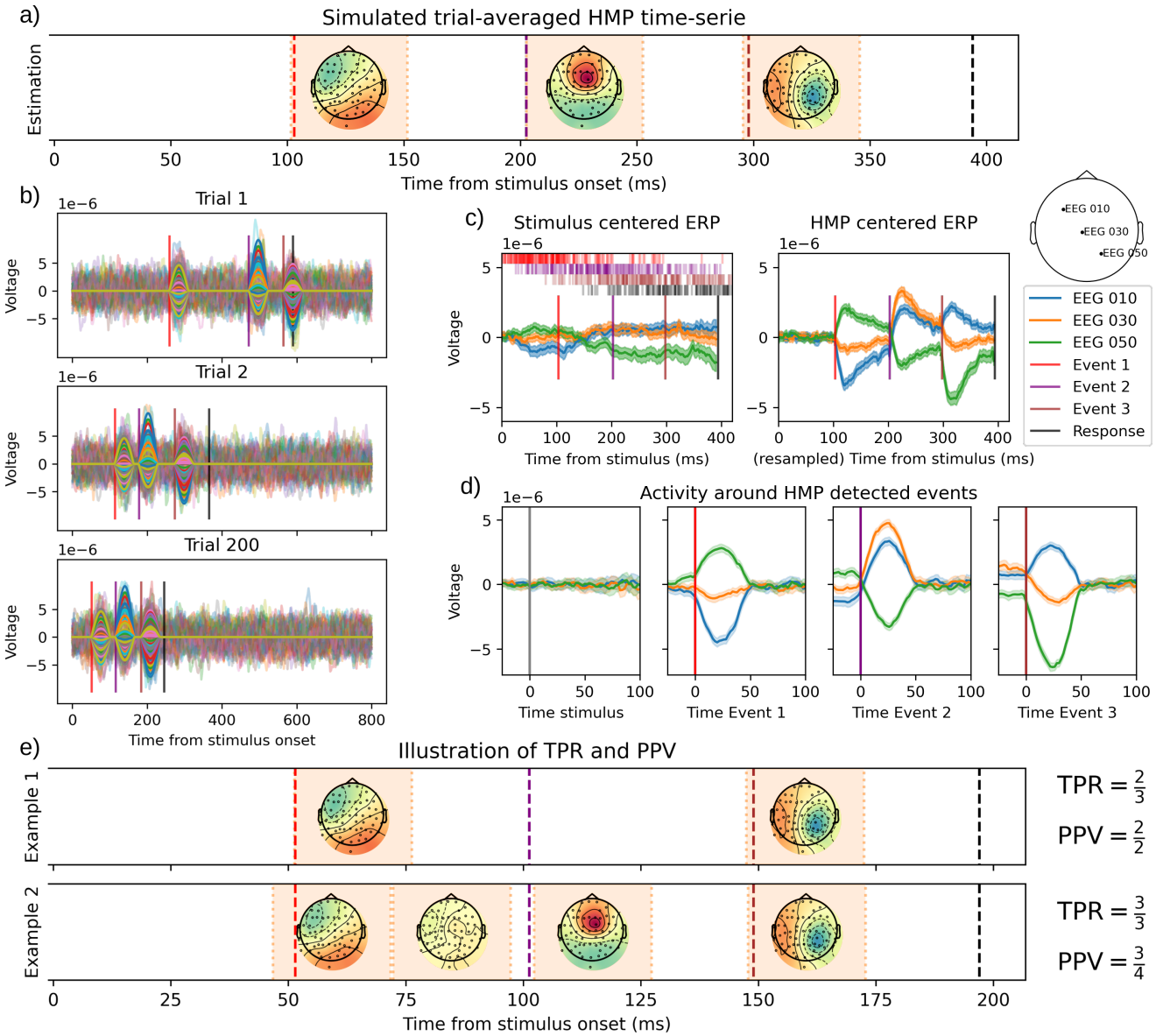

Figure 2

a) Average HMP time-serie representation of the simulated data for 200 trials. Each trial is composed of three events and a simulated response event per trial. The between event-duration has an average around 100ms. b) Illustration of the by-trial activities for the first two and last of the 200 simulated trials for all 59 electrodes. The vertical lines indicate at which times, in each trial, the 4 events (three source activity and a simulated time for the response in black) occurred. The single-trial event-related potential observed at the channel level is displayed transparently, the projection of the underlying source activity to the channel without noise is displayed as solid lines. c) Left panel, average ERP for three selected electrodes (with confidence interval drawn from 1000 bootstraps), transparent vertical lines show when in time each event for all 200 trials were simulated while vertical plain line represent the average onsets of the corresponding events. The right panel show the same representation except that each trial was centered around its estimated most likely onset time and the signal of each trial was resampled to match the average duration between each corresponding events. d) This alternative representation shows the average activity of the three selected electrodes in the vicinity (25ms before and 100 ms after) of each centering event, Stimulus, Event 1, Event 2 and Event 3. e) Two examples of estimated HMP models. For the purpose of illustration, these estimation were done with ill starting values to show the example of a missed event and an additional event detected by HMP. For Example 2, computing the electrode activity distances between the simulated events and the recovered ones suggest three correctly recovered events and one extra event resulting in a True positive rate of 1 (as all simulated events are found) and a positive predictive value of 0.75 as one of the events found doesn't match (or less so than the other) the simulated events.

the data has been centered on its generated onset time without resampling. This representation clearly illustrate that we can recover the underlying events provided that we have the correct timing of each event.

2.5.3 Simulation sensitivity measures

To understand under what circumstances HMP can successfully recover events in this simulation procedure, we rely on the match of the ground truth of each simulated dataset in terms of per-trial event times and topologies of each event with the estimated ones. To achieve this, we first fit an HMP model with a given target pattern width and an expected distribution shape, to each simulated dataset using the cumulative event fit described in Section 2.3. The events estimated this way are then compared to all simulated topologies by computing Euclidian distance between the activity in the channels (here principal components) at the true peak times and at the estimated peak times. We then categorized the events that are closest to a generating event as successfully recovered events (i.e., hits), and all remaining estimated events (e.g. with a higher distance to generating ones or closer to a null channel event) as false alarms. From this we can express the success of the recovery for each simulation by the true positive rate (TPR, *i.e.,* ratio of hits over number of true positives) and the positive predictive value (PPV, *i.e.,* ratio of hits over the number of events found). Figure 2e illustrates two hypothetical examples with either a lower TPR (top panel) or a lower PPV (bottom panel).

3 Results

To illustrate the usefulness of HMP models to analyze neural time-series we present results on both simulated and real EEG data. On the simulated data, we first show how the signal-to-noise ratio, the number of events and the chosen tolerance level (as used in Equation 12) influences the recovery of the generative HMP model. We then move to the model's parametrization by testing how the shape of the expected pattern and the chosen probability distribution changes the model recovery probability. In a final section we then apply HMP, based on the results from the simulations, to the EEG data from Boehm et al. (2014), previously analyzed by Van Maanen et al. (2021) using the original HsMM-MVPA method (Anderson et al., 2016). All presented regression models were estimated using the bambi (Capretto et al., 2022, version 0.13.0) Python package with default priors weakly informed by the data, 4 MCMC chains, and 2000 samples. The graphical summary of the regression model coefficients was done through the Arviz Python package (Kumar et al., 2019, version 0.17.0). Posterior distribution are summarized by the point estimate of the maximum a posteriori and 95% credible interval (CrI) computed based on the highest density interval Kruschke, 2010. All reported confidence interval through bootstraps were estimated either using the bias-corrected and accelerated bootstrap confidence interval from the scipy python package (Virtanen et al., 2020, version 1.12.0) or the seaborn python package (Waskom, 2021, version 0.13.2).

3.1 Simulations

All simulations share the same data-generating and preprocessing steps as described in Section 2.5. Each simulation consists of 500 simulated EEG datasets analyzed independently. One dataset consists of 100 trials across 59 EEG channels. For each of the simulations reported below, except in the first simulation we selected three as the number of simulated events to be found by HMP. The simulated events were defined with a width of 50ms (i.e. a 10Hz half-sine) except in the second simulation. The by-trial onset time of each event relative to the previous one is sampled from a gamma distribution with a shape of 2 and a scale of 100 (hence an average duration of 200 ms) for the first two simulations, and varying probability distribution and average durations in the last simulation. The specific parameters for each simulation are summarized in Table 1.

The simulated datasets were first epoched from 250 ms prior to the stimulus trigger to 5000ms after, then all samples after the RT of a given trial were removed. We then performed dimension reduction on the data by applying PCA on the average variance-covariance matrix of the electrodes (Cohen, 2014) and keeping the 5 first components (therefore applying HMP on 5 virtual channels).

The effect of the dimensions manipulated in each of the simulations below is assessed by fitting binomial logistic regression models to TPR and PPV (see Section 2.5.3 and Figure 2e). All coefficients of these regression models are expressed on the logit scale.

3.1.1 Simulation 1: Signal-to-noise ratio and number of events

In Simulation 1, we investigated the effects and interactions between different signal-to-noise ratios, numbers of to-be-estimated events, and the setting of the tolerance level for the EM algorithm. To this end, we manipulated the amplitude of the events by drawing a source activation from a log-uniform distribution between 10^{-8} and 10^{-6} ampère (A) for each of the 500 simulated datasets. The noise was generated using MNE's dedicated functions and drawn from a multivariate Gaussian with variances of 0.2 μV and covariances among electrodes given by the sample participant. The signal for a chosen source and amplitude varied depending on the sources location in the brain and the electrodes. To give a sense of the signal-to-noise ratio used, Figure 3a illustrates, for the three first sources displayed in 3b, the average activity for all electrodes for lowest, middle and highest values of the source amplitude levels: $10^{-8}A$ $10^{-7}A$ and $10^{-6}A$. These were chosen because they illustrate scenarios where

Simulation #	N events	Amplitude (A)	Event width	Distribution	Shape	Duration	SF
1 - SNR	1:10	-8:-6	50	gamma	2	200	100
2 - Event	3	-6	10:200	gamma	2	200	500
3 - Distribution	3	-7	50	variable	variable	1-500	100

Table 1

Summary of the simulation parameters for the three simulation sections with respect to source amplitude, number of events, event duration, by-trial peak time distribution, shape, mean duration and data sampling frequency. Amplitude is expressed as log_{10} Ampère, event width and mean duration to onset in ms. Sampling frequency is expressed in hertz (Hz).

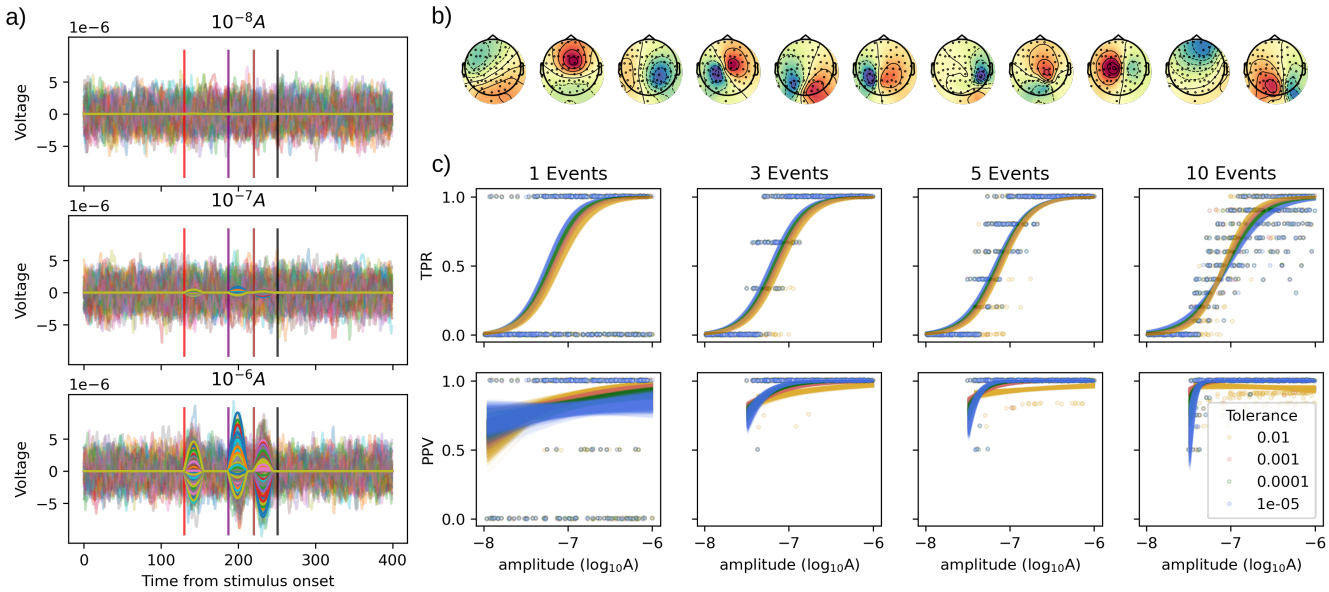

Figure 3

a) Illustration of the single-trial electrode activities yielded by three sources which onset were simulated at the time points indicated by the vertical lines. The top plot shows the single-trial event-related potential for all electrodes of a three-event simulated dataset at $10^{-8}A$, where the single-trial ERP is displayed transparently and the underlying source activity without noise as solid lines. The middle panel represent an amplitude of $10^{-7}A$ while the bottom plot show the event-related potential of a source amplitude of $10^{-6}A$. b) Topological representation of the 11 sources used in the simulation section 1 as events; points represents electrodes and color the sign and strength of topological contribution (red positive, blue negative). c) Results on the TPR (top row) and PPV (bottom row), the tolerance levels are indicated by colors, fitted lines represent 1000 regression lines drawn from the posterior distribution of the fitted binomial regression.

HMP still reliably recovers the simulated events with relatively high accuracy, even though the SNR is lower than typically found in M/EEG data. In Figure 3a, the transparent lines depict typical ERPs, while the solid lines ERPs based only on the source activity while leaving out the noise.

For each of the amplitudes we simulated a sequence of 1, 3, 5 or 10 HMP events (and a response production event), drawn from the source set displayed in 3b, and fitted HMP models using four different tolerance levels. Figure 3c shows the TPR and PPV that we obtained for each of these combinations, with two binomial regression models to illustrate the trend. Both regression models take signal amplitude, number of events and tolerance and their interactions as predictors with a logit link, and TPR and PPV, respectively, as dependent variables.

Given the coding applied to the predictors, the intercept of the binomial regression is the expected TPR or PPV, on the logit scale, for datasets with 3 simulated events, for the lowest tolerance, and for the middle of the signal amplitude range ($10^{-7}A$, as illustrated in the middle plot of Figure 3a). Slopes can be interpreted as the difference when increasing in one $\log_{10}$ unit in tolerance (e.g. from 10^{-3} to 10^{-4}), in signal's amplitude (e.g. from $10^{-7}A$ to $10^{-6}A$) or adding one event in the generated dataset.

For the TPR, the value of the intercept is equal to 1.230 (95% CrI = 1.079, 1.374) on the logit scale, meaning a TPR

of 77%. The slope of the amplitude is equal to 5.632 (95% CrI = 5.207, 6.048) which means that the predicted TPR at a signal amplitude of $10^{-6}A$, illustrated in the lower panel of Figure 3a, is higher than 99.99%. Decreasing the tolerance leads to an increase in TPR (0.193, CrI= [0.145, 0.248]), hence a stricter termination of the EM algorithm leads to increased detection rate. This effect of tolerance does however interact with the number of events (-0.106, CrI= [-0.135, -0.079]) meaning that a higher number of events is more easily recovered with a more liberal criterion on the EM. The number of events decreases the TPR (-0.179, CrI= [-0.206, -0.154]) but interacts with the amplitude of the signal (-0.323, CrI= [-0.396, -0.249]), meaning that to obtain a same TPR, a higher signal amplitude is required for a higher number of events. The interaction between signal amplitude and tolerance (0.078, CrI= [-0.072, 0.229]) is uncertain. The three-way interaction with the number of event (-0.041, CrI= [-0.051, -0.032]) does not change conclusions from the lower level interactions and simple effects.

With respect to the PPV, the value of the intercept is equal to 4.212 (CrI= [3.845, 4.614]), meaning a PPV of 99%. All probability mass of the posterior distributions for the slopes are positive, meaning that increasing tolerance, the number of events, or the signal's amplitude increases the PPV (see Appendix A for the full table of coefficients).

To sum-up, Simulation 1 showed that HMP is able to recover the ground truth of a sequence of events, with few false-alarms, even in a time-series with very low signal-to-noise ratio. If a high number of events is expected (e.g. tasks with long RTs), this signal-to-noise ratio should be higher for HMP to perform well. Interestingly, the TPR is lower for a 1 event generated dataset. The reason for this harder recovery is linked to the fact that multiple events aid at precisely locating the individual events as the other events constrain the parameter space. A stricter tolerance for the EM increases the likelihood of finding the underlying events although a very high number of events benefits from a more liberal tolerance level.

3.1.2 Simulation 2: Event width

In theory, any pattern defined in the sampling frequency of the data can be used for HMP analysis (see Section 2.1.1). In the case of EEG or MEG however, sine or half-sine waves are the expected patterns. As sine-waves can be decomposed into half-sines, these appear to be the minimal unit for an event in E/MEG. The expected duration of those half-sine templates for cognitive events is however unknown. In E/MEG, plausible durations include theta and alpha-band frequencies (Makeig, 2002; Yamanaka & Yamamoto, 2010). In the HsMM-MVPA method, Anderson et al. (2016) used a 10Hz, hence 50ms duration half-sine. In Simulation 2, we vary the simulated events by simulating half-sines from a continuous range of 10 to 200ms and try to recover the ground truth with target pattern durations of 25, 50, 75, and 100 ms (Figure 4a). Based on Simulation 1 we generated 3 events with a signal amplitude of $10^{-6}A$, the aim being to have a high recovery when the expected duration matches the generating event duration. For the purpose of illustration we show in Figure 4b the resulting HMP solutions for data simulated with a 50ms half-sine and with target pattern widths of 100, 75, 50 and 25 ms. It appears that there is a certain flexibility with respect to the difference in expected vs generated event duration. All four solutions capture the generated 50 ms events, showing a high TPR regardless of the difference between expected and generated event durations. However on the PPV, expecting shorter events than the generated ones results in an increase in the number of events detected by HMP (e.g. expecting 25 ms event duration as in the bottom of Figure 4b results in finding 2 or 3 events per generated event). Both of these observations are confirmed across all simulations in Figure 4c.

Similar to Simulation 1, we model the TPR and the PPV by a binomial regression with a logit link using the generated duration, the target pattern duration and their interaction as predictors. In this section however, the data shows a strong deviation from linearity as raw data alone strongly suggests that recovering shorter durations than the one expected is nearly impossible while recovering longer durations from shorter expected duration seems easy for HMP. Therefore we first convert the generated duration to frequencies and use a cubic polyniomal regression on the generated durations (the widely applicable information criterion, WAIC, gives the cubic model as preferred compared to a linear or squared model). The raw data and the fitted lines are represented in Figure 4c.

For this simulation, the interpretation of the coefficients is complicated by the tranformation to frequency domain, the cubic terms, and their interactions with the predictor of the expected duration. A full description of the curves through their coefficients is beyond the scope of this paper (but see Appendix A for the full coefficient table); therefore we rely on the predictions drawn from the posterior distribution of the binomial regressions. On the TPR, averaging over all generated durations, the highest TPR is achieved with the shortest target pattern of 25 ms (84%), the TPR decreases for longer target patterns (e.g. 38% at 50 ms and to 12% at 100 ms). When taking into account the duration of the generated event, we observe that the maximum TPR is achieved close to each of the target pattern width, 25 (TPR = 95% at a generated duration of 39 ms), 50 (TPR = 89% at a generated duration of 42 ms), 75 (TPR = 86% at a generated duration of 66 ms) and 100 ms (TPR = 72% at a generated duration of 95 ms).

When factoring in the generated event width we see that all tested duration perform over 50% on the TPR when the width of the generated pattern exceeds 58ms. However the

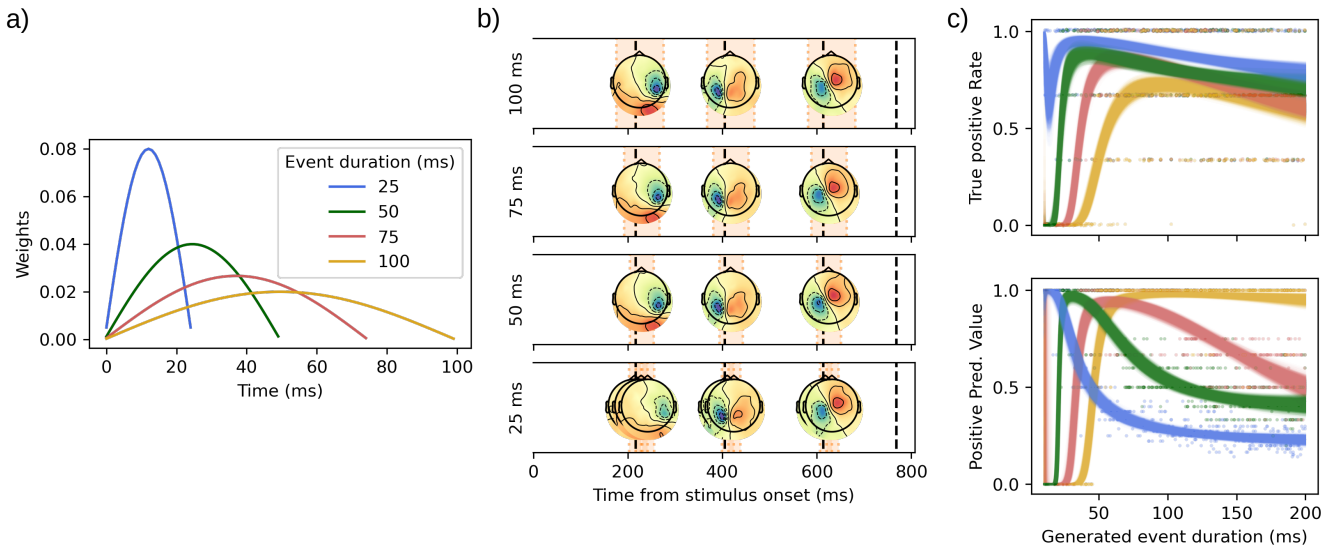

Figure 4

a) Representation of the four event widths used as target pattern in HMP. b) example of solutions found when generating from a 50 ms event width for a 100ms, 75 ms, 50 ms and 25 ms target pattern. c) Results from the 500 simulations for true-positive rate (top) and positive predictive value (bottom), colored lines are drawn from the predictions of 1000 samples from the posterior distribution of the binomial regression for the different expected event width.

shortest generated pattern at which this 50% criterion is achieved varies, with 58ms for a test pattern of 100ms, 37ms for 75 ms, 23ms for 50, and 10ms for 25ms. As suggested by Figure 4b, the PPV is more discriminant on the congruence between target pattern width and generated event duration. Using a target pattern width of 25 ms, a PPV higher than 50% is only observed for event width generated between 10 and 44 ms. For 50ms test duration this range increases from 21 to 110ms, 34 to 200 for 75ms test pattern and 47 to 200 for a 100 ms expected duration.

To sum-up, without knowing the underlying duration of the events, expecting short event widths in HMP yields the best recovery performance, however, at the cost of increasing false positives. If researchers have expectations regarding duration of the events, then the best performance in terms of PPV will be achieved by aiming around the underlying true duration, TPR however will always be high as long as the expected duration is higher than the true underlying one.

3.1.3 Simulation 3: Probability distributions

For Simulation 3 we manipulate the generative probability distribution as well as the average duration between events in the case of three sequential events. Because of the convolutions in Equation 5 and 6, if the events have a strong signal to noise ratio, the weight of the distribution will not strongly influence the result as long as credible values are not excluded (e.g., non-centered positive defined distributions). However, the expected distribution matters when signal-to-noise ratio is low. Therefore, we choose to simulate these datasets with a source amplitude of $10^{-7}A$ (middle panel of Figure 3a). Simulation 1 showed that this level of signal generated a TPR of around 77%.

We simulated event times from the Weibull and the Wald distributions as those are often associated with reaction times (De Boeck & Jeon, 2019; Wagenmakers & Brown, 2007), which are directly related to event times. Additionally we also simulated from a gamma distribution as this was the original distribution used in (Anderson et al., 2016). For each of these distributions we selected 3 different shape values (Figure 5a). The choice of the shape values was made to represent a variety of moments and probability density spreads. For each of the 500 datasets, we first draw four average event times, for the three events and the response event, from a uniform distribution between 1 and 500 ms. We then simulate 100 trial event relative onset times from each of the 9 distributions with the drawn average between event duration (three families x three shapes). For a given dataset, we then try to recover the generative events using 9 HMP models all initialized with the selected distributions. As before we measured TPR and PPV to test the performance of HMP.

As first observation, Figure 5b shows the TPR (color scale) and PPV (dot size) when data were generated with a given distribution (bottom row) and recovered with another or the same distribution (left most column). From this plot is can be seen that HMP performs best when the expected distribution corresponds to the generative one as shown by

the higher TPR and PPV along the diagonal. It is also clear that the Wald and Gamma distributions seem to be most effective at recovering events generated with different distributions, and the Weibull distribution with a shape of 4 the worst.

In real data we do not know what generative distribution is underlying neural-time series in RT tasks. We therefore collapse HMP performance across all generative distributions and only look at the performances as a function of the expected distribution for HMP. As before, we use two binomial regression for both TPR and PPV, with the average reaction time (the sum of all distances between events), the expected distribution and the interaction between RT and distribution. The predictor of the average RT was centered and distributions were treated as a categorical factor with gamma 2 as a reference.

For the TPR, the intercept (1.110, CrI = [1.072, 1.151]) shows that that the recovery of the true underlying events at the average RT for a gamma 2 is around 75%. This is comparable to the result we obtained in Simulation 1 with the same signal amplitude and distribution. When comparing the gamma 2 to other distributions, only the gamma with a shape of 4 (0.082, CrI = [0.030, 0.142]) and a shape of 6 (0.079, CrI = [0.019, 0.133]) have evidence towards a higher TPR. For a worse performance on TPR some evidence is found for Wald with a shape of 1 (-0.136, CrI = [-0.192, -0.084]), Weibull with a shape of 3 (-0.095, CrI = [-0.154, -0.043]) and 4 (-0.514, CrI = [-0.566, -0.460]). The effect of the other three distributions do not provide enough evidence to conclude on a better or worse performance. The average RT does not seem to have any impact on the likelihood of recovering the true sequence for a gamma-2 (-0.015, CrI = [-0.151, 0.119]). On the interaction between distribution and average RT, only evidence for a positive relationship is found for Weibull-3 (0.310, CrI = [0.133, 0.516]) and Weibull-4 (0.575, CrI = [0.391, 0.755]).

The average PPV for a gamma-2 is around 90% (2.049, CrI = [1.992, 2.107]). All tested distributions show evidence toward a lower performance than the gamma-2 with the exception of the Wald-1, which does not show evidence toward a better or worse performance (0.048, CrI = [-0.041, 0.126]). Contrary to the TPR we clearly observe a positive relationship between RT and PPV (-0.675, CrI = [-0.874, -0.460]) for the gamma-2. However, the interaction between distribution and average RT show that this negative relation does not hold, or less so, for the Weibull-2 (0.305, CrI = [0.018, 0.584]), Weibull-3 (0.673, CrI = [0.420, 0.957]), Weibull-4 (0.634, CrI = [0.369, 0.889]), Wald-0.25 (0.624, CrI = [0.364, 0.898]), and the Wald-0.5 (0.512, CrI = [0.238, 0.814]). The negative relationship of RT on PPV is however even more negative for the gamma-4 (-0.431, CrI = [-0.505, -0.354]) and the gamma-6 (-0.770, CrI = [-0.842, -0.699]).

In summary, compared to the gamma with a shape of 2 chosen by Anderson et al. (2016) only a gamma with a shape of 4 or 6 appears to have a slightly higher TPR. However, these latter distributions also have one of the lowest PPV. Based on these results the original gamma 2 does seem like the best option, as it strikes the optimal balance between identifying events while not overestimating the number of events. In general, the likelihood of recovering the sequence of events does not appear to be influenced by the average RT duration, although the PPV decreases for higher average RTs. This means that when analyzing time series with long RTs (e.g., > 1 second) researchers can expect to have some false alarms when fitting an HMP model. This can potentially be solved by applying LOOCV afterwards.

3.2 Example application: HMP and decision-making

To illustrate the usefulness of HMP on real data we chose to replicate the study by Van Maanen et al. (2021). In this task, participants were asked to indicate whether the majority of points in a cloud of moving dots (random dot kinematogram) was moving to the left or or to right with the corresponding hand. The instruction manipulated the speed accuracy trade-off by asking the participants to either favor the speed or the accuracy of their response (see Boehm et al., 2014, for further details on the experimental protocol).

Using a LOOCV procedure on a varying number of events with HsMM-MVPA, Van Maanen et al. (2021) found that a two-event model fit the speed instruction data best while a three events model fit the accuracy instructions best. The authors interpreted this as evidence that an additional event was present in the accuracy condition. The original study used the HsMM-MVPA method described in Section 2.4 but on 100 Hz downsampled EEG data. Here, we attempt to replicate their findings using HMP with the sampling frequency used at the data recording (500 Hz)[1]. For the parametrization we chose to do a direct replication by choosing pattern width of 50 ms, a gamma distribution with a shape of 2 and a minimum inter-event duration of 50 ms, similar to the original study.

3.2.1 Data

The data provided by Boehm et al. (2014) was already artifact-rejected and epoched. To conform with the study of Van Maanen et al. (2021) we applied an offline bandwidth filter between 1 and 35Hz (one-pass, zero-phase, non-causal bandpass IIR filter, -6 dB cutoff frequency at 0.50 Hz and 39.38 Hz) using MNE. Noisy electrodes were marked as bad and interpolated after the independent component analysis (ICA). Electrodes were re-referenced to the average of all EEG electrodes. We then applied an ICA to repair eye-blink

[1]The solution resulting from HMP is not sensitive to the sampling rate in this case; downsampling to 100 Hz gives highly similar estimations.

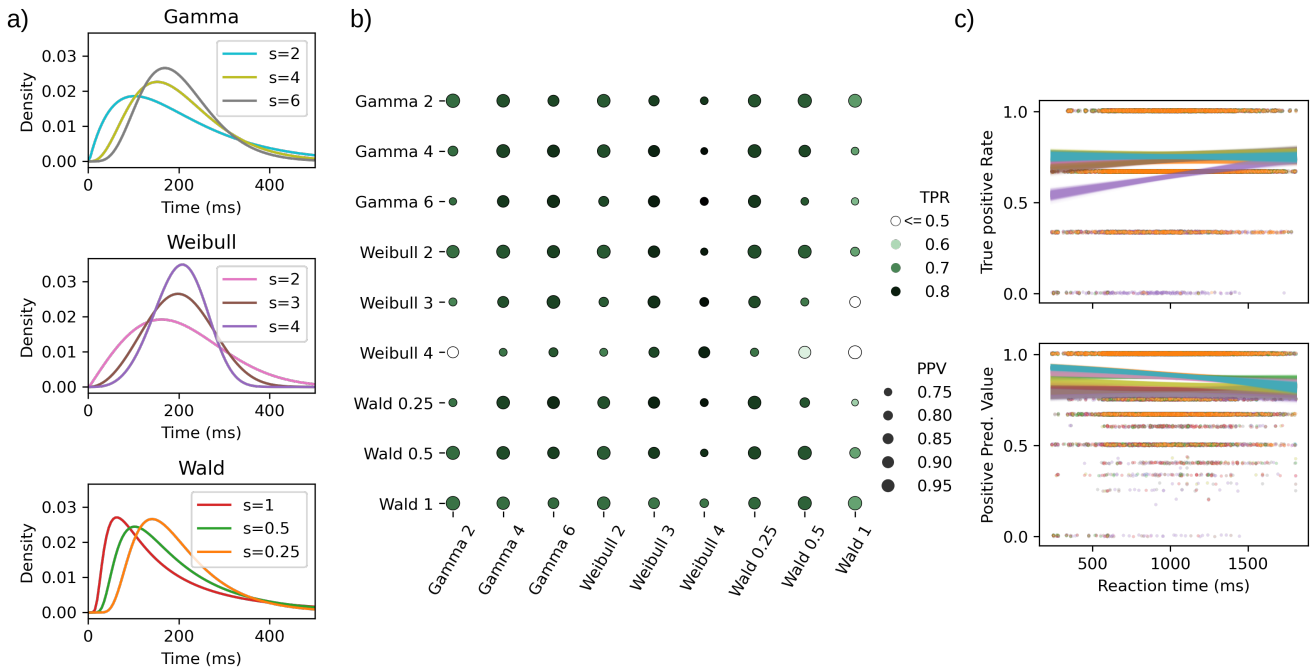

Figure 5

a) Representation of the different distribution used in the simulations, all have a mean of 200 ms for illustration purposes. b) Recovery matrix, the columns show the generative distribution while the row show the distribution used in HMP. TPR is represented by a color scale while PPV by the size of the dots. c) TPR (top) and PPV (bottom) for the 500 datasets with the different average RTs on the x-axis (ms) and colored by distribution used in HMP. Lines show the predicted values for 1000 draws from the posterior of each binomial regression.

artifacts and used auto-reject (v0.4.2) to automatically interpolate epoch-wise bad channels or reject epochs (Jas et al., 2017).

The data are composed of 30 electrodes with 4547 trials for 25 participants, an average of 182 trials per participant. Most rejected trials were rejected based on autoreject analysis, a minor part of trials (< 1%) had too short (<300 ms) or too long RTs (> 3000 ms). Before the fit, the data was truncated to the duration (RT) of each trial. As Van Maanen et al. (2021), a spatial PCA was then applied to the average variance-covariance matrix of all trials, from which we selected the first 10 principal components, explaining more than 99% of the variance.

3.2.2 Results

We fit two separate HMPs to the two speed accuracy trade-off conditions in the experiment. Figure 6a illustrates the resulting solution on the average RT, the times and topological maps were computed based on the by-trial maximum probability for each event as described in Section 2.2.4. After describing the solutions and their ERPs, we analyze the by-trial duration of each stage and the effect of speed accuracy trade-off on each.

3.2.2.1 HMP solutions. As can be seen in Figure 6a, in contrast with the original study, HMP detected 3 events in the speed condition and 4 events in the accuracy condition. Hence, while we replicate the observation of an additional event in accuracy, it does seem like the implementation of HMP is more sensitive to the template than the HsMM-MVPA method as we also find an event preceding the production of a response.

Using HMP changes the possible physiological analysis that can be performed. As an example, the left panel of Figure 6b illustrates the averaged signal of the CPz electrode when time-locked to the stimulus onset. The right panel of Figure 6b represents the same data but with values between each HMP event resampled to the average time between the corresponding events (as in Section 2.5). The comparison of this representation to the standard stimulus-locked ERP in the left panel clearly shows that modeling single-trial event locations resolves the smearage induced by time-jitter of underlying events (see also Berberyan et al., 2021). Especially the additional event detected by HMP in accuracy is extremely visible with the HMP-ERP representation compared to the traditional ERP. Figure 6c shows the same data centered around the HMP events but without resampling. In this

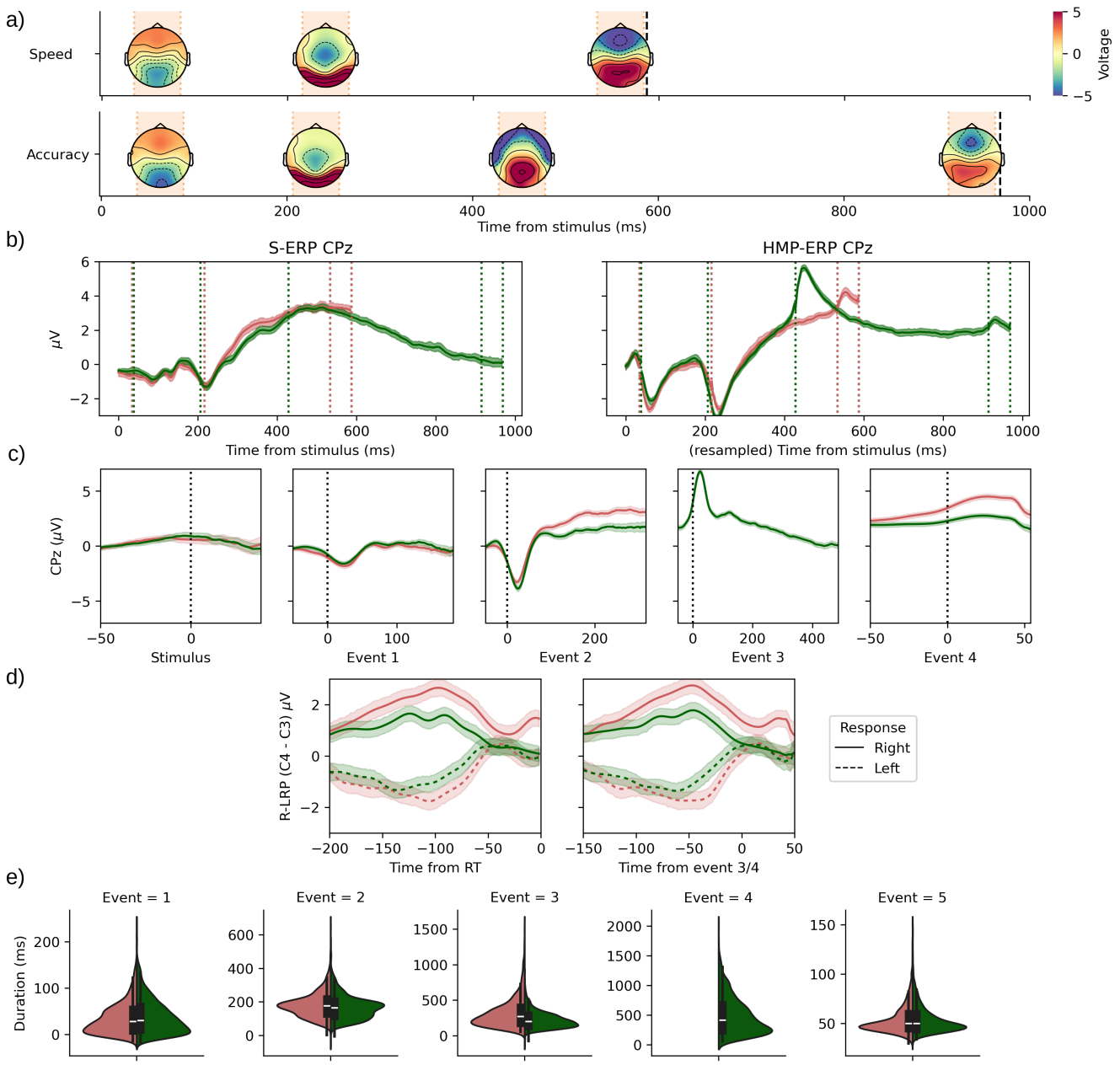

Figure 6

a) Representation of the solutions we find using HMP for for speed (top) and accuracy (bottom) condition. The circles are average topographic maps of the by-trial activity at the likeliest time location of each event. The red rectangle around each head represents the start of the event to the expected duration (50 ms in this application). The color scale represent contribution in μV. b) Left panel: Average event-related potential on the CPz electrode centered on stimulus onset, dotted lines represent the average onset of the different events in each condition. Right panel: Same data but the by-trial values between the events detected by HMP have been resampled to be equal to the average duration between events. For both panels the confidence intervals represent 1000 bootstraps. c) Average ERP for the CPz electrode centered around the detected events with 50 ms prior to the event onset and displayed up to the average duration before the onset of the following event. d) Lateralized readiness potential (defined as the difference between electrodes C3/C4), average activity preceding either the Reaction time (left panel) or the last HMP event (right panel). e) Illustration of the distribution in each condition for the different time between event onset (milliseconds). The first panel represent the time from the stimulus onset to the first event, the last panel the time from the last event to the response.

case also the data representation changes drastically compared to the classical ERP.

Based on both the topologies in Figure6a and the electrophysiological profile in Figure6b, it appears that the additional event in the accuracy condition is a late P3 following an early P3. The functional interpretation of this decomposition is still under debate but among the proposals Polich (2007) suggest that the dissociation between P3 early and late is generated when a perceptual stimulus discrimination occurs. Together with the interpretation of this extra-event being the onset of a confirmation stage by Van Maanen et al. (2021), this does suggest that there might be a second controlled decision after a first, fast, stimulus discrimination shared between speed and accuracy conditions. A full analysis of the resulting ERPs and the speed accuracy trade-off is beyond the scope of the current paper, but this does illustrate how an HMP decomposition is interesting for analyzing the component composition and change in associated ERPs.

As a comparison to a well-defined ERP we consider the Lateralized Readiness Potential (LRP), an electrical potential that develops just prior to the response in the ipsi- and contra-lateral motor cortices (Coles, 1989). This component is typically present when centering on the RT but mostly absent from the stimulus centered averaged ERP (Burle et al., 2004). If HMP decomposes the RT well, we would expect the LRP to be visible when the signal is centered on the event that just precedes the response. Figure 6d shows the comparison between the LRP centered on RT or on the last event of HMP. Both estimations are extremely similar despite HMP not centering on the moment of the response but on the last event before the response. To analyze times between the onset of this component and the response time, researchers have typically relied on arbitrary rules (e.g. a μV threshold, J. Miller et al., 1998) on the averaged LRP waveform. With HMP, the onset of this last event is given by-trial and can be readily used to test hypotheses such as the time between this last event and the recording of a response.

3.2.2.2 By-trial time analysis. Apart from event-onset times, it is also possible to analyze the relative onset of events by taking the difference, for each trial between the most likely time locations of two consecutive events (see Figure 6e). In the current application, the two conditions differ by the insertion of an additional event. To easily analyze the difference in event durations we take the time between the events 2 and 4 in the accuracy condition, hence ignoring the 3rd event in this condition.To draw statistical inferences from the time difference between event onsets we fit a linear mixed model on the duration in milliseconds between events without intercept and with factors event number (categorical) and their interaction with speed-accuracy conditions as predictor. This mixed effect model had the maximal random-effect structure possible given the fixed effects (see Appendix B for the full coefficient table).

The posterior distribution of the effect of SAT on the time preceding event 1 shows that there is no evidence for a decrease in duration with speed instructions (-3.92, CrI = [-11.42, 3.15]). This effect is different for the time between event 1 and 2 (13.63, CrI = [5.01, 22.72]) where substantial evidence is found in favor of a positive SAT effect on the duration between those events. The time between event 2 and 3 displays the highest SAT modification of the effect (-387.83, CrI = [-447.94, -328.39] suggesting that most of the speed up in the RT when asking participant to speed up is contained in this interval. Finally the time between the last event and the time of a response doesn't display any evidence towards an effect of SAT (-0.05, CrI = [-7.59, 7.13). The observation that one inter-event duration drives most of the SAT effect was already reported by Van Maanen et al. (2021), in our modelisation we observe a weak positive effect of SAT instruction on the time between event 1 and 2 (also observable in 6b). This analysis further shows how the time between events detected by HMP can be used to test for differences in conditions.

4 Discussion

In this paper, we introduced HMP, a method intended to detect sequential events in neural time-series.We believe that this approach is a significant step forward to resolving distortions related to averaging in neural time-series (see Luck, 2005; Mouraux & Iannetti, 2008, and Section 2.5). In simulation, we showed the method's ability to recover cognitive processing events under various levels of signal strength and different parameter settings of the optimization method we used. Moreover, we showed that HMP is robust to violations of assumptions about the data-generating process. That is, HMP still recovers the correct cognitive events if inappropriate assumptions are made regarding the timing, target pattern and distributional shape of the data-generating events.

In addition to simulation studies to show the robustness of the method, we also illustrated how one can use the onset of an event in an experimental context. The last Results section shows how one can use the onset of the events to center on particular events and electrodes of interest and how between-event durations derived from event onsets can be used to test hypotheses about the impact of an experimental manipulation on a given temporal event in the brain. The powerful nature of HMP lies in the estimation of by-trial event onsets. Thanks to this property, in this study we could center the ERPs on the corresponding event onsets and analyze between event durations with a linear mixed model at the trial level (hence accounting for missing trials by e.g. EEG exclusion and the full random effect structure). However, this is barely scratching the surface of the possible applications. As an example, Van Maanen et al. (2021) successfully fitted an evidence accumulation model to the by-trial times extracted using HsMM-MVPA. As another example, with these by-trial

onsets it is also possible to use single-trial predictors (e.g. a wide range of continuous stimulus intensity) to estimate the effect of that manipulation.

Note that contrary to most classical methods from mental chronometry (Brown & Heathcote, 2005; Donders, 1868; Ratcliff, 1978; Stone, 1960), HMP does not imply seriality in cognitive processes. This means that a particular process is not expected to end at the start of the following event. Instead, the time between two events at best predicts the time needed for the following cognitive process to start. In other words, common assumptions about parallel processing or co-activation (McClelland, 1979; J. O. Miller, 1993; Townsend & Nozawa, 1995) are perfectly compatible with HMP as long as the onsets of the events follow each other. For example in decision making, the minimal assumption is that a decision follows sensory encoding of the stimulus and precedes the response execution. There is no strict need to assume that a decision process cannot continue during response execution, compatible with observations from a growing number of studies (Buc Calderon et al., 2017; Selen et al., 2012; Servant et al., 2021; Thura et al., 2022; Weindel et al., 2022).

The assumption and development of HMP has been driven by work on mental chronometry and cognitive modelling but it is important to stress that the benefits of HMP go beyond these application domains. Indeed, many applications in cognitive science suffer from jitter in the time of the underlying cognitive-related events in EEG, MEG, and other related measures. By explicitly making the assumption of sequential events in the brain following estimated probability distributions, HMP is able to recover the underlying sequence. This allows researchers to center the signal around the event of interest and perform any type of analysis, whether it is about connectivity (Portoles et al., 2022), time or voltage as in the present example application, spectral composition, or otherwise.

The question of the application of HMP is then reduced to its core assumption: is the sequential nature of processing events reasonable given the task of the participants? If so, and if the data have a proper signal-to-noise ratio similar, if not lower, to one would require for traditional ERPs (see Section 3.1), HMP will work well. Of course, the ancillary assumptions of probability distribution and pattern width expectation explored in Section 3.1 can have an impact on the successful identification of the cognitive events. In particular, violation of the expected pattern duration can have consequences: If (and only if) the actual underlying pattern is shorter than the one that is expected, the TPR decreases, meaning that not all events are identified. We believe that the initial duration of 50 ms chosen by Anderson et al. (2016) is reasonable given the other studies that successfully applied it. Nevertheless, the width of a pattern can change to quite some extent depending on the cells and neurons actually recorded (e.g. MEG vs EEG vs intracranial electrodes). This pattern can also change depending on the stimuli used, for example spatial frequency of visual gratings have been shown to have a broad impact on scalp topologies as seen using EEG (Kenemans et al., 2000). This latter point invites caution when assuming a pattern width even within a given measure for which the usual pattern width work well.

To be sure, running HMP with the minimum possible duration given the sampling frequency will prevent missing events in the data. Researchers could then balance the trade-off between hits and misses of events either by increasing the width of the searched pattern or by relying on alternative, second-step, procedures such as LOOCV (see for example Borst & Anderson, 2024, in the context of HsMM-MVPA). Such an approach would reveal whether the model overfits, and therefore identifies more events than are present in the data.

A final word on the conditions to use HMP is the RT. the method assumes that the RT is informative about the end of the sequence and act as a constraint. Hence, a well defined time for the end of the sequence is a necessity to apply HMP.

4.1 Comparison to other event identification methods

Discovering recurring patterns in the EEG is not a new idea. Lehmann (1971) showed how EEG time-series can be summarized in a small number of so-called micro-states, i.e. quasi-stable states of electrical potential across electrodes, typically 80-150 ms (Brunet et al., 2011; Koenig et al., 2002; Michel & Koenig, 2018; Van de Ville et al., 2010). Each microstate is therefore defined by its EEG scalp topography. In the past, four canonical topographies have been identified, under which most microstates can be classified (Michel & Koenig, 2018). Microstate topographies and their distributions are usually identified based on group-averaged data, after which the samples in the original EEG are labeled accordingly.

This highlights the main difference with HMP, which identifies events on a by-trial basis, and then averages the results (e.g., Figure 6). This is crucial, as averaged data, particularly of longer trials and samples further away from fixed time points, might obscure important events (Borst & Anderson, 2024; Luck, 2005). This is, for example, clearly visible in Figure 2c, where distinct trial-by-trial events result in very weak averaged ERPs. We hypothesize that the events that HMP detects might be the onsets of the microstates found in averaged data (which is the reason they are quasi-stable for some time). Although we think that the by-trial estimation of HMP is an advantage for task-based data, so far HMP cannot be used to analyze resting-state data – an area in which the microstate analysis has been particularly successful (Michel & Koenig, 2018).

Another commonly used method that can be linked to the results of HMP is hidden Markov modeling. Hidden Markov models start to be widely applied in the fields of neuroscience

including on EEG and related measures (Kirchherr et al., 2023; Masaracchia et al., 2023; Quinn et al., 2018). Contrary to HMP, hidden Markov models assume a fixed number of states and generative distributions (e.g. for the electrodes) specific for the states. Despite their resemblance, HMP is not a specification of a hidden Markov model. Their common ground is that both can be described as transitions between states. However, in HMP these transitions are signaled by short-lived multi-variate patterns and can only happen from one state to the following one. Additionally, hidden Markov models can transition to any state but assume that the state's data generating distribution is present during the whole activation of the state.

Although HMMs can be very useful in the context of cognitive neuroscience, their difficult estimation and parametrization can lead to volatility in the estimation and quick erroneous switches between states (Masaracchia et al., 2023). For this reason, we believe that HMP, as long as the assumption of sequential events is met, provides an easier and more robust estimation of a cognitive event sequence in the context of the timing of these events. Moreover, the EM algorithm used also allows to easily fit conditions or participants with lower number of trials. Indeed, in the case of individual participant for example, by providing starting points that are estimated on the group level, it is possible to get reliable estimates even with a low amount of trials. This could be very useful for clinical settings where only a few data points per participant are available but enough participants have been recorded and analyzed to allow for a restricted search in the parameter space.

Other methods have been used to specifically obtain single-trial measurement of EEG or MEG components to achieve RT decomposition. For example, Ruchkin and Glaser (1978, as cited by Smulders et al., 1994) looked for a pattern of a half-sine with a duration varying from 250 to 970 ms to achieve P3 single-trial measurements. For the N2 also single-trial measurement has been proposed by Nunez et al. (2019) this time based on a singular value decomposition of the average ERP. In the case of the detection of a single specific event, such single components methods could outperform HMP as they benefit from looking at known exact properties from a specific component.

4.2 Implementation

A final note on the implementation of HMP is in order. We created an open-source and collaborative Python package (https://github.com/GWeindel/hmp). All results and plots presented in this study are based on this package (v.0.3.1). The repository contains tutorials on the data format, the fit of HMP models and the comparison of experimental conditions. This paper and the online repository should make the method more accessible for a broad range of interested researchers. In terms of computational complexity, fitting an HMP model can be quite expensive in terms of processing time and RAM. Dimension reduction through PCA and resampling have no dire consequences as long as the range of data reduction is in line with the desired accuracy. For example, the TPR in Simulation 1 was comparable to that of Simulation 2, even though the data was sampled at 100Hz in Simulation 1 (vs 500Hz in Simulation 2).

In terms of the number of trials required to fit an HMP model, a precise definition lacks. The optimal number of trials will vary depending on the signal to noise ratio, the number of events expected and therefore the length of the RT (as illustrated in the simulation section, specifically Simulation 1) and the coherence of the sequence across trials. Despite these complex issues to define an optimal number of trials, we can nevertheless provide some guidelines. The experimental data analyzed in the last result section with HMP is based on approximately 2200 trials per condition (both conditions are fitted independently). Per participant, this amounts to around 90 trials, which is usually on the lower end of trial numbers for many standard experiments in cognitive neuroscience. Based on this example and the simulations it appears that HMP does not require an excessive amount of trials.

Although the current paper frames HMP as a method for EEG time-series analysis, analysis of other neural time-series data is also possible. The adaptation to MEG is already implemented within the Python package associated with HMP. Other time and chronometrically honored measures such as intracranial EEG and electro-corticography require, to date, a very small set of adaptations. The main challenge in these applications is the expectations regarding the pattern. In all of those applications, a half-sine seems as justified as for EEG but the duration of the expected event requires some adaptation due to the differences in sensitivity to different type of neural signals and to skin and skull distortions. Other time-series such as electromyography or pupillometry could also be used in an HMP model, but the relative high signal-to-noise ratio and the limited number of channels usually associated with those techniques, render the HMP method probably less useful than existing methods specifically designed for such measures (e.g. https://github.com/lspieser/myonset for electromyography or Cai et al., 2023, for pupilometry).

4.3 Future work

We plan to increase the flexibility of HMP to allow reseachers to test different hypotheses and use HMP in different scenarios. For example, a parametrization with several expected event widths or probability distributions would allow one to progressively build task models that would capture complex sequences of events. Ultimately, these could be used as priors for HMP when analyzing similar tasks. This could even be taken to another level by directly integrating

cognitive models such as ACT-R or evidence accumulation models within the estimation procedure (Turner et al., 2017). Another addition that is currently being studied is the possibility to fit a hierarchical version of the model where the estimation of group parameters inform individuals and vice-versa (Lee & Wagenmakers, 2014). This would allow a cumulative approach to HMP applications by using priors and population distributions to constrain individual estimations. Lastly, for now HMP requires an RT for each trial in order to extract the event sequence, but our intention is to allow the method to accomodate different scenarios without requiring a response from a subject. This includes passive stimulus exposure, decisions without overt actions, or resting state data. This would open up the possibility to further the knowledge gained from classical mental chronometry to new applications.

4.4 Conclusion

By building on the HsMM-MVPA method developed by Anderson et al. (2016) we show how HMP fitted on time-based neural signals can accommodate many different scenarios and robustly determine a sequence of neural events. The sequential nature as well as the use of probability distributions to enhance portion of the signal gives a high granularity in the interpretation of a series of stimulus and response related events in the brain. Given its qualities, this method represents a new venue for fundamental and applied research relying on reaction time based tasks, not solely for mental chronometry related questions, but also for a wide variety of brain-related signal analysis as advocated by the modern mental chronometry (Meyer et al., 1988).

Data and Code Availability

All the data and code used to perform the analysis and generate the figures can be found at https://osf.io/29tgr/.

Author Contributions

GW, LvM and JPB conceptualized the method. GW and JPB implemented the method in the associated python package. GW wrote the first draft of the present manuscript. LvM and JPB reviewed and edited the manuscript.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101066503. JPB and LvM are funded through the Air Force Research Laboratory grant No EOARD FA8655-22-1-7003.

Declaration of Competing Interests

The authors declare that they have no competing interest.

Acknowledgments

We thank John Anderson and Joshua Krause for the discussions about the method and beyond. We also wish to thank Leon Kenemans for providing insightful comments on a first draft of this manuscript. Finally, the authors wish to thank the COBRA and CogMod groups, respectively from Utrecht University and Groningen University, for early feedback on the development of the method.

References

Anders, R., Alario, F., Van Maanen, L., et al. (2016). The shifted wald distribution for response time data analysis. *Psychological methods*, *21*(3), 309.

Anderson, J. R., Borst, J. P., Fincham, J. M., Ghuman, A. S., Tenison, C., & Zhang, Q. (2018). The common time course of memory processes revealed. *Psychological science*, *29*(9), 1463–1474.

Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The discovery of processing stages: Extension of sternberg's method. *Psychological review*, *123*(5), 481.

Berberyan, H. S., van Rijn, H., & Borst, J. P. (2021). Discovering the brain stages of lexical decision: Behavioral effects originate from a single neural decision process. *Brain and Cognition*, *153*, 105786.

Boehm, U., Van Maanen, L., Forstmann, B., & van Rijn, H. (2014). Trial-by-trial fluctuations in cnv amplitude reflect anticipatory adjustment of response caution. *NeuroImage*, *96*, 95–105.

Borst, J. P., & Anderson, J. R. (2024). Discovering cognitive stages in m/eeg data to inform cognitive models. In B. Forstmann & B. Turner (Eds.), *An introduction to model-based cognitive neuroscience*. Springer.

Botwinick, J., & Thompson, L. W. (1966). Premotor and motor components of reaction time. *Journal of Experimental Psychology*, *71*(1), 9–15. https://doi.org/10.1037/h0022634

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, *112*(1), 117–128. https://doi.org/10.1037/0033-295X.112.1.117

Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal analysis of multichannel eeg: Cartool. *Computational intelligence and neuroscience*, *2011*, 1–15.

Buc Calderon, C., Dewulf, M., Gevers, W., & Verguts, T. (2017). Continuous track paths reveal additive evidence integration in multistep decision making. *Proceedings of the National Academy of Sciences*, *114*(40), 10618–10623. https://doi.org/10.1073/pnas.1710913114

Burle, B., Roger, C., Vidal, F., & Hasbroucq, T. (2008). Spatio-temporal dynamics of information processing in the brain: Recent advances, current limitations and future challenges. *International Journal of Bioelectromagnetism*, *10*, 17–21.

Burle, B., Possamaï, C.-A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002). Executive control in the Simon effect: an electromyographic and distributional analysis. *Psychological Research*, *66*(4), 324–36. https://doi.org/10.1007/s00426-002-0105-6

Burle, B., Vidal, F., Tandonnet, C., & Hasbroucq, T. (2004). Physiological evidence for response inhibition in choice reaction time tasks. *Brain and cognition*, *56*(2), 153–164.

Cai, Y., Strauch, C., Van der Stigchel, S., & Naber, M. (2023). Open-dpsm: An open-source toolkit for modeling pupil size changes to dynamic visual inputs. *Behavior Research Methods*, 1–17.

Callaway, E., Halliday, R., Naylor, H., & Thouvenin, D. (1984). The latency of the average is not the average of the latencies. *Psychophysiology*, *21*, 571.

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2022). Bambi: A simple interface for fitting bayesian linear models in python. *Journal of Statistical Software*, *103*(15), 1–29. https://doi.org/10.18637/jss.v103.i15

Christie, L. S., & Luce, R. D. (1956). Decision structure and time relations in simple choice behavior. *The bulletin of mathematical biophysics*, *18*(2), 89–112.

Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. MIT press.

Coles, M. G. (1989). Modern mind-brain reading: Psychophysiology, physiology, and cognition. *Psychophysiology*, *26*(3), 251–269.

De Boeck, P., & Jeon, M. (2019). An overview of models for response times and processes in cognitive tests. *Frontiers in psychology*, *10*, 102.

Donders, F. C. (1868). Die schnelligkeit psychischer processe: Erster artikel. *Archiv für Anatomie, Physiologie und wissenschaftliche Medicin*, 657–681.

Forstmann, B. U., Wagenmakers, E.-J., Eichele, T., Brown, S., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: Opposites attract? *Trends in cognitive sciences*, *15*(6), 272–279.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). Mne software for processing meg and eeg data. *neuroimage*, *86*, 446–460.

Green, D. M. (1971). Fourier analysis of reaction time data. *Behavior Research Methods & Instrumentation*, *3*(3), 121–125.

Groeneweg, E., Archambeau, K., & Van Maanen, L. (2021). Classification of cognitive strategies using EEG time series analyses. *Proceedings of the 19th International Conference on Cognitive Modeling*.

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for meg and eeg data. *NeuroImage*, *159*, 417–429.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *Journal of Neuroscience*, *33*(50), 19434–19441.

Kenemans, J., Baas, J., Mangun, G., Lijffijt, M., & Verbaten, M. (2000). On the processing of spatial frequencies as revealed by evoked-potential source modeling. *Clinical neurophysiology*, *111*(6), 1113–1123.

Kirchherr, S., Mildiner Moraga, S., Coudé, G., Bimbi, M., Ferrari, P. F., Aarts, E., & Bonaiuto, J. J. (2023). Bayesian multilevel hidden markov models identify stable state dynamics in longitudinal recordings from macaque primary motor cortex. *European Journal of Neuroscience*, *58*(3), 2787–2806.

Koenig, T., Prichep, L., Lehmann, D., Sosa, P. V., Braeker, E., Kleinlogel, H., Isenhart, R., & John, E. R. (2002). Millisecond by millisecond, year by year: Normative eeg microstates and developmental stages. *Neuroimage*, *16*(1), 41–48.

Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 658–676.

Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. (2019). Arviz a unified library for exploratory analysis of bayesian models in python. *Journal of Open Source Software*, *4*(33), 1143. https://doi.org/10.21105/joss.01143

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.

Lehmann, D. (1971). Multichannel topography of human alpha eeg fields. *Electroencephalography and clinical neurophysiology*, *31*(5), 439–449.

Luce, R. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford University Press New York. https://doi.org/10.1093/acprof:oso/9780195070019.001.0001

Luck, S. J. (2005). Ten simple rules for designing and interpreting erp experiments. *Event-related potentials: A methods handbook*, *4*.

Makeig, S. (2002). Response: Event-related brain dynamics–unifying brain electrophysiology. *Trends in neurosciences*, *25*(8), 390.

Masaracchia, L., Fredes, F., Woolrich, M. W., & Vidaurre, D. (2023). Dissecting unsupervised learning through

hidden markov modelling in electrophysiological data. *bioRxiv*, 2023–01.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.

Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, *26*(1-3), 3–67.

Michel, C. M., & Koenig, T. (2018). Eeg microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *Neuroimage*, *180*, 577–593.

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring lrp onset latency differences. *Psychophysiology*, *35*(1), 99–115.

Miller, J., Ulrich, R., & Rinkenauer, G. (1999). Effects of stimulus intensity on the lateralized readiness potential. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(5), 1454.

Miller, J. O. (1993). A queue-series model for reaction time, with discrete-stage and continuous-flow models as special cases. *Psychological Review*, *100*(4), 702.

Mouraux, A., & Iannetti, G. D. (2008). Across-trial averaging of event-related eeg responses and beyond. *Magnetic resonance imaging*, *26*(7), 1041–1054.

Noorani, I., & Carpenter, R. (2016). The later model of reaction time and decision. *Neuroscience & Biobehavioral Reviews*, *64*, 229–251.

Nunez, M. D., Gosai, A., Vandekerckhove, J., & Srinivasan, R. (2019). The latency of a visual evoked potential tracks the onset of decision making. *Neuroimage*, *197*, 93–108.

O'connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature neuroscience*, *15*(12), 1729–1735.

Polich, J. (2007). Updating p300: An integrative theory of p3a and p3b. *Clinical neurophysiology*, *118*(10), 2128–2148.

Portoles, O., Blesa, M., van Vugt, M., Cao, M., & Borst, J. P. (2022). Thalamic bursts modulate cortical synchrony locally to switch between states of global functional connectivity in a cognitive task. *PLoS computational biology*, *18*(3), e1009407.

Posner, M. I. (1978). *Chronometric Explorations of Mind*. Lawrence Erlbaum.

Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological review*, *117*(4), 1113.

Quinn, A. J., Vidaurre, D., Abeysuriya, R., Becker, R., Nobre, A. C., & Woolrich, M. W. (2018). Task-evoked dynamic network analysis through hidden markov modeling. *Frontiers in neuroscience*, *12*, 603.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Schall, J. D. (2019). Accumulators, neurons, and response time. *Trends in neurosciences*, *42*(12), 848–860.

Selen, L. P. J., Shadlen, M. N., & Wolpert, D. M. (2012). Deliberation in the Motor System: Reflex Gains Track Evolving Evidence Leading to a Decision. *Journal of Neuroscience*, *32*(7), 2276–2286. https://doi.org/10.1523/JNEUROSCI.5273-11.2012

Servant, M., Logan, G. D., Gajdos, T., & Evans, N. J. (2021). An integrated theory of deciding and acting. *Journal of Experimental Psychology: General*, *150*(12), 2435.

Smith, P. L. (1990). Obtaining meaningful results from fourier deconvolution of reaction time data. *Psychological Bulletin*, *108*(3), 533.

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, *102*(3), 567–593. https://doi.org/10.1037/0033-295X.102.3.567

Smulders, F. T., Kenemans, J., & Kok, A. (1994). A comparison of different methods for estimating single-trial p300 latencies. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *92*(2), 107–114.

Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, *25*(3), 251–260. https://doi.org/10.1007/BF02289729

Tenison, C., Fincham, J. M., & Anderson, J. R. (2016). Phases of learning: How skill acquisition impacts cognitive processing. *Cognitive psychology*, *87*, 1–28.

Thura, D., Cabana, J.-F., Feghaly, A., & Cisek, P. (2022). Integrated neural dynamics of sensorimotor decisions and actions. *PLoS biology*, *20*(12), e3001861.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, *39*(4), 321–359.

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65–79.

Van de Ville, D., Britz, J., & Michel, C. M. (2010). Eeg microstate sequences in healthy humans at rest reveal scale-free dynamics. *Proceedings of the National Academy of Sciences*, *107*(42), 18179–18184.

van der Velde, M., Sense, F., Borst, J. P., Van Maanen, L., & Van Rijn, H. (2022). Capturing dynamic performance in a cognitive model: Estimating act-r memory parameters with the linear ballistic accumulator. *Topics in Cognitive Science*, *14*(4), 889–903.

Van Maanen, L., de Jong, R., & van Rijn, H. (2014). How to assess the existence of competing strategies in cognitive tasks: A primer on the fixed-point property. *PloS one*, *9*(8), e106113.

Van Maanen, L., Portoles, O., & Borst, J. P. (2021). The discovery and interpretation of evidence accumulation stages. *Computational brain & behavior*, *4*(4), 395–415.

Van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture—word interference are two sides of the same coin. *Psychonomic bulletin & review*, *16*, 987–999.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Visser, I. (2011). Seven things to remember about hidden markov models: A tutorial on markovian models for time series. *Journal of Mathematical Psychology*, *55*(6), 403–415.

von Helmholtz, H. (1850). Mittheilung für die physikalische gesellschaft in berlin betreffend versuche über die fortpflanzungsgeschwindigkeit der reizung in den sensiblen nerven des menschen. *Archive of the BerlinBrandenburgische Akademie der Wissenschaften*, 1–4.

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological review*, *114*(3), 830.

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Weindel, G., Anders, R., Alario, F., Burle, B., et al. (2021). Assessing model-based inferences in decision making with single-trial response time decomposition. *Journal of Experimental Psychology: General*, *150*(8), 1528.

Weindel, G., Gajdos, T., Burle, B., & Alario, F.-X. (2022). The decisive role of non-decision time for interpreting decision making models. *PsyArXiv*.

Wu, C. J. (1983). On the convergence properties of the em algorithm. *The Annals of statistics*, 95–103.

Yamanaka, K., & Yamamoto, Y. (2010). Single-trial eeg power and phase dynamics associated with voluntary response inhibition. *Journal of Cognitive Neuroscience*, *22*(4), 714–727.

Yu, S.-Z. (2010). Hidden semi-markov models. *Artificial intelligence*, *174*(2), 215–243.

Zhang, Q., Van Vugt, M., Borst, J. P., & Anderson, J. R. (2018). Mapping working memory retrieval in space and in time: A combined electroencephalography and electrocorticography approach. *NeuroImage*, *174*, 472–484. https://doi.org/10.1016/j.neuroimage.2018.03.039

Appendix A
Coefficient values for the simulations

Coefficient values for the the generalized linear models on the Simulation sections for the True positive Rate:

True positive rate	Mean	SD	2.5%	97.5%
Simulation 1				
Intercept	1.232	0.076	1.085	1.381
Ampere	5.626	0.207	5.220	6.020
Tolerance	0.193	0.027	0.140	0.246
Ampere:Tolerance	0.075	0.075	-0.071	0.226
Nevents	-0.179	0.013	-0.206	-0.153
Ampere:Nevents	-0.322	0.037	-0.392	-0.250
Tolerance:Nevents	-0.041	0.005	-0.051	-0.032
Ampere:Tolerance:Nevents	-0.105	0.014	-0.132	-0.078
Simulation 2				
Intercept	0.358	0.107	0.157	0.572
TargetWidth[25]	0.969	0.155	0.677	1.283
TargetWidth[50]	0.592	0.153	0.282	0.882
TargetWidth[75]	0.122	0.150	-0.186	0.397
bs(GenWidth, 3, degree=3)[0]	6.970	1.289	4.427	9.408
bs(GenWidth, 3, degree=3)[1]	-48.261	6.365	-60.878	-35.963
bs(GenWidth, 3, degree=3)[2]	-24.454	13.662	-50.584	-1.317
TargetWidth:bs(GenWidth, 3, degree=3)[25, 0]	-1.535	1.580	-4.573	1.504
TargetWidth:bs(GenWidth, 3, degree=3)[25, 1]	43.533	6.493	31.400	56.791
TargetWidth:bs(GenWidth, 3, degree=3)[25, 2]	25.998	13.667	2.738	51.983
TargetWidth:bs(GenWidth, 3, degree=3)[50, 0]	-3.679	1.727	-7.171	-0.437
TargetWidth:bs(GenWidth, 3, degree=3)[50, 1]	49.949	7.747	34.987	65.576
TargetWidth:bs(GenWidth, 3, degree=3)[50, 2]	-10.394	17.625	-44.914	24.117
TargetWidth:bs(GenWidth, 3, degree=3)[75, 0]	1.167	1.803	-2.218	4.797
TargetWidth:bs(GenWidth, 3, degree=3)[75, 1]	25.152	8.937	8.064	42.798
TargetWidth:bs(GenWidth, 3, degree=3)[75, 2]	-32.980	31.790	-95.741	22.407
Simulation 3				
Intercept	1.110	0.020	1.071	1.151
RT	-0.021	0.070	-0.150	0.129
Distribution[Gamma 4]	0.082	0.029	0.025	0.137
Distribution[Gamma 6]	0.079	0.029	0.026	0.141
Distribution[Wald 0.25]	0.030	0.029	-0.027	0.085
Distribution[Wald 0.5]	0.017	0.029	-0.044	0.069
Distribution[Wald 1]	-0.137	0.028	-0.190	-0.082
Distribution[Weibull 2]	0.021	0.029	-0.033	0.077
Distribution[Weibull 3]	-0.097	0.028	-0.154	-0.044
Distribution[Weibull 4]	-0.514	0.028	-0.568	-0.460
RT:Distribution[Gamma 4]	0.047	0.101	-0.146	0.246
RT:Distribution[Gamma 6]	0.105	0.102	-0.100	0.301
RT:Distribution[Wald 0.25]	0.052	0.101	-0.153	0.244
RT:Distribution[Wald 0.5]	0.064	0.101	-0.126	0.268
RT:Distribution[Wald 1]	0.089	0.098	-0.108	0.273
RT:Distribution[Weibull 2]	0.131	0.102	-0.064	0.335
RT:Distribution[Weibull 3]	0.316	0.098	0.130	0.512
RT:Distribution[Weibull 4]	0.582	0.096	0.388	0.762

Coefficient values for the GLMs on the Simulation sections for the positive predictive value:

Positive predictive value	Mean	SD	2.5%	97.5%
Simulation 1				
Intercept	4.219	0.203	3.840	4.629
Ampere	7.423	0.551	6.335	8.469
Tolerance	0.489	0.063	0.367	0.609
Ampere:Tolerance	1.537	0.155	1.239	1.841
Nevents	1.554	0.099	1.368	1.754
Ampere:Nevents	3.734	0.250	3.235	4.205
Tolerance:Nevents	0.353	0.026	0.301	0.404
Ampere:Tolerance:Nevents	0.996	0.064	0.872	1.121
Simulation 2				
Intercept	3.057	0.312	2.485	3.706
TargetWidth[25]	-4.272	0.318	-4.919	-3.667
TargetWidth[50]	-3.441	0.321	-4.064	-2.812
TargetWidth[75]	-3.044	0.329	-3.709	-2.442
bs(GenWidth, 3, degree=3)[0]	9.813	2.866	4.208	15.432
bs(GenWidth, 3, degree=3)[1]	-79.282	11.137	-100.379	-56.764
bs(GenWidth, 3, degree=3)[2]	-33.001	17.847	-67.019	2.316
TargetWidth:bs(GenWidth, 3, degree=3)[25, 0]	-8.321	2.923	-14.153	-2.775
TargetWidth:bs(GenWidth, 3, degree=3)[25, 1]	85.090	11.373	64.145	108.207
TargetWidth:bs(GenWidth, 3, degree=3)[25, 2]	48.322	19.257	10.902	85.550
TargetWidth:bs(GenWidth, 3, degree=3)[50, 0]	-8.001	3.013	-13.651	-1.931
TargetWidth:bs(GenWidth, 3, degree=3)[50, 1]	112.099	12.525	88.760	137.469
TargetWidth:bs(GenWidth, 3, degree=3)[50, 2]	-80.822	23.449	-122.815	-28.877
TargetWidth:bs(GenWidth, 3, degree=3)[75, 0]	4.585	3.176	-1.630	10.936
TargetWidth:bs(GenWidth, 3, degree=3)[75, 1]	49.915	13.077	25.176	76.756
TargetWidth:bs(GenWidth, 3, degree=3)[75, 2]	-39.248	39.229	-112.429	38.046
Simulation 3				
Intercept	2.048	0.030	1.991	2.106
RT	-0.681	0.107	-0.883	-0.474
Distribution[Gamma 4]	-0.430	0.038	-0.504	-0.359
Distribution[Gamma 6]	-0.768	0.037	-0.841	-0.696
Distribution[Wald 0.25]	-0.601	0.037	-0.677	-0.530
Distribution[Wald 0.5]	-0.101	0.041	-0.182	-0.022
Distribution[Wald 1]	0.048	0.043	-0.037	0.129
Distribution[Weibull 2]	-0.121	0.041	-0.200	-0.037
Distribution[Weibull 3]	-0.633	0.038	-0.706	-0.558
Distribution[Weibull 4]	-0.805	0.037	-0.876	-0.730
RT:Distribution[Gamma 4]	0.496	0.137	0.239	0.768
RT:Distribution[Gamma 6]	0.697	0.131	0.437	0.944
RT:Distribution[Wald 0.25]	0.628	0.135	0.353	0.883
RT:Distribution[Wald 0.5]	0.518	0.147	0.227	0.802
RT:Distribution[Wald 1]	0.054	0.155	-0.244	0.357
RT:Distribution[Weibull 2]	0.312	0.146	0.016	0.588
RT:Distribution[Weibull 3]	0.681	0.134	0.419	0.940
RT:Distribution[Weibull 4]	0.640	0.136	0.384	0.911

Appendix B
Coefficients for the speed accuracy trade-off data

Coefficient values for the linear mixed models on the speed accuracy trade-off experiment:

	Mean	SD	2.5%	97.5%
Event[1]	38.841	2.844	33.331	44.072
Event[2]	167.729	4.737	158.717	176.531
Event[3]	705.436	34.982	636.796	766.841
Event[4]	53.927	2.725	48.714	58.878
Event:Condition[1, Speed]	-3.923	3.898	-11.422	3.151
Event:Condition[2, Speed]	13.627	4.742	5.014	22.720
Event:Condition[3, Speed]	-387.826	31.904	-447.942	-328.393
Event:Condition[4, Speed]	-0.050	3.884	-7.591	7.126
Duration sigma	126.914	0.673	125.685	128.178
Event|sub. sigma[1]	2.851	2.121	0.002	6.651
Event|sub. sigma[2]	18.469	4.053	10.869	25.957
Event|sub. sigma[3]	90.287	21.607	47.298	131.178
Event|sub. sigma[4]	1.676	1.344	0.000	4.061
Event:Condition|sub. sigma[1, Accuracy]	3.249	2.425	0.003	7.536
Event:Condition|sub. sigma[1, Speed]	3.176	2.515	0.009	7.832
Event:Condition|sub. sigma[2, Accuracy]	5.225	4.084	0.000	12.642
Event:Condition|sub. sigma[2, Speed]	10.471	6.041	0.000	20.217
Event:Condition|sub. sigma[3, Accuracy]	149.087	25.739	104.427	196.973
Event:Condition|sub. sigma[3, Speed]	45.392	27.531	0.003	90.865
Event:Condition|sub. sigma[4, Accuracy]	2.349	1.828	0.000	5.632
Event:Condition|sub. sigma[4, Speed]	2.349	1.854	0.003	5.663