Detecting neural state transitions underlying event segmentation

Linda Geerligs[1], Marcel van Gerven[1] & Umut Güçlü[1]

1. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands

Correspondence to:
Linda Geerligs
Donders Institute for Brain, Cognition and Behaviour
Montessorilaan 3 6525 HR Nijmegen, The Netherlands
+31 24 3616091
l.geerligs@donders.ru.nl

Keywords: Event segmentation, fMRI, Hidden Markov Model, timescales, neural states, greedy search

Abstract

Segmenting perceptual experience into meaningful events is a key cognitive process that helps us make sense of what is happening around us in the moment, as well as helping us recall past events. Nevertheless, little is known about the underlying neural mechanisms of the event segmentation process. Recent work has suggested that event segmentation can be linked to regional changes in neural activity patterns. Accurate methods for identifying such activity changes are important to allow further investigation of the neural basis of event segmentation and its link to the temporal processing hierarchy of the brain. In this study, we introduce a new set of elegant and simple methods to study these mechanisms. We introduce a method for identifying the number of neural states in a brain area, and another one for identifying the boundaries between these states. Furthermore, we present the results of a comprehensive set of simulations and analyses of real fMRI data to provide guidelines for reliable estimation of neural states and show that our proposed methods outperform the current state-of-the-art in the literature. This methodological innovation will allow researchers to make new headway in investigating the neural basis of event segmentation and information processing during naturalistic stimulation.

Introduction

To understand the world around us as it unfolds over time, two processes are essential; information integration and segmentation. We integrate current sensory input with information from the past to make sense of speech or actions that unfold over time (Buonomano and Maass, 2009; Kiebel et al., 2008). We also segment information into distinct events when information from a previous timepoint is no longer relevant for what is occurring now (Kurby and Zacks, 2008; Newtson et al., 1977). Behavioral research has shown that segmenting information into meaningful events enables us to understand ongoing perceptual input (Zacks et al., 2001) and recall distinct events from our past (Flores et al., 2017; Sargent et al., 2013; Zacks et al., 2006). Although segmentation plays a fundamental role in the way we perceive and remember information in daily life, a lot remains unknown about the neural mechanisms that underlie these abilities. Here, we introduce a new method to investigate those mechanisms.

Recent innovations in data-driven analyses of fMRI voxel activity patterns during naturalistic stimulation have shown that event boundaries co-occur with shifts between stable patterns of brain activity (Baldassano et al., 2017). We will refer to these stable time periods as neural states, to distinguish them from the subjectively experienced events that have been described before (Zacks et al., 2007). By studying the duration of neural states, it is possible to see the temporal hierarchy of cortical information processing. High-level brain regions, such as the medial prefrontal cortex, show state durations that are comparable to those of experienced events. However, lower-level brain areas such as visual cortex, show state durations at much shorter time-scales (Baldassano et al., 2017). This observed hierarchy is in line with findings from previous studies that have used different approaches to demonstrate a similar temporal hierarchy for information integration in the cortex (Hasson et al., 2015; Honey et al., 2012; Lerner et al., 2011). State transitions have also been shown to be coupled to a subsequent increase in activity in hippocampus. This boundary-related hippocampal activity predicted reinstatements of brain activity during later recall, suggesting that state boundaries play an important role in memory encoding (Baldassano et al., 2017). In EEG studies, state transitions have been shown to be associated with a rapid reinstatement of the just-encoded movie state which may be associated with memory formation (Silva et al., 2019).

These results show that data-driven state detection methods are a valuable tool for investigating the neural mechanisms underlying event segmentation and memory formation as well as the timescales of cortical information processing. However the state detection method of Baldassano et al. (2017) that was used in these papers has not been systematically validated using simulated data. Here we investigate the accuracy and reliability of this hidden Markov model (HMM)-based state segmentation method and show a number of important limitations. To address these limitations, we introduce a much simpler greedy state boundary search (GSBS) method that outperforms the HMM-based method in terms of accuracy of state boundary detection and computational speed without making any assumptions about when state boundaries should occur. In addition, we introduce a novel t-distance metric for identifying the optimal number of states, which we also validate using simulations. We use empirical data to illustrate how the reliability of state boundary detection is improved with GSBS compared to the HMM-based approach. Empirical data also confirms that our new method for estimating the optimal number of states can recover the expected cortical temporal hierarchy. Finally, we use simulations and real data to explore how noise and data averaging might impact our ability to accurately identify state boundary transitions.

Methods

State segmentation methods

Greedy state boundary search method

The GSBS method we introduce here relies on a simple greedy search algorithm to identify the location of state boundaries (see figure 1A). Let $\mathbf{X}$ be a $T \times V$ matrix of neural timeseries where T is the number of timepoints and V is the number of voxels. Let $\boldsymbol{x}_i$ denote the i-th row of this matrix. The algorithm places state boundaries in an iterative fashion. To determine the location of a boundary t^*, a greedy search is performed over all possible boundary locations. The method searches for transitions between states, but is not designed to identify recurring states. For a given potential boundary location t, the timepoints between the previous boundary and the current boundary $[t^-, \dots, t-1]$ are considered to be one state and the timepoints on and after the boundary $[t, \dots, t^+]$ (up to the following boundary) are another state. For each timepoint, we compute the mean activity pattern over all timepoints within the same state:

$$\hat{\boldsymbol{x}}_t = \frac{1}{M}\sum_{i=t}^{t^+} \boldsymbol{x}_i$$

where M is the number of timepoints between t and t^+. Next, we correlate the brain activity at each timepoint with the activity pattern in its corresponding state before averaging these correlations across all timepoints. That is,

$$\rho_t = \frac{1}{N}\sum_{i=0}^{N} \mathrm{corr}(\hat{\boldsymbol{x}}_i, \boldsymbol{x}_i)$$

where N is the total number of timepoints. This average is our estimate of the boundary fit for one potential boundary location. We repeat this computation for all possible boundary locations and the boundary is set as the timepoint when the boundary fit is the highest:

$$t^* = \underset{t}{\mathrm{argmax}}(\rho_t)$$

This process is repeated to place the next state boundary, keeping the preceding boundaries fixed. A new boundary can be placed at each timepoint as long as it does not overlap with another boundary. This process continues until a predefined number of states has been identified or until the number of states equals the number of timepoints.

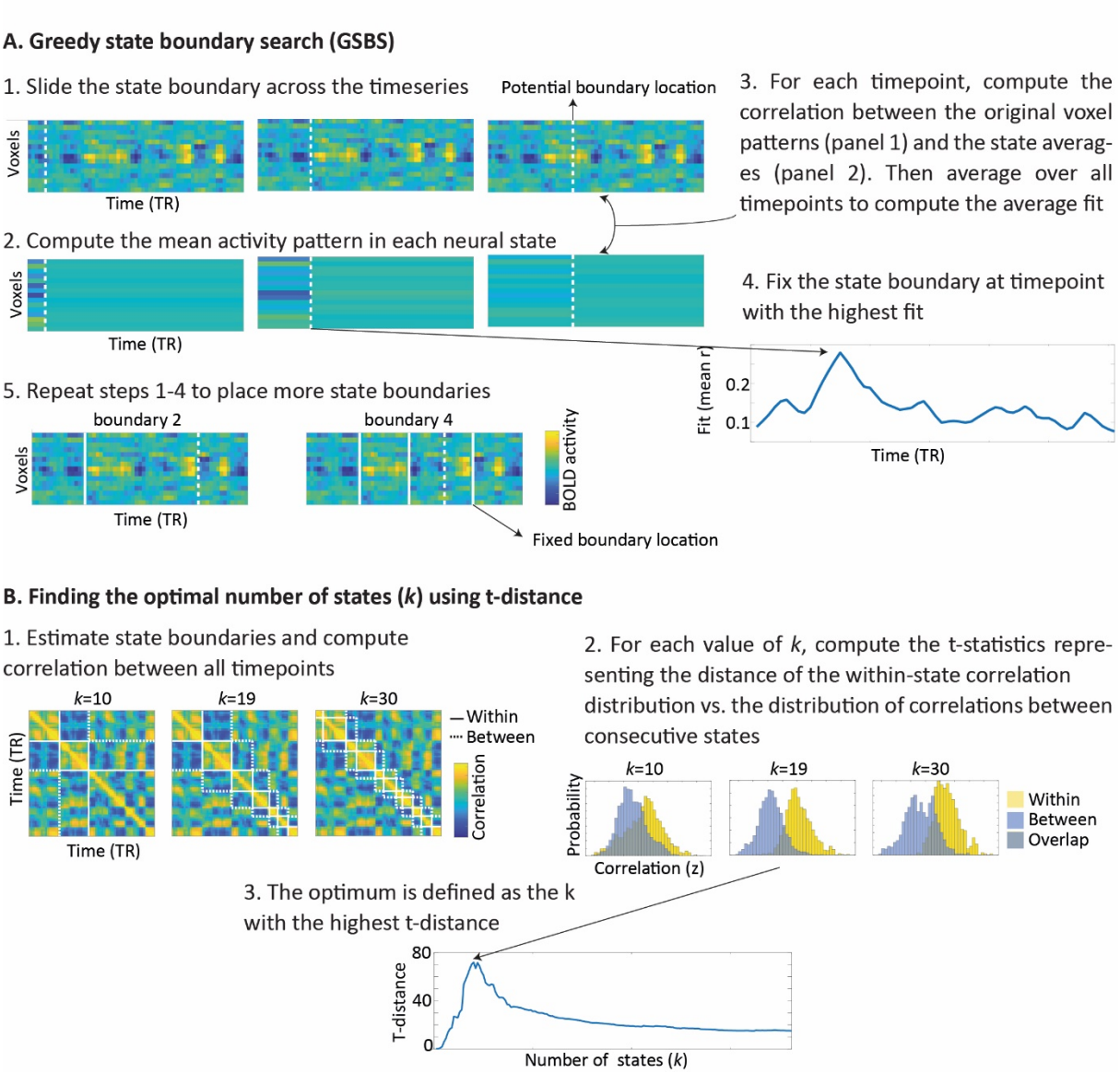


Figure 1: A. Illustration of the GSBS method that we use to identify state boundaries. B. Illustration of the t-distance metric that we use to determine the optimal number of states.

Estimating the number of states with t-distance

While our GSBS method can identify state boundaries, it cannot determine the optimal number of states. Therefore, we designed a separate metric to determine the optimal number of states k^*, which we call t-distance. T-distance is based on maximizing the similarity of timepoints in the same state, while minimizing the similarity of timepoints in consecutive states. The first step to compute the t-distance for a given number of states k is to compute the correlation between the neural activity patterns at each pair of timepoints i, j (see figure 1B, panel 1). That is,

$$\rho_{i,j} = \mathrm{corr}(\boldsymbol{x}_i, \boldsymbol{x}_j) : i \neq j$$

Pairs of timepoints that fall within the same state are considered to be part of the within-state correlation distribution and pairs of timepoints that are in consecutive states are part of the between-consecutive-state correlation distribution. The distance between the distribution of within-state correlation values $\rho_w(k)$ and the distribution of between-consecutive-state correlation values $\rho_{bc}(k)$ is quantified using a T-statistic $t_{\rho_w(k),\rho_{bc}(k)}$ (see figure 1B, panel 2). This t-distance is computed for each possible number of states and the optimal number of states is defined as the one with the highest t-distance. That is,

$$k^* = \underset{k}{\mathrm{argmax}}\left(t_{\rho_w(k),\rho_{bc}(k)}\right)$$

Code that implements our methods in Python is available at https://github.com/lgeerligs/State-segmentation-GSBS

Baseline methods

To evaluate the performance of our GSBS-method for detecting state boundaries and the t-distance metric for determining the optimal number of states, we compare it with existing methods.

The baseline we use for the GSBS-method is the 'event segmentation model' created by Baldassano et al. (2017), which is a variant of a Hidden Markov Model (HMM). We used the Python implementation in the Brain Imaging Analysis Kit (version 0.9.1). This HMM-based state boundary method models the brain activity as a sequence of hidden (unobserved) states. Each state is characterized by a specific mean activity pattern across voxels. In contrast to regular HMMs, this variant is constrained such that there are no recurring states (i.e. the HMM is *null recurrent*). Therefore, the first timepoint of a brain activity timecourse is always in state 1 and the final timepoint is always is state *k*, where *k* is the total number of states. From one timepoint to the next, a brain region can either stay in the same state or jump to the next state. In this implementation, all states are fixed to have the same probability of staying in the same state versus jumping to the next state. The inputs to the HMM-based method consists of a set of z-scored voxel timecourses and a value for *k* (the number of states that needs to be estimated).

To evaluate the performance of our t-distance metric, we compare it to the metric that was used by Baldassano et al (2017). This metric is based on subtracting the average correlation between all states from the average correlation within states. The optimal number of states is defined as the number of states with the largest difference of within- versus across-state correlations (we will refer to this as WAC). The two crucial differences between the t-distance and WAC metrics are 1) t-distance uses the t-statistic while WAC uses the average difference and 2) t-distance only considers the correlations between consecutive states, while WAC averages correlations between all states.

Simulation design

In analyses with real data, it is not possible to know the ground truth of state boundary locations. Therefore, we used simulations to determine how accurate our method is in recovering the simulated state boundaries under different circumstances. These simulations were an extended version of the toy simulations performed by Baldassano et al. (2017). In particular, we constructed state-structured datasets with $V = 50$ voxels and $T = 200$ timepoints and a TR of 2.47 seconds. The number of

timepoints and the TR were selected to (approximately) match our real data. The number of voxels was set at 50 to ensure that there were enough voxels to compute a reliable correlation coefficient. The number of states was varied between $k = 5$, $k = 15$ and $k = 30$ (default $k = 15$). The length of each state was sampled (from first to last) from a normal distribution with a mean $\mu = k/N$ and a standard deviation of μS, where S varied between 0.1, 0.5 and 1 (default $S = 0.5$). A mean pattern was drawn for each state from a standard normal distribution and the resulting voxel time courses were convolved with the canonical HRF from the SPM software package (default HRF peak delay = 6 s, peak dispersion = 1 s). To account for the initialization of the HRF, we initially created a longer timeseries (202 timepoints), these last two timepoints were always in the final state. After convolution, we removed the first 2 TRs from the time series. The simulated data for each timepoint and each voxel was the sum of the convolved state patterns plus randomly distributed noise with zero mean and standard deviation varying between $SD = 0.1$, $SD = 1$, $SD = 5$ and $SD = 10$ (default $SD = 0.1$). To quantify the similarity between the estimated and simulated state boundaries, we computed the Pearson correlation between the two state boundary time courses (with 0's for timepoints in which the state is the same as the previous timepoint and 1's for a state change). Each simulation was repeated 100 times, with different (randomly generated) state structures.

To investigate how the state boundary detection and the estimation of the optimal number of states is affected by noise in the data, individual differences, and averaging, we also simulated a group study. In these stimulations, we created datasets in which a group of 20 participants shared (some of) the states. In the first simulation, participants shared all states and state transitions and we varied the amount of random noise that was added to the data (between $SD = 1$, $SD = 5$ and $SD = 10$, see above). Then we investigated the effect of noise on state boundary detection and the estimation of the optimal number of states. We also investigated whether these effects of noise could be mitigated by averaging the data, or by using cross-validation, such that the state boundaries are defined in the training set and the optimal number of states are defined based on the test set. To investigate how averaging and cross–validation affects the results when not all states are shared between participants, we simulated data in which there is a specific probability that a given state transition that was present in the group, was not present in an individual, with $p \in [0.1, 0.2, 0.4]$. At the same time, we randomly added new state transitions to individuals, such that on average, across simulations, the number of states in each individual was the same as the number of states on the group-level. When a state transition that was present in the group disappeared in an individual, we modeled this by continuing the voxel pattern of the previous state. When we added individual-specific states, we generated a new unique mean activity pattern for each new state in each participant.

Dataset

We also used a real dataset to investigate performance of the different state boundary detection methods. In particular, we used 265 adults (131 female) who were aged 18–50 (mean age 36.3, SD = 8.6) from the healthy, population-derived cohort tested in Stage II of the Cam-CAN project (Shafto et al., 2014; Taylor et al., 2017). Participants had normal or corrected-to-normal vision and hearing, were native English-speakers, and had no neurological disorders. Ethical approval for the study was obtained from the Cambridgeshire 2 (now East of England - Cambridge Central) Research Ethics Committee. Participants gave written informed consent.

Participants were scanned using fMRI while they watched a shortened version of a black and white television drama by Alfred Hitchcock called 'Bang! You're Dead'. In previous studies, a longer version

of this movie has been shown to elicit robust brain activity, synchronized across younger participants (Hasson et al., 2009). Because of time constraints, the full 25-minute episode was condensed to 8 minutes with the narrative of the episode preserved (Shafto et al., 2014). Participants were instructed to watch, listen, and pay attention to the movie.

fMRI data acquisition & pre-processing

The details of the fMRI data acquisition are described in (Geerligs et al., 2018). In short, 193 volumes of movie data were acquired with a 32-channel head-coil, using a multi-echo, T2*-weighted EPI sequence. Each volume contained 32 axial slices (acquired in descending order), with slice thickness of 3.7 mm and interslice gap of 20% (TR = 2470 ms; five echoes [TE = 9.4 ms, 21.2 ms, 33 ms, 45 ms, 57 ms]; flip angle = 78 degrees; FOV = 192mm x 192 mm; voxel-size = 3 mm x 3 mm x 4.44 mm), the acquisition time was 8 minutes and 13 seconds. In addition, 261 volumes of resting state data were acquired with a TR of 1970 ms and a TE of 30 ms (other specifications were the same as for the movie data). High-resolution (1 mm x 1mm x 1 mm) T1 and T2-weighted images were also acquired.

The initial steps of data preprocessing were the same as in Geerligs et al. (2018) and are described there in detail. In short, the preprocessing steps included deobliquing of each TE, slice time correction and realignment of each TE to the first TE in the run, using AFNI (version AFNI_17.1.01; https://afni.nimh.nih.gov). Then multi-echo independent component analysis (ME-ICA) was used to denoise the data for each participant, facilitating the removal of non-BOLD components from the fMRI data, including effects of head motion. Co-registration followed by DARTEL intersubject alignment was used to align participants to MNI space using SPM12 software (http://www.fil.ion.ucl.ac.uk/spm).

Hyperalignment

To optimally align voxels across participants in the movie dataset, we used whole-brain searchlight hyperalignment as implemented in the PyMVPA toolbox (Guntupalli et al., 2016; Hanke et al., 2009). Hyperalignment aligns participants to a common representational space based on their shared responses to the movie stimulus. Because the state boundary detection methods are applied to group averaged voxel-level data, good inter-subject alignment is essential. Hyperalignment uses Procrustes transformations to derive the optimal rotation parameters that minimize intersubject distances between responses to the same timepoints in the movie.

A common representational space was derived by applying hyperalignment iteratively. The first iteration started by hyperaligning one participant to a reference participant. This reference participant was chosen as the participant with the highest level of inter-participant synchrony across the whole cortex (i.e. strongest correlations between the participants' timecourses and the average timecourses from the rest of the group, averaged across voxels). Next, a third participant was aligned to the mean response vectors of the first two participants. This hyperalignment and averaging alternation continued until all participants were aligned. In the second iteration, the transformation matrices were recalculated by hyperaligning each participant to the mean response vector from the first iteration. In a third iteration, the mean response vector was recomputed and this mean was defined as the common space. We then recalculated the transformation matrices for each participant to this common space. To align the whole cortex, hyperalignment was performed in overlapping searchlights with a

radius of three voxels and a stepsize of two voxels between each of the searchlight centers. The individual searchlights were aggregated into a single transformation matrix by averaging overlapping searchlight transformations. These aggregated transformation matrices were used to project each participant's movie fMRI data into the common representational space.

Real data analyses

To investigate the performance of the different state segmentation methods on our data, we selected five regions of interest in the left hemisphere; V1 (MNI, x=-4, y=-90, z=-2), V5 (MNI, x=-44, y=-72, z=2, inferior temporal cortex (IT, MNI x=-50, y= -52, z=-8), angular gyrus (AG, MNI x=-42,y=-64,z=40) and medial prefrontal cortex (mPFC, MNI x=0, y=54, z=22). These peak coordinates were based on a search of these brain regions in the Neurosynth database (Yarkoni et al., 2011). Based on previous work, we expected to see a clear temporal hierarchy across these regions of interest, with the largest number of states in V1, which decreased in number as we move up cortical hierarchy to secondary visual processing areas (V5) and multi-modal association areas (AG & IT) toward the top of the cortical hierarchy in the mPFC (Baldassano et al., 2017; Hasson et al., 2015; Honey et al., 2012; Lerner et al., 2011).

Around each of these peak coordinates, we created spherical searchlights with different sizes (radius 6, 8, 10 or 12 mm, default = 8 mm). We applied the different state segmentation methods in searchlights to the hyperaligned movie data. Voxels were excluded from the analysis if they had an intersubject correlation below r=0.35. To reduce effects of noise prior to running the state segmentation analyses, we averaged the voxel timecourses across groups of participants with varying sizes; data were divided into either 1, 2, 5, 10, 15, 20 or 265 distinct (sets of) participants, resulting in no averaging, or averages of, ~13, ~18, ~26, ~53 ~128 or ~265 participant (default = 15 sets, ~18 averaged participants).

As in the simulations, we first aimed to establish the impact of the different state segmentation methods on the reliability of the state boundaries. Reliability was estimated by computing the Pearson correlation between the state boundary timecourses of one group of participants (with 0's for timepoints in which the state is the same as the previous timepoint and 1's for a state change), and the average state boundary timecourses across all other groups of participants. We also investigated how the estimated optimal number of states align with the expected cortical hierarchy. In additional analyses, we also investigated the effects of data averaging and searchlight sphere size on the reliability and the estimated optimal number of states.

Results

In the first simulation, we evaluated the performance of the greedy state boundary search (GSBS) in detecting state boundaries in simulated data. As a baseline for our evaluation, we contrasted it with the HMM-based event-segmentation model (Baldassano et al., 2017). In this simulation, we assume that the number of states is known ($k = 15$). To quantify the similarity between simulated and estimated state boundary time series, we used the Pearson correlation coefficient. We found that when the simulated state boundaries are equally spaced (e.g. the standard deviation (SD) of state lengths is low) both methods do a good job of recovering the state boundaries accurately (median $r = 1$). However, when the states have variable lengths, we see that performance drops for both methods,

but much more strongly for the HMM-based method (see figure 2A). To identify the cause of this drop in performance, we identified the simulation run in which the state boundaries were most poorly recovered (separately for the HMM-based and GSBS-methods). Figures 2B&C show the real and recovered state boundaries for these simulations. The results show that the performance drop of the HMM-based method is due to its tendency to recover states with similar lengths. This causes it to detect boundaries in periods where no change is occurring. In addition, we see that both the GSBS and HMM-based methods sometimes identify a state boundary slightly later than it actually occurred. Together, these results show that the GSBS-method performs better in recovering state boundaries, especially when states can vary in length.

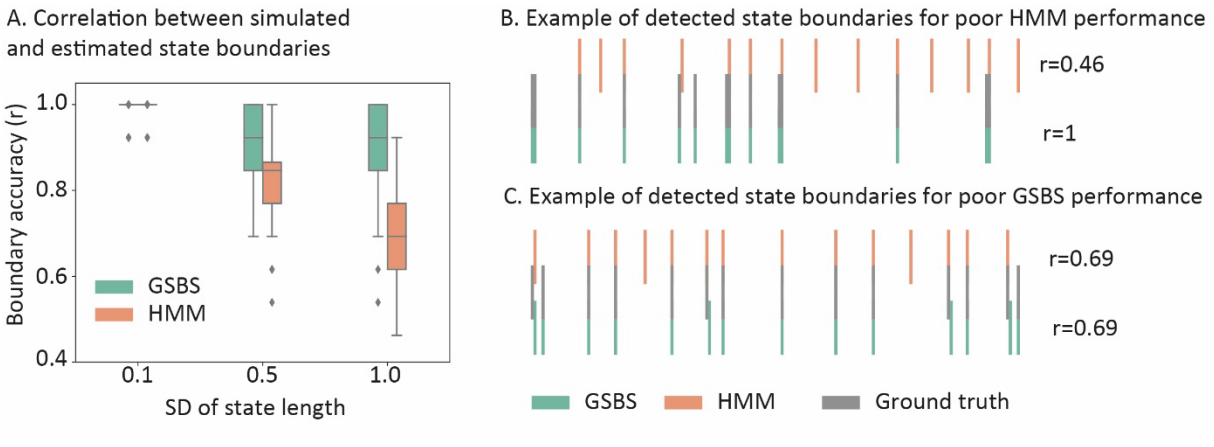


Figure 2: Similarity between simulated and estimated state boundaries for the HMM-based and GSBS methods. In this simulation the number of states is assumed to be known ($k = 15$).

The next simulation was aimed at comparing the metrics used to estimate the optimal number of states. Specifically, we compared the t-distance metric (figure 1B) to the WAC metric (introduced by Baldassano et al., 2017). T-distance uses the t-statistic to optimally separate the distributions of correlations within states and correlations between consecutive states. In contrast, WAC relies on optimizing the differences of within and between state correlations. Figure 3A shows that the t-distance is able to recover the simulated number of states accurately. In contrast, WAC overestimates the number of states when the true number of states is high, while it performs well when the number of states is low. To examine why this is the case, we show the plots of underlying estimates of within and between state correlations in figure 3B. These results demonstrate that WAC is largely driven by within-state similarity, which increases slightly as more states are added even after the simulated number of states is exceeded, potentially due to the autocorrelation caused by the HRF. Indeed, simulations without HRF convolution did not show this behavior. The between-state similarity hovers around zero regardless of the number of states that is estimated and does not have a lot of impact on WAC, at least when the number of states is larger than five (in a simulation with 200 TRs). The overestimation of the number of states for the WAC method occurs even when four TRs around the diagonal are not taken into account (as in Baldassano et al., 2017, see supplementary figure 1). For t-distance, the same within-state similarity is used as for WAC. However, we see that the ongoing increase in within-state similarity as more states are added is offset by the increase in similarity between consecutive states. This allows the method to identify the optimal number of states accurately. Together, these results show that the t-distance metric performs better in recovering the true number of states than the WAC metric.

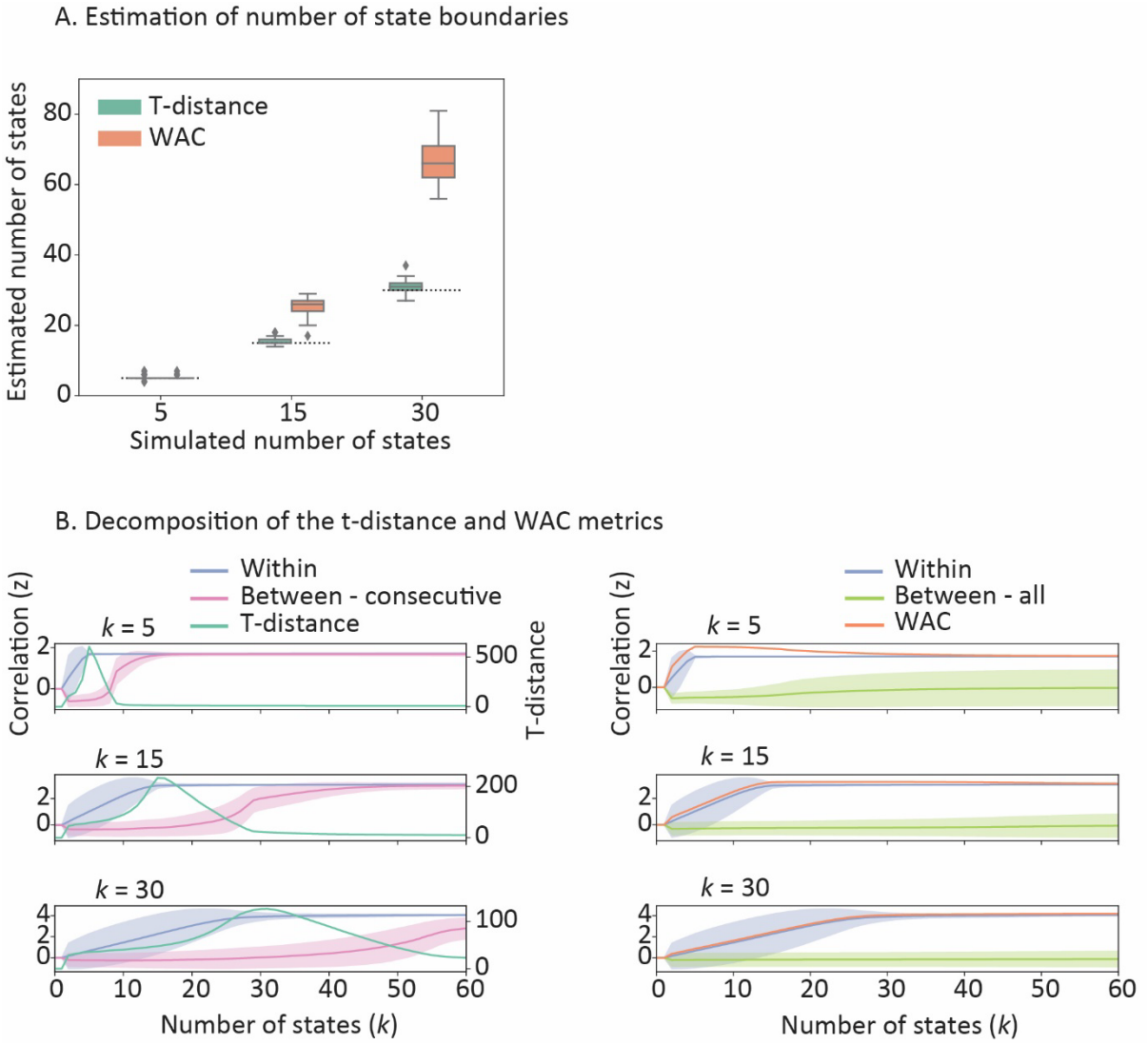


Figure 3. Comparison between the t-distance and WAC metrics for estimating the optimal number of state boundaries. A) Comparison of the estimated and simulated number of states. B) Comparison of the WAC and t-distance metrics and their underlying components for different values of k. The shaded area indicates the standard deviation of the within/between state correlation distributions.

In the next simulation, we compared the computational time that is required to run the GSBS and HMM-based method for different numbers of states. When the number of states that should be estimated is known, the GSBS-method results in a 6- to 15-fold improvement in computational speed (see figure 4A). The differences between methods become even more apparent when the number of states is not known and we need to go through all possible number of states to estimate the optimum. The HMM-based method identifies a new set of states for each value of *k* (the number of states). Therefore, the analyses need to be repeated for each value of *k*, which makes it computationally demanding. In contrast, the GSBS method performs an iterative search, which means that all but one of the boundaries that are detected for *k*=9 are the same as the boundaries that are detected for *k*=10. These differences result in an up to 1400-fold increase in computational speed for the GSBS method, compared to the HMM-based method (figure 4B). The results shown in figure 4 are for one brain region

and one simulated dataset. When we are interested in investigating multiple searchlights across many participants and brain regions, the computational demands of the HMM-based method quickly become prohibitive.

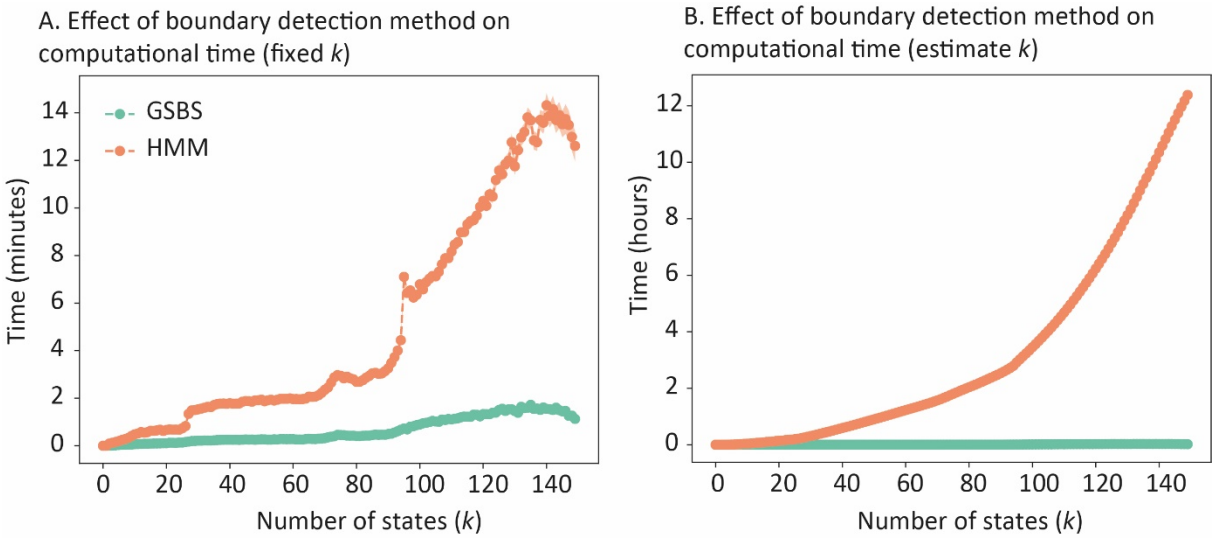


Figure 4. Comparison between the computational time required to run the GSBS and HMM-based boundary detection methods. A) The computational time when we assume that the number of states is known and is fixed at k. B) The computational time when we perform a search through all possible numbers of states, ranging from 2 to k. The shaded area indicates the standard error over 100 repetitions of the computation. These results were obtained using a 3.0 GHz, 32 core CPU.

Empirical data

Next, we compared the reliability of the state boundary detection for the GSBS and HMM-based methods for real fMRI data that was recorded while participants were watching a short movie. To be able to investigate the similarity of the results across participant groups as an estimate of reliability, participants were randomly divided in 15 groups of 17-18 participants. Within each group, the voxel timecourses were averaged within each ROI (the impact of averaging data will be explored in the next section). We investigated the reliability of the methods by correlating the state-boundary timecourses in each group of participants with the average across all other participant groups (similar to inter-subject synchronization). In addition, we estimated the ability of the t-distance and WAC metrics to identify a plausible number of state boundaries. Based on previous work, we expected to see a clear temporal hierarchy across our five regions of interest (Baldassano et al., 2017; Hasson et al., 2015; Honey et al., 2012; Lerner et al., 2011). In particular, we expected the largest number of states in V1, which decreased in number as we move up the visual processing hierarchy (V5) and decreased further in multi-modal association areas, such as the angular gyrus and inferior temporal cortex (IT) and medial prefrontal cortex (mPFC).

Figure 5 shows the results of this analyses. Figure 5A shows that the GSBS method resulted in more reliable state boundaries across the 15 participant groups than the HMM-based method. To investigate the significance of these differences, we performed t-tests on the Fisher-transformed correlation values. We found that the differences between methods were significant for IT ($T(14)=3.83$, $p=0.002$) and mPFC ($T(14)=4.19$, $p<0.001$). To derive the reliability estimates in figure 5A, we used the optimal

number of states for each ROI, as determined by the t-distance and shown in figure 6A. When we estimated different numbers of states (k=10, 20, 30 or 40), we found that the reliability of the GSBS method was always higher than the reliability of the HMM-based method. The difference was statistically significant for 11 out of the 20 comparisons (see supplementary figure 2). This is in line with the simulation results, suggesting that the GSBS method outperforms the HMM-based method in terms of reliably estimating state boundaries.

Simulation results suggested that the poorer performance of the HMM-based method is because it tends to estimate states that have a similar duration. If that is true, we would expect that states are more similar in length for the HMM-based than the GSBS method. Indeed, when we investigate the standard deviation of state lengths, we found that this was significantly higher for GSBS than the HMM-based method for each of the ROIs (all p<0.001, figure 5B), suggesting that the HMM-based method is biased toward finding states with similar durations.

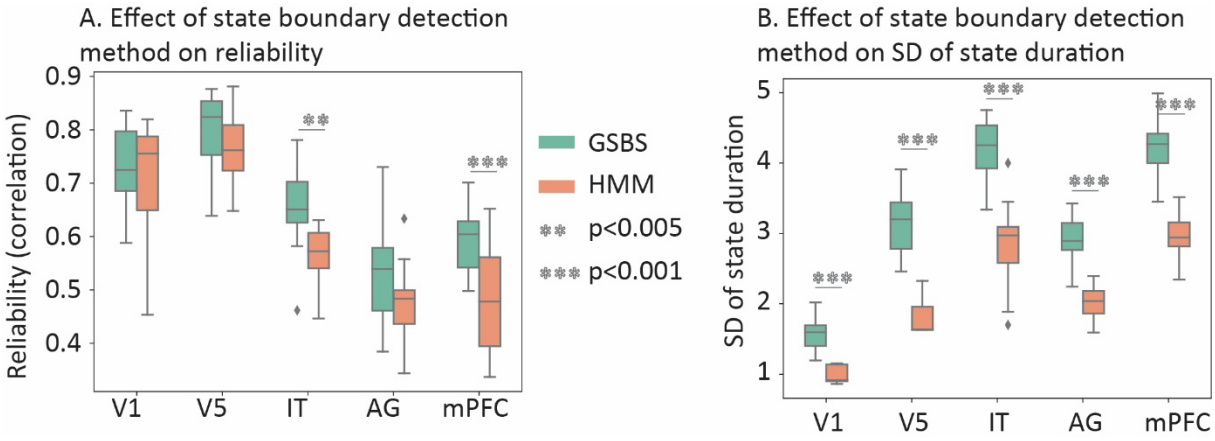


Figure 5. Analyses of real data. A) Reliability of state boundaries detected using GSBS or HMM-based method. B) SD of state boundary lengths for the GSBS and HMM-based methods.

Figure 6A shows the estimates of the number of state boundaries, where the state boundaries were based on GSBS and the number of states were determined using the t-distance. The estimated optimal number of states aligns well with the expected cortical hierarchy of timescales. In contrast, when we estimated the number of state boundaries using the WAC method (see figure 6B), we found an optimum of 150 states for each ROI (150 was the upper limit of the number of states we estimated). The results that are shown in Figure 6C illustrate why this happens. For WAC, the fit is only driven by the within-state similarity, while the between-state similarity stays around zero. Because the within-state similarity keeps increasing as more states are added, the optimal number of states is overestimated. For t-distance, the increase in within-state similarity is countered by the increase in the similarity between sequential states, as the number of states increase. These results are in line with the results for the simulated data, showing that t-distance provides a more accurate estimate of the optimal number of states than WAC.

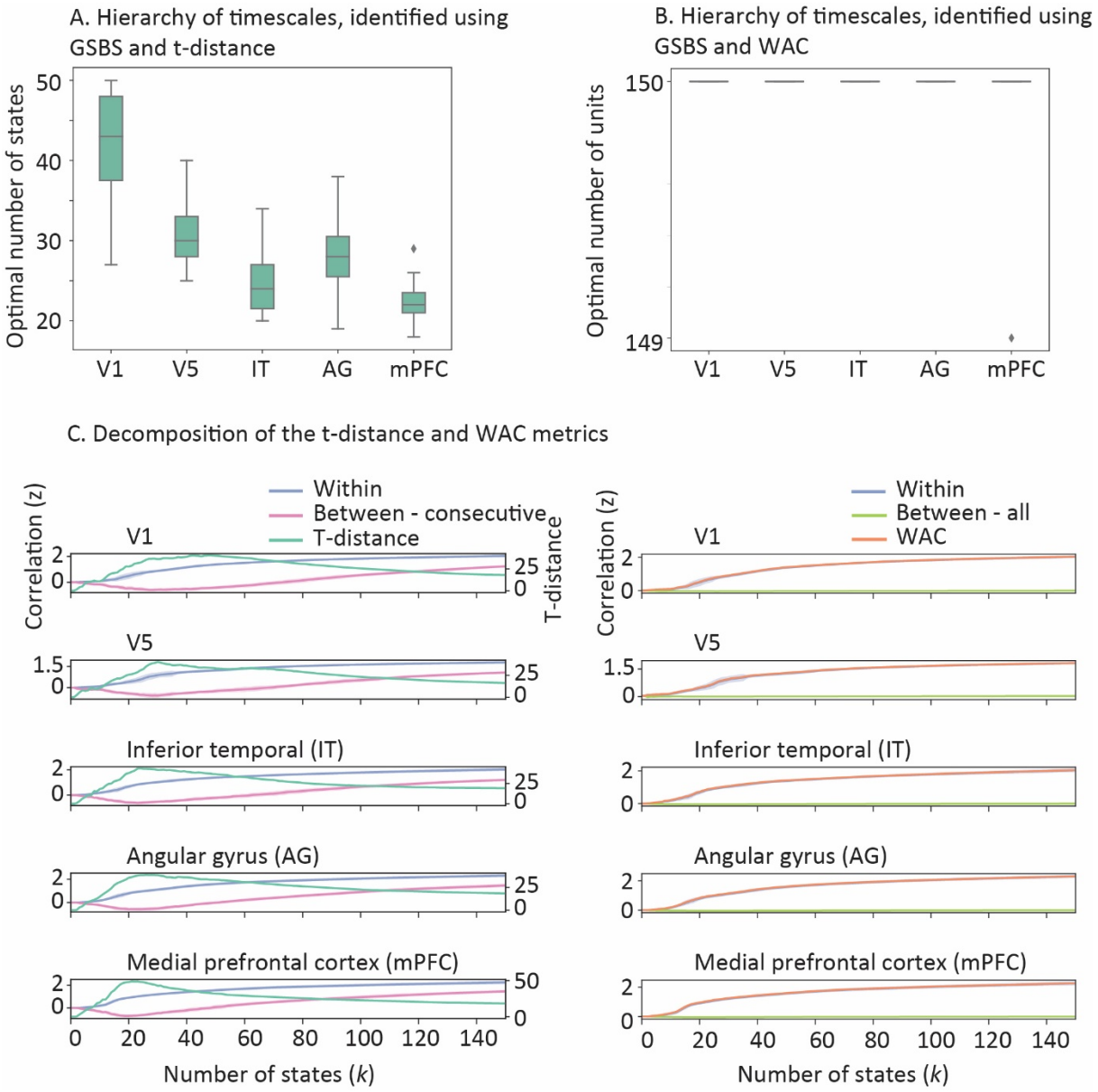


Figure 6. Comparison between the t-distance and WAC metrics for real data. A) Optimal number of states for t-distance. B) Optimal number of states for WAC. WAC is unable to identify the number of states in this dataset and results in an estimate of k=150 for each of the brain regions. C) Fit curves for both methods, showing how the fit estimates are driven by changes in their underlying components.

Simulations – Assumptions, averaging and noise

Now that we have established that the GSBS method combined with the t-distance metric are the best tools to estimate state boundaries in fMRI data, our next step is to investigate the role of potential confounds on the estimation of the location and the number of state boundaries. First, we investigate the role of noise, which we simulate here as BOLD responses generated by brain activity without a particular state structure and we investigate how noise affects the estimation of the number of state boundaries. We simulate data from 20 participants who each move through the same 15 states and we add varying levels of noise to each participants' data. When we identified state boundaries in each participant separately, we found that as the noise increases, the number of states was initially underestimated. However, as it increases further, the number of states was strongly overestimated (see light green bar in figure 7A). We also found that an increase in noise leads to a steady decline in the correlation between the simulated and estimated state boundary timecourses, when we assume that the number of states is known (see light green bar in figure 7C).

Next, we investigated two methods for reducing the impact of noise on the estimate of the number of states. One is to use cross-validation (as in Baldassano et al, 2017), in which the state boundaries are detected in one set of participants and the fit to determine the optimal number of states is derived from another set of participants. This reduces overfitting because the estimate of the number of states needs to be based on state boundaries that are shared across participants. The other method is to average the data across participants before estimating state boundaries and the optimal number of states. Both of these methods assume that state boundaries are the same across participants. This assumption is true in the current simulation.

We observed that both methods (averaging and cross-validation) resulted in more accurate estimations of the number and the location of state boundaries. However, with high amounts of noise, leave-one-out cross validation still led to a strong overestimation of the number of state boundaries. Using 2-fold cross-validation or averaging half of the data resulted in similar levels of performance. Averaging the data across all 20 simulated participants showed the best performance, both for estimating the number of state boundaries and for identifying the location of state boundaries. When all data were averaged, the state boundaries could be detected accurately even when the noise SD was 10x higher than the SD of the neural patterns that defined the states.

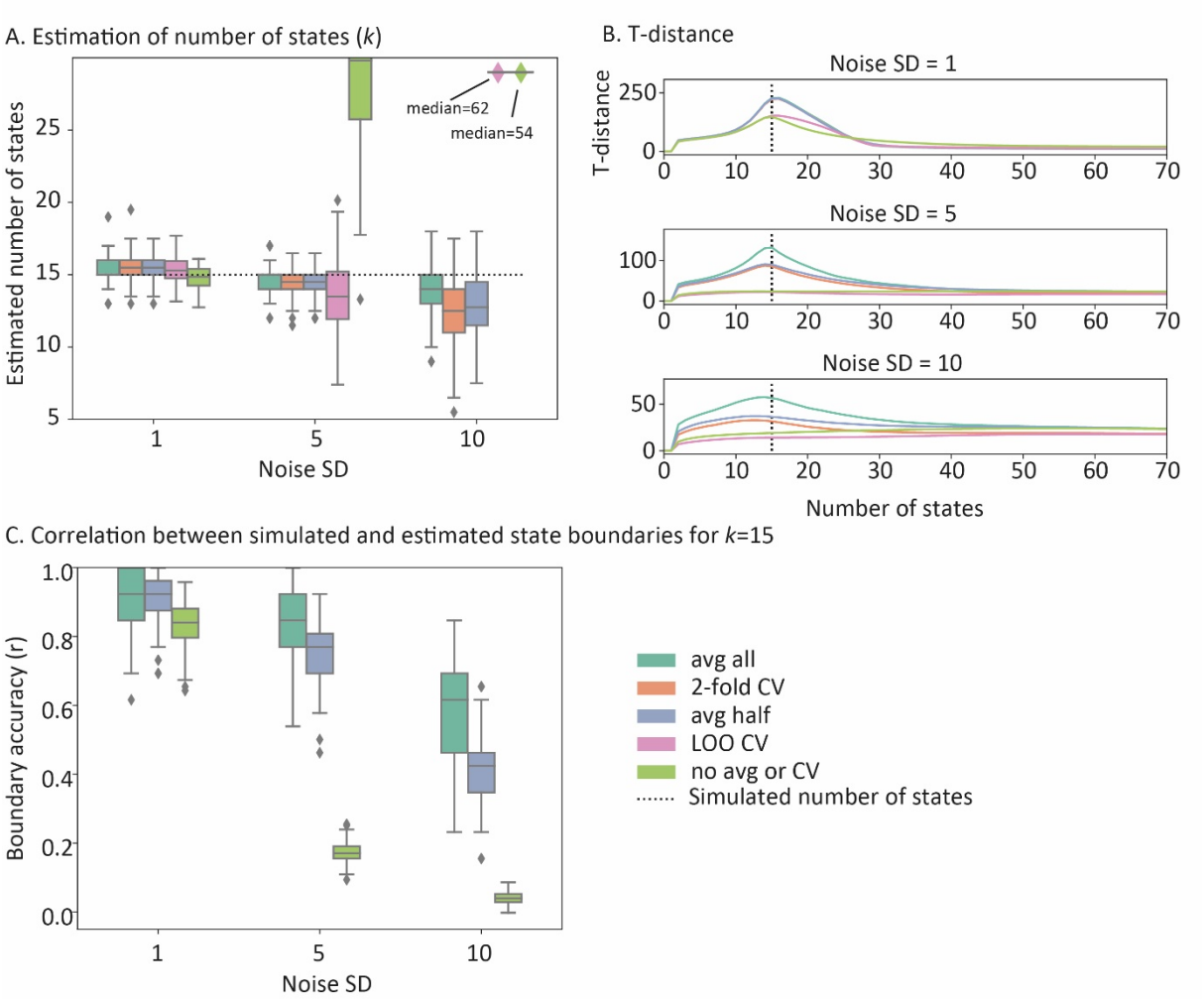


Figure 7. Comparing the effects of noise and data-averaging/cross validation on A. estimating the number of states B. The t-distance that is used to estimate the number of states C. the correlation between simulated and estimated state boundary timecourses. avg=average; LOO=leave-one-out; CV=cross validation.

When data are averaged across multiple participants, or when we use cross-validation, we assume that neural states are shared across participants. However, that assumption might not always be valid. In the next set of stimulations, we have examined whether it is possible to recover shared states when each subject also has some states that are not shared with the other participants. In particular, we simulated data in which participants always traversed 15 states. However, the proportion of states that was shared with other participants could vary. Specifically, we removed 10%, 20% or 40% of the group-level state boundaries in each participant and we randomly added the same number participant-specific states with unique activity patterns.

We found that leave-one-out CV performed poorly when the proportion of participant-specific states increased. In contrast, when we averaged the data across all participants, we were able to recover the simulated number of states correctly and we found a high correlation between the group-level simulated and estimated state boundaries (see figure 8).

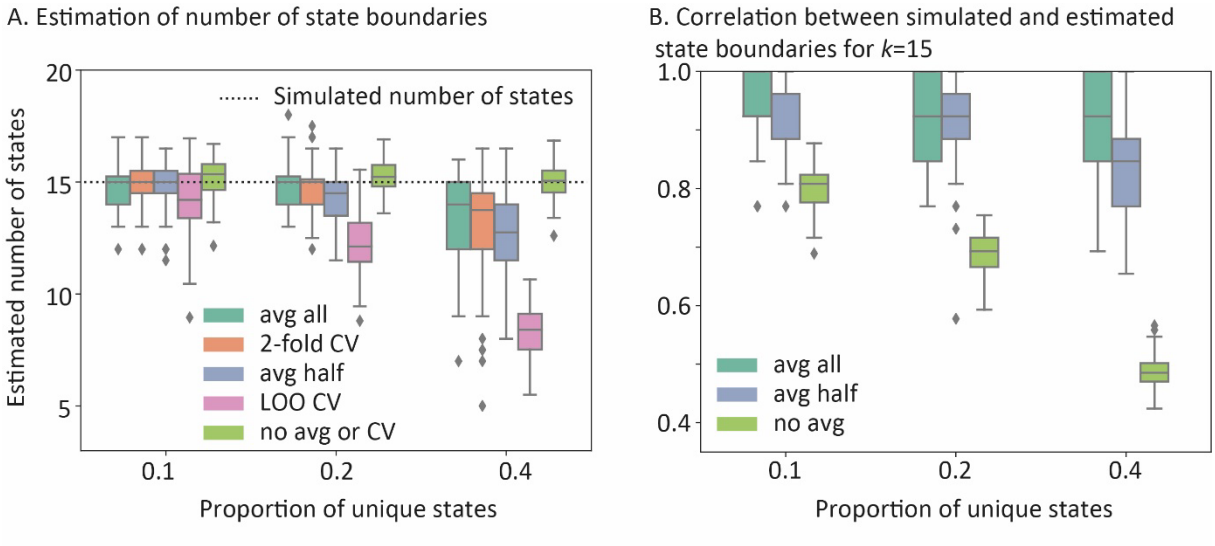


Figure 8. Comparing the effects of participant-specific states and data-averaging/cross validation on A. estimating the number of states B. the correlation between simulated and estimated state boundaries. avg=average; LOO=leave-one-out; CV=cross validation.

To investigate whether differences in the shape of the HRF might bias estimates of the number of neural states we performed an additional simulation. In particular, we varied the HRF peak (4, 6 or 8 s) and the dispersion of the HRF (0.5, 1 or 2 s). We found that differences in the HRF shape did not affect the estimation of the number of states (see supplementary figure 3).

Real data – data averaging

To get more insight into the optimal approach for analyzing real data, we compare different analysis choices in this final section. First, we investigated how data averaging affects the reliability of the recovered state boundaries and the estimated optimal number of states. Second, we investigated how these outcome measures are affected by the number of voxels in the searchlight.

In line with the results of our simulations, we found that the reliability of the state boundaries increased as the voxel timecourses were averaged over more participants (see figure 9A). For area IT and V1, we found that the reliability increase tapered off when more than 25 participants were averaged. For AG, the mPFC and area V5, we found that the reliability increase tapered off when more than 50 participants were averaged. For single-participant data, the similarity between each participant and the average of all other participants was low ($r = 0.1 - 0.3$), suggesting that in this dataset, state boundaries cannot be estimated reliably in single participant data. We found that the estimate of the optimal number of states is stable across the number of averaged participants, as long as that number is around 20 or higher. Below that, we observe an increase in the estimated optimum (see figure 9B). This is similar to the results we observe in our simulations for data with low signal to noise, suggesting that the method starts to overestimate the number of state boundaries due to noisy data. This is particularly clear when data from single participants is used.

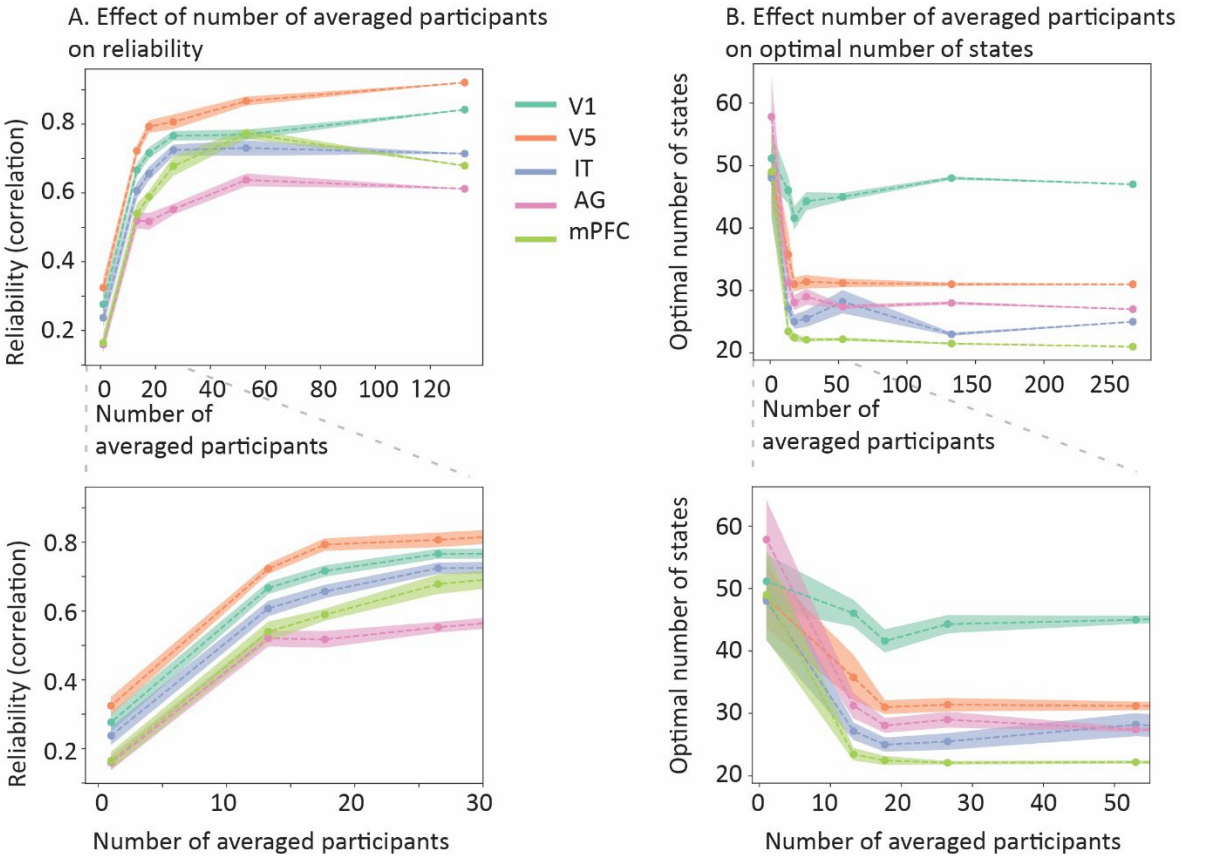


Figure 9: Investigating the effect of data averaging on A) the reliability of state boundaries and B) the estimated optimum number of states. The shaded area indicates the standard error across the independent (groups of) participants.

Averaging the voxel timecourses across participants allows us to isolate the BOLD signals that are evoked by the movie and shared across participants, from signals that have a non-neural original (e.g. head motion and respiration) as well as neural signals that are not evoked by the movie. However, the disadvantage of this approach is that noise is introduced when voxels are not correctly aligned across participants. By using hyperalignment, we attempted to optimally align voxels. However, this alignment will never be perfect. Therefore, we also explored what happens when, instead of averaging voxel timecourses, we average the temporal correlation matrices, shown in figure 1B (panel 1). This would reduce effects of voxel misalignment, but also decrease the noise reduction affects that we get from voxel timecourse averaging. To compare the two methods, we investigated the reliability of the temporal correlation matrices across 15 independent samples of the data (groups of participants). We found that in each ROI, the reliability was higher when we averaged the voxel timecourses than when we averaged the temporal correlation matrices (see supplementary figure 4). These results support our interpretation that averaging voxel timecourses and focusing on the movie-evoked signals allows us to more accurately identify shared state boundaries across participants.

Finally, we explored how different searchlight sizes affect the reliability and the optimal number of states (see supplementary figure 5). We found that the reliability is similar across different searchlight sizes. The estimate of the optimal number of states was stable for searchlights with a radius of 8 mm (+/- 80 voxels) or larger.

Discussion

Event segmentation is an important mechanism that allows us to understand and remember and organize ongoing sensory input. Recent work has suggested that event segmentation can be linked to regional changes in neural activity patterns (neural state boundaries). Accurate methods for identifying neural state boundaries are important to allow further investigation of the neural basis of event segmentation and its link to the temporal processing hierarchy of the brain. In this paper, we have introduced simple and effective new methods for identifying the number of neural states in a brain area, as well as the location of the boundaries between those states; greedy state boundary search (GSBS) and t-distance. We have used a comprehensive set of simulations as well as analyses of real fMRI data to show that these methods outperform an existing method based on an adapted version of HMMs (Baldassano et al., 2017). In addition, we have investigated the impact of noise of the estimation of the location and number of state boundaries, and how this can be mitigated by data averaging.

The GSBS method we introduce here, differs from the HMM-based state segmentation method in a number of important ways. First, the HMM-based method makes implicit assumptions about the duration of neural states. In particular, the method assumes that all states are fixed to have the same probability of staying in the same state versus jumping to the next state. This causes the method to tend to identify states with the same duration, as we observed in our simulations as well as our real data analyses. The GSBS method performs an exhaustive search, in which no assumptions are made about the location of state boundaries. Second, the HMM-based method identifies a new set of states for each value of *k* (the number of states), while the GSBS method perform an iterative search, such that all but one of the boundaries that are detected for *k* states are the same as the boundaries that are detected for *k+1* states. Combined with the simplicity of the GSBS method, this results in an up to 1400-fold increase in computations speed. This makes the GSBS-method very suitable to detect state boundaries across many different brain regions in a whole-brain searchlight approach. The iterative nature of the GSBS method also results in an automatic ordering of state boundaries, as the strongest boundaries (the biggest change in neural patterns) will be detected first.

For both methods, a separate measure is needed to identify the optimal number of states. Here we introduced the t-distance metric, which maximizes a t-statistic that reflects the distance between the within-state correlations and the correlations between subsequent states. The t-distance was able to accurately recover the number of simulated states in the simulated data and also resulted in the expected temporal hierarchy for the real data. In contrast, the original WAC metric by Baldassano et al. (2017) tended to overestimate the number of states unless the number of states was much smaller than the number of timepoints. Most likely, this is due to the autocorrelation introduced by the HRF. Note that in the real data analyses that were presented by Baldassano et al. (2017), the number of states was constrained to be much smaller than the number of timepoints (max. 120 states in 50 minutes), which is perhaps the reason that they did not encounter this problem. Ideally, we would use a single method to identify the number of states as well as the location of state boundaries. One such method is the sticky hierarchical Dirichlet process HMM (Fox et al., 2009), which allows for differing state durations and includes an estimation for the number of states. However, we found that this approach did not work well for our simulated datasets. An exhaustive comparison between all methods of state (boundary) detection and change point detection (Aminikhanghahi and Cook, 2017; e.g. Cribben et al., 2012; Taghia et al., 2018; Xu and Lindquist, 2015) was beyond the scope of the current paper.

In addition to introducing new methods, we also performed extensive simulations and empirical analyses to investigate the effect of noise on the state boundary estimations and to examine how

averaging the data can mitigate this effect. Our simulations showed that high levels of noise (signal to noise ratio of 1/10) result in an overestimation of the number of states and poor reconstruction of the state boundary locations. When we assume that states are shared between participants, e.g. because they are watching the same movie, we find that averaging the data allows us to estimate the number and location of state boundaries correctly even in these very high noise regimes. Our empirical data analyses showed overestimation of the number of states when boundaries were estimated on non-averaged single subject data, but not when data were averaged over 17 or more participants. This suggest that the state-changes we identified were driven by the neural signal that was evoked by the movie rather than the 'background' neural signals that were not shared between participants. Another set of simulations showed that state changes can be detected reliably on the group-level even if there is some inter-individual variability in the states that are visited by participants.

The methods we introduced here were optimized specifically for estimating regional state boundaries in fMRI data. However, they are also be applicable in other settings, such as investigating functional connectivity states (Allen et al., 2014; Damaraju et al., 2014; Wang et al., 2016) or investigating state boundaries in electrophysiological data (Borst and Anderson, 2015; Silva et al., 2019; Vidaurre et al., 2016).

To conclude, we have introduced a set of simple and computationally fast new methods that allow researchers to estimate state boundaries in fMRI data. These methods were validated using real and simulated data, giving us good insights in how they should be used to answer empirical questions. These methods will give researchers new, well-validated tools to investigate state-boundaries in neural data and to investigate the neural underpinnings of event segmentation.

Acknowledgements

LG is supported by a Veni grant [451-16-013] from the Netherlands Organization for Scientific Research. MVG is supported by a Vidi grant [639.072.513] from the Netherlands Organization for Scientific Research.

References

Allen, E. a, Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., Calhoun, V.D., 2014. Tracking whole-brain connectivity dynamics in the resting state. Cereb. Cortex 24, 663–76. https://doi.org/10.1093/cercor/bhs352

Aminikhanghahi, S., Cook, D.J., 2017. A Survey of Methods for Time Series Change Point Detection. Knowl. Inf. Syst. 51, 339–367. https://doi.org/10.1007/s10115-016-0987-z

Baldassano, C., Chen, J., Zadbood, A., Pillow, J.W., Hasson, U., Norman, K.A., 2017. Discovering Event Structure in Continuous Narrative Perception and Memory. Neuron 95, 709-721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Borst, J.P., Anderson, J.R., 2015. The discovery of processing stages: Analyzing EEG data with hidden semi-Markov models. Neuroimage 108, 60–73. https://doi.org/10.1016/j.neuroimage.2014.12.029

Buonomano, D. V., Maass, W., 2009. State-dependent computations: spatiotemporal processing in cortical networks. Nat. Rev. Neurosci. 10, 113–125. https://doi.org/10.1038/nrn2558

Cribben, I., Haraldsdottir, R., Atlas, L.Y., Wager, T.D., Lindquist, M. a, 2012. Dynamic connectivity regression: determining state-related changes in brain connectivity. Neuroimage 61, 907–20. https://doi.org/10.1016/j.neuroimage.2012.03.070

Damaraju, E., Allen, E.A., Belger, A., Ford, J.M., McEwen, S., Mathalon, D.H., Mueller, B.A., Pearlson, G.D., Potkin, S.G., Preda, A., Turner, J.A., Vaidya, J.G., Van Erp, T.G., Calhoun, V.D., 2014. Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. NeuroImage Clin. 5, 298–308. https://doi.org/10.1016/j.nicl.2014.07.003

Flores, S., Bailey, H.R., Eisenberg, M.L., Zacks, J.M., 2017. Event segmentation improves event memory up to one month later. J. Exp. Psychol. Learn. Mem. Cogn. 43, 1183–1202. https://doi.org/10.1037/xlm0000367

Fox, E.B., Sudderth, E.B., Jordan, M.I., Willsky, A.S., 2009. The sticky hdp-hmm: bayesian nonparametric hidden markov models with persistent states. arXiv. https://doi.org/10.1214/10-AOAS395

Geerligs, L., Cam-CAN, Campbell, K.L., 2018. Age-related differences in information processing during movie watching. Neurobiol. Aging 72, 106:120. https://doi.org/10.1016/j.neurobiolaging.2018.07.025

Guntupalli, J.S., Hanke, M., Halchenko, Y.O., Connolly, A.C., Ramadge, P.J., Haxby, J. V, 2016. A Model of Representational Spaces in Human Cortex. Cereb. Cortex 26, 2919–2934. https://doi.org/10.1093/cercor/bhw068

Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J. V., Pollmann, S., 2009. PyMVPA: a Python Toolbox for Multivariate Pattern Analysis of fMRI Data. Neuroinformatics 7, 37–53. https://doi.org/10.1007/s12021-008-9041-y

Hasson, U., Chen, J., Honey, C.J., 2015. Hierarchical process memory: Memory as an integral component of information processing. Trends Cogn. Sci. 19, 304–313. https://doi.org/10.1016/j.tics.2015.04.006

Hasson, U., Malach, R., Heeger, D.J., 2009. Reliability of cortical activity during natural stimulation. Trends Cogn. Sci. 14, 40–48. https://doi.org/10.1016/j.tics.2009.10.011

Honey, C.J., Thesen, T., Donner, T.H., Silbert, L.J., Carlson, C.E., Devinsky, O., Doyle, W.K., Rubin, N., Heeger, D.J., Hasson, U., 2012. Slow cortical dynamics and the accumulation of information over long timescales. Neuron 76, 423–34. https://doi.org/10.1016/j.neuron.2012.08.011

Kiebel, S.J., Daunizeau, J., Friston, K.J., 2008. A Hierarchy of Time-Scales and the Brain. PLoS Comput. Biol. 4, e1000209. https://doi.org/10.1371/journal.pcbi.1000209

Kurby, C.A., Zacks, J.M., 2008. Segmentation in the perception and memory of events. Trends Cogn. Sci. 12, 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Lerner, Y., Honey, C.J., Silbert, L.J., Hasson, U., 2011. Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. J. Neurosci. 31, 2906–2915. https://doi.org/10.1523/JNEUROSCI.3684-10.2011

Newtson, D., Engquist, G.A., Bois, J., 1977. The objective basis of behavior units. J. Pers. Soc. Psychol. 35, 847–862. https://doi.org/10.1037/0022-3514.35.12.847

Sargent, J.Q., Zacks, J.M., Hambrick, D.Z., Zacks, R.T., Kurby, C.A., Bailey, H.R., Eisenberg, M.L., Beck, T.M., 2013. Event segmentation ability uniquely predicts event memory. Cognition 129, 241–255. https://doi.org/10.1016/j.cognition.2013.07.002

Shafto, M.A., Tyler, L.K., Dixon, M., Taylor, J.R., Rowe, J.B., Cusack, R., Calder, A.J., Marslen-Wilson, W.D., Duncan, J., Dalgleish, T., Henson, R.N., Brayne, C., Cam-CAN, Matthews, F.E., 2014. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: A cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. BMC Neurol. 14. https://doi.org/10.1186/s12883-014-0204-1

Silva, M., Baldassano, C., Fuentemilla, L., 2019. Rapid Memory Reactivation at Movie Event Boundaries Promotes Episodic Encoding. J. Neurosci. 39, 8538–8548. https://doi.org/10.1523/JNEUROSCI.0360-19.2019

Taghia, J., Cai, W., Ryali, S., Kochalka, J., Nicholas, J., Chen, T., Menon, V., 2018. Uncovering hidden brain state dynamics that regulate performance and decision-making during cognition. Nat. Commun. 9, 1–19. https://doi.org/10.1038/s41467-018-04723-6

Taylor, J.R., Williams, N., Cusack, R., Auer, T., Shafto, M. a., Dixon, M., Tyler, L.K., Cam-CAN, Henson, R.N., 2017. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. Neuroimage 144, 262–269. https://doi.org/10.1016/j.neuroimage.2015.09.018

Vidaurre, D., Quinn, A.J., Baker, A.P., Dupret, D., Tejero-Cantero, A., Woolrich, M.W., 2016. Spectrally resolved fast transient brain states in electrophysiological data. Neuroimage 126, 81–95. https://doi.org/10.1016/j.neuroimage.2015.11.047

Wang, C., Ong, J.L., Patanaik, A., Zhou, J., Chee, M.W.L., 2016. Spontaneous eyelid closures link vigilance fluctuation with fMRI dynamic connectivity states. Proc. Natl. Acad. Sci. U. S. A. 113, 9653–9658. https://doi.org/10.1073/pnas.1523980113

Xu, Y., Lindquist, M.A., 2015. Dynamic connectivity detection: an algorithm for determining functional connectivity change points in fMRI data. Front. Neurosci. 9. https://doi.org/10.3389/fnins.2015.00285

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8, 665–670. https://doi.org/10.1038/nmeth.1635

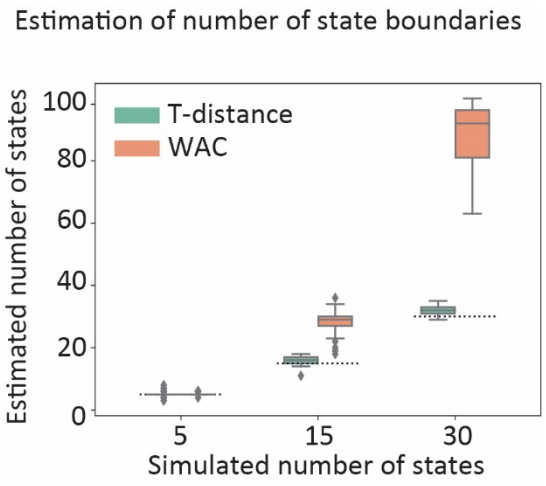
Zacks, J.M., Speer, N.K., Swallow, K.M., Braver, T.S., Reynolds, J.R., 2007. Event perception: A mind-

brain perspective. Psychol. Bull. 133, 273–293. https://doi.org/10.1037/0033-2909.133.2.273

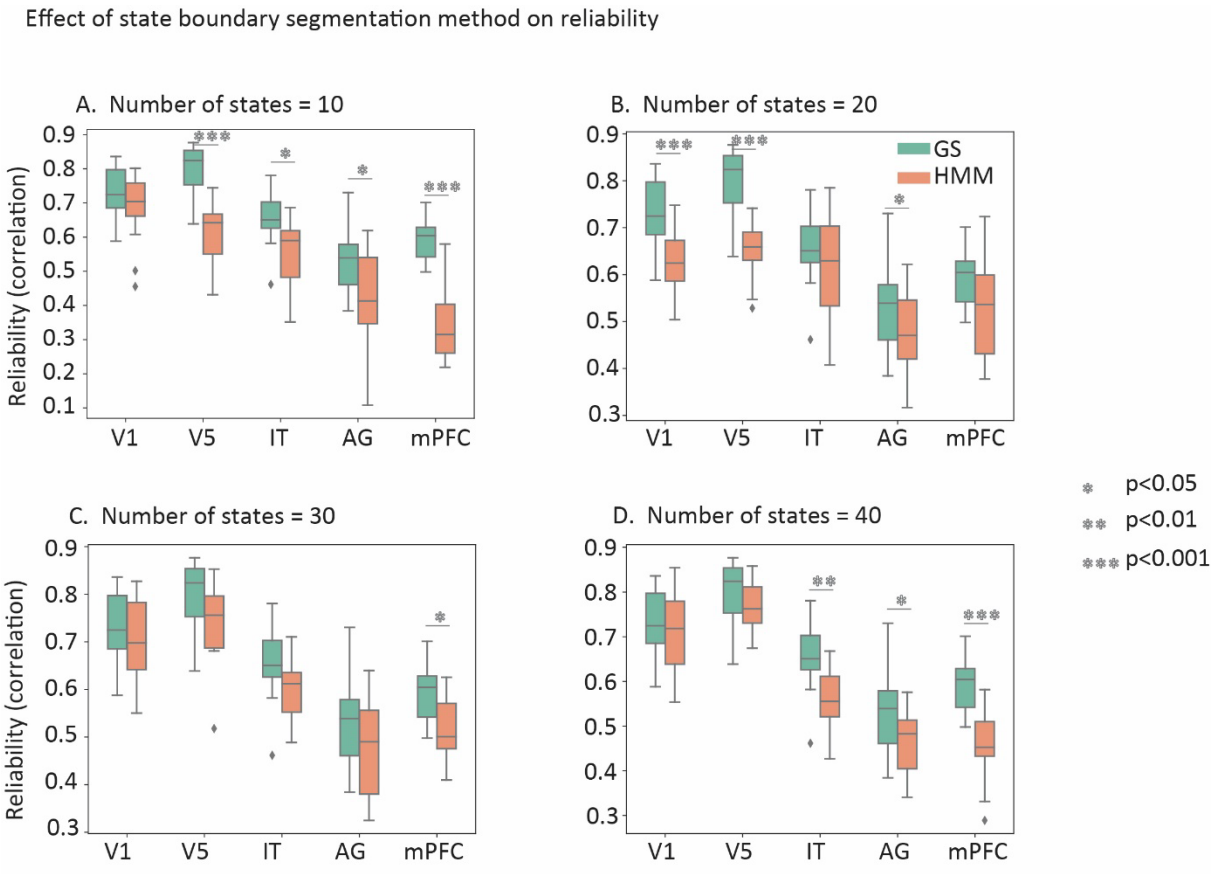
Zacks, J.M., Speer, N.K., Vettel, J.M., Jacoby, L.L., 2006. Event understanding and memory in healthy aging and dementia of the Alzheimer type. Psychol. Aging 21, 466–482. https://doi.org/10.1037/0882-7974.21.3.466

Zacks, J.M., Tversky, B., Iyer, G., 2001. Perceiving, remembering, and communicating structure in events. J. Exp. Psychol. Gen. 130, 29–58.

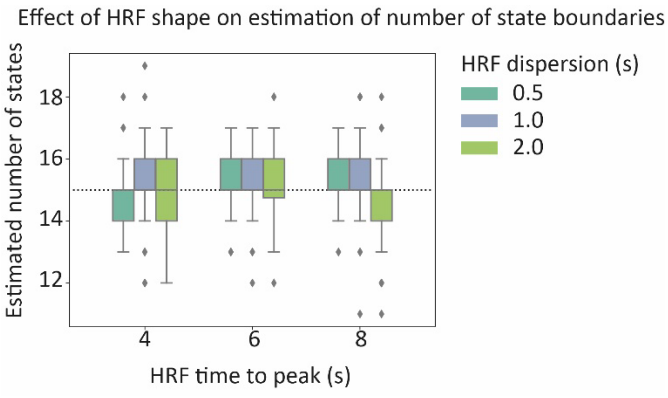
Supplementary figures



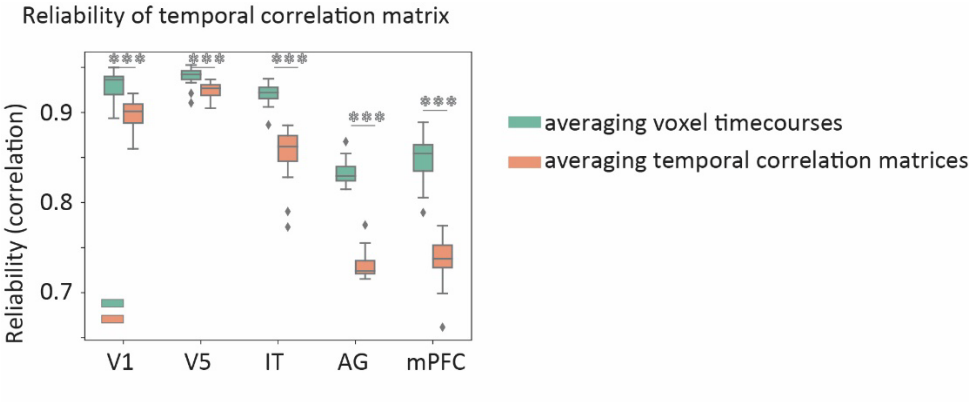
Supplementary figure 1: Comparison of the estimated and simulated number of states for the two fit methods. In this analysis, the 4 TRs around the diagonal of the correlation matrix were not taken into account in the computation of the WAC and t-distance metric (as in Baldassano et al. 2017).



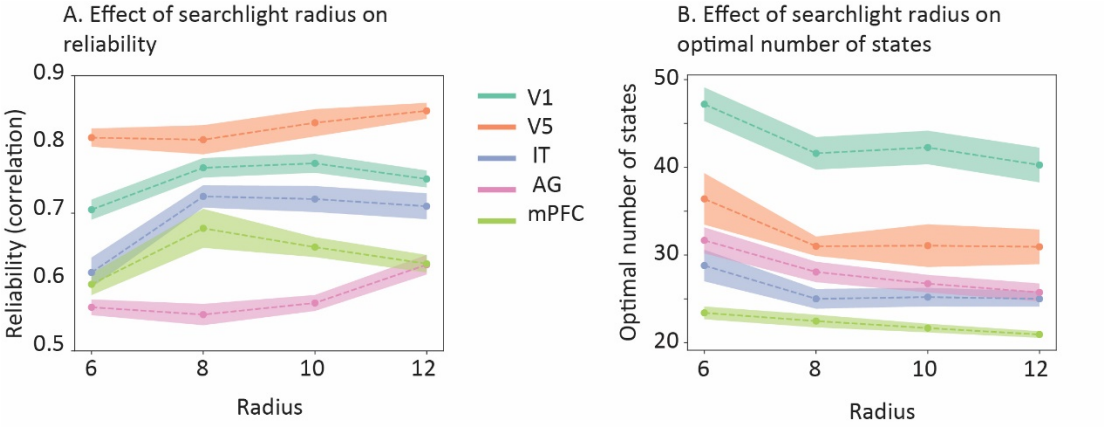
Supplementary figure 2: Reliability of state boundaries detected in real data using GSBS or HMM-based method for different numbers of states.



Supplementary figure 3. Effect of the HRF shape on the estimation of the number of state boundaries.



Supplementary figure 4. Effect of averaging voxel timecourses versus averaging temporal correlation matrices of the reliability of the temporal correlation matrix.



Supplementary figure 5. Effect of searchlight radius on 1) reliability of state boundaries and 2) optimal number of states.