

Timescales and functional organization of neural event segmentation in the human brain

Linda Geerligs¹, Marcel van Gerven¹, Karen Campbell² and Umut Güçlü¹

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

2. Department of Psychology, Brock University, St. Catharines, Canada

Correspondence to:

Linda Geerligs

Donders Institute for Brain, Cognition and Behaviour

Montessorilaan 3 6525 HR Nijmegen, The Netherlands

+31 243616091

lindageerligs@gmail.com

Keywords: Event segmentation, naturalistic viewing, fMRI, Hidden Markov Model, timescales, functional networks, connectivity, fMRI

Abstract

During continuous perception, humans automatically segment their experience into discrete events. At the neural level, each brain region is thought to segment ongoing experience at its own preferred timescale. Previous studies have shown that event segmentation follows a temporal hierarchy in the brain, with regions higher in the hierarchy segmenting experience into events with a longer duration. The assumption is that incoming information is processed and then passed on to higher order regions as discrete units (or chunks), but little is known about how event boundaries are shared both within and across networks. To this end, we use Hidden Markov Models (HMMs) and functional connectivity analysis to study the timescales of event segmentation over the entire cortex and investigate how event boundaries are shared across regions and networks. We demonstrate that there is a distinct temporal gradient of information processing over the entire cortex, with particularly fast events in primary sensory regions and long periods of information integration in the precuneus and medial prefrontal cortex. We also show that event boundaries are shared between regions within functional networks, as well as across the temporal hierarchy between distinct networks. Finally, we observed that the traditional default mode and fronto-parietal networks fractionate into subnetworks with fast and slow event timescales, possibly reflecting functional specialization within these networks during naturalistic viewing. Together, these results provide the first complete overview of how event segmentation is organized in the human brain.

Introduction

Segmentation of information into meaningful units is a fundamental feature of our conscious experience in real life contexts. Spatial information processing is characterized by chunking spatial regions into objects (e.g. DiCarlo and Cox, 2007). In a similar way, temporal information processing is characterized by segmenting our ongoing experience into separate events (Kurby and Zacks, 2008; Newton et al., 1977). Segmentation improves our understanding of ongoing perceptual input (Zacks et al., 2001) and allows us to recall distinct event from our past (Flores et al., 2017; Sargent et al., 2013; Zacks et al., 2006). Recent work has shown that the end of an event triggers an evoked response in the hippocampus (Baldassano et al., 2017; Ben-Yakov and Henson, 2018), suggesting that events form the basis of long-term memory representations. Events that are identified in movie or auditory narratives are often very similar across individuals and can be segmented hierarchically on different timescales (Newton and Rindner, 1979; Zacks et al., 2001). According to event segmentation theory (EST) perceptual systems spontaneously segment activity into meaningful events as a side-effect of predicting future information (Zacks et al., 2007). An event boundary occurs when the predictions become less accurate, which can be due to a change in motion or features of the situation such as characters, causes, goals and spatial location (Zacks et al., 2010).

While much is known about temporal event segmentation on a behavioral level, its neural underpinnings remain largely unknown. Recent work aimed at examining event segmentation at the neural level has suggested that each brain region chunks experience at its own preferred timescale (Baldassano et al., 2017). This is in line with predictions from the hierarchical process memory framework, which posits that all brain regions accumulate information over time, but the timescale of this accumulation varies hierarchically (Hasson et al., 2015). Early sensory regions have short processing timescales (as fast as 10-100 ms) and along the cortical hierarchy, the timescales gradually lengthen up to the scale of seconds or minutes in areas like the angular gyrus and precuneus.

This hierarchy of information accumulation has been demonstrated using movies and auditory narratives that have been scrambled at different timescales, for instance at the word, sentence, or

paragraph level (Hasson et al., 2008; Honey et al., 2012b; Lerner et al., 2011). This work has shown that higher cortical areas (e.g., precuneus, temporal parietal junction, and frontal eye fields) require more temporal context to respond reliably across participants. These regions also show slower resting-state fluctuations in electrocorticography (Honey et al., 2012b) and functional MRI data (Stephens et al., 2013), suggesting that a region's preferred timescale of information processing is an intrinsic property of that region (i.e., not stimulus dependent). More recently, Baldassano and colleagues (2017) examined this temporal hierarchy using a different approach; they employed a modified version of Hidden Markov Models (HMM) to identify event structure in fMRI data of continuous movie viewing using a data-driven approach (Baldassano et al., 2017). They showed that neural event segmentation occurs on a timescale predicted by the hierarchical process-memory framework, ranging from short events in sensory regions to long events in higher order areas. However, they were only able to investigate event segmentation in a subset of cortical areas that responded consistently to the movie (across participants). Based on an analysis of four regions of interest they suggested that brain regions might share event boundaries in a hierarchical fashion, with the boundaries of faster events in regions lower in the cortical hierarchy nested within the boundaries of slower regions higher up the hierarchy.

The aim of the current study was to characterize the temporal hierarchy of event segmentation over the entire cortex and to use network analysis to determine how events are shared across distributed networks throughout the cortex. We did this by using the HMM approach proposed by Baldassano et al. (2017) and adjusting their method for detecting the number of events that best characterizes each brain region (see figure 1). The employed dataset was part of the Cam-CAN project and consists of fMRI data from a large number of subjects who were presented with an engaging movie stimulus (Shafto et al., 2014). This dataset, in combination with the application of hyperalignment to optimize functional alignment (Guntupalli et al., 2016), allowed us to study event segmentation across the entire cortex for the first time, because this dataset shows reliable stimulus-driven activity (significant inter-subject correlations) over nearly all cortical brain regions (Geerligs et al., 2018). In addition, we were able to investigate whether neural events are shared across brain regions that are part of the same functional network, as well as across different networks at different levels of the temporal hierarchy. We hypothesized that events identified by participant ratings should be related to neural events at specific levels of the hierarchy, particularly in the default mode network (Baldassano et al., 2017; Zacks et al., 2001; Zacks et al., 2010). Finally, we expected that event timescales are associated with the neural dynamics of brain regions, as measured by their power spectrum (Honey et al., 2012b; Stephens et al., 2013) and we investigated whether the observed event timescales during movie watching are an intrinsic property of these brain regions that is also present in the resting state.

Methods

Participants

We used the 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 (Shafto et al., 2014). 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.

Movie

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

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) and high-resolution (1 mm x 1mm x 1 mm) T1 and T2-weighted images were acquired.

Data pre-processing

The initial steps of data preprocessing for the movie data 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. ME-ICA is a very promising method for removal of non-BOLD like 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>). Pre-processing for the resting state data included slice time correction, realignment, co-registration and DARTEL intersubject alignment.

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 event segmentation model described below was applied to group averaged voxel-level data, good inter-subject alignment was 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.

Event segmentation Model

To identify discrete events in the fMRI data, we used a variant of a Hidden Markov Model (HMM), created by Baldassano et al. (2017). The HMM models the brain activity as a sequence of hidden (unobserved) states. Each state is characterized by a specific mean activity pattern across voxels within a searchlight volume. In contrast to regular HMMs, this variant is constrained such that there are no recurring states. From one time point to the next, a brain region can either stay in the same state or jump to the next state. Therefore, the first time point 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 events. The inputs to the HMM consists of a set of z-scored voxel timecourses within a searchlight and a value for K (the number of states that needs to be estimated). In the context of the current project, we refer to these HMM states as events.

Timescales of event segmentation

We applied the event segmentation model in a searchlight to the hyperaligned data within the Harvard-Oxford cortical mask. Spherical searchlights were scanned throughout the brain with a step size of two voxels and a radius of three voxels. This resulted in searchlights with an average size of 97 voxels (max: 123, IQR: 82-115), this variation in searchlight size was due to the exclusion of out-of-brain voxels. Voxels were also excluded from the analysis if they had an intersubject correlation below $r=0.35$. Only searchlights with more than 15 voxels were included in the analysis.

In each searchlight, the model was fit for varying numbers of events, ranging from $K=2$ to $K=90$. To determine the optimal number of events in each searchlight, we used 10-fold cross validation. In each fold, 9/10th of the data was used as the training set. These data were averaged across participants and the event segmentation model was applied to the averaged data to identify the event sequence for each value of K. Prior to averaging, the hyperaligned data were highpass-filtered with a cut-off of 0.008 Hz. The final 1/10th of the data was used as a test set to determine the fit of the identified event sequence.

The following procedure was used to quantify the fit (see figure 1). First, we computed time-by-time correlation matrices for both the training and the test datasets. Each index in this matrix reflects the similarity of a searchlight's activity pattern between two time points. However, for the training dataset, we first represented each timepoint within an event by the events' mean activity pattern across voxels within the searchlight. Next, the correlation between the flattened training and test matrices was computed, excluding timepoints with a distance of 5 TRs or less. This correlation tells us how well the identified event structure in the training data represents the similarity between time points in the test data for a given searchlight. We will refer to this correlation between the training and test set as the 'real boundary fit' (blue line in Figure 1).

As more events are modeled, the real boundary fit tends to increase. Therefore, to determine the optimal value of K for a given searchlight, the fit was compared to a baseline. This baseline was created by randomly reordering the event boundaries, while preserving the distribution of event lengths, and repeating the computation of the fit. This was repeated 500 times and an average baseline fit was computed (red line in Figure 1). This is referred to as the 'random boundary fit'. After averaging the results across each of the 10 folds, the number of events with the highest difference between the real and random boundary fits (green line in Figure 1) was chosen as the optimal number of events for this searchlight.

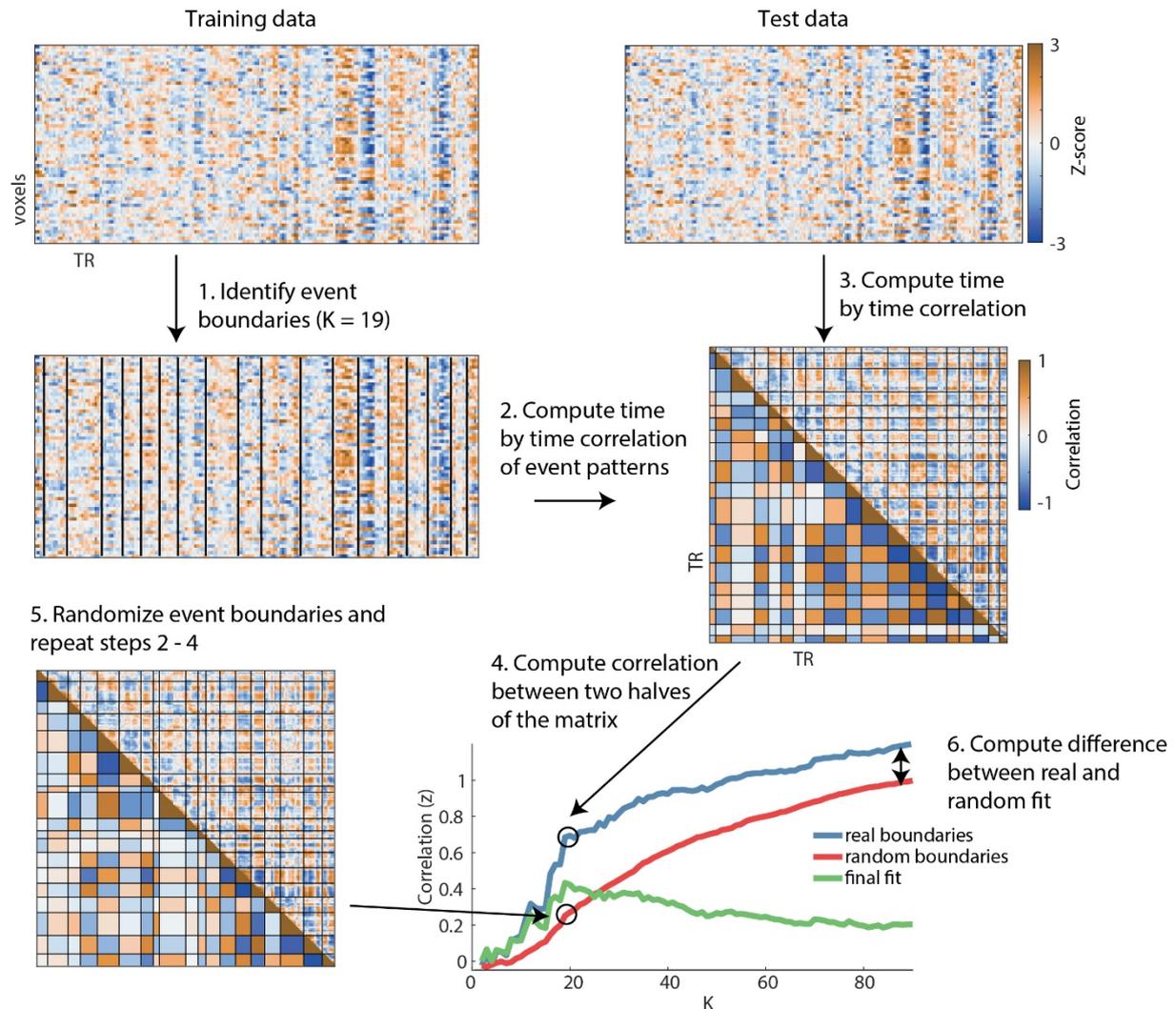


Figure 1: Depiction of the event segmentation and the procedure that was used to determine the optimal number of events for each region. In step 1, the training data is used to determine the event boundaries for a specific value of K , using the HMM event segmentation model. In step 2 the average event patterns are computed based on the training data and the correlation between the event patterns at each time point is computed, resulting in a time by time correlation matrix. In step 3, the time by time correlation matrix of the test set data is computed. In step 4, the correlation is computed between the two flattened halves of the correlation matrices computed in steps 2 and 3. Correlations between timepoints less than 5 TRs apart were not taken into account. In step 5, a baseline fit is computed by repeating steps 2 and 4 for a set of randomized event boundaries. In step 6, the

difference between the real and the random fit is computed (green line). The optimum of this difference is used as the optimal timescale of this searchlight.

To assess whether the detected even boundaries explained the data significantly better than expected by chance, we computed pseudo z-values for each value of K and each searchlight. This was done by subtracting the average random boundary fit from the real boundary fit, and dividing it by the standard deviation of the random boundary fit. This pseudo z-value was averaged across the different cross validation folds. Then it was converted to a p-value, using the standard normal distribution. P-values were Bonferonni corrected for all possible values of K. Across searchlights, we assessed significance by using the "two-stage" procedure by Benjamini et al. (2006) for controlling the false discovery rate (FDR) of a family of hypothesis tests. In all of the searchlights, the real event boundaries explained the data significantly better than the random event boundaries.

Fast and slow regions

To investigate whether some regions had significantly shorter or longer event durations than the average duration across the brain, we computed the optimal number of events within each of the 10-folds for each searchlight. This set of 10 optima was the input to a Wilcoxon rank sum test, which was used to determine whether the observed optima in each searchlight were significantly different from the mean optima across all of the searchlights. P-values were corrected for multiple comparisons using FDR correction.

Identification of functional networks

In order to identify networks of regions that contained the same event structure, and not simply the same duration of events, we first calculated a matrix of Pearson correlations between the boundary vectors in each of the searchlights. Functional networks were then detected using a consensus partitioning algorithm (Lancichinetti and Fortunato, 2012), as implemented in the Brain Connectivity Toolbox (Rubinov and Sporns, 2010). An initial partition into functional networks was created using the Louvain modularity algorithm (Blondel et al., 2008), which was refined using a modularity fine-tuning algorithm (Sun et al., 2009). This partitioning was repeated 1000 times, and all repetitions were combined into a consensus matrix. Each element in the consensus matrix indicates the proportion of repetitions in which the corresponding two searchlights were assigned to the same network. This matrix was used as the input for a new partitioning, until the algorithm converged to a single partition (such that the final consensus matrix consisted only of ones and zeroes). The procedure described above was applied for multiple resolutions (varying gamma between 1 and 3; Reichardt and Bornholdt, 2006). The most stable partitioning (highest normalized mutual information between solutions at different resolutions) was used as our final set of functional networks (gamma=1.8). For gamma values over 2.3, the modularity algorithm resulted in many single region partitions so these were not considered.

We also investigated whether events may be shared in a hierarchical fashion, across different functional networks. To this end, we computed partial correlations between the averaged boundary vectors within each functional network, controlling for the boundary vectors in all other networks. P-values were corrected for multiple comparisons using FDR correction.

Comparison of event boundaries across regions and to human annotations

Subjective annotations on the occurrence of event boundaries in the Cam-CAN movie dataset were collected by Ben-Yakov and Henson (2018). These annotations were based on sixteen observers who watched the movie outside the scanner and indicated with a keypress when they felt “one event (meaningful unit) ended and another began”. Only boundaries identified by a minimum of five observers were included, resulting in a total of 19 boundaries separated by 6.5–93.7 s.

To compare the event boundaries across regions and to the human annotations, we computed Pearson correlations between boundary vectors. These boundary vectors contained zeros for within-event time points and ones for moments where the current event did not match the subsequent event (i.e., event boundaries). For the comparison to human annotations, we refit the event segmentation model for each cortical region using the same number of events as in the human annotation (rather than using each region’s optimal timescale) to ensure that differences between regions were not driven by timescale differences. P-values were corrected for multiple comparisons using FDR correction.

Power spectra

To investigate whether the event timescales could be predicted from regional differences in neural dynamics (i.e., the speed at which populations of neurons within a given region oscillate; Stephens et al., 2013) we performed a spectral analysis. The voxel time series were normalized to unit variance to ensure that the spectra have the same integrated area. The power spectral density (PSD) was estimated using Welch’s method (Welch, 1967) with a Hamming window of width 50 TRs and 50% overlap. Next, we used the PSD to predict the event timescales across all searchlights (the lowest included frequency was 0.01 Hz). Because the PSD estimates were highly correlated across frequencies we decomposed the data using PCA. The PSD estimates were normalized to zero mean and unit variance within each searchlight before the PCA was performed. We used leave-one-searchlight-out cross validation to determine how many principal components needed to be included in the regression to optimally predict the event timescales. In this cross validation, searchlights with voxels close to the left-out searchlight (with a Euclidean distance of three voxels or less) were excluded from the regression model estimation. PCA components were included in the regression model in the order of the percentage of variance they explained in the PSD data. We found an optimum of four principal components. These components were used to estimate a regression model based on the complete movie dataset. The obtained beta estimates were projected back to the original space by multiplying them with the PCA coefficients. Next, to investigate whether the estimated timescales were specific to the movie dataset, we used the same beta weights to predict the event timescales in the movie from the PSD in a resting state dataset.

Data visualization

In all analyses, p-values from the searchlights were projected to the voxel level and averaged across the searchlights that overlapped each voxel before they were thresholded using the FDR-corrected critical p-value (Benjamini et al., 2006). When projecting the results of the analyses to the voxel-level, we excluded voxels for which less than half of the searchlights that covered that voxel were included in the analysis. The excluded searchlights either had too few voxels or too few voxels with sufficient inter-subject correlation (see section Timescales of event segmentation). Data were projected to the surface for visualization, using the Pysurfer and Caret toolboxes.

Results

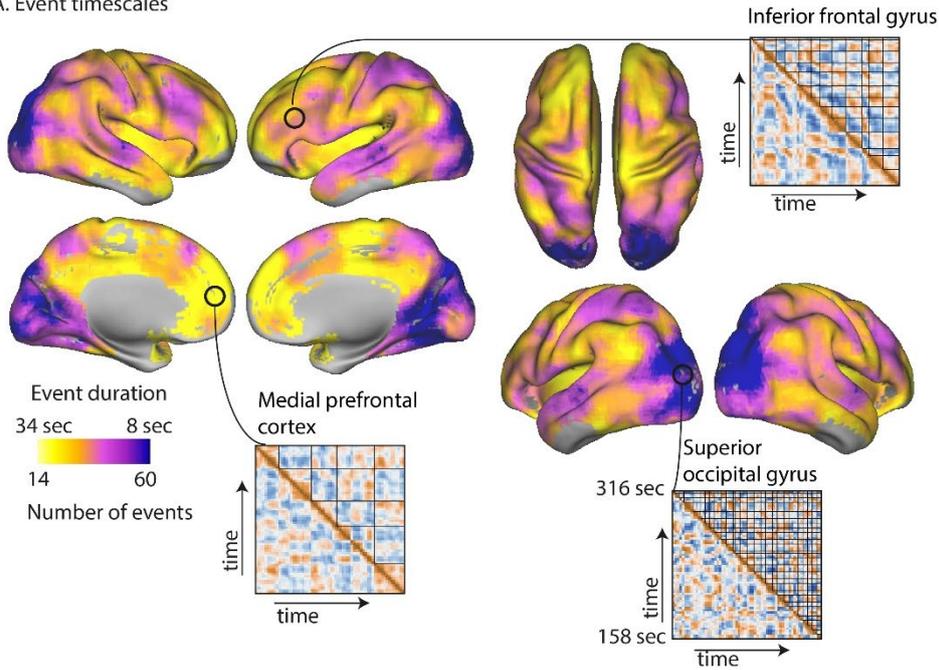
Temporal hierarchy

The optimal number of events differed greatly between searchlights, from as few as 14 to as many as 60 in the fastest regions (see figure 2A). Most searchlights showed average event durations between 10 and 30 seconds per event. However, we also observed that even regions with a slow timescale could show fast event transitions in some cases (see figure 2C). To determine whether regional differences in event duration were reliable, we performed statistical tests, comparing each searchlight's optimal number of events to the average number of events across searchlights. The results show that event duration estimates are highly similar across cross validation folds and that there is reliable variability in the number of events across the brain (see figure 2B).

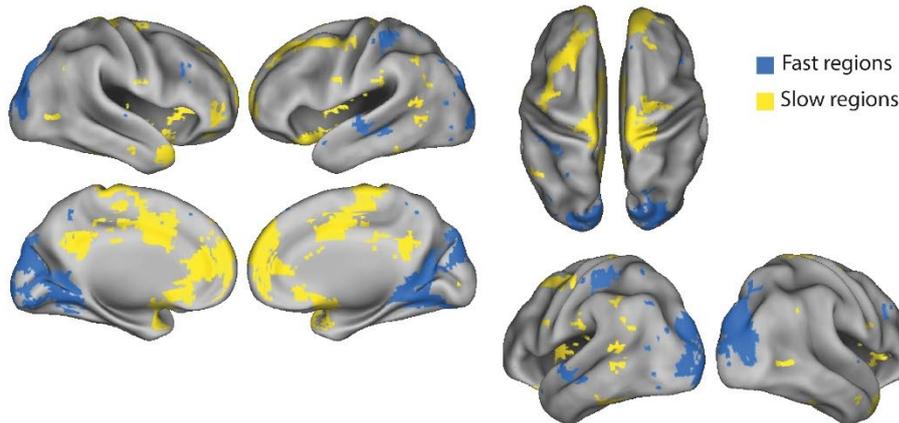
Figure 2A shows that there were particularly fast transitions between stable activity patterns in visual cortex and less so in auditory cortex. Event transitions were slower in areas further up the ventral visual stream, such as the inferior occipital gyrus and the fusiform gyrus, as well as down the dorsal visual stream in the intraparietal sulcus. Similar event durations were found in the inferior frontal gyrus. Particularly slow event transitions were observed in high level-regions such as the medial prefrontal gyrus and the temporal pole, precuneus, the inferior temporal gyrus, the insula as well as medial parts of the precentral gyrus.

For the subset of regions investigated in previous work, we observe a strong correspondence to our findings (Baldassano et al., 2017; Hasson et al., 2008; Lerner et al., 2011). Some brain areas showed speeds that did not align with the expected hierarchy. For example, in V1 searchlights were slightly slower than other searchlights in visual cortex. This is visible on the medial plane of the left hemisphere in figure 2A. This effect was also observed by Baldassano et al. (2017). It seems likely that this is due to restrictions in the frequency of event changes that can be observed with our long interscan interval (2.47 seconds) or with fMRI data in general, due to the slow BOLD response. Therefore the slower timescale in V1 may reflect aliasing of faster neural events.

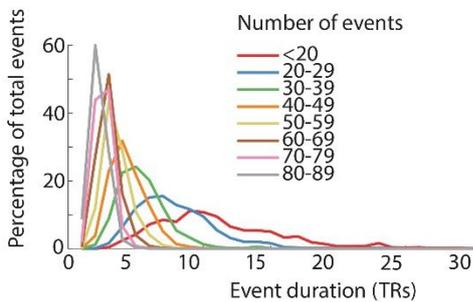
A. Event timescales



B. Regions with significantly slower or faster events than the average timescale across the brain



C. Distribution of event durations for groups of regions with different event timescales



D. Percentage of searchlights with event boundaries over the timecourse of the movie

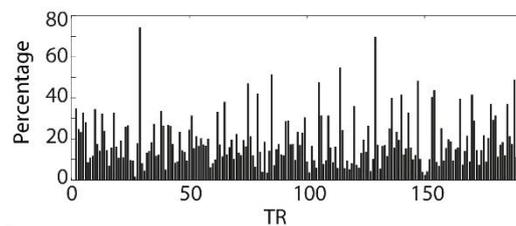


Figure 2. A. The optimal number of events varied greatly across regions, with many events in the visual and auditory cortices and few events in the association cortex, such as the medial prefrontal gyrus and precuneus. Parts of the correlation matrices for three selected searchlights are shown in the insets, representing approximately 2.5 minutes of the movie. The black lines in the upper triangular part of these insets are the event boundaries that are detected by the model. B. The regions with significantly faster or slower than average timescales are depicted (FDR-corrected p -value < 0.05). C. Shows the distribution of event durations for regions grouped according to the

optimal number of events detected in each region. D. Shows the percentage of searchlights that show a neural event simultaneously at each TR over the time course of the movie.

Event networks

Next, we investigated whether event boundaries (and not just event durations) are shared across regions. The results in figure 2D suggest that at some points in time, event boundaries are shared between large numbers of brain regions across the temporal hierarchy. To identify networks of brain regions that shared event boundaries, we computed the Pearson correlation between the boundary vectors of each searchlight and used modularity maximization to identify networks. We found that event boundaries are shared within long-range networks that resemble the functional networks that are typically identified based on (resting state) timeseries correlations (see figure 3A). These include the auditory network, the dorsal attention network, the motor network and a lower-order and higher-order visual network. The higher order visual network consisted of higher order visual regions in the dorsal as well as the ventral visual stream, extending into inferior temporal as well as superior occipital and parietal gyri. We also observed a network composed of the left and right temporo-parietal junctions, which is not typically observed as a separate functional network. In addition, we identified the fronto-parietal (FPN) and default mode networks (DMN), both of which were fractionated into two parts with differing event durations. These fractionated components have different spatial topographies but also show a lot of spatial overlap. This is possible because we used overlapping searchlights and each searchlight was part of a particular functional network. Figure 3a visualizes for each voxel, which functional network label occurs most frequently for searchlights overlapping that voxel. In contrast, the full extent of each of the functional networks can be seen in figure 4.

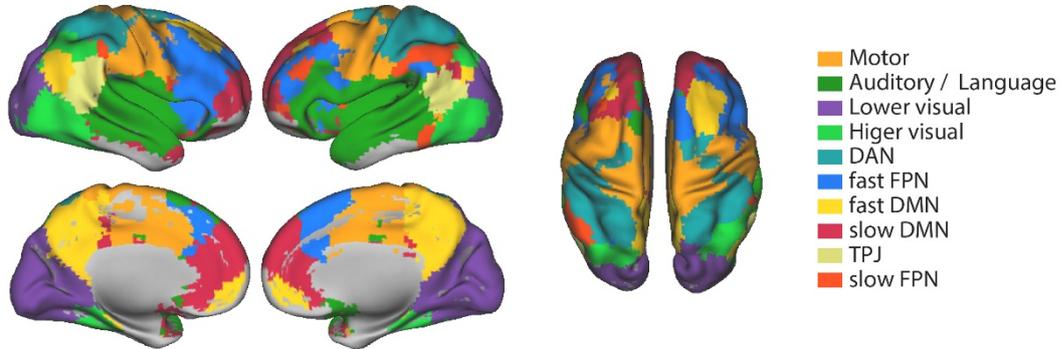
The fractionation of the DMN into slow and fast subnetworks, closely resembles the previously observed decomposition into an anterior and a posterior DMN (Andrews-Hanna et al., 2010; Damoiseaux et al., 2008). The slow-DMN resembles the anterior DMN and is most strongly present in the medial prefrontal cortex. The fast-DMN resembles the posterior DMN and is most strongly present in the precuneus and angular gyri. The difference in event durations between the slow and fast FPN was even more pronounced than for the slow and fast DMN (see figure 3C). The fast-FPN overlaps with the salience network, in the anterior insula and the anterior cingulate cortex, but extends further into the dorsolateral prefrontal cortex, while also encompassing the left superior parietal lobule. The slow-FPN is located in the inferior parietal lobule and supramarginal gyrus, while also extending into the middle frontal gyrus and posterior insula.

The spatial overlap between the fast- and slow DMN and the fast and slow FPN (see figure 4) suggests that overlapping searchlights can belong to different functional networks with different event timescales. This may indicate that brain regions can have multiple preferred timescales for event segmentation. This is supported by the final fit plots shown in figure 3B. In the FPN both the slow and the fast subnetwork show the same peaks at $K=21$ and $K=40$. However, in the fast FPN the latter peak is more prominent while the opposite is true for the slow FPN.

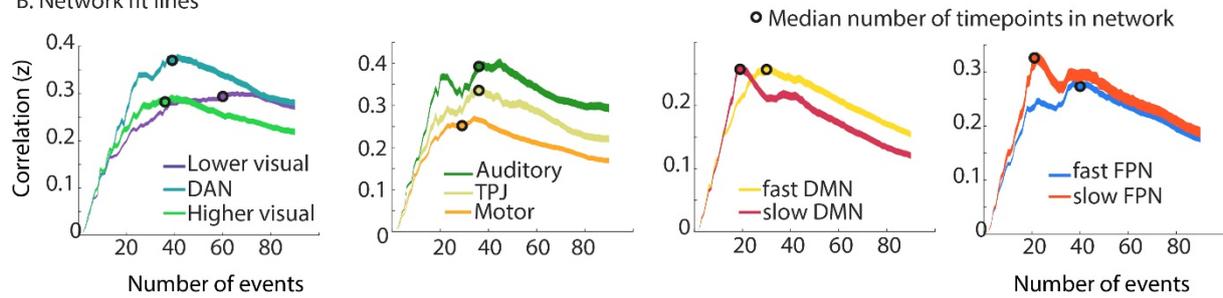
Although regions within functional networks tend to operate on a similar temporal scale, we see that there is also a lot of variability in event duration within networks (see figure 3C). This suggests that the correlations between event boundaries are not simply driven by a similarity in the number of events, but rather by a similarity in the event timings. Searchlights within the auditory and lower-level visual networks show particularly high variability in time-scales. As expected, the higher-order

visual network shows longer event durations than the lower-order visual network. Strikingly, events in the motor network are also rather slow.

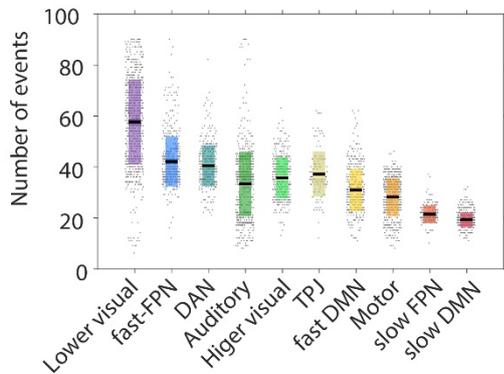
A. Functional network parcellation



B. Network fit lines



C. Timescales of searchlights within each network



D. Graph visualization of partial correlations between networks

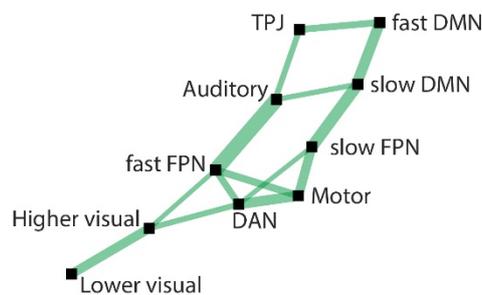


Figure 3. A. Visualization of the detected functional networks. The network label at each voxel is determined by the functional network that occurs most often in all the searchlights that overlap with that voxel. B. The final fit lines of each of the functional networks. The line thickness indicates the 95% confidence interval around the fits of all the searchlights in the network. The median number of timepoints in each network is indicated by the circle. C. Visualization of the event durations within each network. Each searchlight is shown as a dot. The bars show the mean and standard deviation around the mean for each network. D. A graph visualization of the partial correlation between networks is shown (controlling for the boundary vectors in all other networks). The thickness of the line corresponds to the strength of the partial correlation. All lines are significant after FDR-correction for multiple comparisons. TPJ=temporo-parietal junction, DMN=default mode network, FPN=fronto-parietal control network, DAN=dorsal attention network

We also investigated whether events may be shared in a hierarchical fashion, across different functional networks. This was done by computing partial correlations between the average boundary vectors in each network, controlling for the boundary vectors in all other networks. We found evidence that events were indeed shared across different networks, in a way that closely aligns with the hierarchical organization of these different networks (see figure 3D). For example, the lower-level visual network shares boundaries with the higher-level visual network, which in turn connects to the fast-FPN and the dorsal attention network. The fast-FPN also receives input from the auditory network. Also, the fast and slow-DMN are connected to each other, and the fast-DMN is in turn connected to the TPJ network.

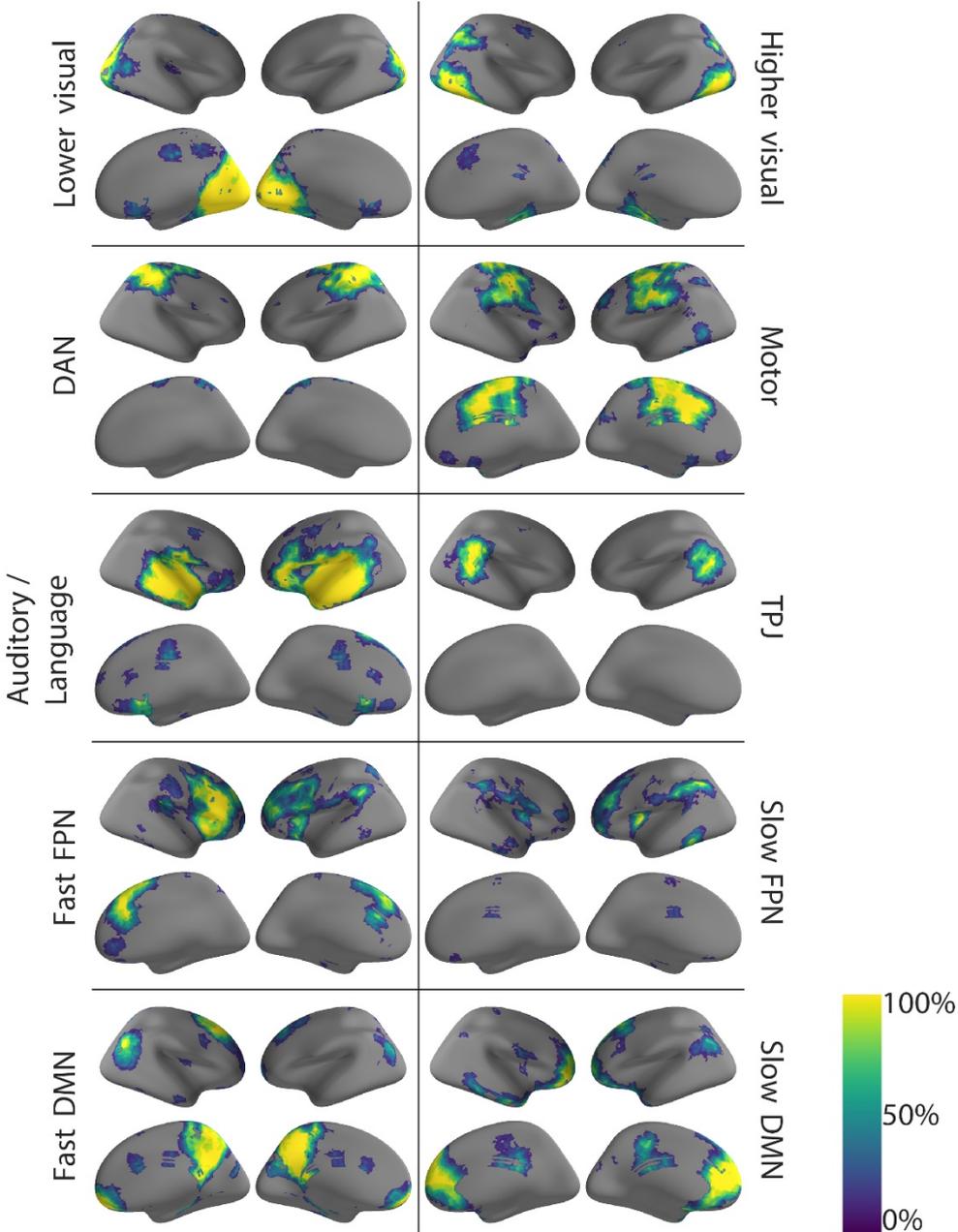


Figure 4: Separate visualizations for each of the identified functional networks. The color indicates the percentage of overlapping searchlights at each voxel that is assigned to that particular functional network. TPJ=temporo-parietal junction, DMN=default mode network, FPN=fronto-parietal control network, DAN=dorsal attention network

($r=0.48$, $p<0.001$). Regions with faster event timescales showed higher power in the frequency range between 0.03-0.045 Hz (see figure 6A).

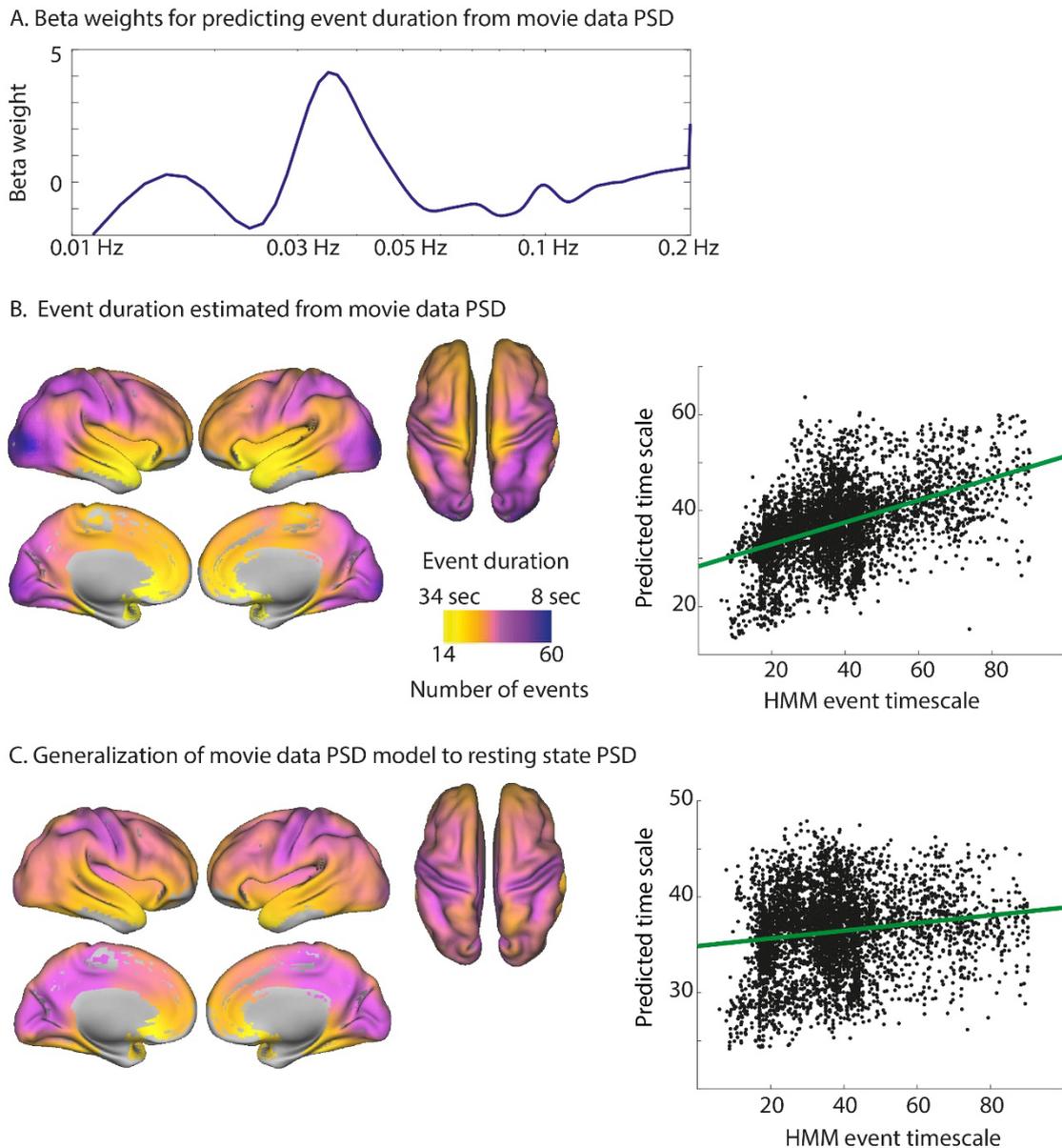


Figure 6. A. Beta values for the prediction of event duration from PSD in the movie dataset (see methods section on power spectra). B. Estimated event timescales based on PSD data in the movie dataset. C. Generalization of movie PSD model to resting state, by applying the beta estimates shown in A.

Figure 6B shows the event timescale estimates based on the PSD data. There are many similarities to the results shown in figure 2A, such as the fast event timescales in the visual system and the slow timescales of the medial prefrontal cortex and anterior temporal lobes. However, we also see some striking differences, primarily in the rather fast timescale of the precuneus and slow timescales in the auditory cortex. These results suggest that regional differences in power spectra may contribute to the observed event timescales, but do not fully explain them.

To determine to what extent timescale estimates depend on the stimuli that were presented, we also applied the same beta estimates to a resting state dataset from the same set of participants. Figure 6C shows the resulting event timescale estimates. The estimates in resting state showed a weak but significant association with the original event timescales ($r=0.14$, $p<0.001$) and a moderate correlation with the estimated event timescales in the movie data ($r=0.53$, $p<0.001$). The hierarchy from fast events in primary sensory regions to slow events in association cortex was not evident when the timescales were estimated based on the resting state PSD. Regional differences in estimated event timescales were also less pronounced in resting state. However, we do still see particularly slow events in the medial prefrontal cortex and anterior temporal lobes. These results suggest that visual and auditory stimulation are a prerequisite for observing the temporal hierarchy we describe in this paper and that this hierarchy only partly reflects an intrinsic property of brain function that is also present in the resting state.

Discussion

We used a narrative stimulus to investigate the time scale and the nature of event segmentation in the brain. Our results demonstrate that there is a distinct temporal gradient of information processing over the entire cortex, with particularly fast events in primary sensory regions and long periods of information integration in the precuneus and medial prefrontal cortex. Event boundaries are shared across long-range functional networks and some events are shared across the temporal hierarchy between distinct functional networks. The validity of the neural events we identified was supported by their correspondence to subjective event boundaries in the precuneus, angular gyrus, medial prefrontal cortex and area V5.

Event timescales across the entire cortex

While previous studies have been able to show regional differences in the time scale of information processing across part of the cortex (Baldassano et al., 2017; Hasson et al., 2008; Honey et al., 2012a; Lerner et al., 2011; Stephens et al., 2013), here we were able to reveal event timescales across the entire cortex for the first time. Our results provide converging evidence for the hierarchical process memory framework (Hasson et al., 2015), showing that the timescales of information accumulation vary along the cortical hierarchy from early sensory regions on one extreme and the default mode network on the other extreme. The slowest timescales were observed in the anterior default mode network, particularly the medial prefrontal cortex. The default mode network has tight links with episodic memory systems in the medial temporal cortex and is involved in social reasoning and processing of an unfolding narrative (Andrews-Hanna, 2012; Van Kesteren et al., 2012; Simony et al., 2016; Spreng et al., 2009; Spreng et al., 2010). In line with previous work, we found that event boundaries in many parts of the default mode network overlapped with subjective events identified by participants (Baldassano et al., 2017; Speer et al., 2007). Together, these results suggest that this slowest timescale of event segmentation in the default mode network aligns with the mental narrative that participants create based on the movie, which is consequently stored in episodic memory (Ben-Yakov et al., 2013; Ben-Yakov and Henson, 2018).

Another region that showed particularly slow information processing is the insula. This area has previously been shown to integrate information about temporal duration over multi-second intervals (Wittmann et al., 2010). It has been hypothesized that the insula creates the experience of duration

by integrating the representations of successive emotional moments over time with interoception (Craig, 2009).

Events are shared across brain regions

Our results show for the first time that neural events are shared across brain regions in distinct functional networks. These networks show a strong resemblance to the functional networks that are typically identified using regular functional connectivity analyses (c.f. Power et al., 2011; Yeo et al., 2011). Because neural event boundaries are defined as a regional shift in brain activity patterns, these results suggest that different brain regions that are involved in similar (cognitive) functions tend to show such shifts at the same time. Critically, network analyses also revealed that the DMN and FPN fractionate into faster and slower sub-networks, with differing spatial topographies.

The slow DMN we identified resembles the anterior DMN (Andrews-Hanna et al., 2010; Lei et al., 2014) and is particularly prominent in the medial prefrontal cortex. The medial prefrontal cortex has been related to self-referential mental thought, affective processing and integrating current information with prior knowledge (Benoit et al., 2014; Van Kesteren et al., 2012; Northoff et al., 2006). Therefore, in the context of the movie stimulus, this subnetwork is most likely involved in building a narrative by integrating incoming information with previous knowledge/schemas, which may be a slow process. In contrast, the fast DMN resembles the posterior DMN which engages in episodic memory retrieval through its connectivity with the hippocampal formation (Andrews-Hanna et al., 2010). It has been proposed to be involved in forming mental scenes or situation models (Ranganath and Ritchey, 2012). Thus, faster events in the posterior portions of the DMN may reflect the construction of mental scenes of the movie and/or retrieval of related episodic memories.

The fast FPN covers the salience network and extends beyond that into the dorsolateral prefrontal cortex. It is connected most to the DAN and the higher visual network, as well as the auditory/language network. Given the role of the FPN in attentional and cognitive control (Shulman and Corbetta, 2012; Vincent et al., 2008) and the role of the salience network in the selection of relevant, rewarding or emotional stimuli and switching between cognitive resources (Seeley et al., 2007; Uddin, 2014), our results suggest that the fast FPN is contributing to maintaining attention to the auditory and visual stimuli in the external environment. In contrast, the slow FPN is more dominant in lateral parietal areas, but also covers the dorsolateral prefrontal cortex. It is interposed between the DAN and slow DMN, suggesting that it may be more involved in inward attention and forming event models or an understanding of the film.

Another interesting difference between our event-based functional networks and the typical resting state functional architecture, was the separate network involving the right and left TPJ that is typically not observed. The TPJ is thought to be responsible for mentalizing or theory of mind, which involves making inferences about the beliefs, desires, and emotions of other people and narrative comprehension (Schurz et al., 2014; Xu et al., 2005; Young et al., 2010). The precuneus and medial prefrontal cortex are also involved in these functions, in line with the connection we observed between the TPJ network and the posterior DMN network.

Events were not only shared within networks but also between networks. This suggests that some events occur across the temporal and cortical hierarchy, with a subset of the event boundaries in low-level processing areas also shared with high-level processing regions. At the highest levels of this hierarchy, the event boundaries align with subjective experience. Our analyses revealed that event boundaries are shared between networks in close alignment with interconnected functions

between those particular networks. For example, the low-level visual network shared event boundaries with the higher-level visual network, which in turn shared event boundaries with the dorsal attention network and the fronto-parietal network.

Event timescales and power spectra

In correspondence with earlier work (Honey et al., 2012b; Stephens et al., 2013), we showed that regional differences in event timescales were related to regional differences in the power spectra during movie watching. However, much of the variance in event timescales could not be explained by the power spectra and for some regions, such as the precuneus and the auditory cortex, predictions from the power spectra diverged strongly from the actual event timescales. One reason for this discrepancy may be that even regions with a very slow timescale can show rapid event transitions. The timescales of event processing are not dominated by one particular frequency, but rather occur within a distribution, which is quite narrow in primary sensory regions and broad in areas like the medial prefrontal cortex, where both long and short events can occur.

We also investigated whether the association between power spectra and event timescales generalized to resting state data. The predicted event timescales were significantly correlated between the movie and the resting state data and we found that the resting state data power spectra were also predictive of the event timescales during movie watching. This is line with previous work that has also shown that timescales of information processing generalize to resting conditions (Honey et al., 2012b; Stephens et al., 2013). However, the correlation between the event timescales and resting state power spectra was weak and inspection of the predictions from resting state showed that the expected fast information processing of primary sensory regions was lacking in rest. In fact, the fastest information processing in rest was observed over the motor cortex and in medial occipital and parietal areas (including the precuneus). This suggests that the timescales of information accumulation may be driven to a large extent by the nature of sensory input and the cognitive demands that are placed on the participant.

Across all brain regions, the event timescales we observed in the current study were faster than those observed by Baldassano et al. (2017). While they observed average event durations in the range of 40 seconds to 5 minutes, average durations were in the range of 8 to 34 seconds in our study. This may be due to the different fitting procedure that was used. While Baldassano et al. (2017) used a procedure that aimed to maximize within-event similarity and minimize between-event similarity, we aimed to identify events such that the correlation between the event signatures were most representative of the full temporal correlation matrix (see figure 1). Therefore, our fitting procedure did not penalize correlations between distinct events and may therefore have resulted in a larger number of events. The nature and duration of the stimulus (50 minutes in Baldassano et al. and 8 minutes in our study) may also have played a role.

Real life experience

Although event segmentation is thought to be a pivotal aspect of how information is processed in real life (Zacks et al., 2007), it is often not considered in experimental settings, where events are predetermined by the trial or block structure. This study and previous work (Baldassano et al., 2017) shows that we are now able to investigate brain activity as it unfolds over time in a way that is much more similar to our daily life experience than typical cognitive neuroscience experiments. This opens

the door for investigations of neural differences during narrative comprehension between groups of participants, such as participants with autism who may have trouble distinguishing events that require them to infer the state of mind of others (Baron-Cohen, 2000; Hasson et al., 2009), or participants with Alzheimer's disease, who may have trouble encoding particular events in memory (Zacks et al., 2006).

This approach also allows us to show at which levels of the cortical hierarchy neural events align with subjective event experience or with objectively defined events. This occurs not only in the default mode network regions discussed previously, but also in area V5, suggesting that movement cues may be one important indicator of perceived event boundaries (in line with Zacks et al., 2001).

Conclusion

Using a unique dataset in combination with a novel model fitting procedure, we were able to show the timescales of neural events across the entire cortex. These timescales were not fully predicted by the regions' power spectra and generalized poorly to resting state data. We found that neural events are shared across brain regions that are typically part of the same functional network, while specific events are also shared across the temporal hierarchy between distinct functional networks. In addition, we observed that known functional networks such as the DMN and FPN fractionate due to differences in the timescales of event segmentation. These results shed new light on how the brain processes information in real life contexts.

Acknowledgements

LG was supported by a Veni grant [451-16-013] from the Netherlands Organization for Scientific Research. MVG was supported by a Vidi grant [639.072.513] from the Netherlands Organization for Scientific Research. KC was supported by the Natural Sciences and Engineering Research Council of Canada (Grant RGPIN-2017-03804 to KC) and the Canada Research Chairs program.

References

- Andrews-Hanna JR (2012): The brain's default network and its adaptive role in internal mentation. *Neuroscientist* 18:251–270.
- Andrews-Hanna JR, Reidler JS, Sepulcre J, Poulin R, Buckner RL (2010): Functional-anatomic fractionation of the brain's default network. *Neuron* 65:550–62.
- Baldassano C, Chen J, Zadbood A, Pillow JW, Hasson U, Norman KA (2017): Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* 95:709–721.e5.
- Baron-Cohen S (2000): Theory of mind and autism: A review. *Int Rev Res Ment Retard* 23:169–184.
- Ben-Yakov A, Eshel N, Dudai Y (2013): Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. *J Exp Psychol Gen* 142:1255–1263.
- Ben-Yakov A, Henson RN (2018): The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. *J Neurosci* 38:10057–10068.
- Benjamini Y, Krieger AM, Yekutieli D (2006): Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93:491–507.

- Benoit RG, Szpunar KK, Schacter DL (2014): Ventromedial prefrontal cortex supports affective future simulation by integrating distributed knowledge. *Proc Natl Acad Sci U S A* 111:16550–16555.
- Craig AD (2009): Emotional moments across time: A possible neural basis for time perception in the anterior insula. *Philos Trans R Soc B Biol Sci* 364:1933–1942.
- Damoiseaux JS, Beckmann CF, Arigita EJS, Barkhof F, Scheltens P, Stam CJ, Smith SM, Rombouta SAR, Rombouts S, Raaijmakers RB (2008): Reduced resting-state brain activity in the “default network” in normal aging. *Cereb Cortex* 18:1856–1864.
- DiCarlo JJ, Cox DD (2007): Untangling invariant object recognition. *Trends Cogn Sci* 11:333–341.
- Flores S, Bailey HR, Eisenberg ML, Zacks JM (2017): Event segmentation improves event memory up to one month later. *J Exp Psychol Learn Mem Cogn* 43:1183–1202.
- Geerligs L, Cam-CAN, Campbell KL (2018): Age-related differences in information processing during movie watching. *Neurobiol Aging* 72:106:120.
- Guntupalli JS, Hanke M, Halchenko YO, Connolly AC, Ramadge PJ, Haxby J V (2016): A Model of Representational Spaces in Human Cortex:2919–2934.
- Hanke M, Halchenko YO, Sederberg PB, Hanson SJ, Haxby J V., Pollmann S (2009): PyMVPA: a Python Toolbox for Multivariate Pattern Analysis of fMRI Data. *Neuroinformatics* 7:37–53.
- Hasson U, Avidan G, Gelbard H, Vallines I, Harel M, Minshew N, Behrmann M (2009): Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Res* 2:220–231.
- Hasson U, Chen J, Honey CJ (2015): Hierarchical process memory: Memory as an integral component of information processing. *Trends Cogn Sci* 19:304–313.
- Hasson U, Yang E, Vallines I, Heeger D, Rubin N (2008): A hierarchy of temporal receptive windows in human cortex. *J Neurosci* 28:2539–2550.
- Honey CJ, Thesen T, Donner TH, Silbert LJ, Carlson CE, Devinsky O, Doyle WK, Rubin N, Heeger DJ, Hasson U (2012a): Slow Cortical Dynamics and the Accumulation of Information over Long Timescales Figure S1 : Reliability of band-passed voltage traces . *J Neurosci* 32:423–434.
- Honey CJ, Thesen T, Donner TH, Silbert LJ, Carlson CE, Devinsky O, Doyle WK, Rubin N, Heeger DJ, Hasson U (2012b): Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* 76:423–34.
- Van Kesteren MTR, Ruitter DJ, Fernández G, Henson RN (2012): How schema and novelty augment memory formation. *Trends Neurosci* 35:211–219.
- Kurby CA, Zacks JM (2008): Segmentation in the perception and memory of events. *Trends Cogn Sci* 12:72–79.
- Lei X, Wang Y, Yuan H, Mantini D (2014): Neuronal oscillations and functional interactions between resting state networks. *Hum Brain Mapp* 35:3517–28.
- Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011): Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. *J Neurosci* 31:2906–2915.
- Newton D, Engquist GA, Bois J (1977): The objective basis of behavior units. *J Pers Soc Psychol* 35:847–862.
- Newton D, Rindner RJ (1979): Variation in behavior perception and ability attribution. *J Pers Soc Psychol* 37:1847–1858.

- Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J (2006): Self-referential processing in our brain—A meta-analysis of imaging studies on the self. *Neuroimage* 31:440–457.
- Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church J a, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE (2011): Functional network organization of the human brain. *Neuron* 72:665–78.
- Ranganath C, Ritchey M (2012): Two cortical systems for memory-guided behaviour. *Nat Rev Neurosci* 13:713–726.
- Reichardt J, Bornholdt S (2006): Statistical Mechanics of Community Detection. *Phys Rev E* 74:16110.
- Rubinov M, Sporns O (2010): Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* 52:1059–1069.
- Sargent JQ, Zacks JM, Hambrick DZ, Zacks RT, Kurby CA, Bailey HR, Eisenberg ML, Beck TM (2013): Event segmentation ability uniquely predicts event memory. *Cognition* 129:241–255.
- Schurz M, Radua J, Aichhorn M, Richlan F, Perner J (2014): Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neurosci Biobehav Rev* 42:9–34.
- Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD (2007): Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci* 27:2349–56.
- Shafto MA, Tyler LK, Dixon M, Taylor JR, Rowe JB, Cusack R, Calder AJ, Marslen-Wilson WD, Duncan J, Dalgleish T, Henson RN, Brayne C, Cam-CAN, Matthews FE (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.
- Shulman GL, Corbetta M (2012): Two attentional networks: Identification and function within a larger cognitive architecture. In: Posner, MI, editor. *Cognitive Neuroscience of Attention* 2nd ed. New York: The Guilford Press. pp 113-128.
- Simony E, Honey CJ, Chen J, Lositsky O, Yeshurun Y, Wiesel A, Hasson U (2016): Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat Commun* 7:12141.
- Speer NK, Reynolds JR, Zacks JM (2007): Human brain activity time-locked to narrative event boundaries. *Psychol Sci* 18:449–455.
- Spreng RN, Mar RA, Kim ASN (2009): The Common Neural Basis of Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the Default Mode: A Quantitative Meta-analysis. *J Cogn Neurosci* 21:489–510.
- Spreng RN, Wojtowicz M, Grady CL (2010): Reliable differences in brain activity between young and old adults: a quantitative meta-analysis across multiple cognitive domains. *Neurosci Biobehav Rev* 34:1178–94.
- Stephens GJ, Honey CJ, Hasson U (2013): A place for time: the spatiotemporal structure of neural dynamics during natural audition. *J Neurophysiol* 110:2019–2026.
- Uddin LQ (2014): Salience processing and insular cortical function and dysfunction. *Nat Rev Neurosci* 16:55–61.
- Vincent JL, Kahn I, Snyder AZ, Raichle ME, Buckner RL (2008): Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *J Neurophysiol* 100:3328–42.

- Welch P (1967): The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans Audio Electroacoust* 15:70–73.
- Wittmann M, Simmons AN, Aron JL, Paulus MP (2010): Accumulation of neural activity in the posterior insula encodes the passage of time. *Neuropsychologia* 48:3110–3120.
- Xu J, Kemeny S, Park G, Frattali C, Braun A (2005): Language in context: emergent features of word, sentence, and narrative comprehension. *Neuroimage* 25:1002–1015.
- Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, Fischl B, Liu H, Buckner RL (2011): The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* 106:1125–65.
- Young L, Dodell-Feder D, Saxe R (2010): What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia* 48:2658–2664.
- Zacks JM, Tversky B, Iyer G (2001): Perceiving, remembering, and communicating structure in events. *J Exp Psychol Gen* 130:29–58.
- Zacks JM, Speer NK, Swallow KM, Braver TS, Reynolds JR (2007): Event perception: A mind-brain perspective. *Psychol Bull* 133:273–293.
- Zacks JM, Speer NK, Swallow KM, Maley CJ (2010): The brain's cutting-room floor : segmentation of narrative cinema. *Front Aging Neurosci* 4:1–15.
- Zacks JM, Speer NK, Vettel JM, Jacoby LL (2006): Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychol Aging* 21:466–482.