# Information flow across the cortical timescales hierarchy during narrative comprehension

Claire H. C. Chang[1], Samuel A. Nastase[1], and Uri Hasson[1]

1. Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey, 08540, USA.

## Abstract

When listening to spoken narratives, we must integrate information over multiple, concurrent timescales, building up from words to phrases to sentences to a coherent narrative. Recent evidence suggests that the brain relies on a chain of hierarchically organized areas with increasing temporal receptive windows to process naturalistic narratives. In this study, we use inter-subject functional connectivity to reveal a stimulus-driven information flow along the cortical hierarchy. Using cross-correlation analysis to estimate the time lags between six functional networks, we found a fixed temporal sequence of information flow, starting in early auditory areas, followed language areas, the attention network, and lastly the default mode network. This gradient is consistent across eight distinct stories but absent in resting-state and scrambled story data, indicating that the lag gradient reflects the construction of narrative features. Finally, we simulate a variety of narrative integration models and demonstrate that nested narrative structure along with the gradual accumulation of information within the boundaries of linguistic events at each level of the processing hierarchy is sufficient to reproduce the lag gradient. Taken together, this study provides a computational framework for how information flows along the cortical hierarchy during narrative comprehension.

## Introduction

Narratives are composed of nested elements that must be continuously integrated to construct a meaningful whole. As a linguistic narrative unfolds, phonemes must be integrated into words, words must be integrated into sentences, sentences must be integrated into paragraphs, paragraphs must be integrated into coherent stories—and these integration processes must occur simultaneously over time (Figure 1a; Christiansen & Chater, 2015). Recent evidence suggests that the human brain relies on a chain of hierarchically organized brain areas with increasing temporal receptive windows (TRWs) to process this temporally evolving, nested structure (Figure 1b). This cortical hierarchy was first revealed by studies manipulating the temporal coherence of naturalistic narratives to show the topography of processing timescales along the cortical hierarchy (Hasson et al., 2008; Lerner et al., 2011). These studies reported a topography of processing timescales where early auditory areas respond reliably to rapidly-evolving acoustic features, adjacent areas along the superior temporal gyrus respond reliably to information at the word level, and nearby language areas respond reliably only to coherent sentences. Finally, areas at the top of the processing hierarchy in the default mode network (DMN) seem to integrate slower-evolving semantic information over many minutes (Yeshurun et al., 2021).

This cortical hierarchy of increasing temporal integration windows recapitulates temporal structures in the external world is thought to be a fundamental organizing principle of the brain (Hasson et al., 2015; Kiebel et al., 2008). The cortical hierarchy of TRWs in humans has been described using fMRI (Chien & Honey, 2020; Hasson et al., 2008; Lerner et al., 2011; Yeshurun et al., 2017) and ECoG (Honey et al., 2012). Recent work has shown that deep language models also learn a gradient or hierarchy of increasing TRWs (Dominey, 2021; Peters et al., 2018; Vig & Belinkov, 2019), and that manipulating the temporal coherence of narrative input to a deep

language model yields representations closely matching the cortical hierarchy of TRWs in the human brain (Caucheteux et al., 2021). Furthermore, the cortical hierarchy of TRWs matches the intrinsic processing timescales observed during rest in humans (Honey et al., 2012; Raut et al., 2020; Stephens et al., 2013) and monkeys (Murray et al., 2014). This cortical topography also coincides with anatomical and functional gradients such as long-range connectivity and local circuitry (Baria et al., 2013; Changeux et al., 2020; Huntenburg et al., 2018), which have been shown to yield varying TRWs (Chaudhuri et al., 2015; Demirtaş et al., 2019).

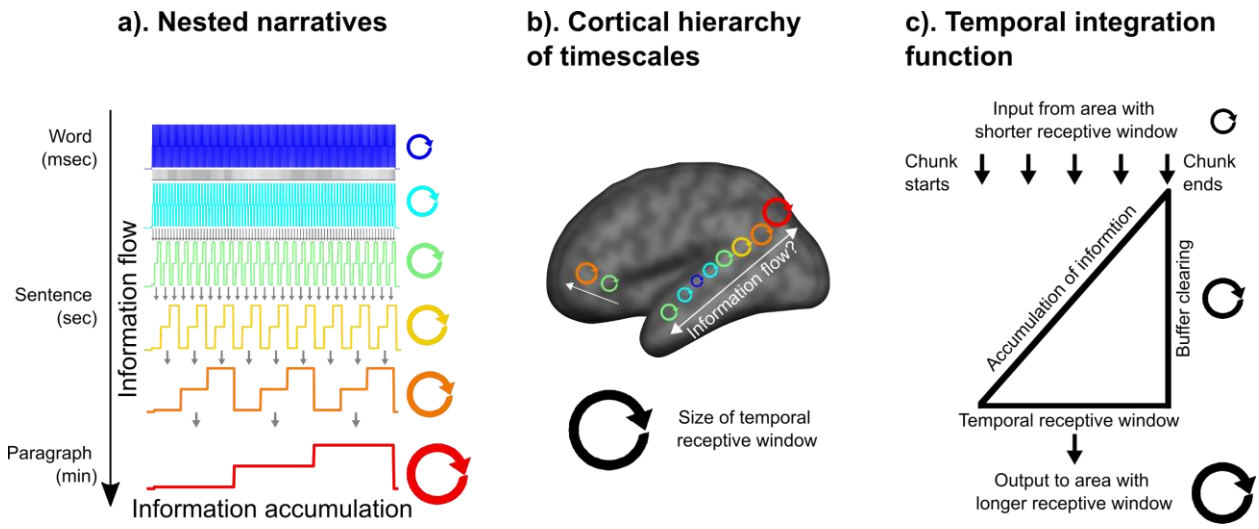



*Figure 1. Information flows along the hierarchy of increasing temporal receptive windows. (a) Narratives are composed of nested units of increasing granularity. Each level of the narrative provides the building blocks for the next level. For example, phrases built over words are constructed into sentences. (b) The cortical hierarchy of increasing temporal receptive windows (adapted from Hasson et al., 2015), corresponding to linguistic units of different sizes, implies a fixed order of information flow across brain regions. (c) At each level of the processing hierarchy, information continuously accumulates over inputs from the preceding level. The accumulated information is continuously transferred to the next level and flushed at structural boundaries.*

The proposal that the cortex is organized according to a hierarchy of increasing TRWs implies that each area chunks and integrates information at its preferred temporal integration window and that information flows from lower- to higher-level areas along the cortical hierarchy. For example, an area that processes phrases receives information from areas that process words (Figure 1c), which are further transmitted along the processing hierarchy to areas that integrate phrases into sentences. At the end of each linguistic or narrative event (e.g. phrase in our

example), information is rapidly cleared to allow for real-time processing of the next phrase (Chien & Honey, 2020; Christiansen & Chater, 2015). The chunking of information at varying granularity is supported by recent studies that used data-driven methods to detect boundaries as shifts between stable patterns of brain activity. These studies revealed a nested hierarchy of cortical events, from brief events in sensory regions to longer-duration events in high-order areas comprising the DMN (Baldassano et al., 2017; Geerligs et al., 2021).

The cascade of information transformations that gives rise to narrative features of increasing granularity suggests that there is a wave of activity that propagates at fixed temporal sequence along the cortical hierarchy. In the current study, we provide empirical evidence for the propagation of information along the processing hierarchy. To that end, we looked at sequential processing along the cortical hierarchy during naturalistic narrative comprehension. We predicted that the temporal lag in response fluctuations among adjacent areas along the processing hierarchy will be smaller than regions further apart in the cortical hierarchy. The temporal difference among brain areas was calculated using lag correlation. To test this hypothesis, we identified six intrinsic functional networks, ranging from early auditory areas to language areas to higher-order attention and default mode networks (Figure 2a). These six intrinsic networks align with the previously documented TRW hierarchy (Figure S1a and S1b).

To focus on neural responses to linguistic and narrative information, we used inter-subject functional connectivity (ISFC) analysis (Nastase et al., 2019; Simony et al., 2016). Unlike traditional within-subject functional connectivity (WSFC), ISFC captures network connectivity driven by the shared stimulus. This connectivity is abolished in the absence of a shared stimulus (i.e. during rest). In other words, ISFC effectively filters out the idiosyncratic fluctuations that drive intrinsic functional correlations within subjects.

Isolating stimulus-locked neural activity from intrinsic neural activity allows us for the first time to observe the propagation of linguistic information across the cortical hierarchy. We aggregated eight functional magnetic resonance imaging (fMRI) story-listening datasets (Nastase et al., 2021) to find the reliable, core sequence of information flow across a variety of diverse spoken story stimuli. Note that linguistic and narrative information unfolds over relatively long timescales: for example, single phonemes and words span hundreds of milliseconds, while phrases and paragraphs unfold over seconds and even minutes (Honey et al., 2012; Stephens et al., 2013). Our findings demonstrate that neural response lags locked to a narrative can be detected in the relatively slow-varying hemodynamic signals measured by fMRI. Finally, we use a simulation to demonstrate how the hierarchical integration of nested narrative features coupled with a hemodynamic response function (HRF) is sufficient to fully reproduce the observed lag gradient.

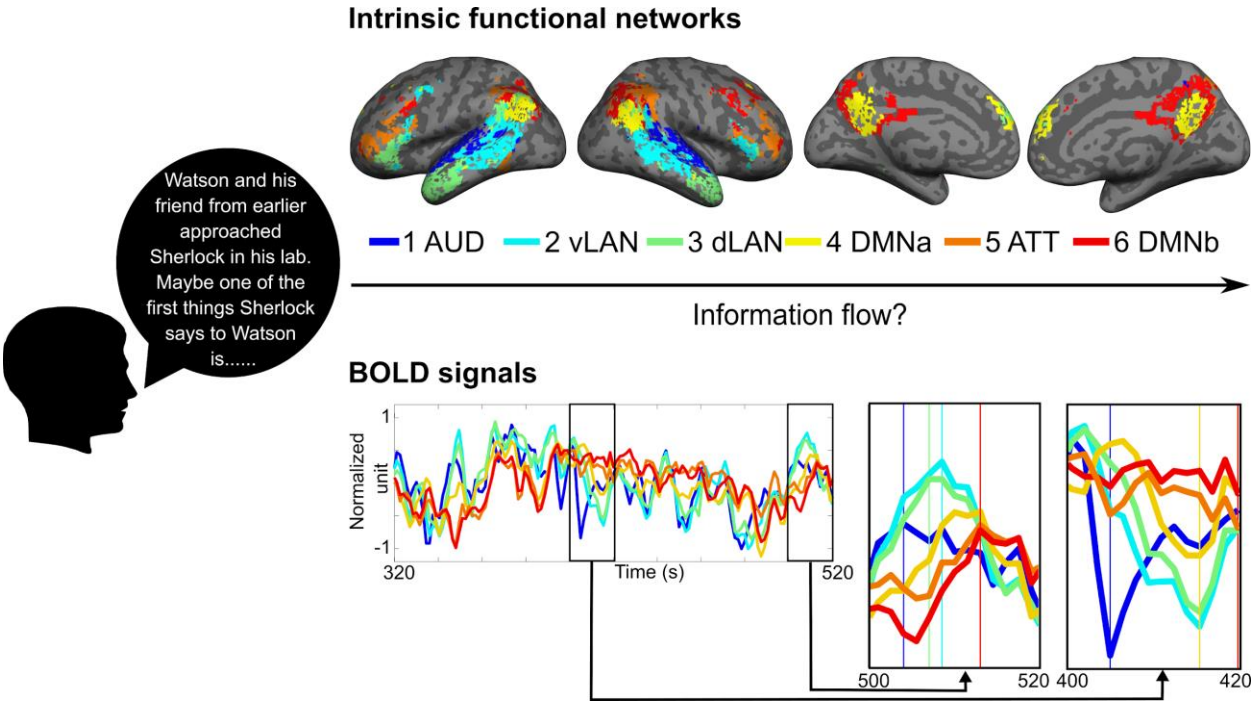



*Figure 2. Averaged fMRI response time series for six intrinsic functional networks while subjects listened to a spoken story ("Sherlock"; AUD: auditory; vLAN: ventral language; dLAN: dorsal language; DMN: default mode network; ATT: attention network; see "Methods" for details). Two example segments of the response time series are highlighted at the bottom right. The peaks of the fluctuations in a given window are indicated by colored vertical lines. Note the stereotyped lag in both positive and negative BOLD fluctuations across networks; e.g. signal deflections in the dark blue auditory network tend to precede deflections in the cyan/green language networks, which tend to precede deflections in the yellow/red default mode networks.*

## Results

To test the hypothesis that narrative information propagates across brain regions in a fixed order, we first divided the neural signals into six networks by applying k-means clustering to WSFC measured during rest (Figure 2): AUD, vLAN, dLAN, ATT, DMNa, and DMNb. We computed

lag-ISFC (i.e. cross-correlation) at varying temporal lags between all pairs of networks (Figure 3a and Figure S2). The lags with maximum ISFC (i.e. “peak lag”) for each seed-target pair were extracted as an index for the temporal gaps in the stimulus-driven processing between each pair of networks. The extracted peak lags were color-coded to construct the network × network peak lag matrix (Figure 3b and 3c). In the following, we describe the observed lag gradient in detail, as well as several control analyses. Finally, we simulated the nested narrative structure and the corresponding brain responses to explore how information integration functions at different timescales could give rise to the observed lag gradient.

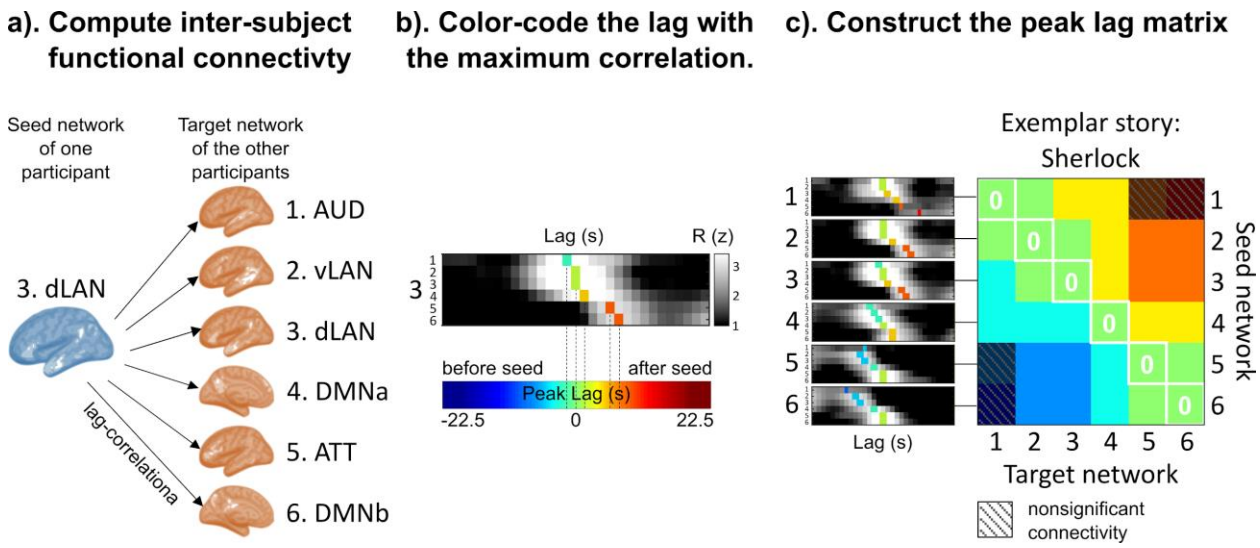



*Figure 3. Construction of the inter-network peak lag matrix. (a) Lag-ISFCs (cross-correlation) between seed-target network pairs were computed using the leave-one-subject out method. The dLAN network is used as an example seed network for illustrative purposes. (b) The matrix depicts ISFC between the dLAN seed and all the target networks at varying lags. The lag with the peak correlation value (colored vertical bars) was extracted and color-coded according to lag. For visualization, the lag-ISFCs were z-scored across lags. (c) The network × network peak lag matrix ($p < .05$, FDR corrected). Warm colors represent peak lags following the seed network, while cool colors represent peak lags preceding the seed network; zeros along the diagonal capture the intra-network ISC. An example story ("Sherlock") is shown for illustrative purposes; see Figure 4 for the peak lag matrix across all stories.*

***Fixed lag gradient across cortical networks***

The average lag-ISFC across stories was computed for each seed network (Figure 4a, left). The lag-ISFC between a seed network and the same network in other subjects always peaked at lag 0, reflecting the strong stimulus-locked within-network synchronization reported in the ISC literature (Hasson et al., 2004; Kauppi et al., 2010; Lerner et al., 2011) (Figure S2). Meanwhile,

non-zero peak lags were found between different networks. Relative to a low-level seed, putatively higher-level networks showed peak connectivity at increasing lags. For example, the stimulus-induced activity in dLAN lagged 1 TR (1.5 s) behind activity in AUD, whereas the activity in DMNb lagged 4 TRs (6 s) behind activity in dLAN. Importantly, regardless of the choice of seed, the target networks showed peak connectivity in a fixed order progressing through AUD, vLAN, dLAN, DMNa, ATT, and DMNb.

To summarize the findings, we color-coded the peak lags and collated them into a peak lag matrix where each row corresponds to a seed network and each column corresponds to a target network (Figure 4a, right). The green diagonal indicates a peak at zero lag within each area, reflecting the intra-network synchronization across subjects (i.e. ISC) (Figure S2), while the cool-to-warm color gradient indicates a fixed order of peak lags. For example, the first row shows a green-to-warm gradient, reflecting that when AUD served as the seed, other networks were either synchronized with or followed AUD, but never preceded it. Conversely, the cool-to-green gradient of the last row indicates that all other networks preceded the DMNb seed. The lag gradient can also be observed in individual stories (Figure S3), although these patterns are noisier than the averaged results. The lag gradient proceeded in a fixed order across all networks, suggesting a propagation of stimulus information along the cortical hierarchy from AUD up to DMNb. Similar results were obtained when we defined the ROIs using the TRW hierarchy (Figure S1c).

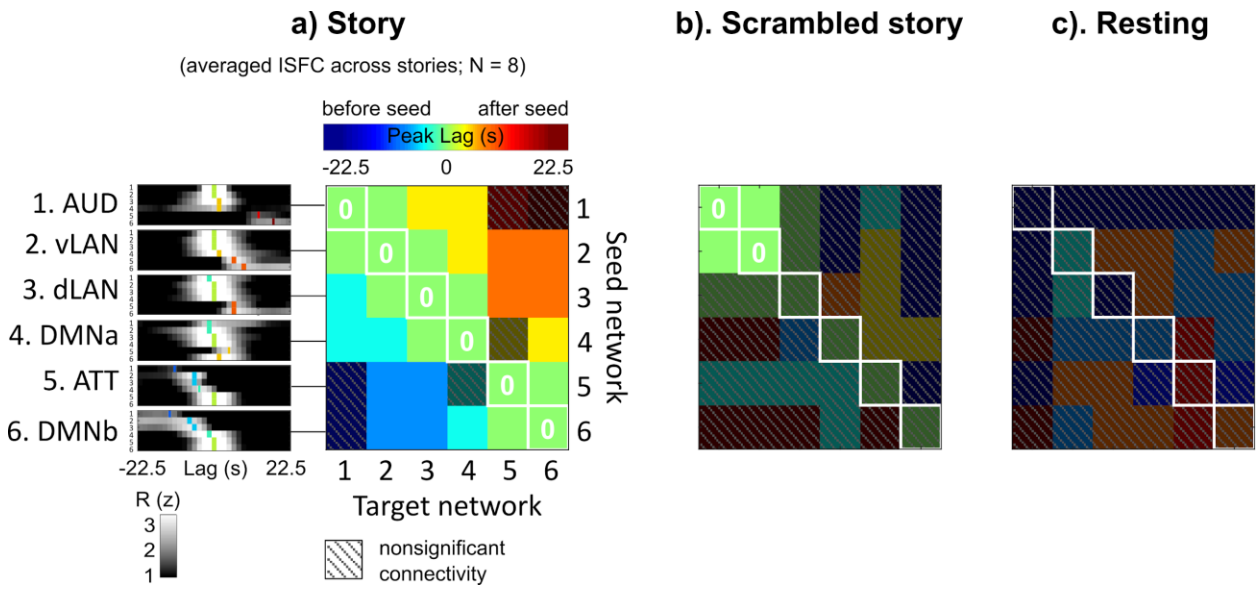



*Figure 4. Peak lag matrix across eight stories reveals a fixed lag gradient across networks, which is abolished during scrambled narratives and rest. (a) The network × network peak lag matrix is based on the averaged lag-ISFC across eight stories. For visualization, lag-ISFC curves at left were z-scored across lags. (b) Peak lag matrix based on responses to a scrambled story stimulus (scrambled words). (c) Peak lag matrix based on resting-state data. Peak lag matrices are thresholded at $p < .05$ (FDR corrected).*

***Temporal scrambling abolishes lag gradient***

We hypothesized that the lag gradient reflected the emergence of macroscopic story features (e.g. narrative situations or events) integrated over longer periods of time in higher-level cortical networks (Baldassano et al., 2017; Geerligs et al., 2021). To support this point, we next used the same procedure to compute the peak lag matrix for a temporally scrambled version of one story ("Pie Man"). In this dataset, the story stimulus was spliced at the word level and scrambled, thus maintaining similar low-level sensory statistics while abolishing the slower-evolving narrative content. The peak lag matrix for the scrambled story revealed synchronized responses at lag 0

both within and between the AUD and vLAN networks, but no significant peaks within or between other networks (Figure 4b). This reflects low-level speech processing limited to the word level and indicates that disrupting the narrative structure of a story abolishes the temporal propagation of information to higher-level cortical areas.

***No lag gradient during rest***

As an additional control, we also examined whether the lag gradient observed during the intact story could be detected during rest. The resting state is dominated by intrinsic fluctuations and there is no external stimulus to drive synchronized brain activity across subjects as well as propagation of activity across cortical areas. As expected, no significant ISFC peaks were found (Figure 4c).

***Idiosyncratic within-subject fluctuations obscure the lag gradient***

We next asked whether the inter-network lag gradient observed during spoken stories can be observed using traditional WSFC. As expected, WSFC analysis revealed a strong peak correlation at lag zero within each network, but also a peak correlation at lag zero across all networks such that no gradient was observed (Figure S4). This result supports the claim that ISFC analysis filters out intrinsic signal fluctuations that propagate across brain areas, revealing the propagation of shared story information across networks (Nastase et al., 2019; Simony et al., 2016). This result also verifies that inter-network differences in hemodynamic responses cannot account for the lag gradient; otherwise, WSFC should show a similar lag pattern as ISFC.

***Lag gradient across fine-grained subnetworks***

To verify that the peak lag gradient could also be observed at a finer spatial scale, we further divided each of the six networks into ten subnetworks, again by applying k-means clustering to resting-state WSFC (k = 10 within each network). The peak lag matrix between the sixty

subnetworks was generated using the same methods as in the network analysis (Figure S5a). We also visualized the brain map of lags between one selected seed (posterior superior/middle temporal gyrus) and all the target subnetworks (Figure S5b). Similar to the network level analysis, the peak lag between the subnetworks revealed a gradient from the early auditory cortex to the language network (auditory association cortex), then to the DMN.

***Dominant path of information flow across networks***

We adopted the method introduced by Mitra et al. (2015) to discern whether there are multiple parallel paths for information flow between networks. We applied principal component analysis (PCA) to the inter-network peak lag matrix (Figure 4a) and examined the cumulative variance accounted for across principal components. Our results revealed that, at the coarse level of the cortical networks used here, the first principle component explains 88.8% of the variance in our lag matrix (Figure S6). This suggests that the lag gradient reflects a single, unidirectional information flow across networks.

***Lag gradient is not driven by transient linguistic boundary effects***

Prior work has reported that scene/situation boundaries in naturalistic stimuli elicit transient brain responses that vary across regions (Ezzyat & Davachi, 2011; Speer et al., 2007; Whitney et al., 2009; Yarkoni et al., 2008; Zacks et al., 2001, 2010). To test whether this transient effect could drive the gradient observed in our lag matrix, we computed lag-ISFC after regressing out the effects of word, sentence, and paragraph boundaries in two stories with time-stamped annotations. As shown in Figure S7, the regression model successfully removed transient effects of the boundaries from the time series. Critically, however, the lag gradient remained qualitatively similar when accounting for boundaries, indicating that the observed lag gradient does not result from transient responses to linguistic boundaries in the story stimulus.

***Simulating the nested narrative structure that drives the lag gradient***

Narratives have a multi-level nested hierarchical structure (Willems et al., 2020). To better understand how the temporal structure of the narrative stimulus could give rise to the observed inter-network lag gradient, we created a simulation capturing the hierarchy of nested narrative structures and the corresponding hierarchy of cortical responses (Figure 5). The simulation hypothesizes that story features emerge across six distinct timescales, which roughly correspond to words, phrases, sentences, 2–3 sentences, paragraphs. The initial level of the simulated narrative hierarchy was populated with relatively brief low-level units, with boundary intervals sampled from actual word durations in a spoken story (Figure S8). These simulated "words" were integrated into "phrases" with a mean length of three words to obtain second-level boundaries. All phrase-level boundaries were also word-level boundaries. A six-level structure was ultimately generated by recursively applying this integration procedure. Since in real stories, paragraphs are often separated by longer silent periods (Figure S9), we inserted pauses at top-level (sixth-level) boundaries.

The hierarchical processing framework postulates that narrative representations of increasing complexity are processed in different brain regions (Changeux et al., 2020; Hasson et al., 2015; Kiebel et al., 2008). Therefore, we simulated brain responses to the six levels of the narrative-like nested structure separately.

The simulated responses were generated using a linearly increasing temporal integration function (Figure 5a), based on prior work showing that information accumulation is accompanied by gradually increasing activation within phrases/sentences (Chang et al., 2020; Fedorenko et al., 2016; Giglio et al., 2021; Matchin et al., 2017; Nelson et al., 2017; Pallier et al., 2011) and paragraphs (Ezzyat & Davachi, 2011; Yarkoni et al., 2008) (a similar sentence/paragraph length

effect was also observed in our data; see Figure S10). The linearly increasing temporal integration function accumulates activity within the interval between linguistic boundaries at given levels and flushes out the accumulated activity at linguistic boundaries.

To account for hemodynamic lag in the fMRI signal, we applied a canonical hemodynamic response function (HRF) (Figure 5b). We averaged the inter-level lag correlations across thirty different simulated structures (equivalent to 30 different stories) and extracted the peak lags. This peak-lag analysis parallels the analysis previously applied to the fMRI data.

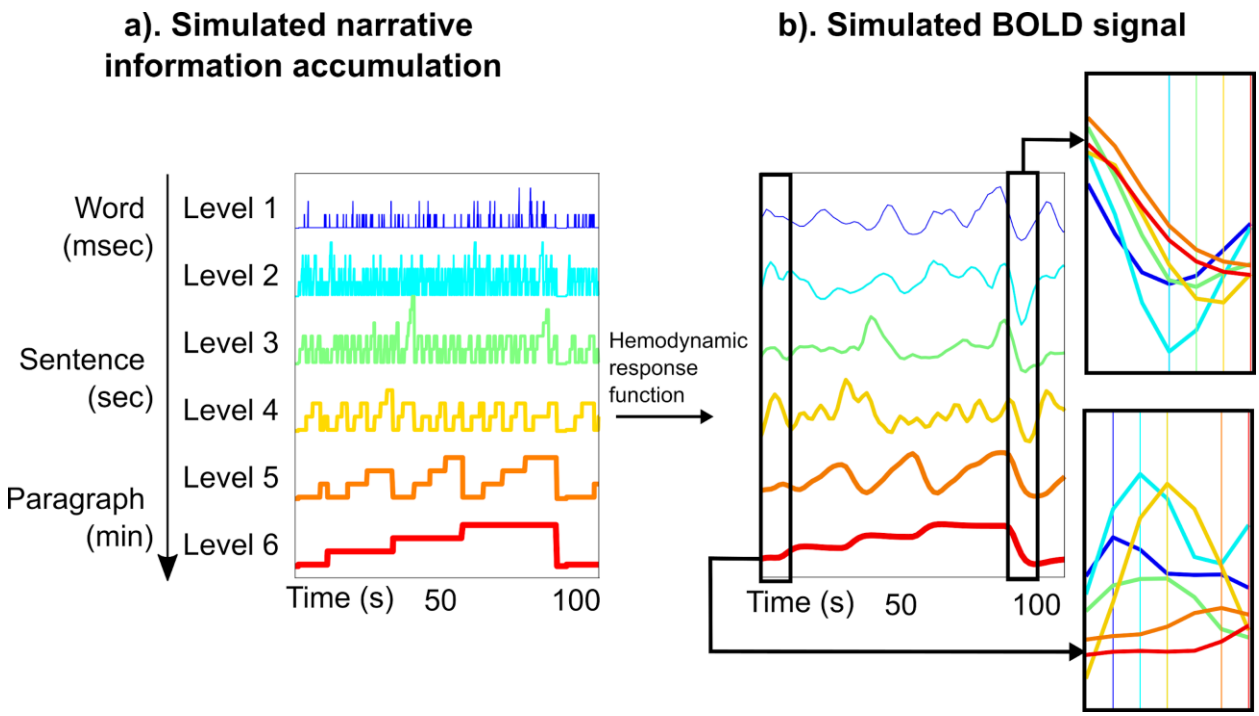



*Figure 5. Simulating information accumulation in nested narrative structures and the corresponding brain responses. (a) Information accumulation within a simulated narrative structure is generated by a linearly increasing temporal integration function. We postulated that information accumulation is accompanied by increased activity. (b) BOLD responses generated by HRF convolution. This visualization is based on parameters estimated from a spoken story stimulus (Table 1).*

The simulation allows us to systematically adjust the narrative structure and the temporal integration function to reveal the conditions under which the lag gradient emerges. We first performed the simulation with a set of "reasonable" parameters roughly motivated by properties of the narrative stimuli and a simple temporal integration function reflecting linear temporal accumulation (Table 1).

| **Exemplar simulation parameters** | |
|---|---|
| speech rate (relative to “Sherlock”) | 1 |
| mean unit length | 3 |
| unit length variance | 0.5 |
| temporal integration function | linearly increasing |
| mean pause length | 3 sec |
| pause length effect size (in SD of the simulated activity) | 0.1 |

*Table 1. A set of exemplar stimulation parameters motivated by a spoken story (“Sherlock”). See the Methods section for a detailed description of each parameter. SD: standard deviation.*

This simple simulation was sufficient to reproduce the inter-network lag gradient observed in the fMRI data (Figure 6; as well as the ISFC at lag zero; Figure S11). In addition, we also compared the spectral properties of the simulated and real BOLD signals (Figure S12). We first computed the average power spectral density (PSD) across stories. Replicating results reported by Stephens and colleagues (2013), we found stronger low-frequency fluctuations in regions with longer TRWs. Computing the PSD of the simulated brain responses similarly revealed increased low-frequency power in responses to high-level structure with longer intervals between boundaries. We then adjusted one parameter at a time to explore the parameter space constrained by natural speech.

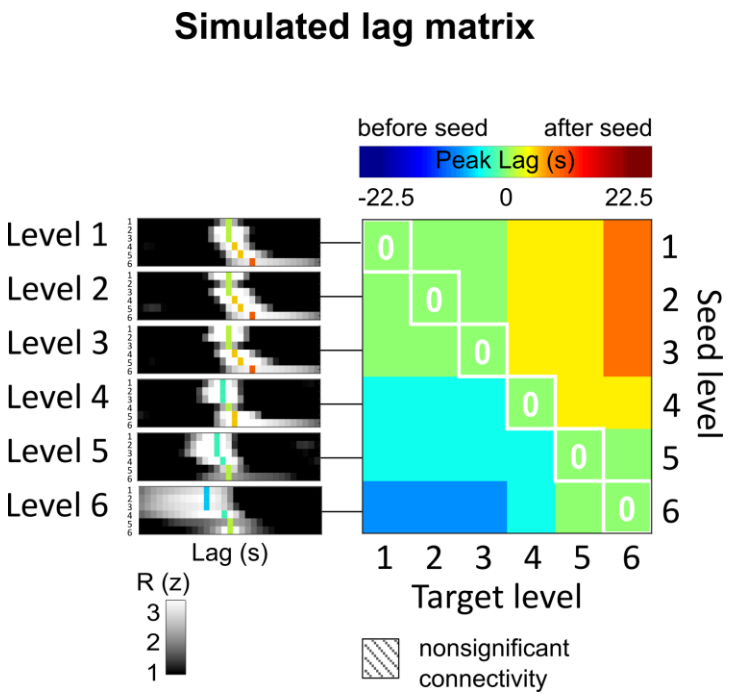



*Figure 6. The simulated inter-level lag matrix recapitulates the peak lag matrix observed during story-listening fMRI (Figure 4a). The same simulation parameters were used as described in Table 1 ($p < .05$, FDR corrected).*

***Key parameters for the emergence of a lag gradient***

Within the bounds of natural speech, we observed that the simulated inter-network lag gradient is robust to varying lengths of linguistic/narrative units (mean: 2–4). Longer length generated longer units, often with the top layers exceeding the length of the simulated story (i.e. 3000 words; variance: 0.1–1). The duration of inter-paragraph pauses was estimated from two stories (mean length: 1.5–3 sec, estimated from "Sherlock" and "Merlin" datasets; pause effect size: 0.01–1 SD of simulated activity). We also found that the model, similar to neural responses as observed in Lerner et al. (2014), was robust to variations in speech rate (0.5–1.5, relative to "Sherlock" speech rate) (Figure S13).

On the other hand, we found that the nested structure is crucial to generate the observed lag gradient. We computed inter-level lag-correlation using simulated responses to different nested structures, which preserved the spectral properties of individual time series but disrupted their nesting relationship. No significant lag correlation was found when violating the nested structure of the narrative stimulus (Figure 7).

In addition to the aforementioned linearly increasing integration function, we also explored several other temporal integration functions. We found that linearly and logarithmically increasing functions both yielded the inter-network lag gradient, but not the symmetric triangular or boxcar functions. The linearly decreasing function resulted in a reversed lag gradient (Figure 7). These results suggest that the nested structure of naturalistic stimulus and a monotonically increasing integration function are key to give rise to the lag gradient.

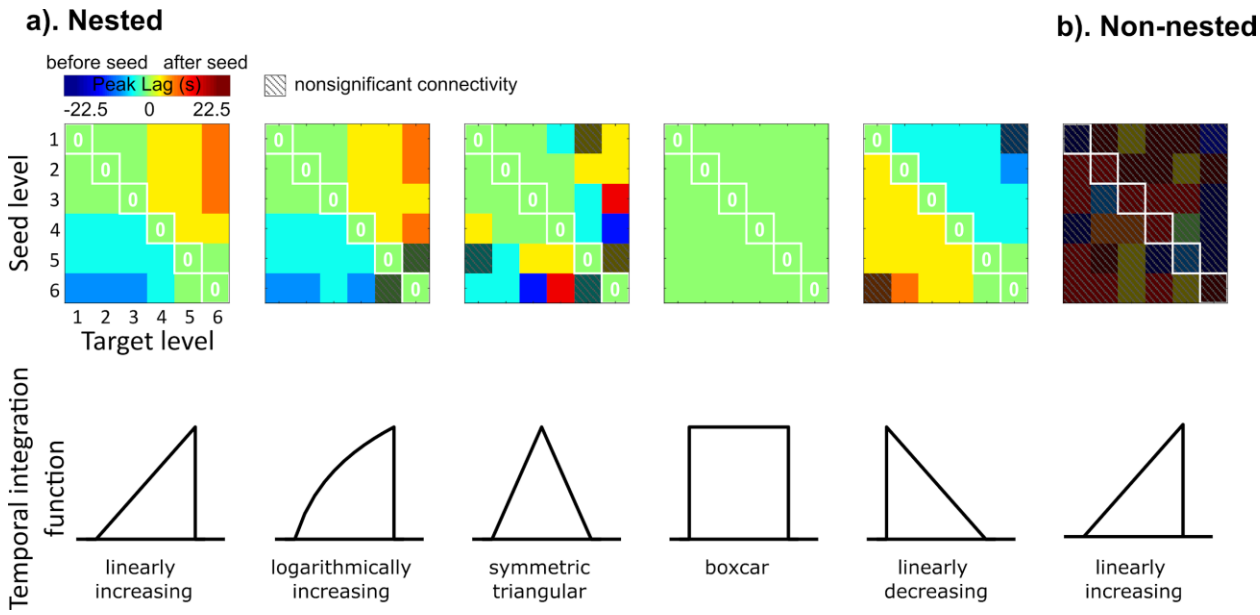



*Figure 7. Key simulation parameters for the emergence of the lag gradient. (a) Lag matrices were generated using different temporal integration functions. (b) Lag matrix from the non-nested structure. The non-nested structure was created by combining levels extracted from independently generated nested structures (the same simulation parameters as in Table 1 unless otherwise indicated) ($p < .05$, FDR correction).*

## Discussion

By applying lag-ISFC to a collection of fMRI datasets acquired while subjects listened to spoken stories, we revealed a temporal progression of story-driven brain activity along a cortical hierarchy for narrative comprehension (Figure 4a). The temporal cascade of cortical responses summarized by the inter-network lag gradient was consistent across stories as well as at a coarse- and fine-grained functional network definition (Figure S5). The results provide evidence for the gradual integration of information along the processing hierarchy, as information flows from early sensory areas into higher-order cortical areas (Figure 1). In support of our interpretation, we found that the lag gradient is absent during rest, when there is no stimulus-evoked processing

(Figure 4b), and also when the temporal structure of the story is disrupted due to word scrambling (Figure 4c).

Our results cannot be explained by regional variations in neurovascular coupling (Rangaprakash et al., 2018) or transient activity impulse at event boundaries. If the lag-gradient only reflects variations in neurovascular coupling across regions, it should be present both when we isolate stimulus-driven activity using ISFC and when we examine idiosyncratic neural responses using WSFC. Instead, however, the information flow was detected only with ISFC, but not WSFC (Figure S4). Furthermore, differences in the hemodynamic response function across brain areas are usually reported at shorter timescales of ~1–2 seconds (Bright et al., 2009; Handwerker et al., 2004) than the 0–9-second inter-network lag differences observed here in the context of narrative comprehension (Figure 4). In addition, we found that transient event boundaries (Ezzyat & Davachi, 2011; Speer et al., 2007; Whitney et al., 2009; Yarkoni et al., 2008; Zacks et al., 2001, 2010) did not account for the lag gradient (Figure S7).

Our simulation provides a simple yet effective account for how information accumulation within linguistic/narrative units can give rise to the inter-network lag gradient by identifying three conditions: (1) nested narrative structure in which the duration of linguistic/narrative units increased from lower to higher levels (Figure 1 and Figure 7); (2) a cortical hierarchy of increasing timescales such that different levels of the narrative are processed in different brain areas (Figure 1 and Figure 5); and (3) information accumulation within the boundaries of events at each processing level, combined with a reset of activity (buffer clearing) at event boundaries (see temporal integration function in Figure 1c and Figure 7a). This simple model was sufficient to explain how information is integrated at varying granularity (e.g. word, sentence, and paragraph) to yield the inter-network lag gradient and spectral properties observed in the fMRI

data (Figure 6, Figure S12). The simulation indicates that the nested structure and the increased activity within varying temporal receptive windows were key to the emergence of the lag gradient (Figure 7). In contrast, adjustments of the other parameters within the bounds of natural speech (i.e. speech rate, silent pause, and length of linguistic/narrative unit) did not change the gradient pattern (Figure S13).

The simulation provides a simple model which bridges the discovery of TRWs using natural stimuli (Hasson et al., 2008; Lerner et al., 2014) and the accumulation of activity within linguistic units found using simple, well-controlled stimuli (e.g. sentences with similar structures) (Chang et al., 2020; Fedorenko et al., 2016; Giglio et al., 2021; Matchin et al., 2017; Nelson et al., 2017; Pallier et al., 2011)al., 2017; Pallier et al., 2011). Importantly, we note that our model is not the only one that could generate the lag gradient. Our aim is to combine separate findings that point to the same cortical hierarchy with the simplest model possible. In addition, narrative processing is unlikely to be purely unidirectional (Pickering & Gambi, 2018). The lag gradient only captures the dominant bottom-up information flow (Figure S6). More studies are needed to examine recurrent or bidirectional connectivity, causal relations between networks, and nonstationary information flow over time.

Our results are also consistent with reports on the spatiotemporal dynamics of brain responses to naturalistic stimuli. A hierarchically nested spatial activation pattern has been revealed using movie, spoken story, and music stimuli (Baldassano et al., 2017; Geerligs et al., 2021; Williams et al., 2021). Chien et al. (2020) reported a gradual alignment of context-specific spatial activation patterns, which was rapidly flushed at event boundaries, similar to the temporal integration function we adopted here. Taken together, the empirical findings, combined with our simulation, indicate that the spatiotemporal neural dynamics reflect the structure of naturalistic,

ecologically-relevant inputs (Kiebel et al., 2008) and that such information is preserved even with the poor temporal resolution of fMRI.

Our results demonstrate both the importance of using inter-subject methods to isolate stimulus-driven signals and the value of data aggregation. The fact that we obtained non-zero inter-network lag only with ISFC but not WSFC (Figure S4) indicates that stimulus-driven network configuration may be masked by the idiosyncratic fluctuations that dominate WSFC analyses (Nastase et al., 2019; Simony et al., 2016). Furthermore, although the inter-network lags could be observed within individual stories (Figure 3, Figure S3), the gradient pattern is much clearer after aggregating across stories (Figure 4). Data aggregation is particularly important when using naturalistic stimuli because it is impossible to control the structure of each narrative (e.g. speaking style, duration, complexity, and content; (Hamilton & Huth, 2018; Lee et al., 2020; Sonkusare et al., 2019, 2019; Willems et al., 2020). Further work will be needed to decode the content of narrative representations—specific to each story—as they are transformed along the cortical hierarchy.

## Materials and Methods

### *fMRI datasets*

This study relied on eight openly available spoken story datasets. Seven datasets were used from the "Narratives" collection (OpenNeuro: https://openneuro.org/datasets/ds002245) (Nastase et al., 2021), including “Sherlock” and “Merlin” (18 participants, 11 females) (Zadbood et al., 2017), “The 21$^{st}$ year” (25 participants, 14 females) (Chang et al., 2021), “Pie Man (PNI)”, “I Knew You Were Black”, “The Man Who Forgot Ray Bradbury”, and “Running from the Bronx (PNI)” (48 participants, 34 females). One dataset was used from Princeton Dataspace: “Pie Man”

(36 participants, 25 females)

(https://dataspace.princeton.edu/jspui/handle/88435/dsp015d86p269k) (Simony et al., 2016). Two non-story datasets were also included as controls: a word-scrambled "Pie Man" (36, participants, 20 females) dataset and a resting-state dataset (36 participants, 15 females) (see the Princeton DataSpace URL above) (Simony et al., 2016).

All participants reported fluency in English and were 18–40 years in age. The criteria of participant exclusion have been described in previous studies for "Sherlock", "Merlin", "The 21st year", and "Pie Man." For "Pie Man (PNI)", "I Knew You Were Black", "The Man Who Forgot Ray Bradbury", and "Running from the Bronx (PNI)," participants with comprehension scores 1.5 standard deviations lower than the group means were excluded. One participant was excluded from "Pie Man (PNI)" for excessive movement (translation along the z-axis exceeding 3 mm).

All participants provided informed, written consent, and the experimental protocol was approved by the institutional review board of Princeton University.

***fMRI preprocessing***

fMRI data were preprocessed using FSL (https://fsl.fmrib.ox.ac.uk/), including slice time correction, motion correction, and high-pass filtering (140 s cutoff). All data were aligned to standard 3 × 3 × 4 mm Montreal Neurological Institute space (MNI152). A gray matter mask was applied.

***Functional networks***

Following Simony et al. (2016), we defined 6 intrinsic connectivity networks within regions showing reliable responses to spoken stories. Voxels showing top 30% ISC in at least 6 out of the 8 stories were included. Using the k-means method (L1 distance measure), these voxels were

clustered according to their within-subject functional connectivity with all the voxels during resting. We refer to these functional networks as the auditory (AUD), ventral language (vLAN), dorsal language (dLAN), attention (ATT), and default mode (DMNa and DMNb) networks (Figure 2). To ensure that our results hold for finer-grained functional networks, we further divided each of the six networks into ten subnetworks, again by applying k-means clustering to resting-state WSFC (k=10 within each superordinate network) (Figure S5).

To compare these intrinsic functional networks to the TRW hierarchy, we computed the TRW index (i.e. intact > word-scrambled story ISC) following (Yeshurun et al., 2017) for voxels within regions showing reliable responses to spoken stories, using the intact and word-scrambled Pie Man. Six TRW networks were then generated by splitting the TRW indices into six bins by five quantiles (Figure S1).

***WSFC, ISFC, and ISC***

In this study, within-subject functional connectivity (WSFC) refers to within-subject inter-region correlation, while inter-subject functional connectivity (ISFC) refers to inter-subject inter-region correlation. Inter-subject correlation (ISC) refers to a subset of ISFC, namely, ISFC between homologous regions (Figure S2). ISFC and ISC were computed using the leave-one-subject-out method, i.e. correlation between the time series from each subject and the average time series of all the other subjects (Nastase et al., 2019).

Before computing the correlation, the first 25 and last 20 volumes of fMRI data were discarded to remove large signal fluctuations at the beginning and end of time course due to signal stabilization and stimulus onset/offset. We then averaged voxelwise time series across voxels within network/region masks and z-scored the resulting time series.

Lag-correlations were computed by circularly shifting the time series such that the non-overlapping edge of the shifted time series was concatenated to the beginning or end. The left-out subject was shifted while the average time series of the other subjects remained stationary. Fisher's z transformation was applied to the resulting correlation values prior to further statistical analysis.

***ISFC lag matrix***

We computed the network × network × lag-ISFC matrix (Figure S2) and extracted the lag with peak ISFC (correlation) value for each network pair (Figure 3). The peak ISFC value was defined as the maximal ISFC value within the window of lags from -15 to +15 TRs; we required that the peak ISFC be larger than the absolute value of any negative peak and excluded any peaks occurring at the edge of the window.

To obtain the mean ISFC across stories (Figure 4), we applied two statistical tests. Only ISFC that passed both tests were considered significant. First, we performed a parametric one-tailed one-sample t-test to compare the mean ISFC against zero (N = 8 stories) and corrected for multiple comparisons by controlling the false discovery rate (FDR; Benjamini & Hochberg, 1995; 6 seed × 6 target × 31 lags; $q < .05$).

Second, to exclude ISFC peaks that only reflected shared spectral properties, we generated surrogates with the same mean and autocorrelation as the original time series by time-shifting and time-reversing. We computed the correlation between the original seed and time-reversed target with time-shifts of -100 to +100 TRs. The resulting ISFC values were averaged across stories and served as a null distribution. A one-tailed z-test was applied to compare ISFCs within the window of lag -15 to +15 TRs against this null distribution. The FDR method was used to control for multiple comparisons (seed × target × lags; $q < .05$). When assessing ISFC for each

story (Figure 3c and Figure S3), only this second test was applied and all possible time-shifts were used to generate the null distribution.

***Principal component analysis of the lag matrix***

We examined whether multiple information flows similarly contributed to the lag matrix, using the method introduced by Mitra et al. (2015). We applied PCA to the lag matrix obtained from the averaged ISFC across stories (Figure 4a), after transposing the matrix and zero-centering each column. Each principal component represents a pattern of relative lags, in other words, information flow. We computed the proportion of overall variance in the lag matrix accounted for by each component in order to determine whether more than one component played an important role (Figure S6).

***Word/sentence/paragraph boundary effect***

To test the transient effect of linguistic boundaries on inter-network lag, we computed the lag-ISFC after regressing out activity impulses at boundaries (Figure S7). A multiple regression model was built for each subject. The dependent variable was the averaged time series of each network, removing the first 25 scans and the last 20 scans as in the ISFC analysis. The regressors included an intercept, the audio envelope, and three sets of finite impulse functions (-5 to +15 TRs relative to boundary onset), corresponding to word, sentence, and paragraph (event) boundaries. We then recomputed lag-ISFC based on the residuals of the regression model.

***Word/sentence/paragraph length effect***

We replicated the sentence length (Chang et al., 2020; Fedorenko et al., 2016; Giglio et al., 2021; Matchin et al., 2017; Nelson et al., 2017; Pallier et al., 2011) and paragraph length (Ezzyat & Davachi, 2011; Yarkoni et al., 2008) effect with the “Sherlock” and “Merlin” datasets, which were collected from the same group of participants. The onsets and offsets of each word,

sentence, and paragraph (event) were manually time-stamped. Given the difficulty of labeling the onset/offset of each syllable, they were estimated by dividing the duration of each word by the number of syllables it contains.

We built individual GLM models that included regressors corresponding to the presence of syllable, word, sentence, and paragraph respectively, accompanied by three parametric modulators: accumulated syllable number within words, accumulated word number within sentences, and accumulated sentence number within paragraphs. These parametric regressors were included to test whether brain activations accumulate toward the end of word/sentence/paragraph; the longer the word/sentence/paragraph the stronger the activations. In addition to the regressors of interest, one regressor was included for speech segments without clear paragraph labels. We did not orthogonalize the regressors to each other.

Effect maps of the three parametric modulators (i.e. word length, sentence length, and paragraph length) from the individual level models of both stories were smoothed with a Gaussian kernel (FWHM = 8 mm) and input to three group-level models to test the word, sentence, and paragraph length effects respectively (flexible factorial design including the main effects of story and participant; $p < .005$, not corrected). The sentence and paragraph length effects are shown in Figure S10. Using the same threshold, no word length effect was observed,

***Simulating the BOLD response to nested narrative structures***

To test whether information accumulation at different timescales could account for the inter-network lag during story-listening, we simulated the nested narrative structures closely following the statistical structure of real spoken stories and generated BOLD responses to each narrative level (Figure 5). To build the first level of a nested structure, we sampled a sequence of 3000 word durations with replacement from “Sherlock,” which is the longest example of spontaneous

speech among our datasets, recorded from a non-professional speaker without rehearsal or script (Figure S8). Boundaries between units at the first level were set up accordingly.

*Unit length*

First-level units were integrated into units of the next level with a lognormal distributed unit length (Figure S8); e.g. integrating three words into a phrase (unit length = 3). Boundaries between second-level units were inserted accordingly. Second-level units were integrated into the third-level units following the same method. A nested structure of six levels was thus generated.

*Temporal integration function*

Postulating that information accumulation is accompanied by increased activity, brain responses to each level of the nested structure were generated as a function of unit length. For example, a linear temporal integration function generates activity [1 2 3] for a “phrase" (i.e. a Level 2 unit) consisting of three "words" (i.e. Level 1 units). The first (word) level integration was computed based on syllable numbers sampled from “Sherlock” along with word durations.

*Pause length and pause effect size*

In natural narrative, boundaries between high-level units were often accompanied by silent pauses (Figure S9). Therefore, we inserted pauses with normally distributed length at the boundaries of the highest level units (Figure S8). Activity during the pause period was set as 0.1 standard deviations below the minimum activity of each level.

To account for HRF delay in fMRI signals, we applied the canonical hemodynamic response function provided by the software SPM (https://www.fil.ion.ucl.ac.uk/spm/) (Penny et al., 2007) and resampled the output time series from a temporal resolution of 0.001 sec to 1.5 sec to match the TR in our data. We ran 30 simulations for each set of simulation parameters. Each simulation

produced different narrative structures (equivalent to different stories). The peak lag of the mean inter-level correlation across simulations was extracted and thresholded using the same method as in the ISFC analysis (Figure 6).

We started with a set of reasonable parameters (speech rate = 1, relative to "Sherlock"; unit length mean = 3; unit length variance = 0.5; temporal integration function = linearly increasing; mean pause length = 3 sec; pause effect size = 0.1 SD of the simulated activity) (Table 1) and explored alternative parameter sets within the bound of natural speech to test whether inter-level lag was robust to parameter changes (Figure 7 and Figure S13).

***Power-spectral density analysis***

To examine whether the simulated and real fMRI signals shared similar power spectra, we performed spectral analyses following Stephens et al. (2013) (Figure S12). For the real fMRI data, we estimated the power spectrum of the primary auditory area and a DMN area (precuneus). As for the connectivity analysis, we cropped the first 25 and last 20 scans and z-scored the time series. For each story, the resulting time series were averaged across subjects and normalized across time. The power spectrum of the group-mean time series was estimated using Welch's method with a Hamming window of width 99 sec (66 TRs) and 50% overlap (based on the parameters from Stephens et al., 2013). The power spectra of individual voxels were averaged within the anatomical masks of left Hesch's gyrus and left precuneus from the AAL atlas. The mean spectra across stories were then computed. The same analyses were applied to the simulated BOLD responses at each of the six levels and averaged across thirty simulations.

## Acknowledgment

This study is supported by the National Institute of Mental Health (R01-MH112357 and DP1-HD091948).

## Competing interest

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

## References

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., Norman, K. A., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, *95*(3), 709-721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Baria, A. T., Mansour, A., Huang, L., Baliki, M. N., Cecchi, G. A., Mesulam, M. M., & Apkarian, A. V. (2013). Linking human brain local activity fluctuations to structural and functional network architectures. *NeuroImage*, *73*, 144–155. https://doi.org/10.1016/J.NEUROIMAGE.2013.01.072

Bright, M. G., Bulte, D. P., Jezzard, P., & Duyn, J. H. (2009). Characterization of regional heterogeneity in cerebrovascular reactivity dynamics using novel hypocapnia task and BOLD fMRI. *NeuroImage*, *48*(1), 166–175. https://doi.org/10.1016/j.neuroimage.2009.05.026

Caucheteux, C., Gramfort, A., & King, J.-R. (2021). Model-based analysis of brain activity reveals the hierarchy of language in 305 subjects. *Conference on Empirical Methods in Natural Language Processing*. https://hal.archives-ouvertes.fr/hal-03361430/document

Chang, C. H. C., Dehaene, S., Wu, D. H., Kuo, W. J., & Pallier, C. (2020). Cortical encoding of linguistic constituent with and without morphosyntactic cues. *Cortex*, *129*, 281–295. https://doi.org/10.1016/j.cortex.2020.04.024

Chang, C. H. C., Lazaridi, C., Yeshurun, Y., Norman, K. A., & Hasson, U. (2021). Relating the past with the present: Information integration and segregation during ongoing narrative processing. *Journal of Cognitive Neuroscience*, *33*(6), 1106–1128. https://doi.org/10.1162/jocn_a_01707

Changeux, J.-P., Goulas, A., & Hilgetag, C. C. (2020). A connectomic hypothesis for the hominization of the brain. *Cerebral Cortex*, 1–25. https://doi.org/10.1093/cercor/bhaa365

Chaudhuri, R., Knoblauch, K., Gariel, M. A., Kennedy, H., & Wang, X.-J. (2015). A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron*, *88*(2), 419–431. https://doi.org/10.1016/j.neuron.2015.09.008

Chien, H. Y. S., & Honey, C. J. (2020). Constructing and forgetting temporal context in the human cerebral cortex. *Neuron*, *106*(4), 675-686.e11. https://doi.org/10.1016/j.neuron.2020.02.013

Christiansen, M. H., & Chater, N. (2015). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*. https://doi.org/10.1017/S0140525X1500031X

Demirtaş, M., Burt, J. B., Helmer, M., Ji, J. L., Adkinson, B. D., Glasser, M. F., Van Essen, D. C., Sotiropoulos, S. N., Anticevic, A., & Murray, J. D. (2019). Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. *Neuron*, *101*(6), 1181-1194.e13. https://doi.org/10.1016/j.neuron.2019.01.017

Dominey, P. F. (2021). Narrative event segmentation in the cortical reservoir. *BioRxiv*.

https://doi.org/10.1101/2021.04.23.441090

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological Science*, *22*(2), 243–252. https://doi.org/10.1177/0956797610393742

Fedorenko, E., Scott, T. L., Brunner, P., Coon, W. G., Pritchett, B., Schalk, G., & Kanwisher, N. (2016). Neural correlate of the construction of sentence meaning. *Proceedings of the National Academy of Sciences*, *113*(41), E6256–E6262. https://doi.org/10.1073/pnas.1612132113

Geerligs, L., van Gerven, M., Campbell, K. L., & Güçlü, U. (2021). A nested cortical hierarchy of neural states underlies event segmentation in the human brain. *BioRxiv*, 2021.02.05.429165. https://doi.org/10.1101/2021.02.05.429165

Giglio, L., Ostarek, M., Weber, K., & Hagoort, P. (2021). Commonalities and asymmetries in the neurobiological infrastructure for language production and comprehension. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhab287

Hamilton, L. S., & Huth, A. G. (2018). The revolution will not be controlled: Natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, 1–10. https://doi.org/10.1080/23273798.2018.1499946

Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, *21*(4), 1639–1651. https://doi.org/10.1016/j.neuroimage.2003.11.029

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences*, *19*(6), 304–313. https://doi.org/10.1016/j.tics.2015.04.006

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization

of cortical activity during natural vision. *Science*, *303*(5664), 1634–1640. https://doi.org/10.1126/science.1089506

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience*, *28*(10), 2539–2550. https://doi.org/10.1523/JNEUROSCI.5487-07.2008

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*(2), 423–434. https://doi.org/10.1016/j.neuron.2012.08.011

Huntenburg, J. M., Bazin, P. L., & Margulies, D. S. (2018). Large-scale gradients in human cortical organization. *Trends in Cognitive Sciences*, *22*(1), 21–31. https://doi.org/10.1016/j.tics.2017.11.002

Kauppi, J. P., Jääskeläinen, I. P., Sams, M., & Tohka, J. (2010). Inter-subject correlation of brain hemodynamic responses during watching a movie: Localization in space and frequency. *Frontiers in Neuroinformatics*, *4*(MAR), 5. https://doi.org/10.3389/fninf.2010.00005

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, *4*(11), e1000209. https://doi.org/10.1371/journal.pcbi.1000209

Lee, H., Bellana, B., & Chen, J. (2020). What can narratives tell us about the neural bases of human memory? *Current Opinion in Behavioral Sciences*, *32*, 111–119. https://doi.org/10.1016/j.cobeha.2020.02.007

Lerner, Y., Honey, C. J., Katkov, M., & Hasson, U. (2014). Temporal scaling of neural responses to compressed and dilated natural speech. *Journal of Neurophysiology*, *111*(12), 2433–

2444. https://doi.org/10.1152/jn.00497.2013

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, *31*(8), 2906–2915. https://doi.org/10.1523/JNEUROSCI.3684-10.2011

Matchin, W., Hammerly, C., & Lau, E. (2017). The role of the IFG and pSTS in syntactic prediction: Evidence from a parametric study of hierarchical structure in fMRI. *Cortex*, *88*, 106–123. https://doi.org/10.1016/J.CORTEX.2016.12.010

Mitra, A., Snyder, A. Z., Blazey, T., & Raichle, M. E. (2015). Lag threads organize the brain's intrinsic activity. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(52), E7307. https://doi.org/10.1073/pnas.1523893113

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., & Wang, X.-J. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience*, *17*(12), 1661–1663. https://doi.org/10.1038/nn.3862

Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience*, *14*(6), 667–685. https://doi.org/10.1093/scan/nsz037

Nastase, S. A., Liu, Y.-F., Hillman, H., Zadbood, A., Hasenfratz, L., Keshavarzian, N., Chen, J., Honey, C. J., Yeshurun, Y., Regev, M., Nguyen, M., Chang, C. H. C., Baldassano, C., Lositsky, O., Simony, E., Chow, M. A., Leong, Y. C., Brooks, P. P., Micciche, E., … Hasson, U. (2021). The "Narratives" fMRI dataset for evaluating models of naturalistic language comprehension. *Scientific Data*, *8*(1), 250–250. https://doi.org/10.1038/s41597-021-01033-3

Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., Cash, S. S., Naccache, L., Hale, J. T., Pallier, C., & Dehaene, S. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, *114*(18), E3669–E3678. https://doi.org/10.1073/pnas.1701590114

Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, *108*(6), 2522–2527. https://doi.org/10.1073/pnas.1018711108

Penny, W., Friston, K., Ashburner, J., Kiebel, S., & Nichols, T. (2007). Statistical parametric mapping: The analysis of functional brain images. In *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. https://doi.org/10.1016/B978-0-12-372560-8.X5000-1

Peters, M. E., Neumann, M., Zettlemoyer, L., & Yih, W. T. (2018). Dissecting contextual word embeddings: Architecture and representation. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, 1499–1509. https://doi.org/10.18653/v1/d18-1179

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Raut, R. V., Snyder, A. Z., & Raichle, M. E. (2020). Hierarchical dynamics as a macroscopic organizing principle of the human brain. *Proceedings of the National Academy of Sciences*, *117*(34), 20890–20897. https://doi.org/10.1073/pnas.2003383117

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension.

*Nature Communications*, *7*, 12141. https://doi.org/10.1038/ncomms12141

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic stimuli in neuroscience: Critically acclaimed. *Trends in Cognitive Sciences*, *23*(8), 699–714. https://doi.org/10.1016/j.tics.2019.05.004

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science*, *18*(5), 449–455. https://doi.org/10.1111/j.1467-9280.2007.01920.x

Stephens, G. J., Honey, C. J., & Hasson, U. (2013). A place for time: The spatiotemporal structure of neural dynamics during natural audition. *Journal of Neurophysiology*, *110*(9), 2019–2026. https://doi.org/10.1152/jn.00268.2013

Vig, J., & Belinkov, Y. (2019). Analyzing the structure of attention in a transformer language model. *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 63–76. https://doi.org/10.18653/v1/W19-4808

Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., & Kircher, T. (2009). Neural correlates of narrative shifts during auditory story comprehension. *NeuroImage*, *47*(1), 360–366. https://doi.org/10.1016/J.NEUROIMAGE.2009.04.037

Willems, R. M., Nastase, S. A., & Milivojevic, B. (2020). Narratives for neuroscience. *Trends in Neurosciences*, *43*(5), 271–273. https://doi.org/10.1016/j.tins.2020.03.003

Williams, J. A., Margulis, E. H., Nastase, S. A., Chen, J., Hasson, U., Norman, K. A., & Baldassano, C. (2021). High-order areas and auditory cortex both represent the high-level event structure of music. *BioRxiv*, 2021.01.26.428291.

Yarkoni, T., Speer, N. K., & Zacks, J. M. (2008). Neural substrates of narrative comprehension and memory. *NeuroImage*, *41*(4), 1408–1425.

https://doi.org/10.1016/J.NEUROIMAGE.2008.03.062

Yeshurun, Y., Nguyen, M., & Hasson, U. (2017). Amplification of local changes along the timescale processing hierarchy. *Proceedings of the National Academy of Sciences*, 201701652. https://doi.org/10.1073/pnas.1701652114

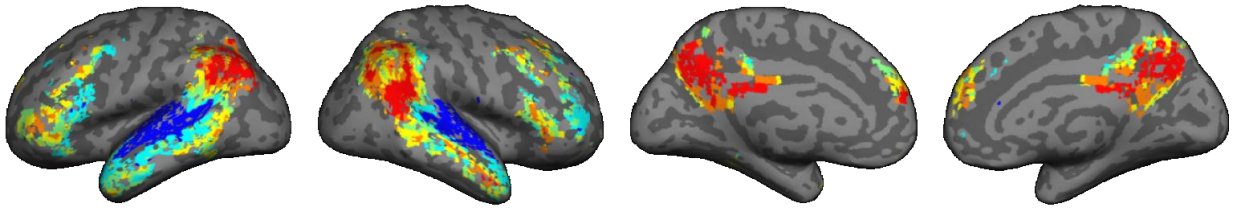
Yeshurun, Y., Nguyen, M., & Hasson, U. (2021). The default mode network: Where the idiosyncratic self meets the shared social world. *Nature Reviews Neuroscience*, *22*(3), 181–192. https://doi.org/10.1038/s41583-020-00420-w

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*(6), 651–655. https://doi.org/10.1038/88486

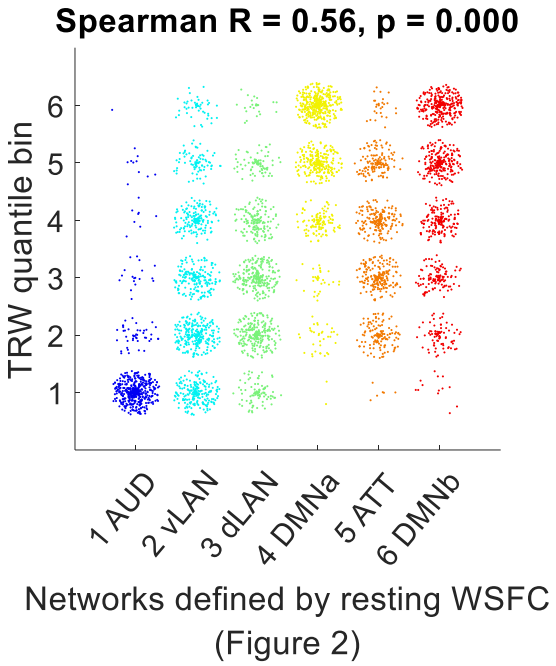
Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience*, *4*(October), 1–15. https://doi.org/10.3389/fnhum.2010.00168

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., & Hasson, U. (2017). How we transmit memories to other brains: Constructing shared neural representations via communication. *Cerebral Cortex*, *27*(10), 4988–5000. https://doi.org/10.1093/cercor/bhx202

**a). Networks defined by TRW index (intact > scrambled story ISC)**



**b).**




**c). Peak lag matrix between networks defined by TRW index**

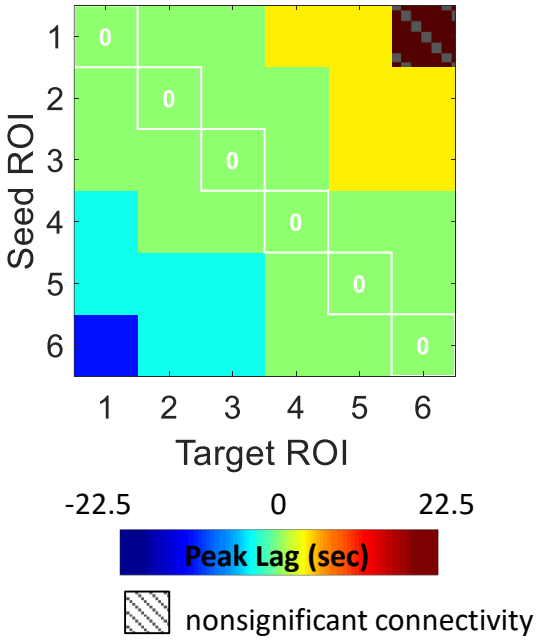



Figure S1. a) Networks generated by splitting the TRW indices (intact > scrambled story ISC) into 6 bins by five quantiles. b). Networks defined by TRW index shows a similar topographic gradient as the networks defined by resting-state WSFC (Figure 2), from the auditory areas to DMN, which is manifested by the significant correlation between the two sets of networks index. Radom jitters are added to better show the overlapped data points. c). Peak lag matrix between networks defined by TRW index across seven stories ($p < .05$, FDR corrected). Pie Man was excluded from this analysis, since it was used to compute the TRW index.

**ISFC**:
seed x target network at lag 0

**Lag-ISFC**:
seed x target x lag

**One seed lag-ISFC**:
target x lag

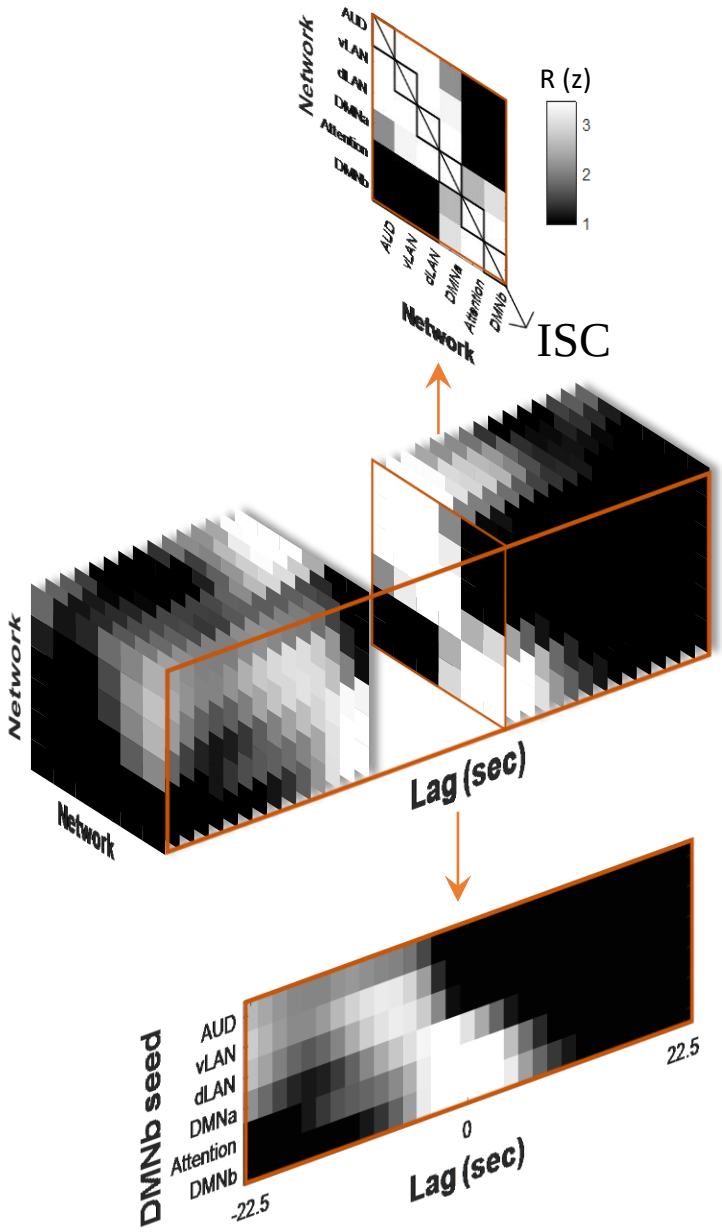



Figure S2. The relationship between inter-subject functional connectivity (ISFC), inter-subject-correlation (ISC), and lag-ISFC. This figure shows real data from “Sherlock.”

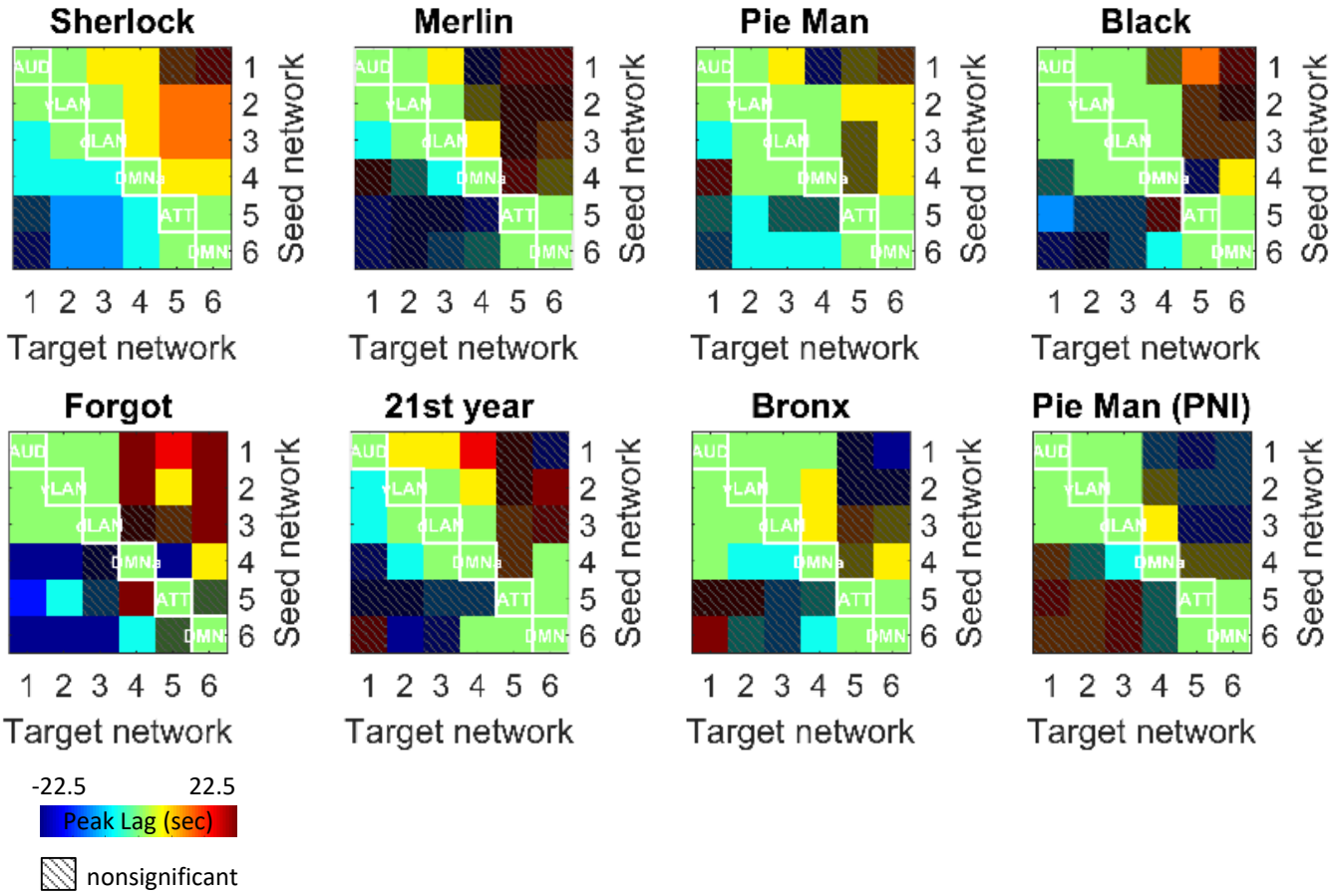



Figure S3. The network × network peak lag matrix based on the lag-ISFC in each individual story ($p < .05$, FDR corrected).

# Within-subject functional connectivity

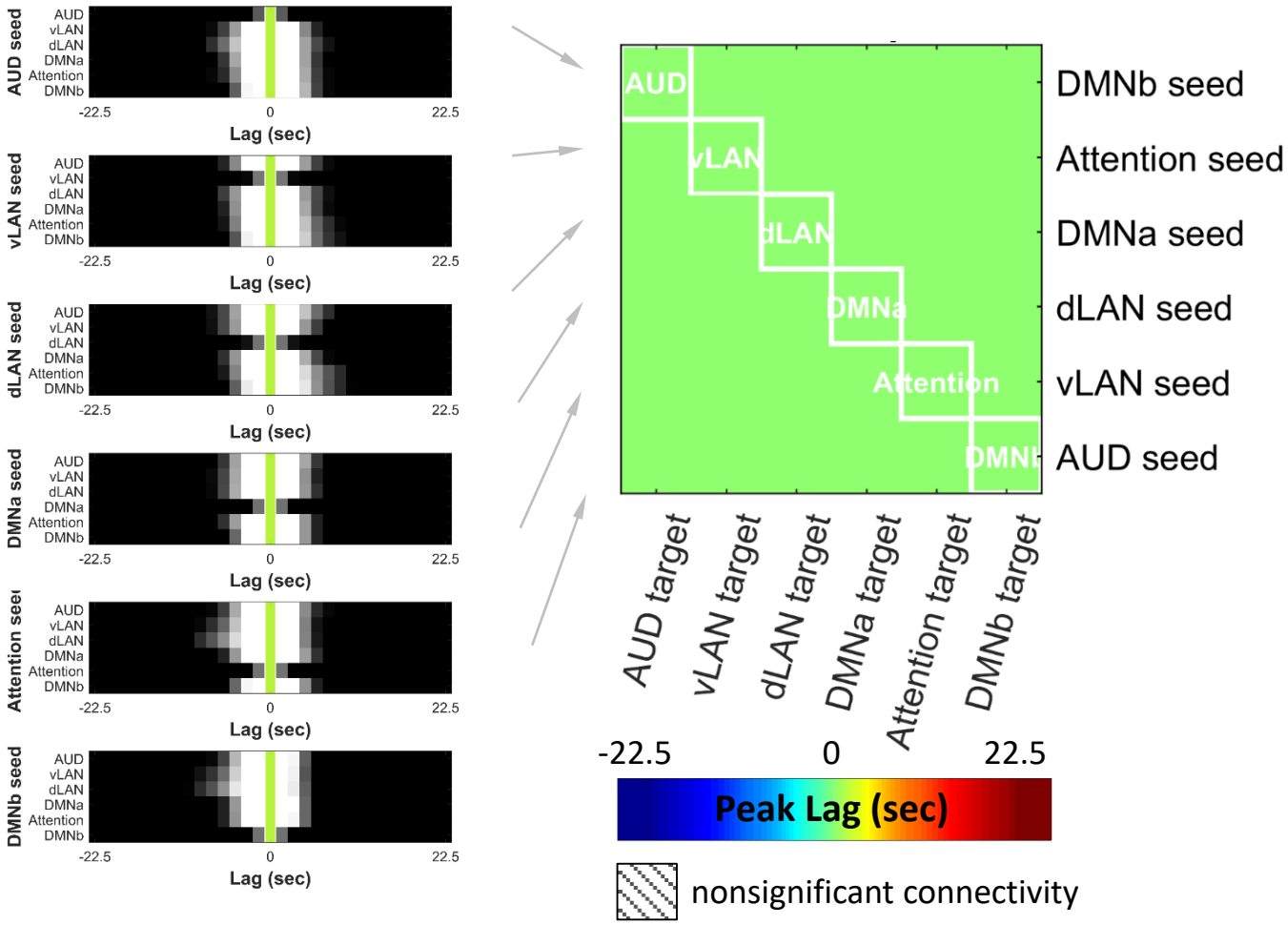



Figure S4. The network × network peak lag matrix based on the averaged lag-WSFC across eight stories ($p < .05$, FDR corrected).

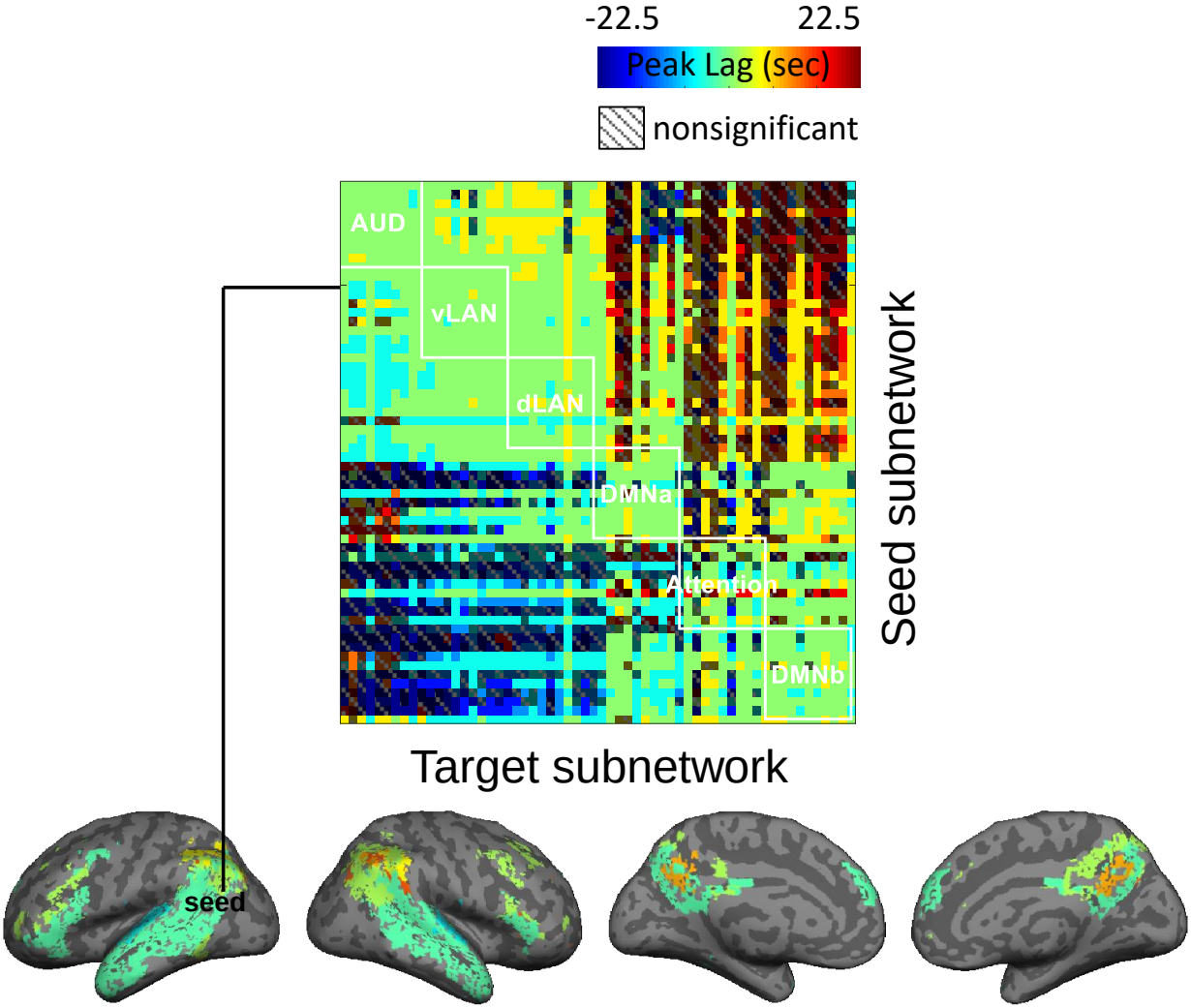



Figure S5. subnetwork x subnetwork peak lag matrix based on the averaged lag-ISFC across eight stories ($p < .05$, FDR corrected). The subnetworks were created by dividing each of the six main functional networks (Figure 2) into 10 subnetworks, applying k-means clustering to resting-state WSFC ($k = 10$ within each network). Lower panel shows the brain map of peak lags between one seed subnetwork (posterior superior/middle temporal gyrus) and all the sixty subnetworks

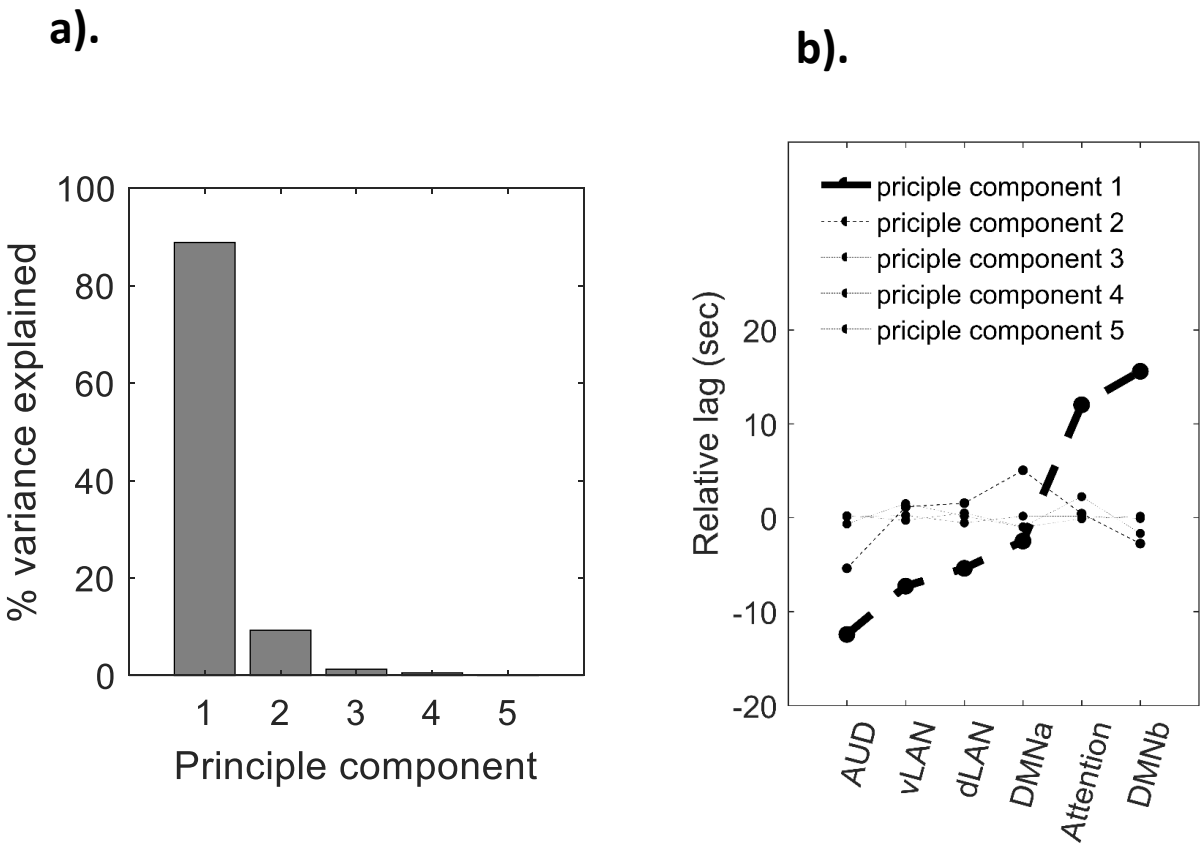
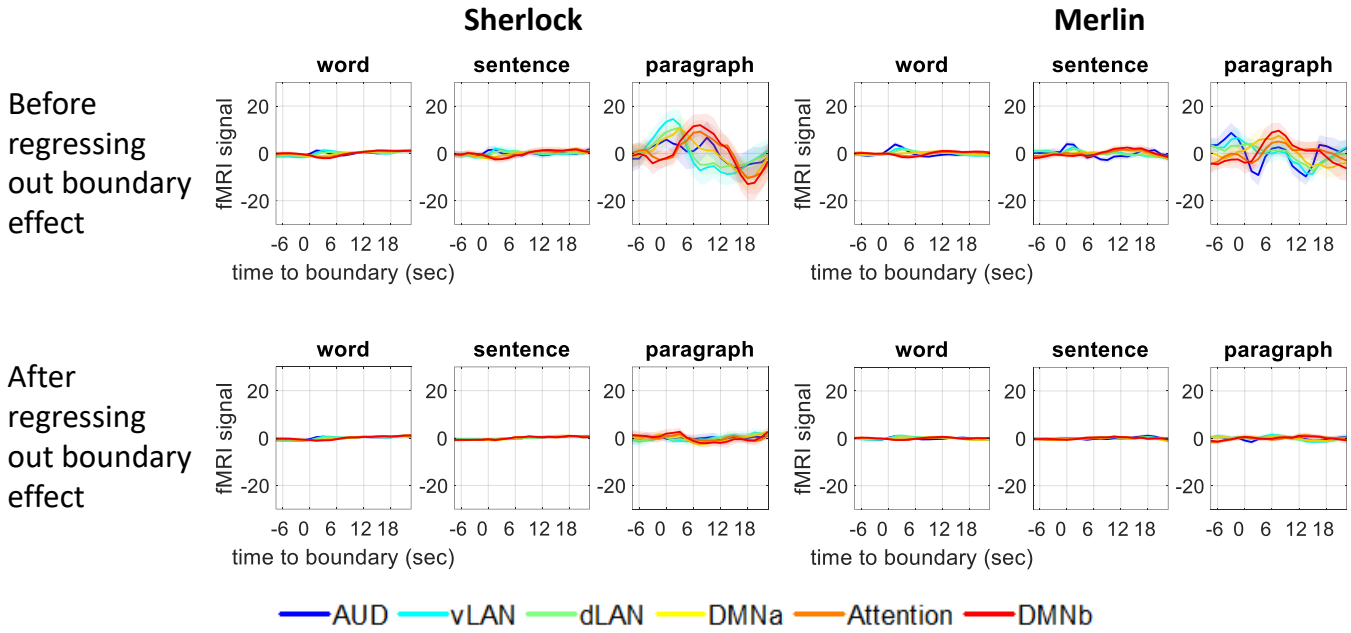



Figure S6. Principle component analysis of the inter-network lag matrix across eight stories (Figure 4a). (a) The percentage of variance explained by each principle component. (b) Relative-lag values from each principle component. Line thickness indicate the percentage of variance explained by that component.

**a). Transient activity impulse at boundaries**




**b). Boundary effect on inter-network ISFC lag matrix**

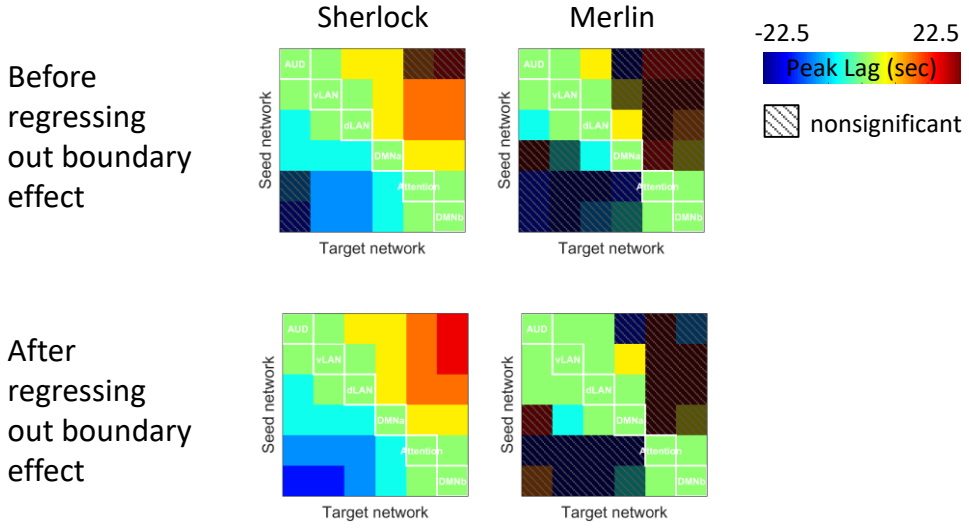



Figure S7. Boundary effect on the network x network peak lag matrix across stories (Figure 4a). (a) The fMRI signals around word, sentence, and paragraph boundaries before and after regressing out the boundary effects. Shaded areas indicate 95% confidence interval across subjects. (b) The peak lag matrix before and after regressing out the boundary effects ($p < .05$, FDR corrected).

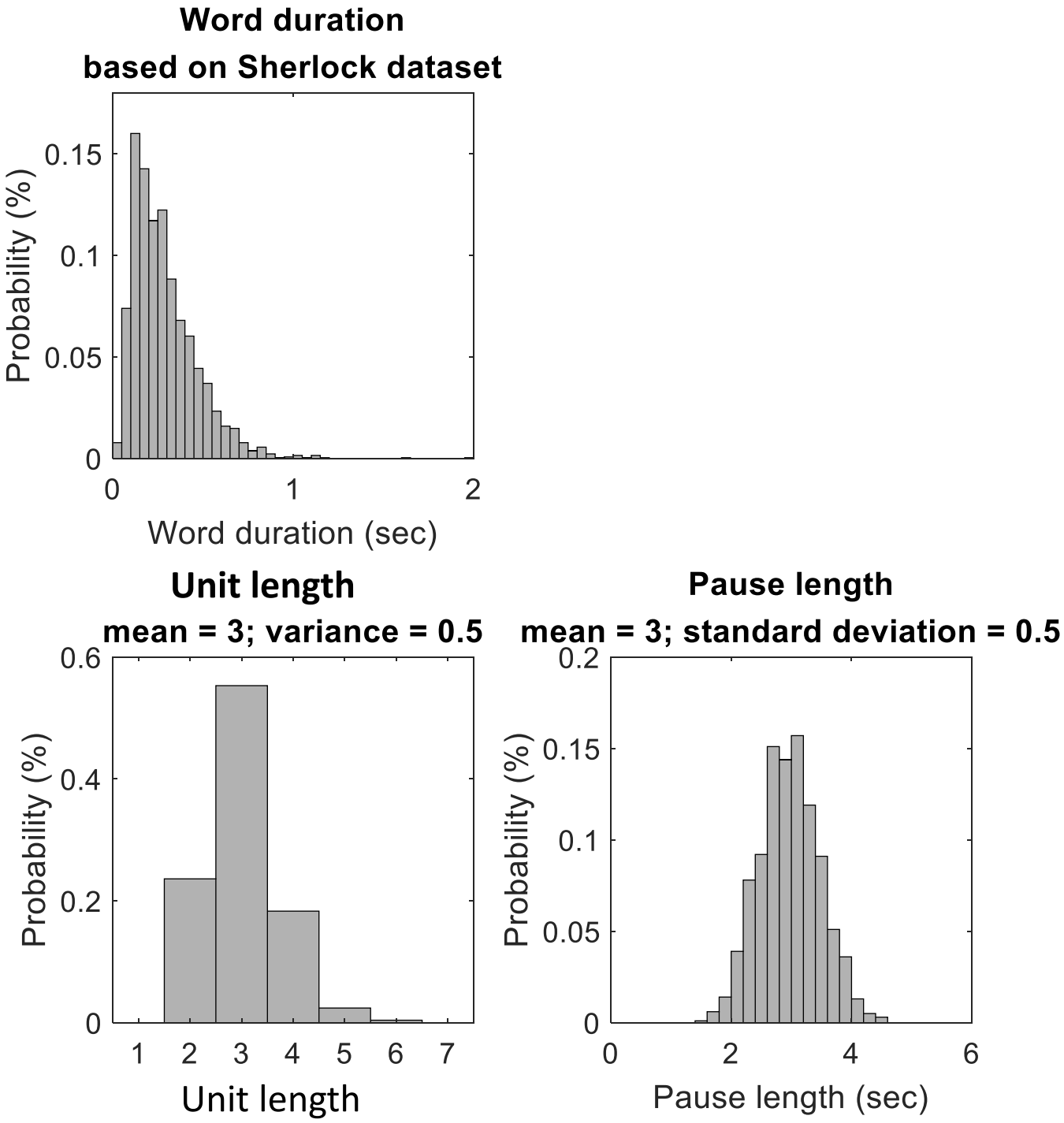



Figure S8. The distributions of word duration, unit length, and pause length used to simulate nested narrative structures (Table 1).

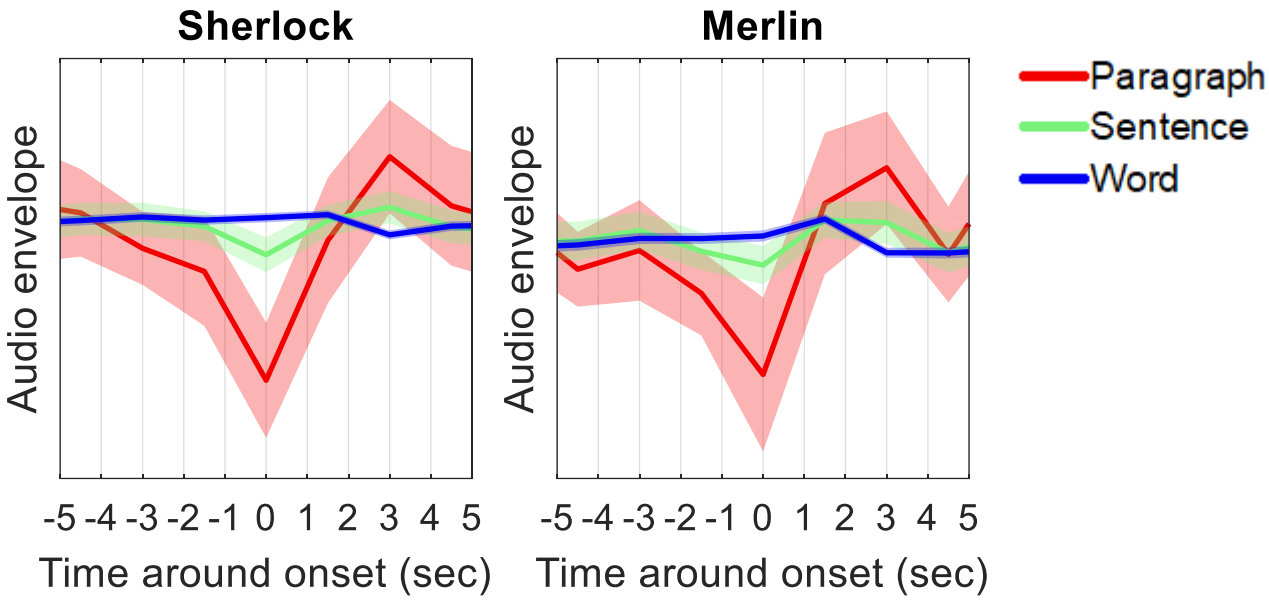



Figure S9. The silent pause between paragraphs shown in real spoken stories. Shaded areas indicate 95% confidence interval.

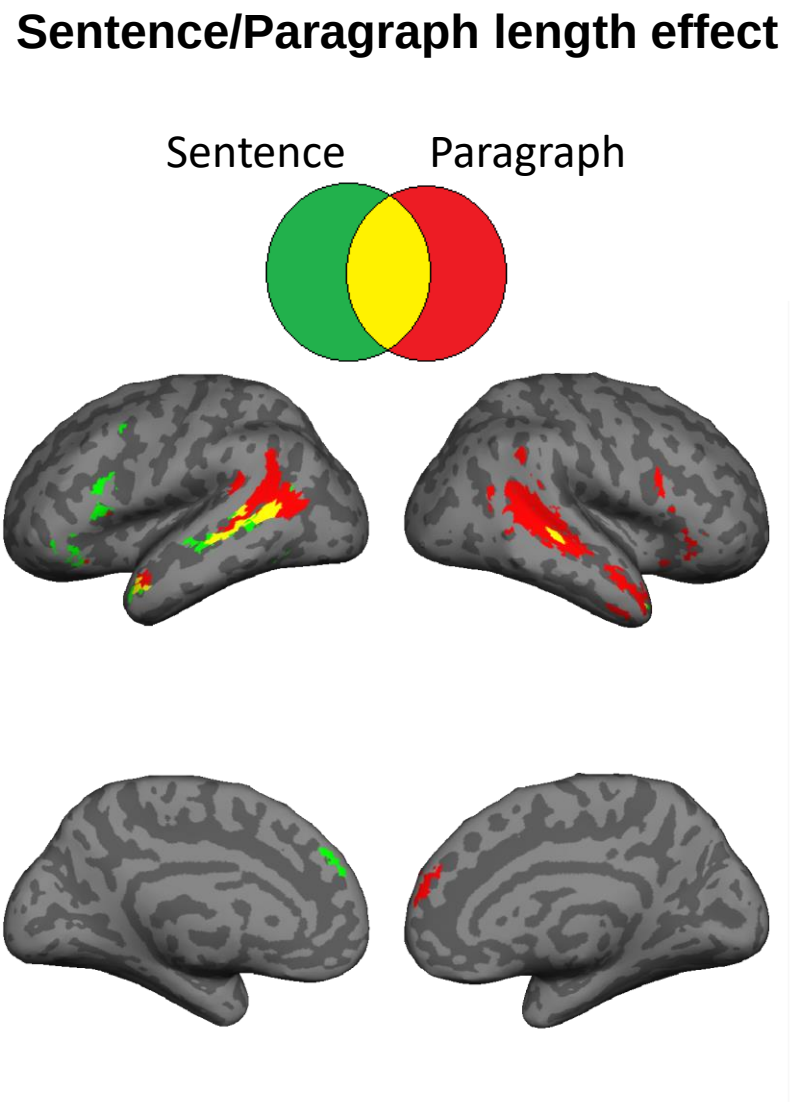



Figure S10. Sentence and paragraph length effects in two time-stamped stories (“Sherlock” & “Merlin”) ($p < .005$, uncorrected). Significant length effect indicates activation accumulation from the start toward the end of sentences or paragraphs.

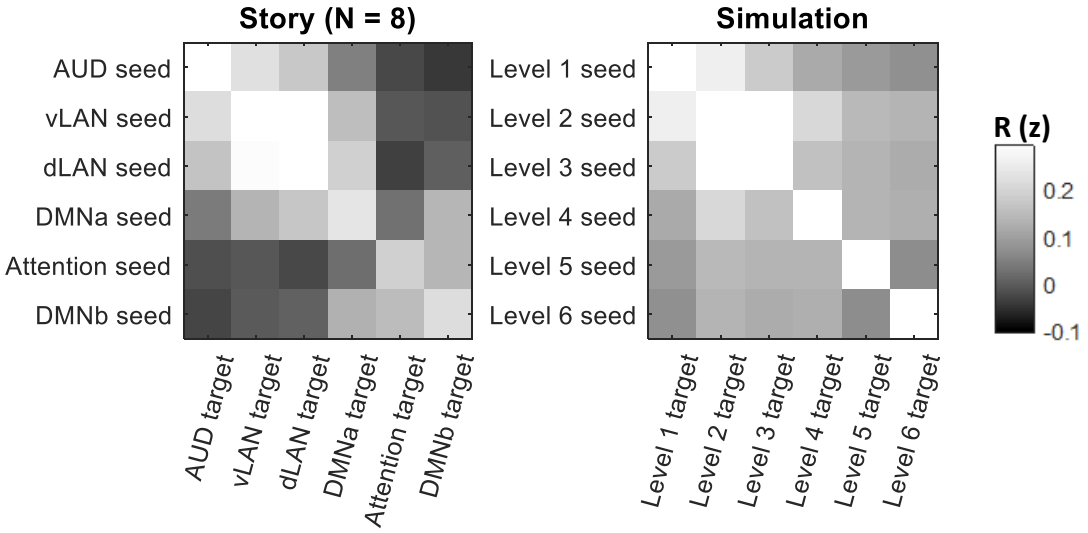



Figure S11. ISFC matrices at lag 0 in real and simulated stories (the same simulation parameters as in Table 1).

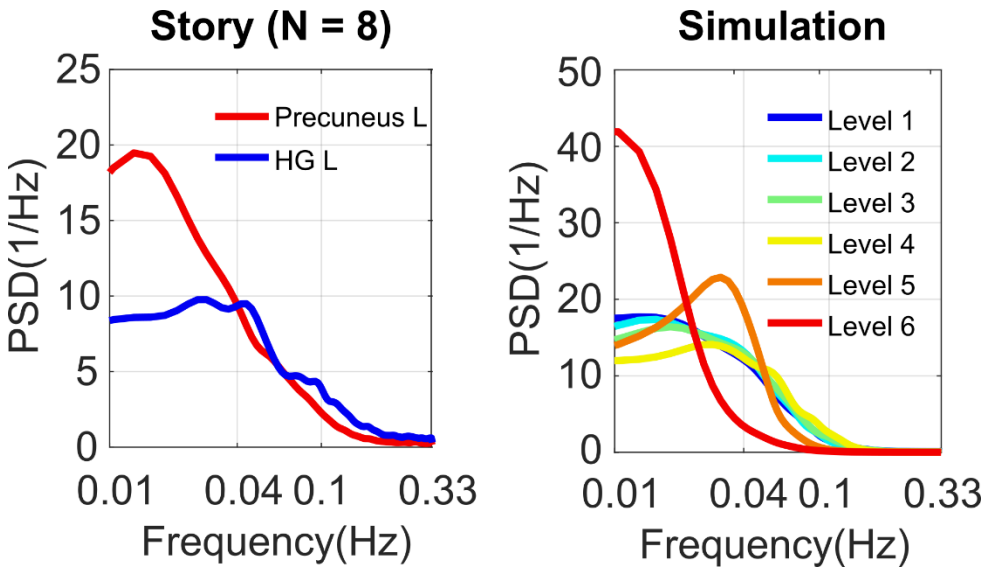
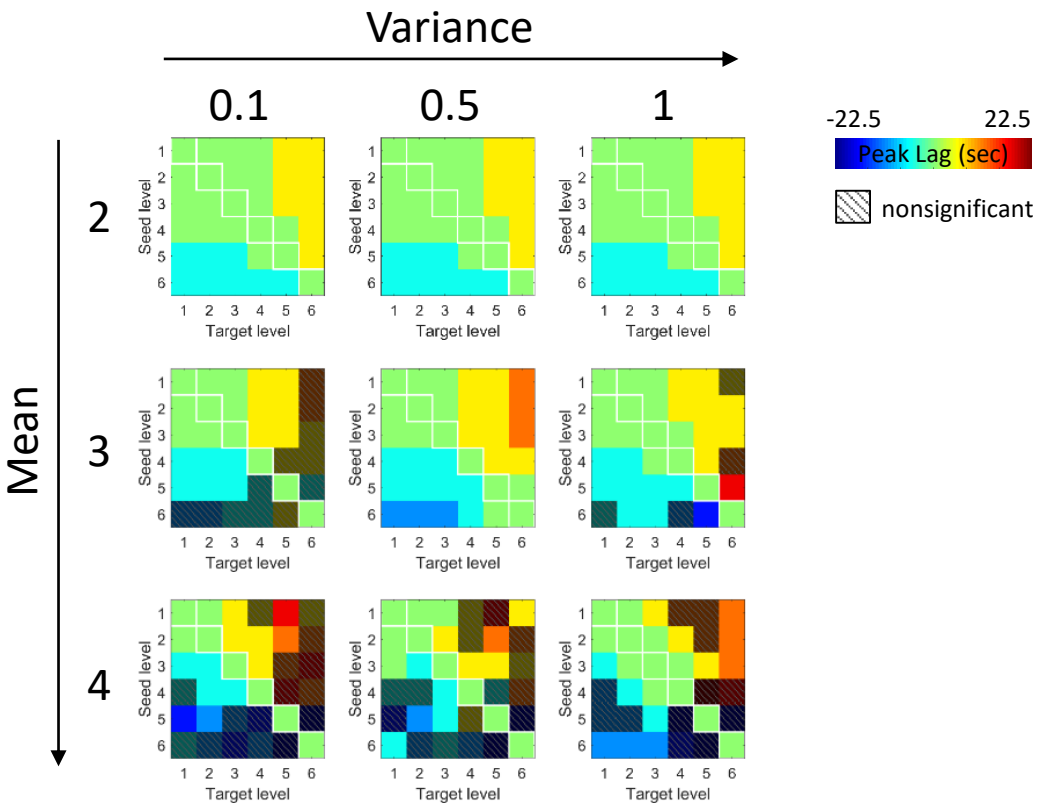
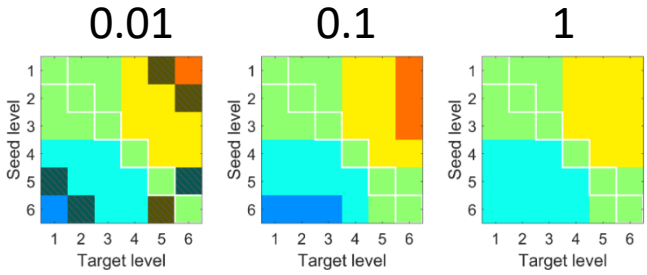



Figure S12. Power spectral densities of real (left) and simulated (right, the same parameter set as Table 1) BOLD responses to stories. PSD of the actual BOLD data exhibited stronger low-frequency fluctuations at regions with longer temporal receptive windows. Simulated BOLD responses to hierarchically nested narrative structures show a similar pattern.

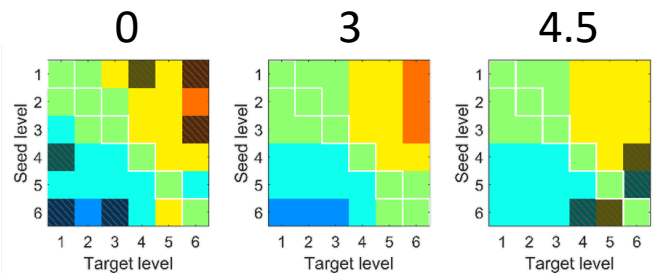
## Unit length




## Pause effect size
(SD of the simulated activity)




## Mean pause length (sec)




## Speech rate
(relative to "Sherlock" speech rate)

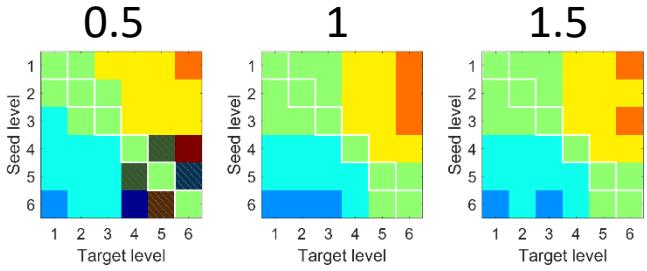



Figure S13. Parameter space within which the lag gradient was found to be robust (the same parameters as in Table 1 unless otherwise indicated) ($p < .05$, FDR correction).