NeuroImage 329 (2026) 121809

Contents lists available at ScienceDirect

NeuroImage

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

Time points or plot points - Are movies processed according to their temporal duration or their underlying content structure?

Falko Mecklenbrauck [a,b,*], Ricarda I. Schubotz [a,b]

[a] Department of Psychology, Biological Psychology, University of Münster, Germany
[b] Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Germany

ARTICLE INFO

Keywords:
Naturalistic movie watching
Hierarchically nested stimuli
Temporal receptive windows
Visual processing hierarchy
fMRI
Inter-subject correlation
Scene network

ABSTRACT

Movie-watching studies have shown that specific cortical areas are tuned to stimulus segments of certain durations. However, increases in stimulus duration naturally co-occur with increases in content complexity. This study aimed to disentangle the effects of stimulus content and duration to determine whether hierarchically nested, complex, naturalistic stimuli, like movies, are processed primarily on the basis of their underlying temporal or content structure. To this end, 48 participants watched six equal-length blocks of movie frames presented at a constant frame rate in an fMRI experiment. Frames were extracted from either movie scenes or movie shots (Content Level) and displayed as continuous segments for 4, 12 or 36 s (Duration). We applied inter-subject correlation and three-dimensional linear mixed-effects modeling with crossed random effects to identify cortical areas selectively modulated by Content Level and Duration. Effects along the visual processing hierarchy were additionally assessed in a ROI analysis. Whole-brain results were located predominately within distinct subnetworks of the scene network: Movie scenes, compared to shots, elicited stronger engagement in the visually-attuned posterior subnetwork containing the occipital and posterior parahippocampal place areas. Longer stimuli whereas additionally engaged the memory-related anterior scene network including the anterior parahippocampal place area, retrosplenial cortex, and caudal inferior parietal lobe. ROI analyses confirmed that temporally extended, content-rich stimuli preferentially engaged hierarchically higher areas. Overall, these findings support a functional differentiation within the scene network while expanding on its relation to temporal receptive windows, demonstrating that Content Level and Duration interact in shaping the cortical processing of naturalistic movie stimuli.

1. Introduction

Hierarchically nested stimuli are all around us (Ludwigs, 2020; Wu, 2013): The steps and individual movements of an action (Fitch and Martins, 2014; Grafton and Hamilton, 2007; Hamilton and Grafton, 2006; Uithol et al., 2012; Wakita, 2014), a text, made of paragraphs, sentences and words (Fitch and Martins, 2014; Jeon, 2014; Lerner et al., 2011) or a movie, that contains scenes, shots and individual captured frames (Haq et al., 2019; Yang et al., 2008c).

This study focuses on the structure of movies, as movies represent a compelling method to investigate cortical responses in a highly engaging, essentially naturalistic setting (Finn and Bandettini, 2021; Landesman et al., 2008b). Additionally, movies not only portray a captivating narrative structure (Cutting, 2016), but can also be organized in a hierarchically nested architecture (e.g., Haq et al. 2019). Here we concentrate on three distinct levels of this hierarchical structure: *frames* (low level), *shots* (intermediate level), and *scenes* (high level).

At the lowest level, a frame refers to a single image captured by the camera (e.g., Yeh et al. 2016). When presented rapidly, frames are perceived as continuous motion rather than individual images (Wilcox et al., 2015).

The intermediate level, a movie shot, is a continuous sequence of frames without a cut, forming the smallest narrative unit (Cutting, 2016). Cuts mark abrupt transitions between shots and are often accompanied by a change in camera angle, thus leading to a jarring visual difference between adjacent frames, that, e.g., automatic annotation programs can use to parcellate different shots (Chakraborty et al., 2021; Mas and Fernandez, 2003; Rui et al., 1998; Zhu et al., 2023). For the observer, however, cuts between shots are often designed to go unnoticed (Magliano and Zacks, 2011), while salient transitions

* Corresponding author at: Department of Psychology, Biological Psychology, University of Münster, Germany.
E-mail addresses: f_meck01@uni-muenster.de (F. Mecklenbrauck), rschubotz@uni-muenster.de (R.I. Schubotz).

https://doi.org/10.1016/j.neuroimage.2026.121809
Received 27 August 2025; Received in revised form 4 November 2025; Accepted 12 February 2026
Available online 13 February 2026

typically announce more narratively significant shifts, i.e., between movie scenes.

Movie scenes represent high-level, visually and narratively complex segments, encompassing multiple shots, i.e., camara perspectives, taking place at a certain location and time, or involving a distinct set of characters (Cutting, 2014). A more technical scene definition relies on screenplay slug lines, which specify the setting and time the scene will be filmed in (Field, 2005). Since definitions of higher-order units (e.g., sequences or action changes) are less consistent (Cutting, 2019a; Magliano and Zacks, 2011), we restrict our analysis to hierarchical levels of frames, shots, and scenes (Fig. 1).

Like all hierarchically nested stimuli, a generalizable way to organize these levels is by their temporal duration (Uithol et al., 2012; Valentine and May 1996; Wu, 2013), since hierarchically higher levels (i.e., scenes) persist longer than hierarchically lower levels (i.e., shots; Cutting et al., 2012; Cutting and Candan, 2015). For the brain to be able to process stimuli according to their temporal durations, it was proposed that different cortical areas contain a range of so-called *temporal receptive windows* (TRWs; Yang, et al., 2008c). In analogy to spatial receptive fields, TRWs integrate or separate incoming information by their temporal duration (Golesorkhi et al., 2021b; Wolff et al., 2022). The existence and range of these TRWs in the cortex was shown by previous findings, demonstrating that the different hierarchically nested levels of the stimulus' content consistently engaged additional areas corresponding to their duration (for a review see Hasson et al., 2015). Specifically, activation related to shorter stimuli (~ 4 s) was found in primary cortices reflecting the presented stimulus modality, while longer stimuli (~ 10 - 40 s) were additionally encoded in the precuneus, the temporoparietal junction, the posterior superior temporal sulcus and areas in the medial frontal cortex (Aberbach-Goodman and Mukamel, 2023; Farbood et al., 2015; Hasson et al., 2008c; Lerner et al., 2011). These results point to an increase in TRWs along the unimodal-to-multimodal gradient (Hasson et al., 2015; Wolff et al., 2022) which also aligns with the distribution of intrinsic cortical time scales observed at rest (Golesorkhi et al., 2021a; Ito et al., 2020; Mecklenbrauck et al., 2024b). Additionally, the retention and recall of longer, temporally coherent stimuli and event structures have been linked to the hippocampus and parahippocampus (Baldassano et al., 2017; Ben-Yakov and Henson, 2018; Chen et al., 2016b; Hasson et al., 2008a; Zadbood et al., 2017), which are functionally connected to areas of longer TRWs (Chen et al., 2017b; Rugg and Vilberg, 2013; E. Yang et al., 2023; Yeshurun et al., 2021).

Overall, the presented results highlight the importance of the stimulus' duration and persistence for its neuronal processing across different cortical areas and networks. However, these previous studies also underscore the inherent link between the temporal structure of a stimulus and its hierarchical organization (Uithol et al., 2012; Valentine and May, 1996; Wu, 2013). In naturalistic stimuli, like the ones applied in the context of TRWs (e.g., Lerner et al. 2011), the duration of a stimulus and the complexity of its content are implicitly confounded. Staying on the topic of movies, a movie scene presents the viewers with a multitude of shots and therefore visually distinct camera angles of possibly different characters or action (Cutting et al., 2012). Showing only one of the shots contained within that scene, however, will not only be temporally shorter but also less visually and narratively diverse.

In the context of this study, we therefore define two distinct factors - the Content Level and the Duration. The Content Level characterizes whether a stimulus covers a higher or lower level within the hierarchical content structure, i.e., in the movie example, whether viewers perceive a single scene or shot of that movie. The differences between the two levels are manifold: Movie scenes are inherently more visually diverse than shots as because they encompass multiple camera angles (Cutting et al., 2012; T.J. Smith et al., 2012). Also, movie scenes define specific narrative events within the story and transitions between scenes mark identifiable narrative shifts (Cutting et al., 2012; Magliano and Zacks, 2011). A movie scene is thus not only visually but also more narratively complex, integrating multiple, less complex subunits linked by a stable location or timeframe (Cutting, 2019b; Cutting et al., 2012; Cutting and Iricinschi, 2015; Hammoud and Kouam, 2000; Markhabayeva and Tseng, 2024). Hence, we define the difference between the two Content Levels as reflecting increases in both visual and narrative complexity, collectively termed *content complexity*.

The Duration in turn describes the time a stimulus remains within one narrative unit, so how long a continuous section of the movie is presented for before switching to a different segment. In the context of TRW studies, these stimulus durations have ranged from very brief segments lasting less than one second (e.g., Honey et al. 2012) to longer events of up to about one minute (e.g., Farbood et al. 2015).

So far, no studies have tried to specifically disentangle potentially separate contributions of the Content Level and Duration. The general procedure of TRW studies is to parcellate a natural stimulus according to its content structure which is naturally confounded with its duration (e.g., Uithol et al. 2012) and then to shuffle the segments before the presentation, so the different conditions vary in temporal as well as content persistency (e.g., Lerner et al. 2011). Areas significantly engaged by only higher levels of stimulus persistency are interpreted to have longer TRWs, although through the time-content-confound these regions could also be excited by higher-levels of content. Thus, it remains an open question whether the neural processing of a stimulus is (1) based only on its Duration, (2) based only on the underlying Content Level, or (3) influenced by both factors. In this study we focused specifically on movie stimuli to address this question within the introduced hierarchically nested framework of scenes, shots and frames.

Earlier movie studies applying the TRW framework predominately focused on the Duration, as they did not divide the stimulus according to narrative structure, but only according to visual or acoustic onsets, creating ever finer segments (Hasson et al., 2008; Honey et al., 2012). Intriguingly, they still found specific brain areas encoding these different segments, corroborating a duration-based processing. A more recent investigation, however, demonstrated two separate systems integrating stimuli over time and a certain amount of information, respectively (Baumgarten et al., 2021), outlining the importance of both the presented stimulus content as well as their duration. These two systems were further suggested to be adaptive, adjusting to the frequency of stimulus presentation (Baumgarten et al., 2021; Rubin and Turano, 1992; Vagharchakian et al., 2012), which follows the theory that neural processing is accommodating to the characteristics of the stimuli and environment (VanRullen and Koch, 2003). This adaptability was also described within the TRW framework, as the TRWs of areas were shown to be modulated by the speed of stimulus presentation (Lerner et al., 2014) as well as by the attention to the stimulus (Shahdloo et al., 2019). Crucially in the context of movies, Geerligs et al. (2022) suggested that TRWs also adjust to the narrative event structure of a stimulus. This proposed interaction between stimulus duration and hierarchical content structure was further supported by findings that a movie's structure is detected and represented across multiple timescales (Baldassano et al., 2017; Oetringer et al., 2025). Interestingly, these studies demonstrated that the scale of content representation, similarly to TRWs (Hasson et al., 2015; Wolff et al., 2022) also follows the cortical hierarchy, as engagement by high-level content elements was observed higher along the visual processing hierarchy (Baldassano et al., 2017; Geerligs et al., 2022; Oetringer et al., 2025). Appropriately, high-level representations were also shown to be modality-independent and shared amongst subjects (Chen et al., 2017b; Zadbood et al., 2017). Overall, the evidence points to both the Content Level and the Duration influencing the processing of naturalistic stimuli across the cortex.

In this study we therefore systematically extracted different number of frames from both shots and scenes. By maintaining a constant frame rate, we could generate three Duration conditions - 4, 12 and 36 s - for the two structural Content Levels of interest. Since this frame-level manipulation strips the movie stimulus of its auditory information, these durations were specifically chosen to mirror the experimental

setup of Hasson et al. (2008c), who comparatively presented silent film stimuli. Thus, replicating the whole-brain engagement presented by Hasson et al. (2008c) would reinforce the validity of the applied stimulus manipulation. Accordingly, we also used the same analysis method - *inter-subject correlation* (ISC), which allows the localization of areas consistently engaged by the manipulated stimuli of different levels of Duration and Content Level (Hasson et al., 2004; Kauppi et al., 2010; Nastase et al., 2019; Pajula et al., 2012). Importantly, the ISC approach enables the flexible analysis of naturalistic stimuli without the need for a specific stimulus model (Nastase et al., 2019; Pajula et al., 2012), which is able to identify processes that remain hidden to classic massed univariate models (Ben-Yakov et al., 2012). It is therefore perfectly suited to answer the question, whether natural, hierarchically nested stimuli like movies are processed by the brain according to their temporal duration, following the content complexity , or based on the combination or interaction of the two. Further, as both TRWs (e.g., Wolff et al. 2022) and processing of content structure (e.g., Baldassano et al. 2017) were shown to extend along the cortical hierarchy, we further examined the effects of Content Level and Duration in established regions of interest representing the visual processing hierarchy (Baldassano et al., 2017). Given the presented theoretical backdrop (e.g., Baldassano et al. 2017; Baumgarten et al., 2021; Geerligs et al., 2022; Hasson et al., 2008c), we expected to find both Duration and Content Level to play a role in processing hierarchically nested, naturalistic stimuli like movies.

2. Materials & methods

2.1. Participants

In total, 56 right-handed participants (48 Female, $M = 22.52$, $SD = 2.44$ years; range = 19 - 29 years) were recruited for this study. Six participants were excluded because of excessive (> 2.5 mm (one voxel size)) movement during the experiment. An additional subject refrained from participation before entering the MR scanner and another missed the scanning session. Finally, data of 48 (40 female, $M = 22.63$, $SD = 2.45$ years; range = 19 - 29 years) subjects entered the analysis. All subjects gave informed written consent approved by the University of Münster Research Ethics Committee in accordance with the declaration of Helsinki. The participants ensured to have no history of neurological or psychiatric diseases or any ferromagnetic material inside their bodies. The Edinburgh Handedness Inventory (Oldfield, 1971) was used to assess right-handedness of the subjects ($M = 85.88$, $SD = 14.30$, range = 40 - 100). All participants had (corrected-to-) normal vision. Upon finishing or aborting the study the participants received course credit or money for their participation.

2.2. Stimuli

As stimuli, the participants saw frames that were extracted from movie shots and movie scenes of the Hollywood movie “Forrest Gump” (Zemeckis, 1994). Previous work as part of the StudyForrest project (https://www.studyforrest.org/) meticulously annotated the movie down to the level of perceptual confounds of individual frames (Häusler et al., 2021). We used these publicly available annotations (Häusler and Hanke, 2016) to parcellate the movie into its scenes and shots. This movie and these annotations in particular already served as the basis in multiple other studies investigating the neural processing of naturalistic movie stimuli (e.g., Ben-Yakov and Henson 2018; Oetringer et al., 2025). In total, “Forrest Gump” is comprised of 205 scenes containing 848 shots, with an average of 4.12 ($SD = 7.04$, range = 1 - 68) shots per scene. Using the shot-wise rating of emotional arousal of 36 labelled characters, we selected only shots and scenes that contained people with a rated emotional arousal (Hanke et al., 2015). We could thus also determine the number of characters and average emotional arousal within each scene and shot.

In contrast to previous research on TRWs, we did not present the participants with the same stimuli across the conditions (e.g., Hasson et al., 2008c) but each condition contained non-repetitive content of the movie to mitigate neural attenuation throughout the experiment (Hasson et al., 2006; Larsson and Smith, 2012). For the Scene conditions we thus selected scenes that depicted at least one character and had at least two shots. Additionally, we excluded scenes that were exceptionally long (> Q3 + 1.5 IQR; > 158.4 s) or shorter than 4 s. This selection left us with 80 possible scenes. For the Shot conditions we also excluded shots without people in them and importantly any shot already included in the selected movie scenes, thus all conditions were created from unique sections of the movie (Fig. 1). This resulted in mainly selecting suitable one-shot scenes (66 scenes only comprised of a single shot) as well as shots from the discarded scenes (109) that were longer than four seconds.

For the experiment, the selected shots and scenes were needed to be balanced and assigned to the following across the six experimental conditions: Scene 4s, with frames extracted from scenes presented for four seconds (12 frames with three frames per second (FPS)), Scene 12s (36 frames), Scene 36s (108 frames), Shot 4s (12 frames), Shot 12s (36 frames), and Shot 36s (108 frames; Fig. 1). As the conditions were presented in blocks of equal length (648 frames), we needed to distribute the scenes and shots differently amongst the conditions. Fifty-four shots and scenes were needed for 4s conditions, 18 for 12s conditions and six for the 36s conditions, so 78 shots and scenes, respectively, were needed in total. For the final selection, we minimized the Euclidian distance of the number of characters, average arousal and average duration in seconds to match 78 scenes to 78 shots. While this matched movie scenes and shots as closely as possible, natural differences in distributions of arousal and duration between scenes and shots made it impossible to perfectly match the material (Table 1).

Subsequently, we assigned the matched shots and scenes to the different conditions while controlling for the average emotional arousal, number of characters, duration and, specifically for scenes, the number of shots. Also, we balanced the number of indoor (INT) or outdoor (EXT) as well as DAY or NIGHT settings amongst the blocks. In this final assignment of conditions, the selected scenes and shots were distributed equally across the movie’s runtime, as assessed by comparing the onset timestamp of each segment between conditions, $F_{\text{Content Level}}(1, 150) = 1.20$, $p = .275$, $\eta_p^2 = 0.01$; $F_{\text{Duration}}(2, 150) = 2.80$, $p = .064$, $\eta_p^2 = 0.04$; $F_{\text{Content Level x Duration}}(2, 150) = 0.98$, $p = .377$, $\eta_p^2 = 0.01$.

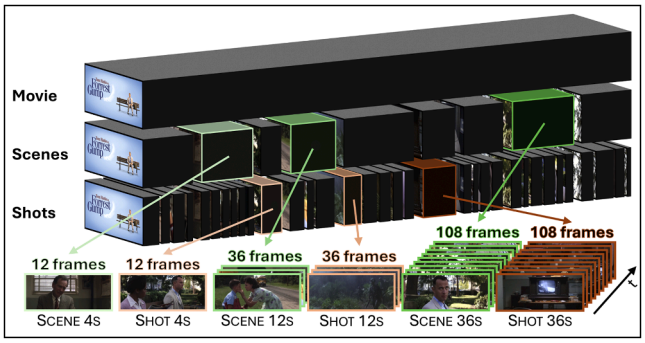

Fig. 1. Hierarchical movie structure and overview over isochronic frame extraction of all conditions.
Note. Movie was annotated into movie scenes and further into movie shots according to Häusler and Hanke (2016). We selected specific movie scenes that had at least two shots and movie shots that either were one-shot scenes or shots from longer scenes not selected for the movie scenes before. Selected movie scenes and shots were assorted into six conditions, balancing the segment’s duration, emotional arousal, number of characters time and location between conditions (Table 1). Depending on the condition 12, 36 or 108 frames were extracted from each movie Scene or Shot, which were presented at a constant frame rate of 3 FPS to manipulate the Duration (4s, 12s, 36s) of each movie segment in the experiment independently from its Content Level.

Table 1
Summary of emotional arousal, video duration, characters per movie segment and scene features of the selected shots and scenes across the levels of Duration.

Average emotional arousal of depicted characters			
Duration		Content Level - Shot	Content Level - Scene
	4s	−0.35 (0.41)	−0.22 (0.44)
	12s	−0.39 (0.30)	−0.21 (0.33)
	36s	−0.52 (0.25)	−0.21 (0.36)
Video duration (in seconds)			
Duration		Content Level - Shot	Content Level - Scene
	4s	19.94 (17.93)	33.26 (21.92)
	12s	20.81 (9.97)	43.04 (22.96)
	36s	17.53 (23.91)	37.59 (20.43)
Characters per movie segment			
Duration		Content Level - Shot	Content Level - Scene
	4s	1.74 (0.73)	1.87 (0.75)
	12s	1.89 (0.76)	2.17 (0.92)
	36s	1.83 (0.98)	2.17 (0.98)
Scene feature - Time: DAY vs. NIGHT			
Duration		Content Level - Shot	Content Level - Scene
	4s	DAY = 33, NIGHT = 21	DAY = 48, NIGHT = 6
	12s	DAY = 12, NIGHT = 6	DAY = 16, NIGHT = 2
	36s	DAY = 4, NIGHT = 2	DAY = 5, NIGHT = 1
Scene feature - Location: EXT vs. INT			
Duration		Content Level - Shot	Content Level - Scene
	4s	EXT = 34, INT = 20	EXT = 35, INT = 19
	12s	EXT = 10, INT = 8	EXT = 11, INT =7
	36s	EXT = 4, INT = 2	EXT = 4, INT = 2

Note. Values were calculated based on publicly available annotations of "Forrest Gump" (Häusler and Hanke, 2016). EXT = *exterior scene*, INT = *interior scene*.

As scenes are generally longer than shots (Cutting and Candan, 2015), compressing them to the same number of frames will shorten scenes disproportionally more than shots, creating differences in the *duration ratio*, i.e., the perceived tempo. We mitigated this bias by repeating the balancing of scenes and shots between conditions 100 times and selecting the distribution that minimized the difference in this duration-shortening ratio. Thus, the final mean ratios for each condition were as follows: $\overline{\text{Ratio}}_{\text{Scene 4s}} = 8.32$, $\overline{\text{Ratio}}_{\text{Scene 12s}} = 3.59$, $\overline{\text{Ratio}}_{\text{Scene 36s}} = 1.04$, $\overline{\text{Ratio}}_{\text{Shot 4s}} = 4.98$, $\overline{\text{Ratio}}_{\text{Shot 12s}} = 1.73$, $\overline{\text{Ratio}}_{\text{Shot 36s}} = 0.49$. The remaining differences will be discussed later.

As a last step, we pseudo-randomly shuffled the movie segments within each condition. The resulting order thereby balanced the chronological leaps forward and backward through the movie, while not allowing for consecutive sections of the movie or repetitions of the same depicted locations.

Using the timestamps, we cut the movie into short clips of each movie scene and shot with ffmpeg (4.2.7; Tomar, 2006). We then applied an isochronic frame extraction approach to select the necessary number of frames from each clip at regular intervals while keeping the relative temporal structure of the movie segment intact. While this frame extraction allowed us to control the temporal duration of each segment, it unfortunately prevented us from presenting the stimuli with their audio. Therefore, the stimuli that the participants saw consisted of a sequence of images depicting the respective movie scenes or shots for 12, 36, or 108 frames at a time, presented without the associated audio track and thus comparable to the silent film stimuli in Hasson et al. (2008c)

After extracting all frames for the experiment, we compared the conditions with regard to the visual features of the frames. The analysis revealed that the conditions did not significantly vary regarding static, low-level visual features, including luminance, color contrast, saturation, and overall image complexity. However, measures that quantified the variations of one frame to the next, like motion energy and normalized perceptual difference, did vary significantly between Content Levels and Durations. Although, this was to be expected, based on the visual and temporal gaps introduced by the frame-wise stimulus manipulation, we will take these differences into account when interpreting and discussing the results (for all details, see Supplemental Material 1).

In preparation of the experiment, we created 48 different orders of the six blocked conditions, counterbalancing across participants that each condition appeared at every block position in the experiment as well as each condition transitioning to each other condition equally often. Consequently, every participant viewed a different order of conditions.

As task during the fMRI measurement, participants were instructed to attentively watch the presented frames. After each block, they were then presented with individual frames of the previous block, that could be presented either correctly or mirrored, flipping left to right. The task was to evaluate if the frames stayed the same (answering "YES") or were mirrored (answering "NO"). For this, we chose six movie segments per block and selected one isochronically extracted frame that resembled the key frame (Katna v0.9.2; Jain et al., 2021) of the respective scene or shot. We made sure that all selected task frames portrayed a person but contained no text. Further, we controlled the overall number of mirrored frames in the experiment and counterbalanced which frames appeared mirrored across participants. For the purpose of this paper, we consider this task as a cover-task, since it just ensured the attention of the participants to the presented frames.

Stimulus selection and balancing was accomplished using RStudio (2024.9.1.394; Posit team, 2024) and MATLAB (9.10.0.1602886 (R2021a); The MathWorks Inc., 2021). Stimuli during the fMRI session were presented using Neurobs Presentation® software (20.3, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Frames were presented at a visual angle of 3.09° on a dark grey background, RGB = (20, 20, 20).

2.3. Procedure

The experiment was divided into two sessions: a group session where multiple participants watched the movie and an individual experimental fMRI session two days later.

For the movie session we invited all participants (three to five at a time) that would partake in the fMRI session two days later to watch the movie. This session was included to familiarize the participants with the material. As some participants had never seen "Forrest Gump" (Zemeckis, 1994) prior to the study while others already had watched it up to five times (62.5 % knew the movie, $M_{\text{Times seen}} = 1.83$, $SD_{\text{Times seen}} = 1.03$), we aligned the different levels of experience by showing each participant the movie in advance. The session started with a short questionnaire assessing the subjects' current mood before the movie using the Self-Assessment Manikin (SAM; Bradley and Lang, 1994; Knejzlíková et al., 2021) and the Positive and Negative Affect Schedule (PANAS; Janke and Glöckner-Rist, 2014; D. Watson et al., 1988), which are used to evaluate the affect of individuals (e.g., Thorenz et al. 2023). Then the participants watched the movie as if they were at the cinema. After the movie the subjects repeated the SAM and PANAS and answered additional questions on their experience with "Forrest Gump", their opinion on the movie as well as if they rated certain scenes especially crucial for the plot, particularly memorable and whether specific scenes where personally relevant to them.

Two days later, the participants returned for individual fMRI sessions. These sessions started with capturing a high-resolution T1-weighted anatomical reference image. After that, we collected eight minutes of open-eyed resting-state fMRI from the participants. The subjects had to focus on a fixation cross for the duration, which is recommended for controlled resting-state measurements (Agcaoglu et al., 2019; Patriat et al., 2013; Van Dijk et al., 2010). The following main experiment then consisted of six blocks, presenting each combination of Content Levels and Durations once. The order of blocks was counterbalanced across participants. Each block started with a fixation cross (2100 - 2600 ms), that blinked three times in rapid succession signaling the start of the frame presentation. The blocks contained 648 frames, that were presented with a frame rate of 3 FPS (Fig. 2a). After this

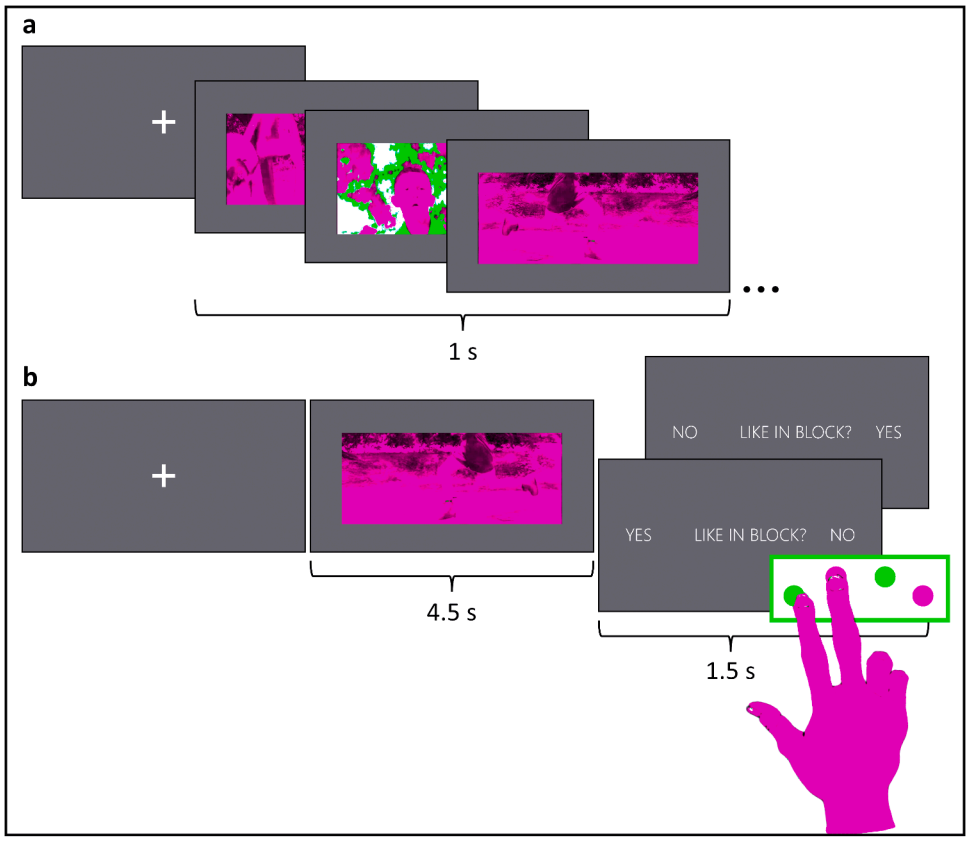

Fig. 2. Overview of experimental procedure in fMRI session.
Note. (a) Example of frame presentation in a block. Frame rate was set at 3 frames per second. (b) Example of a cover-task trial at the end of each block. Participants saw a frame for 4.5 s and had to determine if it was depicted like in the block before ("YES") or mirrored ("NO"). The assignment of the answers to the response buttons could change randomly between trials.

sequence of frames, the participants were noticed about the upcoming task, which had six trials. Each trial stated with a fixation cross (800 - 1300 ms), followed by 4.5 s of frame presentation. This frame could be presented correctly or mirrored. The participants then saw the question "LIKE IN BLOCK?", which they had to answer with "YES" if the frame looked correct or "NO" if it was mirrored. They had 1.5 s to give an answer, skipping to the next trial after they pressed a button. The assignment of the two answer buttons to "YES" and "NO" could randomly switch between trials. The current assignment for each trial was always presented with the question (Fig. 2b). The random switches were implemented to prevent any preparatory activation in the motor areas during the presentation of the frame (Schubotz and von Cramon, 2002). After each block the participants had a break of approximately one minute while the frames for the next block were pre-loading. During this break the experimenters would check in with the subjects, asking them if they were ready to continue. At the end of the experiment, the participants received feedback of how many of the 36 total cover-task trials they answered correctly. The scanning protocol finished with the acquisition of five phase-inverted functional scans that were later used for field map correction. In total the participants spent about 45 min in the MR scanner. After the scanning session the participants filled out a final questionnaire, asking for scenes that were crucial for the plot, especially memorable, or personally important as well as assessing their MRI experience.

2.4. MRI data acquisition

MRI images were recorded using a 3 T Siemens Magnetom scanner (Siemens, Erlangen) equipped with a 20-channel head coil. High-resolution structural reference images were acquired using a standard Siemens 3D T1-weighted magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence (voxel size = 1 × 1 × 1 mm (isotropic); field of view (FOV) = 256 mm; 256×256 data acquisition matrix; 192 slices; 8° flip angle; time of repetition (TR) = 2130 ms; echo time (TE) = 2.28 ms). Resting-state as well as task-related functional blood-oxygen-level-dependent (BOLD) images were acquired parallel to the anterior commissure/posterior commissure line. A T2*-weighted simultaneous multi-slice echo-planar imaging (SMS-EPI; 84×84 matrix; FOV = 210 mm; 71° flip angle; TR = 1500 ms; TE = 30 ms) was applied. Each acquired volumed was made up of 63 axial slices with a slice thickness of 2.5 mm, a distance factor of 5 % (0.125 mm) and a voxel size of 2.5 × 2.5 × 2.4 mm^3. Slice acquisition and multi-slice mode were interleaved with an acceleration factor of 3. The functional data was acquired in one 8 min resting state run and six task-blocks sequences were each 5:10 min long. Additionally, we acquired five supplementary volumes at the end using the same acquisition parameters but flipping the phase encoding direction. The volumes were later used to estimate and correct field inhomogeneities using FSL's topup (Andersson et al., 2003; S.M. Smith et al., 2004).

2.5. fMRI preprocessing and nuisance regression

In preparation of the ISC analysis, the preprocessing of the fMRI task-data was in large parts based on the recommendations in Nastase et al. (2019) for data processing in the Statistical Parametric Mapping software (SPM12; FIL Methods Group, 2017). Additionally, after manual reorientation to MNI space, we applied topup in FSL (6.0.5.1; Andersson et al., 2003; S.M. Smith et al., 2004) to correct for magnetic field inhomogeneities. We then returned to SPM12 for the slice time correction of the data to the temporally first slice. Functional images were then

realigned to the mean image, co-registered with anatomical data and normalized to MNI standard space, before being smoothed using a 6 × 6 × 6 mm³ Gaussian smoothing kernel. This smoothing kernel size was chosen based on recommendations for temporal ISC (Nastase et al., 2019). Further, as we planned to correct for multiple comparisons by controlling the familywise error rate (FWE), a smoothing kernel of at least twice the voxel size was necessary to fulfill the smoothness assumptions of Gaussian random field theory underlying FWE (Poldrack et al., 2011). After smoothing, the signal was the high pass filtered at 0.01 Hz in accordance with Nastase et al.'s (2019) ISC tutorial. To filter any unwanted signal out of the data, we then applied a nuisance regression. As an extension to the guidelines, the applied model included the six motion six motion parameters from the realignment as well as the top five components of both the white matter and cerebrospinal fluid signal (Behzadi et al., 2007). This nuisance regression model was shown to offer the best balance between maximizing ISCs and controlling for adverse influences of movement (Nastase and Hasson, 2024).

2.6. Inter-subject correlation

In general, ISC can be used to determine the reliability of neural activity across a group of subjects who viewed a temporally synchronized stimulus. If the ISC is particularly high in a specific area, it can be assumed that this area encodes information about the presented stimulus (Nastase et al., 2019). In the context of the TRW framework, ISC can identify areas that are specifically engaged by different levels of the hierarchically nested stimulus, as shuffling the stimulus according to its structure will only engage areas that are tuned to specific, maintained aspects of the stimulus (e.g., Lerner et al. 2011).

The following ISC analysis was conducted on the residuals resulting from the completed nuisance regression. These residuals were further *z*-scaled per condition and voxel, and masked using a gray matter mask, averaged across all normalized participants. This gray matter mask included a total of 99,709 voxels. As recommended (Nastase et al., 2019), we excluded the first five images (7.5 s) of each block, which left a time course of 147 images representing 220.5 s of stimulus presentation per condition. ISC was then used to correlate the voxel-wise time courses of all 99,709 voxels across participants (Fig. 3).

We specifically calculated the correlations using the pairwise-ISC approach which correlates the voxel-wise time course of one subject with the time course in the same voxel of every other subject (Hasson et al., 2004) using Pearson correlation:

$$r_{AB} = r(x_A,\ x_B) = \frac{\sum_{t=1}^{T}\left(x_A(t) - \bar{x}_A\right)\left(x_B(t) - \bar{x}_B\right)}{\sqrt{\sum_{t=1}^{T}\left(x_A(t) - \bar{x}_A\right)^2}\sqrt{\sum_{t=1}^{T}\left(x_B(t) - \bar{x}_B\right)^2}} \quad (1)$$

with $x_A(t)$ and $x_B(t)$ being the time course of the same single voxel of subjects A and B, respectively. Applying Eq. (1) to all voxel-wise time courses resulted in $\frac{N*(N-1)}{2} = \frac{48*47}{2} = 1128$ pairwise-ISC whole-brain maps per condition. Each of the 99,709 gray matter voxels in every whole-brain map therefore contained the correlation between time course evoked by a specific condition for a subject pair in this particular voxel. These maps were the basis for the following whole-brain analysis (Fig. 3). The goal of the analysis was to identify clusters of voxels in

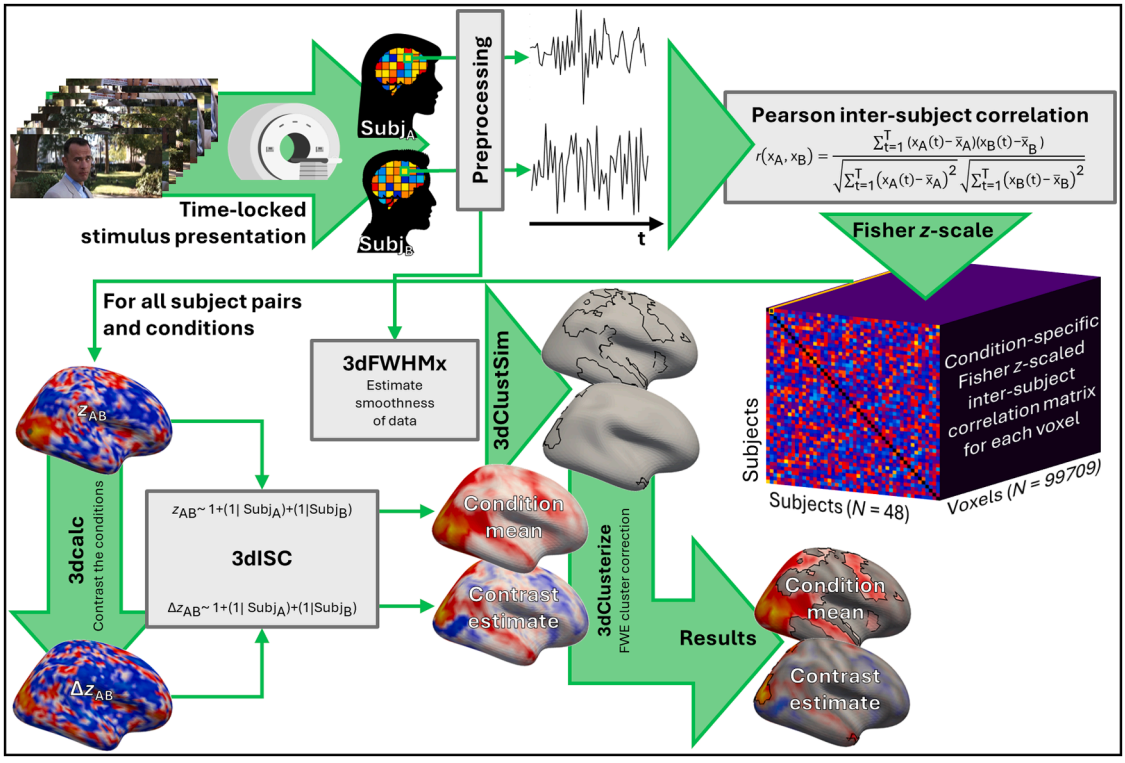

Fig. 3. Analysis pipeline of whole-brain inter-subject correlation (ISC) analysis.
Note. Frames were presented in a time-locked manner for each condition. Cortical time courses were therefore synchronized between participants. Preprocessing included slice time correction, motion correction, normalization to Montreal Neurological Institute (MNI) standard space, smoothing with a 6 × 6 × 6 mm³ Gaussian kernel, as well as nuisance regression of the data. The preprocessed gray matter voxel time courses were then Pearson inter-subject correlated between each pair of subjects and Fisher *z*-scaled for normal distribution (Bond and Richardson, 2004). Condition-specific z-scaled voxel-wise inter-subject correlations (ISCs) then either directly modeled for the condition-wise means or further contrasted between conditions for estimating the effects of Content Level and Duration. Modeling was performed using AFNI's 3dISC (Chen et al., 2016a) specifically modeling the convoluted covariance structure of ISC data via crossed random effects for each subject pair. Data smoothness was estimated (3dFWHMx) and cluster size was determined (3dClustSim) using a $p < .001$ initial cluster forming threshold and representing an FWE-corrected $\alpha = 5$ %. Thresholds were applied to the modeling results through 3dClusterize to produce the final results for the condition-wise means and estimated effects of Content Level and Duration.

which the conditions differed significantly regarding the evoked ISC correlation.

Statistically ISC data have to be handled carefully. Through the correlation of data between participants, the resulting values violate the independence assumption, preventing the use of conventional analysis methods (Chen et al., 2016a; Kauppi et al., 2014; Nastase et al., 2019). A recent set of papers established advanced methods to specifically handle the complex covariance structure amongst pairwise-ISC data in a linear mixed-effect modeling framework by using crossed random effects for both subjects involved in the pairwise correlations (Chen et al., 2016a, 2017a, 2020). Thus, 3dISC (Chen et al., 2017a) in AFNI (25.1.03; Cox, 1996) allowed us to specifically consider the dependencies between participants when modeling and comparing the ISC values in each single voxel to identify cortical areas specifically engaged by different levels of Content Level and Duration in the stimuli.

2.6.1. Mean effect calculation

Before contrasting the conditions to identify the effects of Content Level and Duration, we first calculated the condition-wise average ISC maps. With these ISC maps we wanted to validate the applied manipulation by reproducing the effects detailed for naturalistic movie stimuli before (Hasson et al., 2008c) using just our frame-wise manipulation. Therefore, we Fisher z-transformed every gray matter voxel the individual ISC maps to normally distribute the correlation values (Bond and Richardson, 2004). Then we applied a simple, intercept-only model within the 3dISC framework (Chen et al., 2017a) that estimated z_{AB}, i.e., the Fisher z-transformed correlation in a specific voxel between two subjects, by the grand average of this voxel for the condition and the crossed random effects for the given subject pair (Fig. 3):

$$z_{AB} \sim 1 + (1|\ \mathrm{Subj}_A) + (1|\mathrm{Subj}_B) \quad (2)$$

This model therefore determined the average ISC for each condition in every voxel, which were projected to ICBM152 surface space using BrainNet Viewer (1.7; Xia et al., 2013; http://www.nitrc.org/projects/bnv/) and are depicted in Fig. 4.

2.6.2. Contrast calculation

To then estimate the differences between conditions, we calculated contrasts between the pairs of ISC whole-brain maps. Specifically, we considered six two-sided comparisons to cover all main effects and interactions of interest: Shot vs. Scene for the main effect of Content Level, 4s vs. 12s, 4s vs. 36s, and 12s vs. 36s for the effects of Duration, and Scene × 12s = (Scene 4s > Shot 4s) vs. (Scene 12s > Shot 12s) and Scene × 36s = (Scene 4s > Shot 4s) vs. (Scene 36s > Shot 36s) for the specific effects of the interactions for longer scenes. Additional contrasts that were not part of the main analysis are reported in Supplemental Material 6. All contrasts were determined by applying AFNI's 3dcalc to calculate weighted averages as well as differences between the Fisher's z transformed ISC maps (Fig. 3). We thus ended up with 1128 whole-brain maps of all pairs for each contrast. Using these contrast maps, we could apply a similar model to Eq. (2) just with the pairwise difference Δz_{AB} and estimated the effects of each contrast in each gray matter voxel across all subject pairs (Fig. 3):

$$\Delta z_{AB} \sim 1 + (1|\ \mathrm{Subj}_A) + (1|\mathrm{Subj}_B) \quad (3)$$

Using Monte Carlo simulations, we determined the cluster threshold that corresponded to an FWE-corrected $\alpha < 0.05$ at a cluster forming threshold of $p < .001$ for both the average ISC maps and the contrast maps. Therefore, we first estimated the smoothness of the data from the underlining residuals used in each model (AFNI's 3dFMWHx), before applying AFNI's ClustSim to determine the cluster thresholds (Fig. 3). The resulting thresholds included an average of 35.08 voxels (range = 34 - 36 voxels). The specific thresholds are noted in the respective result tables (Tables 2–4) and figures (Fig. 4–7). Finally, we applied these thresholds to the results of 3dISC using AFNI's 3dCulterize for bisided tests and clusters connected to their neighbors, face-, edge- and cornerwise ($NN = 3$). The procedure was applied to the whole-brain results for both the condition-wise averages as well as the contrasted maps.

2.6.2.1. Whole-brain figure construction. The resulting thresholded whole-brain contrasts are depicted on the brain's surface, mirroring the result graphics of previous TRW studies for easier comparison (e.g., Hasson et al. 2008c). Additionally, we followed the "Highlight, don't Hide" principle (Taylor et al., 2023), i.e., the clusters passing the significance threshold are plotted at full opacity with a black outline, while the opacity of all results below the threshold is scaled down to represent their respective level of significance. This method of plotting improves overall objectivity, reliability, and reproducibility of the result reports (Taylor et al., 2023).

To achieve this portrayal, we plotted a combination of three whole-brain maps resulting from the 3dISC analysis described above (Fig. 3): the thresholded cluster maps created by 3dClusterize, the unthresholded contrast estimates produced by the 3dISC modeling and the t-values associated with these unthresholded voxel-wise maps. All three images were translated from MNI volume-space into fsaverage surface space using nilearn.surface.vol_to_surf (0.11.1; https://nilearn.github.io/dev/modules/generated/nilearn.surface.vol_to_surf.html) which is common practice for this conversion (e.g., Pang et al. 2023). This function sampled the voxel values within a 3 mm radius around each pial surface vertex. A 3 mm sampling radius thereby corresponded to the $6 \times 6 \times 6$ mm^3 smoothing kernel that was applied in preprocessing; thus, no additional blurring was introduced. Having obtained a value for each vertex, we plotted the 3D surfaces using pyvista.Plotter(0.45.2; https://docs.pyvista.org/api/plotting/_autosummary/pyvista.plotter). In this context, we scaled the absolute t-values to values between 0 and 1 and used them as the opacity parameter for the depiction of the unthresholded contrast estimates. Lastly, the cluster maps were used to overlay the thresholded, significant clusters on top of the unthresholded results, plotted at full opacity and outlined in black, making them more discernible against the surface plots. The full code for figure creation together with all the other code from stimulus selection to analysis as well as a link to data for replication can be found at https://github.com/FalkMeck/Time_Points_or_Plot_Points.

2.6.3. Region of interest analysis

As previous studies found a distinction for both the duration and content of the stimulus throughout the visual processing hierarchy (Baldassano et al., 2017; Hasson et al., 2015; Jääskeläinen et al., 2021; Wolff et al., 2022), we additionally looked at the effects of Content Level and Duration not only at whole-brain level but also in specific regions of interest (ROIs) along the cortical hierarchy. For this purpose, we based our selection of ROIs on the areas established by Baldassano et al. (2017), who chose four bilateral ROIs outlining the visual processing hierarchy: the primary visual cortex (V1), an area in the extrastriate visual cortex, namely hV4 (Wade et al., 2002), the angular gyrus (AG) and the posterior medial cluster from the default mode network (pmDMN; Shirer et al., 2012). As a control ROI Baldassano et al. (2017) included the Heschel's Gyrus which represents the primary auditory cortex (A1). We defined the same ROIs using the same methods described in Baldassano et al. (2017): Thus, hV4 was based on the probabilistic atlas introduced in L. Wang et al. (2015) thresholded at a probability of 25 %. AG was obtained from the Anatomy toolbox (Eickhoff et al., 2005) based on the Jülich atlas (Amunts et al., 2020). The cluster pmDMN was extracted from the maps published in accordance with Shirer et al. (2012) and split into a left and right hemispheric cluster. The control area A1/Heschl's Gyrus was pulled from the probabilistic Harvard-Oxford atlas (Desikan et al., 2006; Frazier et al., 2005; Goldstein et al., 2007; Makris et al., 2006) threshold at 25 % probability. Lastly, since the definition of V1 was not closer described, we retrieved V1 from the Anatomy toolbox (Eickhoff et al., 2005).

From these five ROIs we extracted the average, Fisher *z*-transformed pairwise-ISC for each condition, separately for both hemispheres using MarsBaR (Brett et al., 2002).

To analyze the effects of Content Level, Duration and their interaction in the ROIs, we modelled the data in using the lme4 framework (Bates et al., 2015) in RStudio (Posit team, 2024). As the 3dISC tool is also based on lme4, we were able to also model the dependencies in covariance structure using crossed random effects for the involved subjects mirroring the whole-brain model. Hence, we used the same model in all ten ROIs across both hemispheres:

$$\overline{z_{AB_{ROI}}} = 1 + \text{CONTENT LEVEL} * \text{DURATION} + (1|\text{Subj}_A) + (1|\text{Subj}_B) \quad (4)$$

with the Content Levels Scene and Shot, the Durations 4s, 12s, and 36s and $\overline{z_{AB_{ROI}}}$ being the average Fisher *z*-transformed ISC per ROI. As the dependencies between data points complicate the specification of the degrees of freedom, we employed parametric bootstrapping within lme4

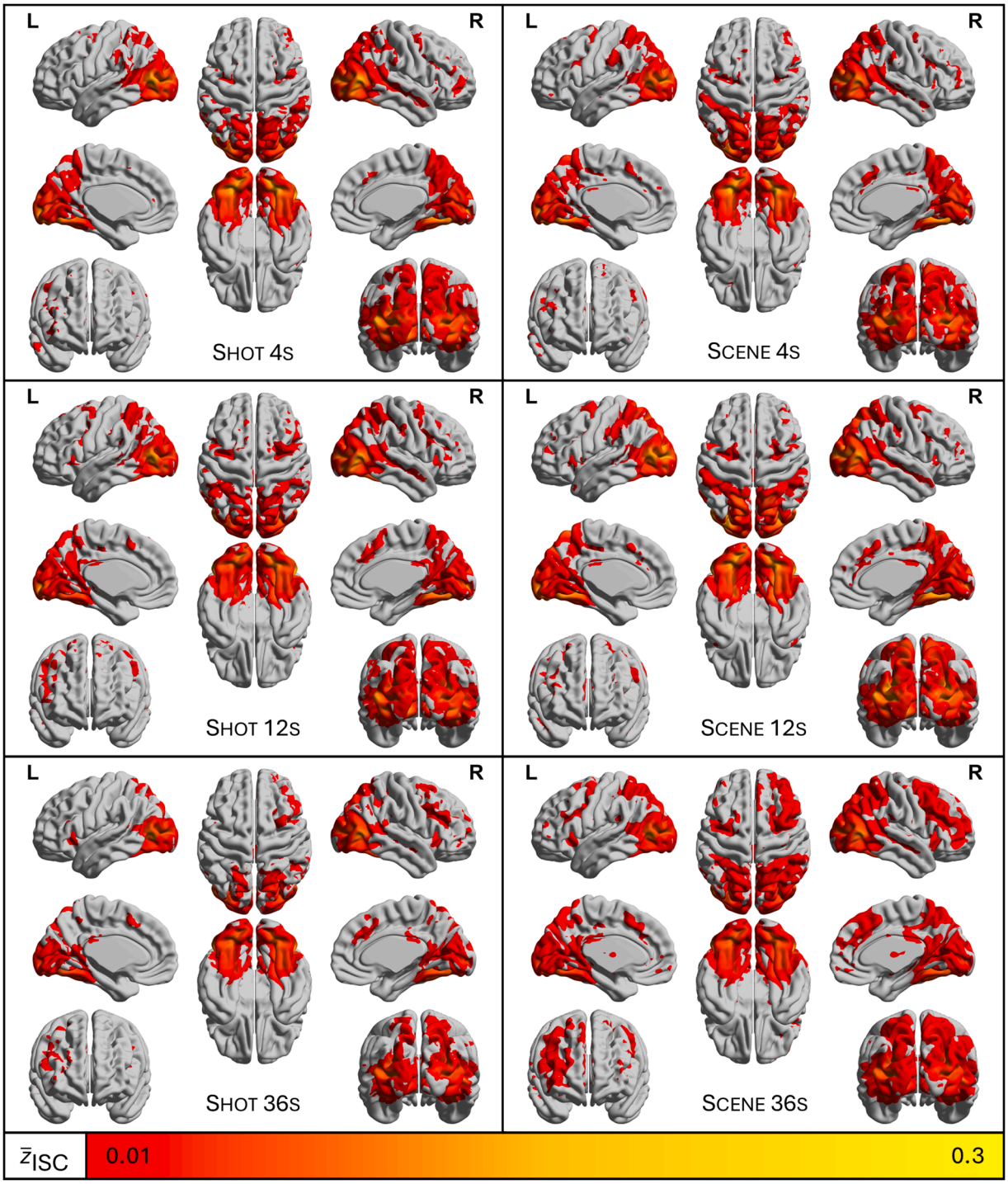

Fig. 4. Average Fisher z-transformed pairwise inter-subject correlation by condition of the Content Level x Duration design.
Note. Maps are FWE-cluster corrected at $\alpha = 0.05$ with a cluster forming threshold of $p < .001$, Monte Carlo simulated cluster extend threshold is are in order top-left to bottom-right 36, 35, 36, 36, 36, and 34 voxels, respectively. Maps are depicted on the ICBM152 template using BrainNet Viewer (1.7; Xia et al., 2013; http://www.nitrc.org/projects/bnv/).

(bootMer, $N_{boot} = 500$) to determine the estimates, confidence intervals and *p*-values for all fixed effects. Lastly, we adjusted the *p*-values for multiple comparisons across all contrasts between conditions and ROIs controlling the false discovery rate (FDR) at $\alpha = 0.05$ (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001). Results are reported with the FDR-corrected *p*-values.

Complementary to the five ROIs related to the visual processing hierarchy, we additionally investigated five more areas that were commonly found in the context of movie watching studies in a supplementary analysis: the parahippocampal gyrus (PHG), the anterior (aHC) and posterior hippocampus (pHC), the temporo-parietal junction (TPJ), and the medial prefrontal cortex (mPFC; for all details see Supplemental Material 4).

3. Results

Before directly comparing the six conditions of the 2 Content Level x 3 Durations design, we looked at the average ISC within each condition across subjects (Fig. 4).

Regardless of the condition, the stimuli consistently engaged the entire occipital lobe, including ventral areas like the fusiform gyrus (FG) and lingual gyrus (LingG). Furthermore, all conditions induced reliable signal in the superior parietal lobe (SPL), engaging areas along the intraparietal sulcus (IPS), as well as precuneal areas. In the temporal lobe, the condition-wise average ISC extended along the right superior temporal sulcus (STS), while only in the Scene 36s condition we found consistent engagement also along the left STS, indicating a difference between the conditions. This difference was also noticeable in the frontal lobe. Although all conditions showed increased ISC values bilaterally along the precentral sulcus as well as the right medial frontal cortex (Brodman area 8), the engagement of additional areas in the lateral portions of the frontal lobe varied between the conditions. For instance, the Scene conditions engaged areas in both the left and right middle frontal gyrus (MFG), while the Shot 36s condition in particular depicted a more lateralized effect, only engaging the right MFG. To specifically test these visually identified differences, we conducted direct comparisons between the conditions to identify the main effects of Content Level and Duration and their interaction.

3.1. Effect of Content Level

Bidirectionally comparing the average ISC between the Scene and Shot conditions, we only found significant differences for the comparison Scene > Shot (Table 2, Fig. 5). Scene stimuli compared to Shots significantly engaged a large cluster in the left middle occipital gyrus (MOG) that overlapped with the occipital place area (OPA; Dilks et al., 2013) and early visual cortices like V3A/B, while also extending to the right precuneus and bilaterally to the SPL. Moreover, we found a cluster in the right MOG which included OPA as well and spread to the right IPS. Additionally, more engagement for Scenes was found bilaterally in the collateral sulcus between the FG and LingG, continuing into the posterior parahippocampal place area (PPA; Steel et al., 2025). Finally, we localized significant clusters in the right anterior middle temporal gyrus (MTG), i.e., the temporal pole, as well as the left supramarginal gyrus (SMG) and cingulate sulcus.

3.2. Effect of Duration

Next, we compared the three levels of Duration against each other while averaging across the two Content Levels. In the bidirectional comparison between the 4s and 12s Duration (Table 3, Fig. 6a) we identified the bilateral collateral sulcus extending into the PHG in the left hemisphere, which overlapped with the anterior PPA (Steel et al., 2025). Further clusters emerged in the bilateral SPL and bilateral superior frontal gyrus (SFG) as well as the right retrosplenial cortex (RSC) and right inferior parietal lobe (IPL), that were significantly more engaged by stimuli of 12s Duration. In the opposite comparison, we found that the left LingG and left FG to be more reliably engaged by 4s Duration stimuli.

When comparing the 36s and 4s Durations (Table 3, Fig. 6b), we located three clusters in the right hemisphere, namely in the RSC, the IPL and the inferior frontal junction, that were significantly more reliably engaged by stimuli of 36s Duration. Areas that were more engaged by 4s Duration stimuli included a large cluster spanning from the left collateral sulcus, representing the posterior PPA, to the left inferior occipital gyrus (IOG), bilateral cuneus and left superior LingG as well as additional clusters in the right LingG, the left SMG and the right STS.

Lastly, contrasting the conditions of 36s and 12s Duration (Table 3, Fig. 6c), we did not identify any clusters that were significantly more engaged for 36s Duration stimuli. But the opposite side of this bidirectional contrast yielded a large cluster covering the left IOG, right anterior LingG and bilateral cuneus. An IOG cluster was also found in the right hemisphere. In addition, clusters emerged in the collateral sulcus in both hemispheres, which comprised the entire PPA in the right and the anterior PPA in the left hemisphere. Finally, the contrast revealed clusters in the right inferior postcentral sulcus, the left RSC, left precuneus, left SMG, and left MOG/MTG overlapping with area MT+ (Dumoulin et al., 2000).

3.3. Interaction between Content Level and Duration

To test for an interaction effect between the two factors, we contrasted the respective differences between Scenes and Shots of 12s (Scene 12s > Shot 12s) and Scenes and Shots of 36s (Scene 36s > Shot 36s) with

Table 2
Significant differences in Fisher *z*-transformed inter-subject correlations between the Scene and Shot conditions averaged across Durations.

Localization	H	Cluster extent (in voxel)	Peak coordinates (MNI space)			*z* value difference
			x	y	z	
Scene > Shot						
Superior partial lobe	R	67	20.5	−74.5	45	0.04
Supramarginal gyrus	L	75	−57	−29.5	40	0.02
Cingulate sulcus	L	69	−12	−27	37.5	0.02
Middle occipital gyrus (OPA)	L	1594	−32	−89.5	12.5	0.09
Superior parietal lobe	L	l.m.	−14.5	−59.5	62.5	0.05
	L	l.m.	−14.5	−74.5	47.5	0.04
Superior parietal lobe	R	l.m.	13	−57	65	0.05
	R	l.m.	25.5	−57	62.5	0.04
Middle occipital gyrus (OPA)	R	588	33	−84.5	10	0.07
Intraparietal sulcus	R	l.m.	30.5	−79.5	20	0.06
Collateral sulcus (posterior PPA)	R	56	30.5	−59.5	−7.5	0.04
Collateral sulcus (posterior PPA)	L	310	−27	−62	−10	0.07
Middle temporal gyrus/ Temporal pole	R	78	55.5	8	−22.5	0.02
Scene < Shot						
No significant clusters						

Note. Results are FWE-cluster corrected at $\alpha = 0.05$ with a cluster forming threshold of $p < .001$, Monte Carlo simulated cluster extend threshold is 36 voxels. *H = Hemisphere; L = left; R = right*; MNI = *Montreal Neurological Institute*; l.m. = *local maximum*; OPA =*occipital place area*; PPA = *parahippocampal lace area.*

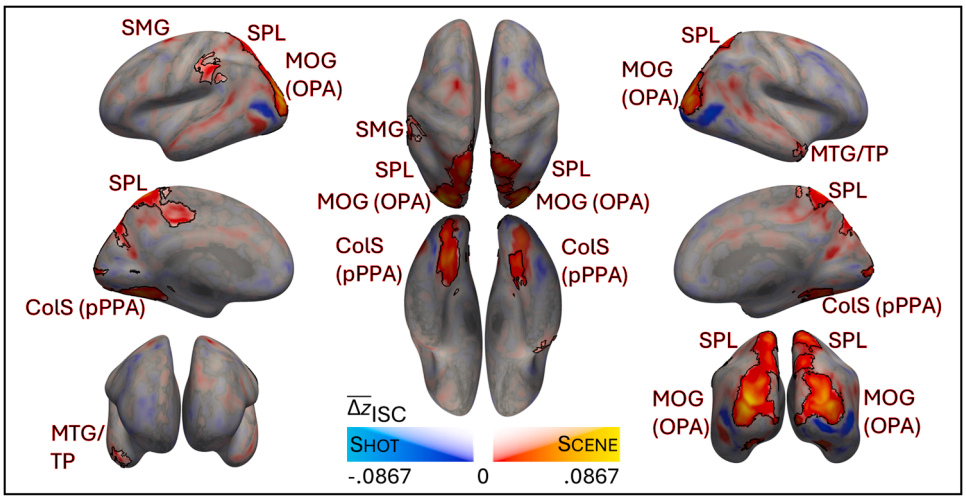

Fig. 5. Whole-brain results of the differences in Fisher z-transformed inter-subject correlations between the SCENE and SHOT conditions averaged across DURATIONS. *Note.* SCENE > SHOT contrast in warm colors, SHOT > SCENE contrast in cold colors. Results are FWE-cluster corrected at $\alpha = 0.05$ with a cluster forming threshold of $p < .001$, Monte Carlo simulated cluster extend threshold was 36 voxels. Significant clusters are depicted with a black border and are labelled accordingly. Red labels describe clusters belonging to the SCENE > SHOT contrast. Results figures applies the "Highlight, don't Hide" principle (Taylor et al., 2023). Thus, clusters with non-significant *p*-values are not omitted but displayed at lower opacity. Whole-brain results are projected on inflated fsaverage pial surface. ColS = *collateral sulcus,* MOG = *middle occipital gyrus*, MTG = *middle temporal gyrus*, OPA = *occipital place area*, pPPA = posterior parahippocampal place area, SMG = supramarginal gyrus, SPL = superior parietal lobe, TP = temporal pole.

the difference of SCENES and SHOTS of 4s DURATION (SCENE 4s > SHOT 4s). With these comparisons we could specifically determine areas that were more reliably engaged by SCENES of a longer DURATION (SCENE × 12s and SCENE × 36s, respectively). The contrast between these two interactions is reported in Figure S6.4. For SCENE × 12s stimuli the engaged areas included a cluster in the left MOG encompassing OPA and extending to the area MT+ and the IPS as well as a cluster in the left collateral sulcus covering PPA that spread laterally into the left temporo-occipital incision. We further located clusters in the left SPL, left RSC and right precuneus as well as the right PHG, which circumscribed the anterior PPA, engaged in the SCENE × 12s condition (Table 4, Fig. 7a). Similarly, for the SCENE × 36s condition we also found a cluster in precuneus but in the left hemisphere as well as clusters in the right LingG and the left IPS, that both expanded into the left and right MOG, respectively, and contained the OPAs. Additionally, we identified the bilateral IOG as well as the bilateral MOG/MTG representing area MT+. Lastly, we found clusters in the bilateral STG and STS stretching anterior to the temporal pole to be specially engaged for this interaction (Table 4, Fig. 7b). The respective opposite contrasts did not yield any significant clusters.

3.4. Region of interest analysis

To investigate the interaction between CONTENT LEVEL and DURATION in more detail, we conducted a ROI analysis. We modelled the average ISC per ROI by their CONTENT LEVEL and DURATION and then used parametric bootstrapping to test the post-hoc comparisons, FDR-corrected for multiple comparisons, and identified significant differences between the conditions (Fig. 6, Table S1.1 in Supplemental Material 1). We analyzed the ISC values from five previously defined bilateral ROIs (see 2.6.3.): A1, V1, hV4, AG, pmDMN describing the visual processing hierarchy (plus A1 as control ROI) according to Baldassano et al. (2017; Fig. 8).

The results were very similar for the left and right ROIs (Table S2.1). Thus, we will summarize the depicted trends for both hemispheres combined. As expected, we did not see many differences between the conditions in the control ROI A1, as just right A1 showed significantly less ISC in the SCENE 36s than the SCENE 12s condition. In V1 we identified a significant decrease in ISC with increasing DURATION of the stimuli while only at DURATION 36s we found a significantly higher ISC for SCENES than for SHOTS. This trend of a decreasing ISC with higher levels of DURATION continued for hV4 although now also featuring a definite distinction between the CONTENT LEVELS: SCENES produced significantly higher ISC than SHOTS for the DURATIONS 12s and 36s, but did not differ at 4s, indicating the interaction between CONTENT LEVELS and DURATIONS that was shown at the whole-brain level already in hV4. Following the trend, in AG we still saw a continuous decrease of the ISC values for the CONTENT LEVEL SHOT, whereas for SCENES only between the 12s and 36s DURATIONS a significant decline emerged. This was complemented by significantly higher ISC for SCENES than SHOTS bilaterally for the 12s DURATION and only in the left AG for the 36s DURATION. Also, for the pmDMN ROI, we saw a significant decrease in ISC between SHOTS of 12s and 36s DURATION, while the ISC for the CONTENT LEVEL SCENE remained constant. Accordingly, ISC was significantly higher in the SCENE 36s than the SHOT 36s condition.

Additionally, we considered five more ROIs, which described areas that were previously related the cortical processing of movies (e.g., Jääskeläinen et al. 2021): TPJ, PHG, aHC, pHC, and mPFC. All results of this ROI analysis are reported in the Supplemental Material 4.

4. Discussion

When we watch rich, naturalistic stimuli like movies, our entire brain is involved in processing everything from low-level visual features to the understanding of the plot (Rajimehr et al., 2024; E. Yang et al., 2023). While some aspects of this processing are idiosyncratic and can be used to differentiate between individuals (Eickhoff et al., 2020; Finn and Bandettini, 2021; Nastase et al., 2019), the majority is shared between subjects when synchronously watching a movie, especially for well-structured and -edited Hollywood movies that guide the viewers' attention (Hasson et al., 2008b). Previous studies used this shared processing to identify cortical areas that are reliably engaged by naturalistic stimuli (e.g., Hasson et al. 2008c), resulting in regions that were additionally engaged when longer, more persistent portions of the movie or other stimuli were presented (Aberbach-Goodman and Mukamel, 2023; Farbood et al., 2015; Hasson et al., 2008c; Lerner et al., 2011). However, because longer durations of naturalistic stimuli coincide with an increase in the presented content, a longer section of a movie likely includes presenting multiple camera shots or even scenes. This confound is not only part of movies, but is a general feature of all naturally occurring hierarchically nested stimuli from language to action sequences (Fitch

Table 3
Significant differences in Fisher *z*-transformed inter-subject correlations between the levels of Duration averaged across the Content Levels.

Localization	H	Cluster extent (in voxel)	Peak coordinates (MNI space)			*z* value difference
			x	y	z	
12s > 4s						
Superior parietal lobe	L	67	−22	−59.5	62.5	0.03
Superior frontal gyrus	L	34	−27	−2	60	0.02
Superior frontal gyrus	R	120	23	3	57.5	0.02
Superior parietal lobe	R	129	18	−64.5	52.5	0.03
Inferior parietal lobe	R	84	33	−77	25	0.05
Retrosplenial cortex	R	326	15.5	−52	7.5	0.05
Collateral sulcus (anterior PPA)	R	179	28	−39.5	−10	0.04
Collateral sulcus/ Parahippocampal gyrus (anterior PPA)	L	137	−27	−44.5	−10	0.05
12s < 4s						
Superior lingual gyrus	L	440	−2	−77	7.5	−0.03
Inferior lingual gyrus	L	l.m.	−7	−72	2.5	−0.02
Fusiform gyrus	L	95	−29.5	−77	−12.5	−0.07
36s > 4s						
Inferior frontal junction	R	48	38	8	30	0.03
Inferior parietal lobe	R	58	35.5	−74.5	27.5	0.04
Retrosplenial cortex	R	67	8	−49.5	7.5	0.02
36s < 4s						
Supramarginal gyrus	L	84	−62	−32	32.5	−0.03
Superior temporal sulcus	L	66	50.5	−37	5	−0.03
Lingual Gyrus	R	347	18	−92	−5	−0.09
Collateral sulcus (posterior PPA)	L	2374	−24.5	−82	−15	−0.09
Cuneus	R	l.m.	0.5	−79.5	37.5	−0.03
	L	l.m.	−2	−77	25	−0.04
	L	l.m.	−2	−92	17.5	−0.03
	R	l.m.	0.5	−74.5	12.5	−0.05
	R	l.m.	0.5	−89.5	10	−0.04
Superior lingual gyrus	L	l.m.	−4.5	−69.5	2.5	−0.04
Inferior occipital gyrus	L	l.m.	−14.5	−92	−5	−0.07
36s > 12s						
No significant clusters						
36s < 12s						
Precuneus	L	35	−7	−44.5	55	−0.02
Supramarginal gyrus	L	73	−59.5	−27	37.5	−0.02
Inferior postcentral sulcus	R	95	60.5	−19.5	37.5	−0.02
Retrosplenial cortex	L	85	−17	−59.5	15	−0.03
Middle occipital gyrus/ Middle temporal gyrus (MT+)	L	54	−49.5	−77	10	−0.05
Inferior occipital gyrus	R	274	18	−92	−2.5	−0.07
Inferior occipital gyrus	L	1197	−14.5	−94.5	−7.5	−0.12
Cuneus	R	l.m.	3	−79.5	32.5	−0.03
	L	l.m.	−4.5	−79.5	25	−0.03
Anterior lingual gyrus	R	l.m.	10.5	−64.5	0	−0.03
Collateral sulcus (anterior PPA)	L	116	−29.5	−47	−7.5	−0.06
Collateral sulcus (anterior and posterior PPA)	R	250	30.5	−44.5	−7.5	−0.07

Note. Results are FWE-cluster corrected at $\alpha = 0.05$ with a cluster forming threshold of $p < .001$, Monte Carlo simulated cluster extend threshold is 34 voxels (12s ≠ 4s), 34 voxels (36s ≠ 4s), and 35 voxels (36s ≠ 12s), respectively. *H = Hemisphere; L = left; R = right*; MNI = *Montreal Neurological Institute*; l.m. = *local maximum*; PPA = *parahippocampal place area*.

and Martins, 2014; Uithol et al., 2012; Valentine and May, 1996; Wu, 2013). Thus, using naturalistic stimuli, it is impossible to answer the question whether the additional areas were engaged simply by a longer stimulus or if the depicted content also played a role in what areas were reliably activated.

To answer this question, the present study was the first to disentangle the dependency between Duration and Content Level of naturalistic stimuli by applying the ISC approach in a controlled experimental setting and identifying areas specifically and separately engaged by different levels of temporal persistency or content complexity. For this, we presented a fixed number of frames extracted from movie Shots or Scenes, thus manipulating the Duration independently of the Content Level of Shots and Scenes. The findings exposed differences in the processing of movie Shots and movie Scenes as well as differences based on the Duration of the content and also areas which were specifically engaged by stimuli of a higher Content Level presented for a longer Duration, i.e., longer movie scenes. The identified areas predominately laid within a visual scene-processing network (Epstein and Baker, 2019). More specifically, while the effects of Content Level were found in the posterior portion of the network, the Duration effects were located in the anterior regions, supporting a functional distinction between the two subnetworks (Baldassano et al., 2016; Steel et al., 2021). Additionally, a ROI analysis was conducted to investigate the differences in ISC along the visual processing hierarchy (Baldassano et al., 2017) and other areas related to the processing of movies (Jääskeläinen et al., 2021; see Supplemental Material 4), highlighting the differences in processing Shots and Scenes at various Duration across the cortex.

4.1. Inter-subject correlation using only frames - a prove of concept

In contrast to the previous studies using the ISC approach with continuous sections of naturalistic stimuli (e.g., Hasson et al. 2008c), we only presented our participants with isochronally extracted frames at a constant frame rate of 3 FPS - lower than the necessary frame rate for motion perception (Wilcox et al., 2015). Thus, our stimulus material differed from naturalistic, well-structured movies that maximize the synchrony of the viewers' reaction (Hasson et al., 2008b; Loschky et al., 2020). Despite the differences, we were able to find significant ISC across our experimental conditions that mirrored the results of Hasson et al. (2008c), who - comparatively to the individual frames without sound - used clips from silent films. Hence, summarized across all six conditions, we also found significant ISC in the entire occipital lobe, including the primary as well as higher order visual cortices like the lateral occipital complex and PPA that Hasson, Yang et al. (2008) pointed out specifically. When looking at the ISC especially found for Scenes of longer Durations, we saw the same spread towards parietal, temporal and frontal regions, including the precuneus, SPL and TPJ as well as significant clusters in the mPFC, SFG and STS, which Hasson et al. (2008c) demonstrated to be engaged particularly for longer stimuli. Thus, we could not only replicate the findings shown specifically for the ISC distribution of movies, while only depicting individual frames, but also reproduced a trend found for the ISC using other stimulus modalities (e.g., Farbood et al. 2015, Hasson et al. 2015, Lerner et al. 2011): While temporally shorter and hierarchically lower situated stimuli only reliably engage the cortices processing the respective modality (visual, auditory, etc.), longer stimuli that involve higher Content Levels and Durations additionally show significant ISC in multimodal integration areas like the precuneus or the mPFC (e.g., Sepulcre et al. 2012). Replicating these results proves that the manipulation worked and that the stimuli were processed similarly enough to naturalistic

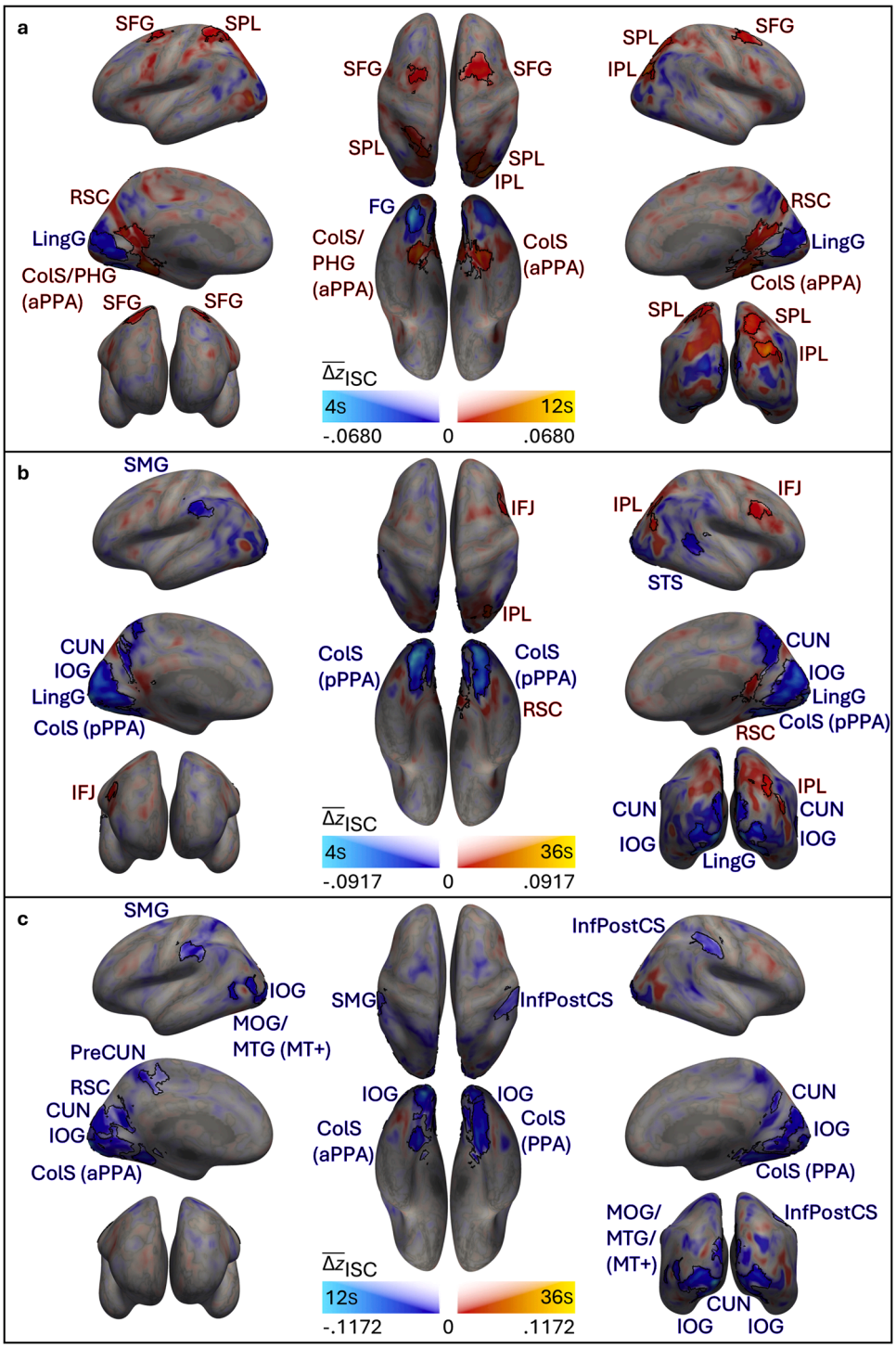

Fig. 6. Whole-brain results of the differences in Fisher z-transformed inter-subject correlations between the levels of Durations averaged across the Content Levels. *Note.* (a) Results of contrasting 12s and 4s Duration. (b) Results of contrasting 36s and 4s Durations. (c) Results of contrasting 36s and 12s Durations. Longer Duration > shorter Duration contrasts in warm colors, shorter Duration > longer Duration contrasts cold colors. Results are FWE-cluster corrected at $\alpha = 0.05$ with a cluster forming threshold of $p < .001$, Monte Carlo simulated cluster extend thresholds were 34 (a), 34 (b) and 35 (c) voxels, respectively. Significant clusters are depicted with a black border and are labelled accordingly. Red labels describe clusters belonging to the longer Duration > shorter Duration contrast, blue labels describe clusters belonging to the short Duration > longer Duration contrast. Results figures applies the "Highlight, don't Hide" principle (Taylor et al., 2023). Thus, clusters with non-significant *p*-values are not omitted but displayed at lower opacity. Whole-brain results are projected on inflated fsaverage pial surface. (a)PPA = (*anterior) parahippocampal place area*, ColS = *collateral sulcus*, CUN = *cuneus*, FG = fusiform gyrus, IFJ = *inferior frontal junction*, InfPostCS = *Inferior postcentral sulcus*, IOG = *inferior occipital gyrus*, IPL = intraparietal lobe, LingG = lingual gyrus, MOG = *middle occipital gyrus*, MTG = middle temporal gyrus, pPPA = *posterior parahippocampal place area*, PreCUN = *precuneus*, PHG = *parahippocampal* gyrus, RSC = retrosplenial cortex, SFG = superior frontal gyrus, SMG = *supramarginal gyrus*, SPL = superior parietal lobe.

Table 4
Significant differences in Fisher *z*-transformed inter-subject correlations from testing the specific interaction effects for Scenes of longer Durations (Scene × 12s and Scene × 36s).

Localization	H	Cluster extent (in voxel)	Peak coordinates (MNI space)			*z* value difference
			x	y	z	
Scene × 12s = (Scene 12s > Shot 12s) > (Scene 4s > Shot 4s)						
Superior parietal lobule	L	91	−17	−72	50	0.08
Precuneus	R	56	3	−54.5	45	0.07
Retrosplenial cortex	L	44	−17	−59.5	15	0.05
Middle occipital gyrus (OPA)	L	462	−29.5	−94.5	0	0.13
Intraparietal sulcus	L	l.m.	−29.5	−84.5	17.5	0.12
Middle occipital gyrus (MT+)	L	l.m.	−47	−74.5	15	0.10
Parahippocampal gyrus (anterior PPA)	R	50	30.5	−42	−7.5	0.08
Collateral sulcus (anterior and posterior PPA)	L	146	−27	−69.5	−12.5	0.09
Temporo-occipital incision	L	l.m.	−47	−67	−10	0.06
Scene × 36s = (Scene 36s > Shot 36s) > (Scene 4s > Shot 4s)						
Precuneus	L	53	−7	−49.5	37.5	0.05
Intraparietal sulcus	L	213	−27	−87	22.5	0.10
Middle occipital gyrus (OPA)	L	l.m.	−17	−97	12.5	0.09
Middle occipital gyrus/ Middle temporal gyrus (MT+)	R	119	50.5	−69.5	10	0.14
Middle occipital gyrus/ Middle temporal gyrus (MT+)	L	87	−49.5	−74.5	10	0.12
Superior temporal gyrus	L	121	−62	−44.5	7.5	0.06
Superior temporal sulcus	R	113	48	−37	5	0.07
Lingual gyrus	R	372	13	−89.5	−2.5	0.14
Middle occipital gyrus (OPA)	R	l.m.	20.5	−92	17.5	0.07
Inferior occipital gyrus	L	182	−42	−82	−5	0.11
Inferior occipital gyrus	R	48	45.5	−77	−5	0.09
Superior temporal gyrus/ Temporal pole	R	77	55.5	5.5	−15	0.04
Superior temporal sulcus/ Temporal pole	L	35	−49.5	3	−22.5	0.03

Note. Results are FWE-cluster corrected at α = 0.05 with a cluster forming threshold of $p < .001$, Monte Carlo simulated cluster extend threshold is 35 voxels (Scene × 12s) and 34 voxels (Scene × 36s), respectively. *H = Hemisphere; L = left; R = right*; MNI = *Montreal Neurological Institute*; l.m. = *local maximum*; OPA = *occipital place area*; PPA = *parahippocampal place area.*

stimuli to allow us to generalize the results of the comparison between conditions in this study to the processing of actual naturalistic stimuli.

4.2. Comparing inter-subject correlation between conditions - revealing the scene network

The investigated contrasts between the Content Levels and the different Durations resulted in significant clusters all across the cortex, including early visual cortices as well as hierarchically higher integration areas like the precuneus or temporal pole. For the sake of brevity, this discussion will focus primarily on the regions related to the scene-processing network (Epstein and Baker, 2019) as our main, novel finding. All other results will be discussed in detail in Supplemental Material 3.

4.2.1. The effect of Content Level

Contrasting the ISC of the Shot and Scene conditions averaged across Duration revealed, clusters in the bilateral MOG including areas V3A/B, the bilateral collateral sulcus, the bilateral SPL as well as the left SMG and right MTG that were additionally more engaged for Scenes.

The clusters in the MOG overlap with the previously described OPA (Dilks et al., 2013), located around the transverse occipital sulcus. Together with the PPA (Epstein and Kanwisher, 1998) and the RSC (Epstein, 2008) it forms the human *scene network* (Epstein and Baker, 2019). It is pivotal to note that a *scene* in the context of the scene network and movie scenes, are not the same. A scene in the more general definition of the network is a "[…] semantically coherent (and often nameable) view of a real-world environment comprising background elements and multiple discrete objects arranged in a spatially licensed manner." (Henderson and Hollingworth, 1999, p.244). In this sense, every frame in our experiment represented a scene, which could explain the reliable engagement of the scene network areas in all six experimental conditions.

More recent evidence however indicates that these three areas, OPA, PPA and RSC, do not capture the full extent of the network. It was suggested that the scene network is divided into two subnetworks, an anterior and a posterior part (Baldassano et al., 2017; Steel et al., 2021, 2023, 2025). While there is no full agreement of all regions included in each subnetwork, it is evident that the posterior portions are linked more closely to the sensory input as it receives signal from the early visual cortices, while the anterior portions of the network integrate memory into the process with their connections to, e.g., the hippocampus (Baldassano et al., 2016; Nasr et al., 2013; Silson et al., 2016). According to this division, the posterior scene network is made up of the OPA, the posterior PPA and possibly a more perceptively attuned part of the RSC which is also called medial place area (MPA; Baldassano et al., 2016; Steel et al., 2025). The anterior network consists of an area in the caudal inferior parietal lobe (cIPL) referenced also as the lateral place memory area (LPMA; Steel et al., 2021), the anterior portions of the PPA as well as the RSC or at least the memory-attuned part of the MPA (Baldassano et al., 2016; Steel et al., 2025). This distinction within the scene network is crucial for interpreting our results since - as will be highlighted throughout the discussion - the different contrasts were associated only with specific subnetworks.

Hence, the contrast of the Content Levels revealed that specifically the OPA as well as the posterior PPA were more reliably engaged by movie Scenes than movie Shots regardless of their Duration. Both these areas are part of the posterior scene network and thus are more involved with perceptual processing than the integration from memory (Baldassano et al., 2016; Steel et al., 2025). Accordingly, the OPA was previously shown to be activated when distinguishing between different sceneries based on the own, *egocentric* perspective representing the sense of left or right and distance (Epstein and Baker, 2019; Kamps et al., 2016). Further, this area is possibly used for the immediate localization of near boundaries or obstacles (Julian et al., 2018). This egocentric perspective taking is especially essential for movie Scenes, because in contrast to a Shot which depicts a single camera perspective, movie Scenes per definition often depict multiple Shots and thus include changes in the perspective taken by the camera, i.e., the viewer, making Scenes more visually variable (Cutting et al., 2011; Haq et al., 2019; Ji et al., 2019), which a comparison the visual features our Scene and Shot stimuli also demonstrated (see Supplemental Material 1). Subsequently, Scenes regardless of the duration have more egocentric views that need

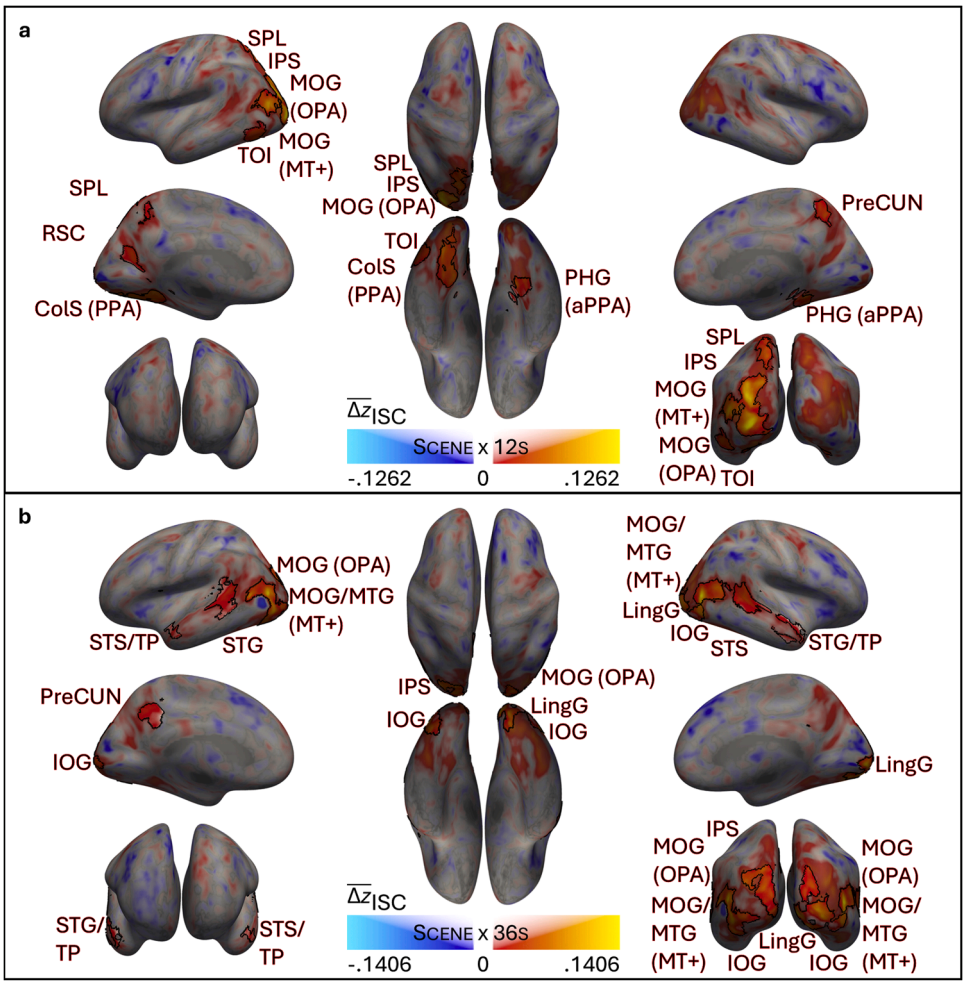

Fig. 7. Whole-brain results of the differences in Fisher z-transformed correlations from testing the specific interaction effects for Scenes of longer Durations.
Note. (a) Results for the Scene × 12s interaction. (b) Results for the Scene × 36s interaction. Scene × 12s > 0 and Scene × 36s > 0 contrasts in warm colors, opposing contrasts in cold colors. Results are FWE-cluster corrected at $\alpha = 0.05$ with a cluster forming threshold of $p < .001$, Monte Carlo simulated cluster extend thresholds were 35 (a) and 34 (b) voxels, respectively. Significant clusters are depicted with a black border and are labelled accordingly. Red labels describe clusters belonging to the Scene × 12s > 0 and Scene × 36s > 0 contrasts. Results figures applies the "Highlight, don't Hide" principle (Taylor et al., 2023). Thus, clusters with non-significant *p*-values are not omitted but displayed at lower opacity. Whole-brain results are projected on inflated fsaverage pial surface. ColS = *collateral sulcus*, IOG = *inferior occipital gyrus*, IPS = *intraparietal sulcus*, LingG = *lingual gyrus*, MOG = *middle occipital gyrus*, MTG = *middle temporal gyrus*, OPA = *occipital place area*, PHG = *parahippocampal gyrus*, PostCS = postcentral sulcus, (a)PPA = *(anterior) parahippocampal place area*, PreCUN = *precuneus*, RSC = *retrosplenial cortex*, SPL= *superior parietal lobe*, STG = *superior temporal gyrus*, STS = *superior temporal sulcus,* TOI = *temporo-occipital incision,* TP = *temporal pole*.

to be processed, which could explain the more reliable engagement of the OPA, as more frequent changes leave less room for idiosyncratic processing, increasing the stimulus-driven ISC (Faber et al., 2018).

Additionally, the Scene > Shot contrast also revealed clusters in the collateral sulcus which anchors the PPA (Epstein and Baker, 2019). The identified clusters specifically overlap with the posterior PPA, which is often identified in the context of visually guided spatial perception (Baldassano et al., 2013; Häusler et al., 2022; Steel et al., 2025). This is supported by findings demonstrating that particularly the posterior PPA is functionally related to spatial relationship (Aminoff and Bar, 2003; Aminoff and Tarr, 2015; Bilalić et al., 2019; Van den Stock et al., 2014), the existence of specific retinotopic maps in the posterior PPA (Arcaro et al., 2009; Beyh et al., 2022) and its general preference for simpler stimuli (Arcaro et al., 2009; Nasr et al., 2014; Rajimehr et al., 2011).

To summarize, the Scene > Shot contrast revealed engagement in a network of regions that process visual scenery. Intriguingly, only the more perceptively specialized posterior portion of this network - consisting of the OPA and the posterior PPA - was involved. This subnetwork is related to collecting and processing environmental features necessary for spatial navigation (e.g., Baldassano et al. 2016). Therefore, the results suggest that the necessary integration across multiple Shots in a Scene is processed as the perception of visual scenery from multiple different egocentric perspectives.

4.2.2. *The effect of Duration*

While clearly engaging the scene network the clusters identified for the effect of Content Level were interestingly only found in the posterior subnetwork. Looking at the effects of Duration however, we identified clusters overlapping with the anterior scene network (Baldassano et al., 2016; Steel et al., 2025). Thus, the contrast of 12s > 4s revealed significant engagement in the bilateral RSC, the anterior PPA as well as the right cIPL for longer stimuli. Also, in the contrast between the 36s > 4s condition we identified clusters in the right RSC and right cIPL. Only the 12s > 36s contrast did not uncover any more reliably engaged areas for the longer stimuli. This null finding could however be related to limitations regarding the 36s conditions, that will be discussed later (also see Supplemental Material 3).

The anterior portion of the scene network is suggested to extend on the function of the posterior part by helping with spatial navigation but also integrating information from memory indicated by its close

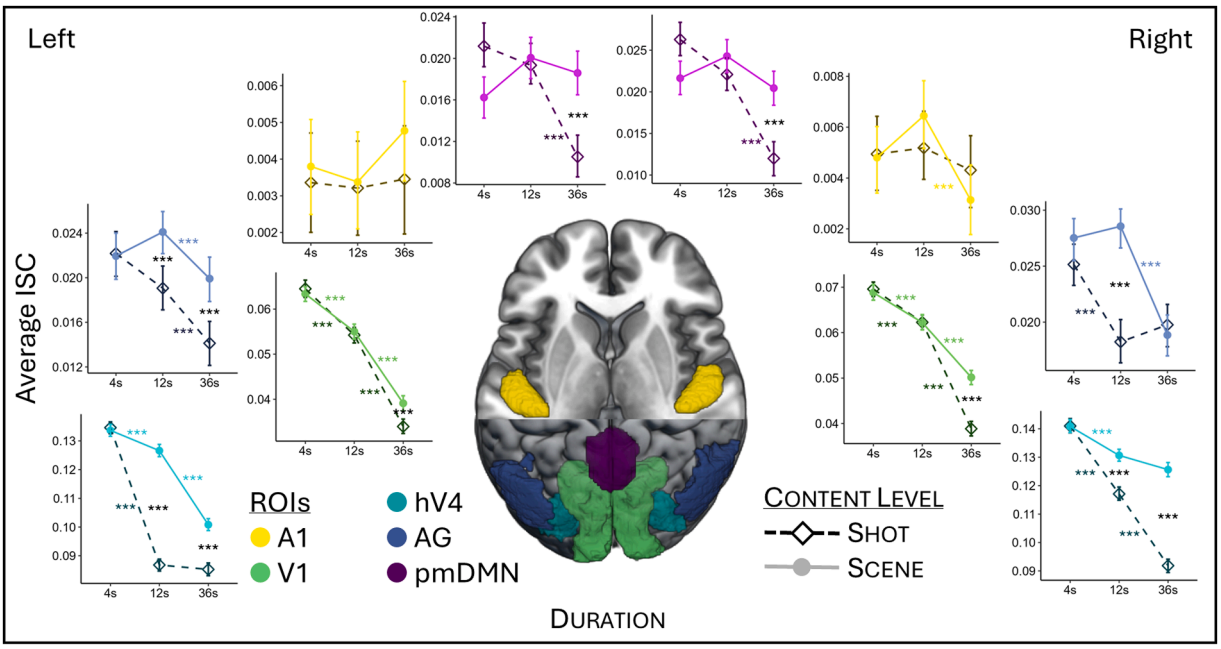

Fig. 8. Post-hoc test results of region of interest analysis for the visual cortical processing hierarchy.
Note. Results for the ROIs based on the visual cortical processing hierarchy (Baldassano et al., 2017). Content Level Shot in darker shade and dashed line, Content Level Scene in lighter shade and solid line. ROI = *region of interest*, A1 = *primary auditory cortex*, V1 = *primary visual cortex*, hV4 = *extrastriate visual cortex*, AG = *angular gyrus*, pmDMN = *posterior medial cluster of the default mode network*. * $p_{FDR\text{-}boot} < 0.05$, ** $p_{FDR\text{-}boot} < 0.01$, *** $p_{FDR\text{-}boot} < 0.001$, $N_{boot} = 500$. ROI masks are depicted on the MNI152 standard template. For a depiction of the ROIs on the pial fsaverage surface space see Figure S6.5a.

connection to the hippocampus (Baldassano et al., 2016; Silson et al., 2016; D. M. Watson and Andrews, 2024) as well as its activity in imagining scenes without actual input (Boccia et al., 2015). This higher level of visual processing is supported by findings about the individual regions that are part of this subnetwork:

Starting with the anterior PPA, its connectivity to the default mode network (DMN) as well as its relatively lower sensitivity to objects, suggest a focus on more global scene features and possibly a recall of features from memory (Aminoff et al., 2007; Baldassano et al., 2013). Moreover, previous studies showed that the anterior PPA is more concerned with non-visual information about scenery, while the posterior PPA processes visual cues (Häusler et al., 2022). Consequently, the anterior PPA is involved in mental imagery (Boccia et al., 2015) and the recall of spatial scenes (Bainbridge et al., 2021; Srokova et al., 2022). Recently, Steel et al. (2025) defined the spatial distinction between memory and perception related scene processing areas, also splitting the PPA into a posterior, visually-attuned and anterior, memory-attuned area. Our results are further support of this distinction of the posterior PPA and anterior PPA possibly linking the anterior PPA to slower intrinsic signals, longer TRWs and integration over longer periods of time.

Along the same line, we also identified the RSC only for stimuli of longer duration (12s > 4s and 36s > 4s). The RSC as part of the anterior scene subnetwork fulfills the role of orientation and navigation of an individual within a scene (Baldassano et al., 2016). This proposed function matches the rich literature on the RSC indicating that it encodes the heading direction of an individual (for reviews see Julian et al. 2018, Vann et al. 2009). In this context, the RSC processes consistent landmarks of the environment to from both egocentric and allocentric perspectives while also transitioning between these reference frames (Auger and Maguire, 2018; Balcerek et al., 2021; Julian et al., 2018; Peer et al., 2021; Sulpizio et al., 2020). Additionally, Alexander et al. (2023) proposed that the RSC is not only representing the spatial environment but also the temporal context, playing a role in episodic memory retrieval, which is supported by its involvement in recalling a previously watched movie (Zadbood et al., 2017). Moreover, recent evidence also suggests a functional distinction within the RSC and neighboring scene processing areas along the parieto-occipital fissure (Balcerek et al., 2021; Silson et al., 2019; Steel et al., 2021), dividing this area into a more visually-guided posterior and more memory-attuned anterior MPA (Silson et al., 2019; Steel et al., 2021, 2025). Based on the recently published coordinates for the MPA, the clusters we identified in the Duration contrasts rather overlap with the memory-attuned anterior part (Steel et al., 2025). This could further indicate that with increasing temporal duration the stimuli are not only visually processed but integrated with information from memory or other high-level processing steps which corresponds to an increase of TRWs along the cortical hierarchy (Golesorkhi et al., 2021b; Hasson et al., 2015; Wolff et al., 2022).

For both the 12s > 4s and 36s > 4s contrasts we also identified a third cluster in the right hemisphere which is part of the anterior subnetwork of the scene network, namely the cIPL (Baldassano et al., 2016) or LPMA (Steel et al., 2021). Located anterior and lateral to the OPA this area was previously related to the processing of familiar scenes regardless of the egocentric view point (Epstein et al., 2007) as well as imagery of familiar places (Scrivener et al., 2025) and the egocentric, orientation in larger environmental contexts (Han and Epstein, 2025; Steel et al., 2023). Thus, the cIPL - like the other anterior scene network regions - is also implied to be involved in a memory-based navigation of the own person in space.

Overall, finding these clusters specifically in longer Duration conditions might suggest that memory integration becomes more reliable with longer time. This is supported by recent results, demonstrating that the signal time course in a recall tasks peaks after four to twelve seconds in all three memory-attuned place areas - cIPL, anterior PPA, and RSC - and therefore later than in their perception-attuned equivalents (Steel et al., 2023).

Interestingly, the contrasts of Duration also revealed unexpected stronger effects for the 12s than the 36s condition (see Supplemental Material 3 for further discussion). Moreover, the supplemental ROI analysis in the PHG, that mainly overlapped with the canonical anterior PPA location (Steel et al., 2025), revealed an attunement to the 12s

Duration for both Shots and Scenes (see Supplemental Material 4). These results might thus indicate that specifically the 12s Duration rather than the 36s Duration was more relevant to the anterior scene network. In this context, studies investigating the shifts in cortical network states while watching a movie identified the brain state dwell times in networks and regions associated with the anterior scene network (Baldassano et al., 2013, 2016; D. M. Watson and Andrews, 2024) in the range of 10 s to 16 s (E. Yang et al., 2023), while no brain state regularly persisted for as long as 36 s (Geerligs et al., 2022; E. Yang et al., 2023). These brain state shifts are thought to coincide with the event boundaries (Baldassano et al., 2017; Geerligs et al., 2022; Sava-Segal et al., 2023) - e.g., changes in location - that are specifically processed in the scene network (Rah et al., 2025; Zacks et al., 2010). Complex stimuli with fewer event boundaries, like the 36s condition, have been shown to produce less synchronized behavioral and neural state boundaries (Sava-Segal et al., 2023). Moreover, longer periods without discernible situational changes increase the chance of mind wandering (Faber et al., 2018). Thus, the regular, noticeable spatial boundary shifts in the 12s condition, identified by the scene network, might have enhanced the ISC for this Duration. Additionally, while the average movie Scene can take somewhere in the range of 20 s to over a minute (Cutting et al., 2012; Zacks et al., 2009, 2010), movie Shots usually only last a few seconds (Cutting et al., 2011; Cutting and Candan, 2015; Pearlman, 2019). Hence, especially for the Shot condition, 12s stimuli could represent a more natural movie watching experience, possibly leading to a more synchronized viewing and thus relatively higher ISCs than in 36s condition. Since this study used three distinct levels of Duration, mirroring Hasson et al. (2008c), future studies could investigate a broader range of time windows, to more precisely estimate the relevant event durations for the anterior and posterior scene networks (e.g., Aberbach-Goodman and Mukamel, 2023) and further distinguish effects of Duration and Content Level. This is especially relevant as the anterior scene network, through its connectivity with the hippocampus, may also play a role in encoding events into memory (Baldassano et al., 2016; Ritchey and Cooper, 2020; Torres-Morales and Cansino, 2024). Using other time frames, previous studies have similarly identified the increase of TRWs along the cortical hierarchy (Honey et al., 2012; Lerner et al., 2011). Since it was also found that the cortical dynamics can, to a certain degree, adjust to the stimulus (Cavanagh et al., 2020; Lerner et al., 2014; Soltani et al., 2021), it is possible that our findings could be replicated at different scales of annotation or presentation frame rate. In contrast, the structural basis of cortical timescales (Chaudhuri et al., 2015; X.-J. Wang, 2020) might only allow for a limited temporal flexibility.

To conclude, Scenes and Shots of longer Duration compared with 4s stimuli specifically engaged the anterior portion of the scene network that was related to memory integration processes (e.g., Baldassano et al. 2017). Thus, these results could indicate that only longer stimulus durations allow for higher levels of processing, enabling deeper understanding of meaning, retention of the content, or prediction of upcoming events (Alexander et al., 2023; Steel et al., 2023). Short stimuli, on the other side of the spectrum, are too transient for these slower processes, only engaging earlier visual processing levels (Geerligs et al., 2022; Oetringer et al., 2025).

4.2.3. Interaction effect - areas reliably engaged by longer scenes

Since the effects of Content Level and Duration both revealed significant but also distinct areas, we also looked at regions that were additionally engaged for Scenes of a longer Duration (Scene × 12s and Scene × 36s). Some identified clusters, especially within the scene network, were shared between the main effects and the interactions, which supports that Content Level and Duration together are relevant for engaging this network. More specifically, across the two interactions of interest we located areas in both the anterior and posterior subnetworks, like the OPA or the anterior PPA and parts of the RSC. So, we saw the expected conjunction of visually-attuned and memory-attuned areas in the processing of the content-rich and temporally persistent Scene conditions. Therefore, even more interesting are areas specifically engaged by the interaction. Here we found clusters in the precuneus, the lateral MOG merging into MTG and along the entire STS specifically for the Scene × 36s contrast.

The precuneus is an area which is commonly found in movie-watching studies (Jääskeläinen et al., 2021; Oren et al., 2016; Zacks et al., 2001). In that context, it was related to the understanding and recalling the temporally consecutive order of events of the movie (Baldassano et al., 2017; Kwok et al., 2012; Lahnakoski et al., 2017; Tylén et al., 2015; Whitney et al., 2009; Yeshurun et al., 2017; Zacks et al., 2010; Zadbood et al., 2017). This level of complex and deeper comprehension could be explained by the precuneus' role as high-level rich and integrative cortical hub (Gollo et al., 2015; Mecklenbrauck et al., 2024a; Riedel et al., 2022). Multiple studies have identified the precuneus and its subfields as parts of different functional networks related to memory integration, working memory, as well as the processing of naturalistic stimuli (Cavanna and Trimble, 2006; Dadario and Sughrue, 2023; Deng et al., 2019; Yeo et al., 2011). Accordingly, we also discovered reliable precuneus engagement for longer Scene stimuli which allow for a more profound understanding of the content.

The other clusters in the lateral MOG/MTG as well as the extension along the STS might be related to a recently described third visual pathway which appropriately involves the human motion selective area MT+ in the lateral MOG/MTG, the posterior STS as well as the anterior STS (Pitcher and Ungerleider, 2021). This network supposedly processes dynamic movements in a social context and aids in the understanding of actions of other individuals (Pitcher and Ungerleider, 2021). This is supported by motion selective regions located along the STS (Allison et al., 2000; Cheng et al., 2025; Grossman et al., 2005; Herrington et al., 2011; McAleer et al., 2014) culminating in a face-focused region in the temporal pole (Duchaine and Yovel, 2015; H. Yang et al., 2016). However, we see significantly reliable engagement in this pathway mainly in the results for the Scene × 36s interaction which contains the contrast of Scenes and Shots of 36s. Hence, this contrast could have been dominated by the perception of an ever-changing, motion-inducing Scene 36s conditions compared to a more constant, still-frame-like Shot 36s condition (see Supplemental Material 3 for further discussion). Yet, in the context of movie watching especially the anterior STS is also a sign for a high-level understanding and retention of the content (AbdulSabur et al., 2014; Akimoto et al., 2014; Hasson et al., 2008a; Lahnakoski et al., 2017; Speer et al., 2007; Yeshurun et al., 2017; Zacks et al., 2010). To investigate the effects of Content Level and Duration along the STS in more detail, we conducted an exploratory ROI analysis of three functionally distinct areas along the STS (Zhen et al., 2015). This analysis demonstrated that the anterior portions of the STS are not only more engaged by the longer, content-richer Scenes but also that the engagement increased significantly with Duration (see Supplemental Material 5), supporting the anterior STS's specific role in higher-level and social aspects of stimulus processing (e.g., Chen et al., 2017b, Lee Masson and Isik 2021, Yeshurun et al. 2017).

In this light, the presence of clusters in the anterior STS of precuneus for longer Scenes is consistent with evidence that extended stimulus segments permit integration with memory content (Steel et al., 2023), thereby supporting higher-level comprehension.

4.3. The effects of content level and duration along the visual cortical hierarchy

There is evidence for both the processing of the stimulus content (Baldassano et al., 2017; Geerligs et al., 2022; Oetringer et al., 2025) and of the temporal organization (Chaudhuri et al., 2015; Golesorkhi et al., 2021a; Hasson et al., 2015) to align with the cortical hierarchy. As discussed, our whole-brain analysis confirmed the relevance of both factors, as longer, more content-rich movie scenes engaged hierarchically higher areas. To examine the relationship between Content Level, Duration and the cortical hierarchy in greater detail, we conducted

a ROI analysis. Thereby, the ROIs were chosen to reflect the visual processing hierarchy described by Baldassano et al. (2017) as well as region commonly implicated in movie perception (e.g., Jääskeläinen et al. 2021). The results for the latter are reported and discussed in Supplemental Material 4.

At the earliest stage of hierarchy, we found no major differences between movie Scenes and Shots. Instead, ISC decreased with longer Duration across all conditions. Given that the early visual cortex is primarily engaged by low-level visual features, such as contrasts, edges or motion (Grill-Spector and Malach, 2004; Seidemann and Geisler, 2018; Tong, 2003), stimulus properties that are largely shared between the Content Levels (see Supplemental Material 1), no difference in engagement was expected (see Supplemental Material 3 for further discussion).

One step higher, in hV4, we observed the first clear separation between the Content Levels. Scenes elicited stronger ISC than Shots at 12s and 36s Duration. This is consistent with hV4's functions not only encompassing color and motion perception (Cohen et al., 2019; Grill-Spector and Malach, 2004; Wade et al., 2002) but also the integration of visual information across larger receptive fields (Grill-Spector and Malach, 2004; Winawer and Witthoft, 2015). Its role as an early integrative hub and first branching point between the dorsal and ventral processing streams (Winawer and Witthoft, 2015) may further explain why differences between Scenes and Shots already emerged at this stage in the processing.

At the upper end of the visual processing hierarchy, the angular gyrus and the posterior-medial DMN cluster (precuneus/posterior cingulate cortex) showed robust differences in the ISC between movie Scenes and Shots, especially at longer Durations. Both regions are well-established as integrative hubs, associated with memory and narrative comprehension (Ben-Yakov and Henson, 2018; Hasson et al., 2015; Raichle, 2015; Rockland and Graves, 2023; Seghier, 2023). The angular gyrus, as part of the DMN, is functionally linked to the hippocampus (Rockland and Graves, 2023) and involved in semantic processing (Ben-Yakov and Henson, 2018; Seghier, 2023). The posterior-medial DMN cluster (Shirer et al., 2012) whereas forms a core network hub (Fransson and Marrelec, 2008; Hagmann et al., 2008) that is part of the hippocampal-parietal memory system (Raichle, 2015). Our ROI results support this high-level role in the integration of temporally extended, content-rich information.

Thus, in line with our whole-brain findings, the ROI results also suggest an interaction of Content Level and Duration. Hierarchically higher areas showed relatively greater ISC for longer Scenes than Shots, consistent with the idea that more complex and persistent stimuli engage regions with longer TRWs that integrate information over longer periods of time (Hasson et al., 2015; Wolff et al., 2022). The functional distinction between the anterior and posterior scene subnetworks further supports this account: Clusters specifically engaged by 12s and 36s stimuli overlapped with the memory-related regions identified by Steel et al. (2021, 2025). Crucially, Steel et al. (2025) demonstrated that the visually attuned posterior regions and memory-related anterior regions occupy distinct positions within the cortical hierarchy (e.g., Margulies et al. 2016), with posterior areas closer to unimodal cortices and anterior areas closer to multimodal integration zones. Fittingly, this uni-to-multimodal gradient was shown to explain both internal timescales and ISC distributions across the cortex (Golesorkhi et al., 2021a; Ito et al., 2020; Mecklenbrauck et al., 2024b; Raut et al., 2020). The anterior portions of the scene network regions are also more strongly connected to the DMN (Baldassano et al., 2013; 2017; D. M. Watson and Andrews, 2024), including the posterior DMN cluster investigated here (Shirer et al., 2012), which has been specifically associated with slow, self-referential, memory-based integration processes (Menon, 2023; Raichle, 2015). Thus, both ROI and whole-brain results suggest that hierarchically higher, more integrative areas are engaged more reliably by longer, more complex stimuli and evidently contain longer TRWs.

However, while our results are broadly consistent with previous TRW findings, the ISC pattern we observed differed in some important respects from previous investigations. Earlier work reported stable ISCs in the primary cortices regardless of the duration and demonstrated an increase in higher-level areas when both duration and content were extended (Aberbach-Goodman and Mukamel, 2023; Farbood et al., 2015; Hasson et al., 2008c; Lerner et al., 2011). In contrast, we discovered a general decline in ISC with Duration, with the only relatively higher values for Scenes reflecting a less steep decrease compared to Shots. This overall decline of ISC might however be an artifact of our stimulus design: Instead of well-structured and continuous naturalistic movie clips, participants viewed only sequences of still frames, which may have reduced inter-subject synchrony (Hasson, Landesman et al., 2008; Loschky et al., 2020).

Moreover, although conditions were carefully balanced, some natural differences remained - particularly at longer Durations, as movie shots are inherently shorter and visually more uniform than movie scenes (Cutting et al., 2011; Cutting and Candan, 2015). We see these differences firstly in the duration-shorting ratio between longer and shorter conditions as well as Shots and Scenes (see 2.2). While neural dynamics were shown to adapt to the environmental and task demands (Cavanagh et al., 2020; Golesorkhi et al., 2021; Lerner et al., 2014; Soltani et al., 2021; Song et al., 2024), a significant acceleration of a stimulus can result in less reliable neural activity across subjects (Lerner et al., 2014). Yet, Scenes, that were overall more accelerated than Shots, still evoked significantly higher ISC, which suggests the effects of Content Level persisted despite this stronger temporal compression.

The other remaining differences concern the higher visual variability within and between movie segments at shorter Durations (see Supplemental Material 1). The fewer, isochronously extracted frames of shorter conditions naturally produced more noticeable visual gaps. Additionally, the shorter the Duration, the more distinct segments - and thus transitions - were presented. Although these differences are inherent to the naturalistic stimulus material and present in other TRW studies as well (e.g., Hasson et al. 2008c), these confounds still have to be considered when interpreting the results. Specifically, the effects in the early visual cortex for the contrast between shorter and longer durations might have been affected by these remaining visual differences (see Supplemental Material 3; Lee and Chen, 2024; Lu et al., 2016; Nunez-Elizalde et al., 2019; Oetringer et al., 2025). Nevertheless, the replication of the key engagement patterns reported by Hasson et al. (2008c) despite these methodological differences, still supports the robustness of our findings and interpretation.

4.4. Conclusions and future research

The present study investigated the processing of movie stimuli. We disentangled the effects of the Content Level (Shots and Scenes) from Duration to determine their respective roles in parsing hierarchically nested, complex, naturalistic stimuli. Overall, our results point to the engagement of a network responsive to visual scenery (Epstein and Baker, 2019) during movie viewing. With higher Content Level, we observed additional engagement of a posterior subnetwork more tightly linked to the incoming visual input and integrating across viewpoints. This is complemented, at longer Duration, by greater engagement of an anterior subnetwork supporting integration - and possible prediction - from memory (Alexander et al., 2023; Baldassano et al., 2016; Steel et al., 2021).

On the one hand, previous studies have suggested processing as a function of the stimulus duration (e.g., Hasson et al. 2015). Based on our results, the Duration of the stimulus does play a crucial role in enabling more in-depth processing. However, in particular the interaction of Content Level and Duration indicates that longer as well as more content-rich stimuli engage higher-level cortical areas. On the other hand, two separate systems have been proposed for integrating across content or across time (Baumgarten et al., 2021). While the focus of effects of Content Level and Duration to the posterior and anterior portions of the scene network, respectively, could suggest a distinction,

these two subnetworks are constantly exchanging information supporting the integration of visual and memory content. In doing so, these networks possibly represent the intersection between externally and internally focused processing (Steel et al., 2024), therefore emphasizing two collaborating rather than distinct cortical systems.

Future studies applying eye-tracking or using actual, naturalistic stimuli are required to further demonstrate the interplay of Content Level and Duration in the scene network when watching a movie. Furthermore, even though Steel et al. (2025) recently demonstrated that the anterior scene subnetwork represents a higher level in the cortical hierarchy than the posterior portion and our results indicate a possible difference in TRWs between the two subnetworks, the precise mechanisms underlying naturalistic stimulus processing remain unresolved. What makes a movie scene a movie scene? What content is integrated from memory? How is the scene network with its connection to both the DMN and the hippocampus as well as the early visual cortices related to the top-down-bottom-up hierarchy in prediction (Alexander et al., 2023)? Beyond the realm of cinema, it would also be interesting to see whether effects of Content Level and Duration can be generalized to other hierarchically nested, natural stimuli like language, music, or actions, and whether the anterior and posterior scene networks operate cross-modally or if similar networks can be found for non-visual modalities as well.

Ultimately, this study demonstrates the importance and the interplay of stimulus' content complexity as well as its temporal duration during movie watching and reveals that the two subnetworks of the scene network - previously primarily linked to spatial navigation - are key to movie comprehension.

Data and code availability statement

All code for stimulus creation, running the experiment as well as data preprocessing, analysis and visualization is available at https://github.com/FalkMeck/Time_Points_or_Plot_Points.

Data of one exemplary, minimally processed participant, all pairwise-ISC maps and the ROI analysis data can be downloaded at https://osf.io/xk6cv/overview?view_only=036a15ee8620401a9b1db5fd6fc25086.

Ethics statement

This study was conducted in accordance with the Declaration of Helsinki and approved by the Local Ethics Committee of the University of Münster. Participants signed an informed consent before participation.

Funding

The authors received no specific funding for this work.

CRediT authorship contribution statement

Falko Mecklenbrauck: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ricarda I. Schubotz:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors have declared that no competing interests exist.

Acknowledgment

The authors thank Monika Mertens, Anna Sophia Asdonk, Thea Maria Bröring, Maike Denguth, Vanessa Klein, Pia Niedrée, Yannah Loïs Schröder, and Zübeyde Nihan Yildirm for their help during data collection, Sebastian Bias, Laura Sophie Petretto, and Zübeyde Nihan Yildirm for help in testing the paradigm, and Lena Charlotte Innig for helpful comments on the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2026.121809.

References

AbdulSabur, N.Y., Xu, Y., Liu, S., Chow, H.M., Baxter, M., Carson, J., Braun, A.R., 2014. Neural correlates and network connectivity underlying narrative production and comprehension: a combined fMRI and PET study. Cortex 57, 107–127. https://doi.org/10.1016/j.cortex.2014.01.017.

Aberbach-Goodman, S., Mukamel, R., 2023. Temporal hierarchy of observed goal-directed actions. Sci. Rep. 13 (1), 1–13. https://doi.org/10.1038/s41598-023-46917-z.

Agcaoglu, O., Wilson, T.W., Wang, Y.P., Stephen, J., Calhoun, V.D., 2019. Resting state connectivity differences in eyes open versus eyes closed conditions. Hum. Brain Mapp. 40 (8), 2488–2498. https://doi.org/10.1002/hbm.24539.

Akimoto, Y., Sugiura, M., Yomogida, Y., Miyauchi, C.M., Miyazawa, S., Kawashima, R., 2014. Irony comprehension: social conceptual knowledge and emotional response. Hum. Brain Mapp. 35 (4), 1167–1178. https://doi.org/10.1002/hbm.22242.

Alexander, A.S., Place, R., Starrett, M.J., Chrastil, E.R., Nitz, D.A., 2023. Rethinking retrosplenial cortex: perspectives and predictions. Neuron 111 (2), 150–175. https://doi.org/10.1016/j.neuron.2022.11.006.

Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: role of the STS region. Trends Cogn. Sci. 4 (7), 267–278. https://doi.org/10.1016/S1364-6613(00)01501-1.

Aminoff, E.M., Bar, M., 2003. Cortical analysis of visual context. Neuron 38 (2), 347–358.

Aminoff, E.M., Gronau, N., Bar, M., 2007. The parahippocampal cortex mediates spatial and nonspatial associations. Cereb. Cortex 17 (7), 1493–1503. https://doi.org/10.1093/cercor/bhl078.

Aminoff, E.M., Tarr, M.J., 2015. Associative processing is inherent in scene perception. PLoS One 10 (6), 1–19. https://doi.org/10.1371/journal.pone.0128840.

Amunts, K., Mohlberg, H., Bludau, S., Zilles, K., 2020. Julich-Brain: a 3D probabilistic atlas of the human brain's cytoarchitecture. Science (1979) 369 (6506), 988–992. https://doi.org/10.1126/science.abb4588.

Andersson, J.L.R., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20 (2), 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7.

Arcaro, M.J., McMains, S.A., Singer, B.D., Kastner, S., 2009. Retinotopic organization of human ventral visual cortex. J. Neurosci. 29 (34), 10638–10652. https://doi.org/10.1523/JNEUROSCI.2807-09.2009.

Auger, S.D., Maguire, E.A., 2018. Retrosplenial cortex indexes stability beyond the spatial domain. J. Neurosci. 38 (6), 1472–1481. https://doi.org/10.1523/JNEUROSCI.2602-17.2017.

Bainbridge, W.A., Hall, E.H., Baker, C.I., 2021. Distinct representational structure and localization for visual encoding and recall during visual imagery. Cereb. Cortex 31 (4), 1898–1913. https://doi.org/10.1093/cercor/bhaa329.

Balcerek, E., Włodkowska, U., Czajkowski, R., 2021. Retrosplenial cortex in spatial memory: focus on immediate early genes mapping. Mol. Brain 14 (1), 1–17. https://doi.org/10.1186/s13041-021-00880-w.

Baldassano, C., Beck, D.M., Fei-Fei, L., 2013. Differential connectivity within the Parahippocampal place area. Neuroimage 75, 228–237. https://doi.org/10.1016/j.neuroimage.2013.02.073.

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

Baldassano, C., Esteva, A., Fei-Fei, L., Beck, D.M., 2016. Two distinct scene-processing networks connecting vision and memory. eNeuro 3 (5). https://doi.org/10.1523/ENEURO.0178-16.2016.

Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1). https://doi.org/10.18637/jss.v067.i01.

Baumgarten, T.J., Maniscalco, B., Lee, J.L., Flounders, M.W., Abry, P., He, B.J., 2021. Neural integration underlying naturalistic prediction flexibly adapts to varying sensory input rate. Nat. Commun. 12 (1), 1–14. https://doi.org/10.1038/s41467-021-22632-z.

Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37 (1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate : a practical and powerful approach to multiple testing. author (s): yoav Benjamini and Yosef Hochberg source : journal of the Royal Statistical Society . Series B (Methodological), Vol . 57, No . 1 (1995), Publi J. R. Stat. Soc. 57 (1), 289–300.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29 (4), 1165–1188. https://doi.org/10.1214/aos/1013699998.

Ben-Yakov, A., Henson, R.N., 2018. The hippocampal film editor: sensitivity and specificity to event boundaries in continuous experience. J. Neurosci. 38 (47), 10057–10068. https://doi.org/10.1523/JNEUROSCI.0524-18.2018.
Ben-Yakov, A., Honey, C.J., Lerner, Y., Hasson, U., 2012. Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. Neuroimage 63 (1), 501–506. https://doi.org/10.1016/j.neuroimage.2012.07.008.
Beyh, A., Dell'Acqua, F., Cancemi, D., De Santiago Requejo, F., ffytche, D., Catani, M., 2022. The medial occipital longitudinal tract supports early stage encoding of visuospatial information. Commun. Biol. 5 (1), 1–10. https://doi.org/10.1038/s42003-022-03265-4.
Bilalić, M., Lindig, T., Turella, L., 2019. Parsing rooms: the role of the PPA and RSC in perceiving object relations and spatial layout. Brain Struct. Funct. 224 (7), 2505–2524. https://doi.org/10.1007/s00429-019-01901-0.
Boccia, M., Piccardi, L., Palermo, L., Nemmi, F., Sulpizio, V., Galati, G., Guariglia, C., 2015. A penny for your thoughts! patterns of fMRI activity reveal the content and the spatial topography of visual mental images. Hum. Brain Mapp. 36 (3), 945–958. https://doi.org/10.1002/hbm.22678.
Bond, C.F., Richardson, K., 2004. Seeing the Fisher Z-transformation. Psychometrika 69 (2), 291–303. https://doi.org/10.1007/BF02295945.
Bradley, M.M., Lang, P.J., 1994. Self-assessment Manikin (SAM). J.Behav. Ther. Exp. Psychiat 25 (1), 49–59.
Brett, M., Anton, J.-L., Valabregue, R., Poline, J.-B., 2002. Region of interest analysis using *An spm* toolbox. In: Proceedings of the 8th International Conferance on Functional Mapping of the Human Brain, 497. https://matthew.dynevor.org/research/abstracts/marsbar/marsbar_abstract.pdf.
Cavanagh, S.E., Hunt, L.T., Kennerley, S.W., 2020. A diversity of intrinsic timescales underlie neural computations. Front. Neural Circuits 14 (December), 1–18. https://doi.org/10.3389/fncir.2020.615626.
Cavanna, A.E., Trimble, M.R., 2006. The precuneus: a review of its functional anatomy and behavioural correlates. Brain 129 (3), 564–583. https://doi.org/10.1093/brain/awl004.
Chakraborty, S., Thounaojam, D.M., Sinha, N., 2021. A shot boundary detection technique based on visual colour information. Multimed. Tools. Appl. 80 (3), 4007–4022. https://doi.org/10.1007/s11042-020-09857-8.
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.
Chen, G., Shin, Y.W., Taylor, P.A., Glen, D.R., Reynolds, R.C., Israel, R.B., Cox, R.W., 2016a. Untangling the relatedness among correlations, part I: nonparametric approaches to inter-subject correlation analysis at the group level. Neuroimage 142, 248–259. https://doi.org/10.1016/j.neuroimage.2016.05.023.
Chen, G., Taylor, P.A., Qu, X., Molfese, P.J., Bandettini, P.A., Cox, R.W., Finn, E.S., 2020. Untangling the relatedness among correlations, part III: inter-subject correlation analysis through bayesian multilevel modeling for naturalistic scanning. Neuroimage 216, 116474. https://doi.org/10.1016/j.neuroimage.2019.116474. December 2019.
Chen, G., Taylor, P.A., Shin, Y.W., Reynolds, R.C., Cox, R.W., 2017a. Untangling the relatedness among correlations, part II: inter-subject correlation group analysis through linear mixed-effects modeling. Neuroimage 147, 825–840. https://doi.org/10.1016/j.neuroimage.2016.08.029. October 2016).
Chen, J., Honey, C.J., Simony, E., Arcaro, M.J., Norman, K.A., Hasson, U., 2016b. Accessing real-life episodic information from minutes versus hours earlier modulates hippocampal and high-order cortical dynamics. Cereb. Cortex 26 (8), 3428–3441. https://doi.org/10.1093/cercor/bhv155.
Chen, J., Leong, Y.C., Honey, C.J., Yong, C.H., Norman, K.A., Hasson, U., 2017b. Shared memories reveal shared structure in neural activity across individuals. Nat. Neurosci. 20 (1), 115–125. https://doi.org/10.1038/nn.4450.
Cheng, Y., Xin, Y., Lu, X., Yang, T., Ma, X., Yuan, X., Liu, N., Jiang, Y., 2025. Shared and unique neural codes for biological motion perception in humans and Macaque monkeys. Adv. Sci. 12 (18), 1–13. https://doi.org/10.1002/advs.202411562.
Cohen, D., Goddard, E., Mullen, K.T., 2019. Reevaluating hMT+ hV4 functional specialization for motion and static contrast using fMRI-guided repetitive transcranial magnetic stimulation. J. Vis. 19 (3), 1–20. https://doi.org/10.1167/19.3.11.
Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29 (3), 162–173. https://doi.org/10.1006/cbmr.1996.0014.
Cutting, J.E., 2014. Event segmentation and seven types of narrative discontinuity in popular movies. Acta Psychol. 149, 69–77. https://doi.org/10.1016/j.actpsy.2014.03.003.
Cutting, J.E., 2016. Narrative theory and the dynamics of popular movies. Psychon. Bull. Rev. 23 (6), 1713–1743. https://doi.org/10.3758/s13423-016-1051-4.
Cutting, J.E., 2019a. Sequences in popular cinema generate inconsistent event segmentation. Atten. Percept. Psychophys. 81 (6), 2014–2025. https://doi.org/10.3758/s13414-019-01757-w.
Cutting, J.E., 2019b. Simplicity, complexity, and narration in popular movies. Narrat. Complex. 2019, 200–222. https://doi.org/10.2307/j.ctvhktjh6.14. August.
Cutting, J.E., Brunick, K.L., Candan, A., 2012. Perceiving event dynamics and parsing Hollywood films. J. Exp. Psychol.: Hum. Percept. Perform. 38 (6), 1476–1490. https://doi.org/10.1037/a0027737.
Cutting, J.E., Brunick, K.L., DeLong, J.E., 2011. How act structure sculpts shot lengths and shot transitions in Hollywood film. Projections 5 (1), 1–16. https://doi.org/10.3167/proj.2011.050102.
Cutting, J.E., Candan, A., 2015. Shot durations, shot classes, and the increased pace of popular movies. Projections (New York) 9 (2), 40–62. https://doi.org/10.3167/proj.2015.090204.
Cutting, J.E., Iricinschi, C., 2015. *Re*-presentations of space in Hollywood movies: an event-indexing analysis. Cogn. Sci. 39 (2), 434–456. https://doi.org/10.1111/cogs.12151.
Dadario, N.B., Sughrue, M.E., 2023. The functional role of the precuneus. Brain 146 (9), 3598–3607. https://doi.org/10.1093/brain/awad181.
Deng, Z.Z., Wu, J.F., Gao, J.Q., Hu, Y., Zhang, Y.W., Wang, Y.S., Dong, H.M., Yang, Z., Zuo, X.N., 2019. Segregated precuneus network and default mode network in naturalistic imaging. Brain Struct. Funct. 224 (9), 3133–3144. https://doi.org/10.1007/s00429-019-01953-2.
Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31 (3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021.
Dilks, D.D., Julian, J.B., Paunov, A.M., Kanwisher, N., 2013. The occipital place area is causally and selectively involved in scene perception. J. Neurosci. 33 (4), 1331–1336. https://doi.org/10.1523/JNEUROSCI.4081-12.2013.
Duchaine, B., Yovel, G., 2015. A revised neural framework for face processing. Annu. Rev. Vis. Sci. 1 (1), 393–416. https://doi.org/10.1146/annurev-vision-082114-035518.
Dumoulin, S.O., Bittar, R.G., Kabani, N.J., Baker, C.L., Le Goualher, G., Pike, G.B., Evans, A.C., 2000. A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. Cereb. Cortex 10 (5), 454–463. https://doi.org/10.1093/cercor/10.5.454.
Eickhoff, S.B., Milham, M., Vanderwal, T., 2020. Towards clinical applications of movie fMRI. Neuroimage 217 (April), 116860. https://doi.org/10.1016/j.neuroimage.2020.116860.
Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25 (4), 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034.
Epstein, R.A., 2008. Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn. Sci. 12 (10), 388–396. https://doi.org/10.1016/j.tics.2008.07.004.
Epstein, R.A., Baker, C.I., 2019. Scene perception in the Human brain. Annu. Rev. Vis. Sci. 5, 373–397. https://doi.org/10.1146/annurev-vision-091718-014809.
Epstein, R.A., Higgins, J.S., Jablonski, K., Feiler, A.M., 2007. Visual scene processing in familiar and unfamiliar environments. J. Neurophysiol. 97 (5), 3670–3683. https://doi.org/10.1152/jn.00003.2007.
Epstein, R.A., Kanwisher, N., 1998. Acortical representation of the local visual environment. Nature 392 (6676), 598–601.
Faber, M., Radvansky, G.A., D'Mello, S.K, 2018. Driven to distraction: a lack of change gives rise to mind wandering. Cognition 173, 133–137. https://doi.org/10.1016/j.cognition.2018.01.007. December 2016.
Farbood, M.M., Heeger, D.J., Marcus, G., Hasson, U., Lerner, Y., 2015. The neural processing of hierarchical structure in music and speech at different timescales. Front. Neurosci. 9 (APR), 1–13. https://doi.org/10.3389/fnins.2015.00157.
Field, S., 2005. Screenplay: The Foundations of Screenwriting. Random House Publishing Group.
FIL Methods Group. (2017). *SPM12 manual*. 15(3), 1–508. https://www.fil.ion.ucl.ac.uk/spm/doc/manual.pdf%0Ahttp://www.fil.ion.ucl.ac.uk/spm/.
Finn, E.S., Bandettini, P.A., 2021. Movie-watching outperforms rest for functional connectivity-based prediction of behavior. Neuroimage 235 (February), 117963. https://doi.org/10.1016/j.neuroimage.2021.117963.
Fitch, W.T., Martins, M.D., 2014. Hierarchical processing in music, language, and action: lashley revisited. Ann. N. Y. Acad. Sci. 1316 (1), 87–104. https://doi.org/10.1111/nyas.12406.
Fransson, P., Marrelec, G., 2008. The precuneus/posterior cingulate cortex plays a pivotal role in the default mode network: evidence from a partial correlation network analysis. Neuroimage 42 (3), 1178–1184. https://doi.org/10.1016/j.neuroimage.2008.05.059.
Frazier, J.A., Chiu, S., Breeze, J.L., Makris, N., Lange, N., Kennedy, D.N., Herbert, M.R., Bent, E.K., Koneru, V.K., Dieterich, M.E., Hodge, S.M., Rauch, S.L., Grant, P.E., Cohen, B.M., Seidman, L.J., Caviness, V.S., Biederman, J., 2005. Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. Am. J. Psychiatry 162 (7), 1256–1265. https://doi.org/10.1176/appi.ajp.162.7.1256.
Geerligs, L., Gözükara, D., Oetringer, D., Campbell, K.L., van Gerven, M., Güçlü, U., 2022. A partially nested cortical hierarchy of neural states underlies event segmentation in the human brain. Elife 11, 1–36. https://doi.org/10.7554/eLife.77430.
Goldstein, J.M., Seidman, L.J., Makris, N., Ahern, T., O'Brien, L.M., Caviness, V.S., Kennedy, D.N., Faraone, S.V., Tsuang, M.T., 2007. Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. Biol. Psychiatry 61 (8), 935–945. https://doi.org/10.1016/j.biopsych.2006.06.027.
Golesorkhi, M., Gomez-Pilar, J., Tumati, S., Fraser, M., Northoff, G., 2021a. Temporal hierarchy of intrinsic neural timescales converges with spatial core-periphery organization. Commun. Biol. 4 (1), 1–14. https://doi.org/10.1038/s42003-021-01785-z.
Golesorkhi, M., Gomez-Pilar, J., Zilio, F., Berberian, N., Wolff, A., Yagoub, M.C.E., Northoff, G., 2021b. The brain and its time: intrinsic neural timescales are key for input processing. Commun. Biol. 4 (1), 1–16. https://doi.org/10.1038/s42003-021-02483-6.
Gollo, L.L., Zalesky, A., Matthew Hutchison, R., van den Heuvel, M.P., Breakspear, M., 2015. Dwelling quietly in the rich club: brain network determinants of slow cortical

fluctuations. Philos. Trans. R. Soc. B: Biol. Sci. 370 (1668). https://doi.org/10.1098/rstb.2014.0165.
Grafton, S.T., Hamilton, A.F.D.C., 2007. Evidence for a distributed hierarchy of action representation in the brain. Hum. Mov. Sci. 26 (4), 590–616. https://doi.org/10.1016/j.humov.2007.05.009.
Grill-Spector, K., Malach, R., 2004. The human visual cortex. Annu. Rev. Neurosci. 27 (June), 649–677. https://doi.org/10.1146/annurev.neuro.27.070203.144220.
Grossman, E.D., Battelli, L., Pascual-Leone, A., 2005. Repetitive TMS over posterior STS disrupts perception of biological motion. Vis. Res. 45 (22), 2847–2853. https://doi.org/10.1016/j.visres.2005.05.027.
Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Van Wedeen, J., Sporns, O., 2008. Mapping the structural core of human cerebral cortex. PLoS Biol. 6 (7), 1479–1493. https://doi.org/10.1371/journal.pbio.0060159.
Hamilton, A.F.D.C., Grafton, S.T., 2006. Goal representation in human anterior intraparietal sulcus. J. Neurosci. 26 (4), 1133–1137. https://doi.org/10.1523/JNEUROSCI.4551-05.2006.
Hammoud, R., Kouam, D.G., 2000. A mixed classification approach of shots for constructing scene structure for movie films. In: Irish Machine Vision and Image Processing Conference, pp. 223–230.
Han, L.T., & Epstein, R.A. (2025). Distinct mechanisms for panoramic and landmark-based view integration in human scene-selective cortex. BioRxiv : the preprint server for biology. https://pubmed.ncbi.nlm.nih.gov/39896651/.
Hanke, M., Labs, A., Reich, T., Schulenburg, H., Boennen, M., Mareike, G., Golz, M., Hartigs, B., Hoffmann, N., Keil, S., Perlow, M., Peukmann, A.K., Rabe, L.N., von Sobbe, F.R., 2015. Portrayed emotions in the movie "Forrest Gump.". F1000Resarch 4. https://doi.org/10.12688/f1000research.6230.1.
Haq, I.U., Muhammad, K., Hussain, T., Kwon, S., Sodanil, M., Baik, S.W., Lee, M.Y., 2019. Movie scene segmentation using object detection and set theory. Int. J. Distrib. Sens. Netw. 15 (6), 0–7. https://doi.org/10.1177/1550147719845277.
Hasson, U., Chen, J., Honey, C.J., 2015. Hierarchical process memory: memory as an integral component of information processing. Trends Cogn. Sci. 19 (6), 304–313. https://doi.org/10.1016/j.tics.2015.04.006.
Hasson, U., Furman, O., Clark, D., Dudai, Y., Davachi, L., 2008a. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. Neuron 57 (3), 452–462. https://doi.org/10.1016/j.neuron.2007.12.009.
Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., Heeger, D.J., 2008b. Neurocinematics: the neuroscience of film. Projections 2 (1), 1–26. https://doi.org/10.3167/proj.2008.020102.
Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science (1979) 303 (5664), 1634–1640. https://doi.org/10.1126/science.1089506.
Hasson, U., Nusbaum, H.C., Small, S.L., 2006. Repetition suppression for spoken sentences and the effect of task demands. J. Cogn. Neurosci. 18 (12), 2013–2029. https://doi.org/10.1162/jocn.2006.18.12.2013.
Hasson, U., Yang, E., Vallines, I., Heeger, D.J., Rubin, N., 2008c. A hierarchy of temporal receptive windows in human cortex. J. Neurosci. 28 (10), 2539–2550. https://doi.org/10.1523/JNEUROSCI.5487-07.2008.
Häusler, C.O., Eickhoff, S.B., Hanke, M., 2022. Processing of visual and non-visual naturalistic spatial information in the "parahippocampal place area. Sci. Data 9 (1), 1–15. https://doi.org/10.1038/s41597-022-01250-4.
Häusler, C.O., Hanke, M., 2016. An annotation of cuts, depicted locations, and temporal progression in the motion picture "Forrest Gump. F1000Resarch 5, 1–9. https://doi.org/10.12688/F1000RESEARCH.9536.1.
Häusler, C.O., Hanke, M., & Waite, A. (2021). *Annotations of low-level perceptual confounds in the research cut of the audio-visual movie "Forrest Gump" and its audio-description*. https://github.com/psychoinformatics-de/studyforrest-data-confoundsannotation.
Henderson, J.M., Hollingworth, A., 1999. High-level scene perception. Annu. Rev. Psychol. 50 (1), 243–271. https://doi.org/10.1146/annurev.psych.50.1.243.
Herrington, J.D., Nymberg, C., Schultz, R.T., 2011. Biological motion task performance predicts superior temporal sulcus activity. Brain Cogn. 77 (3), 372–381. https://doi.org/10.1016/j.bandc.2011.09.001.
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.
Ito, T., Hearne, L.J., Cole, M.W., 2020. A cortical hierarchy of localized and distributed processes revealed via dissociation of task activations, connectivity changes, and intrinsic timescales. Neuroimage 221 (February), 117141. https://doi.org/10.1016/j.neuroimage.2020.117141.
Jääskeläinen, I.P., Sams, M., Glerean, E., Ahveninen, J., 2021. Movies and narratives as naturalistic stimuli in neuroimaging. Neuroimage 224, 117445. https://doi.org/10.1016/j.neuroimage.2020.117445. June 2020.
Jain, M., Katyal, N., & Alok. (2021). *Katna: tool for automating video keyframe extraction, video compression, image autocrop and smart image resize tasks*. https://github.com/keplerlab/katna.
Janke, S., & Glöckner-Rist, A. (2014). Deutsche Version der positive and negative affect schedule (PANAS). Zusammenstellung Sozialwissenschaftlicher Items Und Skalen, 1–15.
Jeon, H.A., 2014. Hierarchical processing in the prefrontal cortex in a variety of cognitive domains. In: Frontiers in Systems Neuroscience, 8. Frontiers Research Foundation, pp. 1–8. https://doi.org/10.3389/fnsys.2014.00223.
Ji, H., Hooshyar, D., Kim, K., Lim, H., 2019. A semantic-based video scene segmentation using a deep neural network. J. Inf. Sci. 45 (6), 833–844. https://doi.org/10.1177/0165551518819964.
Julian, J.B., Keinath, A.T., Marchette, S.A., Epstein, R.A., 2018. The neurocognitive basis of spatial reorientation. Curr. Biol. 28 (17), R1059–R1073. https://doi.org/10.1016/j.cub.2018.04.057.
Kamps, F.S., Lall, V., Dilks, D.D., 2016. The occipital place area represents first-person perspective motion information through scenes. Cortex 83, 17–26. https://doi.org/10.1016/j.cortex.2016.06.022.
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. Front. Neuroinform. 4 (MAR), 1–10. https://doi.org/10.3389/fninf.2010.00005.
Kauppi, J.P., Pajula, J., Tohka, J., 2014. A versatile software package for inter-subject correlation based analyses of fMRI. Front. Neuroinform. 8 (JAN), 1–13. https://doi.org/10.3389/fninf.2014.00002.
Knejzlíková, T., Světlák, M., Malatincová, T., Roman, R., Chládek, J., Najmanová, J., Theiner, P., Linhartová, P., Kašpárek, T., 2021. Electrodermal response to mirror exposure in relation to subjective emotional responses, emotional competences and affectivity in adolescent girls with restrictive anorexia and healthy controls. Front. Psychol. (September), 12. https://doi.org/10.3389/fpsyg.2021.673597.
Kwok, S.C., Shallice, T., Macaluso, E., 2012. Functional anatomy of temporal organisation and domain-specificity of episodic memory retrieval. Neuropsychologia 50 (12), 2943–2955. https://doi.org/10.1016/j.neuropsychologia.2012.07.025.
Lahnakoski, J.M., Jääskeläinen, I.P., Sams, M., Nummenmaa, L., 2017. Neural mechanisms for integrating consecutive and interleaved natural events. Hum. Brain Mapp. 38 (7), 3360–3376. https://doi.org/10.1002/hbm.23591.
Larsson, J., Smith, A.T., 2012. FMRI repetition suppression: neuronal adaptation or stimulus expectation? Cereb. Cortex 22 (3), 567–576. https://doi.org/10.1093/cercor/bhr119.
Lee Masson, H., Isik, L., 2021. Functional selectivity for social interaction perception in the human superior temporal sulcus during natural viewing. Neuroimage 245 (November), 118741. https://doi.org/10.1016/j.neuroimage.2021.118741.
Lee, Y., Chen, J., 2024. The relationship between event boundary strength and pattern shifts across the cortical hierarchy during naturalistic movie-viewing. J. Cogn. Neurosci. 36 (11), 2317–2342. https://doi.org/10.1162/jocn_a_02213.
Lerner, Y., Honey, C.J., Katkov, M., Hasson, U., 2014. Temporal scaling of neural responses to compressed and dilated natural speech. J. Neurophysiol. 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. J. Neurosci. 31 (8), 2906–2915. https://doi.org/10.1523/JNEUROSCI.3684-10.2011.
Loschky, L.C., Larson, A.M., Smith, T.J., Magliano, J.P., 2020. The scene perception & event comprehension theory (SPECT) applied to visual narratives. Top. Cogn. Sci. 12 (1), 311–351. https://doi.org/10.1111/tops.12455.
Lu, K.H., Hung, S.C., Wen, H., Marussich, L., Liu, Z., 2016. Influences of high-level features, gaze, and scene transitions on the reliability of BOLD responses to natural movie stimuli. PLoS One 11 (8), 1–19. https://doi.org/10.1371/journal.pone.0161797.
Ludwigs, M., 2020. Hierarchical thinking, grammatical structures, and the originary scene. J. Gener. Anthropol. 2, 1–15. https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=10837264&AN=143353825&h=Xky38kxzpIfQlKRssQNpxCfWZA4Fq35O1ktR4GhPFwSwk3vRAPCv%2FDBAMzPOBMJRGrLQWFtxcjKRbfM%2FgsppzQ%3D%3D&crl=c.
Magliano, J.P., Zacks, J.M., 2011. The impact of continuity editing in narrative film on event segmentation. Cogn. Sci. 35 (8), 1489–1517. https://doi.org/10.1111/j.1551-6709.2011.01202.x.
Makris, N., Goldstein, J.M., Kennedy, D., Hodge, S.M., Caviness, V.S., Faraone, S.V., Tsuang, M.T., Seidman, L.J., 2006. Decreased volume of left and total anterior insular lobule in schizophrenia. Schizophr. Res. 83 (2–3), 155–171. https://doi.org/10.1016/j.schres.2005.11.020.
Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., Jefferies, E., Smallwood, J., 2016. Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc. Natl. Acad. Sci. U.S.A. 113 (44), 12574–12579. https://doi.org/10.1073/pnas.1608282113.
Markhabayeva, D., Tseng, C.I., 2024. Multimodal cohesion and viewers' comprehension of scene transitions in film: an empirical investigation. Front. Commun. 9. https://doi.org/10.3389/fcomm.2024.1347788.
Mas, J., & Fernandez, G. (2003). Video shot boundary detection based on color histogram. *TRECVID 2003: notebook papers*. https://www-nlpir.nist.gov/projects/tvpubs/tvpapers03/ramonlull.paper.pdf.
McAleer, P., Pollick, F.E., Love, S.A., Crabbe, F., Zacks, J.M., 2014. The role of kinematics in cortical regions for continuous human motion perception. Cogn. Affect. Behav. Neurosci. 14 (1), 307–318. https://doi.org/10.3758/s13415-013-0192-4.
Mecklenbrauck, F., Gruber, M., Siestrup, S., Zahedi, A., Grotegerd, D., Mauritz, M., Trempler, I., Dannlowski, U., Schubotz, R.I., 2024a. The significance of structural rich club hubs for the processing of hierarchical stimuli. Hum. Brain Mapp. 45 (4), 1–28. https://doi.org/10.1002/hbm.26543.
Mecklenbrauck, F., Sepulcre, J., Fehring, J., Schubotz, R.I., 2024b. Decoding cortical chronotopy—Comparing the influence of different cortical organizational schemes. Neuroimage 303 (October), 120914. https://doi.org/10.1016/j.neuroimage.2024.120914.
Menon, V., 2023. 20 years of the default mode network: a review and synthesis. Neuron 111 (16), 2469–2487. https://doi.org/10.1016/j.neuron.2023.04.023.
Nasr, S., Devaney, K.J., Tootell, R.B.H., 2013. Spatial encoding and underlying circuitry in scene-selective cortex. Neuroimage 83, 892–900. https://doi.org/10.1016/j.neuroimage.2013.07.030.

Nasr, S., Echavarria, C.E., Tootell, R.B.H., 2014. Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. J. Neurosci. 34 (20), 6721–6735. https://doi.org/10.1523/JNEUROSCI.4802-13.2014.

Nastase, S.A., Gazzola, V., Hasson, U., Keysers, C., 2019. Measuring shared responses across subjects using intersubject correlation. Soc. Cogn. Affect. Neurosci. 14 (6), 669–687. https://doi.org/10.1093/scan/nsz037.

Nastase, S.A., Hasson, U., 2024. Confound regression models for intersubject correlation analysis with naturalistic stimuli twenty confound regression models were constructed from standard head movie-viewing results [Poster Presentation]. In: Annual Meeting of the Organization for Human Brain Mapping. Seoul, South Korea. https://github.com/snastase/isc-confounds.

Nunez-Elizalde, A.O., Huth, A.G., Gallant, J.L., 2019. Voxelwise encoding models with non-spherical multivariate normal priors. Neuroimage 197, 482–492. https://doi.org/10.1016/j.neuroimage.2019.04.012. August 2018.

Oetringer, D., Gözükara, D., Güçlü, U., Geerligs, L., 2025. The neural basis of event segmentation: stable features in the environment are reflected by neural states. Imaging Neurosci. 3 (February). https://doi.org/10.1162/imag_a_00432.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh Inventory. Neuropsychologia 9, 97–113. https://doi.org/10.1007/978-3-319-55065-7_301010.

Oren, N., Shapira-Lichter, I., Lerner, Y., Tarrasch, R., Hendler, T., Giladi, N., Ash, E.L., 2016. How attention modulates encoding of dynamic stimuli. Front. Hum. Neurosci. 10 (OCT2016), 1–11. https://doi.org/10.3389/fnhum.2016.00507.

Pajula, J., Kauppi, J.P., Tohka, J., 2012. Inter-subject correlation in fMRI: method validation against stimulus-model based analysis. PLoS One 7 (8). https://doi.org/10.1371/journal.pone.0041196.

Pang, J.C., Aquino, K.M., Oldehinkel, M., Robinson, P.A., Fulcher, B.D., Breakspear, M., Fornito, A., 2023. Geometric constraints on human brain function. Nature 618 (7965), 566–574. https://doi.org/10.1038/s41586-023-06098-1.

Patriat, R., Molloy, E.K., Meier, T.B., Kirk, G.R., Nair, V.A., Meyerand, M.E., Prabhakaran, V., Birn, R.M., 2013. The effect of resting condition on resting-state fMRI reliability and consistency: a comparison between resting with eyes open, closed, and fixated. Neuroimage 78, 463–473. https://doi.org/10.1016/j.neuroimage.2013.04.013.

Pearlman, K., 2019. On Rhythm in Film Editing. In: Carroll, N., Di Summa, L.T., Loht, S. (Eds.), The Palgrave Handbook of the Philosophy of Film and Motion Pictures. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-19601-1_7.

Peer, M., Hayman, M., Tamir, B., Arzy, S., 2021. Brain coding of social network structure. J. Neurosci. 41 (22), 4897–4909. https://doi.org/10.1523/JNEUROSCI.2641-20.2021.

Pitcher, D., Ungerleider, L.G., 2021. Evidence for a third visual pathway specialized for social perception. Trends Cogn. Sci. 25 (2), 100–110. https://doi.org/10.1016/j.tics.2020.11.006.

Poldrack, R.A., Mumford, J.A., Nichols, T.E., 2011. Handbook of functional MRI Data analysis. A Practical Approach to Medical Image Processing. Cambridge University Press. https://doi.org/10.1017/CBO9780511895029.

Posit team. (2024). *RStudio: integrated development environment for R*. http://www.posit.co/.

Rah, Y.J., Shin, J.H., Lee, S.A., 2025. The visual scene network and hippocampus represent spatial boundary structures for temporal episodic memory organization. J. Cogn. Neurosci. 1–16. https://doi.org/10.1162/jocn_a_02229.

Raichle, M.E., 2015. The brain's default mode network. Annu. Rev. Neurosci. 38, 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030.

Rajimehr, R., Devaney, K.J., Bilenko, N.Y., Young, J.C., Tootell, R.B.H, 2011. The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. PLoS Biol. 9 (4). https://doi.org/10.1371/journal.pbio.1000608.

Rajimehr, R., Xu, H., Farahani, A., Kornblith, S., Duncan, J., Desimone, R., Rajimehr, R., Xu, H., Farahani, A., Kornblith, S., Duncan, J., Desimone, R., 2024. Functional architecture of cerebral cortex during naturalistic movie watching. Neuron 112 (24), 1–17. https://doi.org/10.1016/j.neuron.2024.10.005.

Raut, R.V., Snyder, A.Z., Raichle, M.E., 2020. Hierarchical dynamics as a macroscopic organizing principle of the human brain. Proc. Natl. Acad. Sci. 117 (34), 20890–20897. https://doi.org/10.1073/pnas.2003383117.

Riedel, L., van den Heuvel, M.P., Markett, S., 2022. Trajectory of rich club properties in structural brain networks. Hum. Brain Mapp. 43 (14), 4239–4253. https://doi.org/10.1002/hbm.25950.

Ritchey, M., Cooper, R.A., 2020. Deconstructing the posterior medial episodic network. Trends Cogn. Sci. 24 (6), 451–465. https://doi.org/10.1016/j.tics.2020.03.006.

Rockland, K.S., Graves, W.W., 2023. The angular gyrus: a special issue on its complex anatomy and function. Brain Struct. Funct. 228 (1), 1–5. https://doi.org/10.1007/s00429-022-02596-6.

Rubin, G.S., Turano, K., 1992. Reading without saccadic eye movements. Vis. Res. 32 (5), 895–902. https://doi.org/10.1016/0042-6989(92)90032-E.

Rugg, M.D., Vilberg, K.L., 2013. Brain networks underlying episodic memory retrieval. Curr. Opin. Neurobiol. 23 (2), 255–260. https://doi.org/10.1016/j.conb.2012.11.005.

Rui, Y., Huang, T.S., Mehrotra, S., 1998. Exploring video structure beyond the shots. In: Proceedings of the IEEE Conference on Protocols for Multimedia Systems and Multimedia Networking, PROMS-MmNet, pp. 237–240. https://doi.org/10.1109/mmcs.1998.693648.

Sava-Segal, C., Richards, C., Leung, M., Finn, E.S., 2023. Individual differences in neural event segmentation of continuous experiences. Cereb. Cortex 33 (13), 8164–8178. https://doi.org/10.1093/cercor/bhad106.

Schubotz, R.I., Von Cramon, D.Y., 2002. Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. Neuroimage 15 (4), 787–796. https://doi.org/10.1006/nimg.2001.1043.

Scrivener, C.L., Teed, J.A., Silson, E.H., 2025. Visual imagery of familiar people and places in category selective cortex. Neurosci. Conscious (1), 2025. https://doi.org/10.1093/nc/niaf006.

Seghier, M.L., 2023. Multiple functions of the angular gyrus at high temporal resolution. Brain Struct. Funct. 228 (1), 7–46. https://doi.org/10.1007/s00429-022-02512-y.

Seidemann, E., Geisler, W.S., 2018. Linking V1 activity to behavior. Annu. Rev. Vis. Sci. 4, 287–310. https://doi.org/10.1146/annurev-vision-102016-061324.

Sepulcre, J., Sabuncu, M.R., Yeo, T.B., Liu, H., Johnson, K.A., 2012. Stepwise connectivity of the modal cortex reveals the multimodal organization of the human brain. J. Neurosci. 32 (31), 10649–10661. https://doi.org/10.1523/JNEUROSCI.0759-12.2012.

Shahdloo, M., Acar, M., & Çukur, T. (2019). *Attention during story listening modulates temporal receptive windows across human cortex*. 168–171. 10.32470/ccn.2019.1039-0.

Shirer, W.R., Ryali, S., Rykhlevskaia, E., Menon, V., Greicius, M.D., 2012. Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb. Cortex 22 (1), 158–165. https://doi.org/10.1093/cercor/bhr099.

Silson, E.H., Gilmore, A.W., Kalinowski, S.E., Steel, A., Kidder, A., Martin, A., Baker, C.I., 2019. A posterior-anterior distinction between scene perception and scene construction in human medial parietal cortex. J. Neurosci. 39 (4), 705–717. https://doi.org/10.1523/JNEUROSCI.1219-18.2018.

Silson, E.H., Steel, A.D., Baker, C.I., 2016. Scene-selectivity and retinotopy in medial parietal cortex. Front. Hum. Neurosci. 10 (August), 17. https://doi.org/10.3389/fnhum.2016.00412.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., de Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., de Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 (SUPPL. 1), 208–219. https://doi.org/10.1016/j.neuroimage.2004.07.051.

Smith, T.J., Levin, D., Cutting, J.E., 2012. A window on reality: perceiving edited moving images. Curr. Dir. Psychol. Sci. 21 (2), 107–113. https://doi.org/10.1177/0963721412437407.

Soltani, A., Murray, J.D., Seo, H., Lee, D., 2021. Timescales of cognition in the brain. Curr. Opin. Behav. Sci. 41, 30–37. https://doi.org/10.1016/j.cobeha.2021.03.003.

Song, M., Shin, E.J., Seo, H., Soltani, A., Steinmetz, N.A., Lee, D., Jung, M.W., Paik, S.-B., 2024. Hierarchical gradients of multiple timescales in the mammalian forebrain. Proc. Natl. Acad. Sci. 121 (51), 2017. https://doi.org/10.1073/pnas.2415695121.

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

Srokova, S., Hill, P.F., Rugg, M.D., 2022. The retrieval-related anterior shift is moderated by age and correlates with memory performance. J. Neurosci. 42 (9), 1765–1776. https://doi.org/10.1523/JNEUROSCI.1763-21.2021.

Steel, A., Billings, M.M., Silson, E.H., Robertson, C.E., 2021. A network linking scene perception and spatial memory systems in posterior cerebral cortex. Nat. Commun. 12 (1). https://doi.org/10.1038/s41467-021-22848-z.

Steel, A., Garcia, B.D., Goyal, K., Mynick, A., Robertson, C.E., 2023. Scene perception and visuospatial memory converge at the anterior edge of visually responsive cortex. J. Neurosci. 43 (31), 5723–5737. https://doi.org/10.1523/JNEUROSCI.2043-22.2023.

Steel, A., Prasad, D., Garcia, B.D., Robertson, C.E., 2025. Relating scene memory and perception activity to functional properties, networks, and landmarks of posterior cerebral cortex—A probabilistic atlas. The J. Neurosci. 45 (23), e0028252025. https://doi.org/10.1523/JNEUROSCI.0028-25.2025.

Steel, A., Silson, E.H., Garcia, B.D., Robertson, C.E., 2024. A retinotopic code structures the interaction between perception and memory systems. Nat. Neurosci. 27 (2), 339–347. https://doi.org/10.1038/s41593-023-01512-3.

Sulpizio, V., Galati, G., Fattori, P., Galletti, C., Pitzalis, S., 2020. A common neural substrate for processing scenes and egomotion-compatible visual motion. Brain Struct. Funct. 225 (7), 2091–2110. https://doi.org/10.1007/s00429-020-02112-8.

Taylor, P.A., Reynolds, R.C., Calhoun, V., Gonzalez-Castillo, J., Handwerker, D.A., Bandettini, P.A., Mejia, A.F., Chen, G., 2023. Highlight results, don't hide them: enhance interpretation, reduce biases and improve reproducibility. Neuroimage 274, 120138. https://doi.org/10.1016/j.neuroimage.2023.120138. October 2022.

The MathWorks Inc, 2021. MATLAB Version: 9.10.0 (R2021a). The MathWorks Inc. https://www.mathworks.com.

Thorenz, K., Berwinkel, A., Weigelt, M., 2023. A validation study for the German versions of the feeling scale and the felt arousal scale for a progressive muscle relaxation exercise. Behav. Sci. (7), 13. https://doi.org/10.3390/bs13070523.

Tomar, S., 2006. Converting video formats with FFmpeg. Linux J. 2006 (146), 10.

Tong, F., 2003. Cognitive neuroscience: primary visual cortex and visual awareness. Nat. Rev. Neurosci. 4 (3), 219–229. https://doi.org/10.1038/nrn1055.

Torres-Morales, C., Cansino, S., 2024. Brain representations of space and time in episodic memory: a systematic review and meta-analysis. Cogn. Affect. Behav. Neurosci. 24 (1), 1–18. https://doi.org/10.3758/s13415-023-01140-1.

Tylén, K., Christensen, P., Roepstorff, A., Lund, T., Østergaard, S., Donald, M., 2015. Brains striving for coherence: long-term cumulative plot formation in the default mode network. Neuroimage 121, 106–114. https://doi.org/10.1016/j.neuroimage.2015.07.047.

Uithol, S., van Rooij, I., Bekkering, H., Haselager, P., 2012. Hierarchies in action and motor control. J. Cogn. Neurosci. 24 (5), 1077–1086. https://doi.org/10.1162/jocn_a_00204.

Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., Dehaene, S., 2012. A temporal bottleneck in the language comprehension network. J. Neurosci. 32 (26), 9089–9102. https://doi.org/10.1523/JNEUROSCI.5685-11.2012.

Valentine, J.W., May, C.L., 1996. Hierarchies in biology and paleontology. Paleobiology 22 (1), 23–33. https://doi.org/10.1017/S0094837300015992.

Van den Stock, J., Vandenbulcke, M., Sinke, C.B.A., Goebel, R., De Gelder, B., 2014. How affective information from faces and scenes interacts in the brain. Soc. Cogn. Affect. Neurosci. 9 (10), 1481–1488. https://doi.org/10.1093/scan/nst138.

Van Dijk, K.R.A., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., Buckner, R.L., 2010. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J. Neurophysiol. 103 (1), 297–321. https://doi.org/10.1152/jn.00783.2009.

Vann, S.D., Aggleton, J.P., Maguire, E.A., 2009. What does the retrosplenial cortex do? Nat. Rev. Neurosci. 10 (11), 792–802. https://doi.org/10.1038/nrn2733.

VanRullen, R., Koch, C., 2003. Is perception discrete or continuous? Trends Cogn. Sci. 7 (5), 207–213. https://doi.org/10.1016/S1364-6613(03)00095-0.

Wade, A.R., Brewer, A.A., Rieger, J.W., Wandell, B.A., 2002. Functional measurements of human ventral occipital cortex: retinotopy and colour. Philos. Trans. R. Soc. B: Biol. Sci. 357 (1424), 963–973. https://doi.org/10.1098/rstb.2002.1108.

Wakita, M., 2014. Broca's area processes the hierarchical organization of observed action. Front. Hum. Neurosci. 7 (JAN). https://doi.org/10.3389/fnhum.2013.00937.

Wang, L., Mruczek, R.E.B., Arcaro, M.J., Kastner, S., 2015. Probabilistic maps of visual topography in human cortex. Cereb. Cortex 25 (10), 3911–3931. https://doi.org/10.1093/cercor/bhu277.

Wang, X.-J., 2020. Macroscopic gradients of synaptic excitation and inhibition in the neocortex. Nat. Rev. Neurosci. 21 (3), 169–178. https://doi.org/10.1038/s41583-020-0262-x.

Watson, D., Clark, L.A., Tellegen, A., 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers. Soc. Psychol. 54 (6), 1063–1070.

Watson, D.M., Andrews, T.J., 2024. Mapping the functional and structural connectivity of the scene network. Hum. Brain Mapp. 45 (3), 1–21. https://doi.org/10.1002/hbm.26628.

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.

Wilcox, L.M., Allison, R.S., Helliker, J., Dunk, B., Anthony, R.C., 2015. Evidence that viewers prefer higher frame-rate film. ACM Trans. Appl. Percept. 12 (4), 1–12. https://doi.org/10.1145/2810039.

Winawer, J., Witthoft, N., 2015. Human V4 and ventral occipital retinotopic maps. Vis. Neurosci. 32. https://doi.org/10.1017/S0952523815000176.

Wolff, A., Berberian, N., Golesorkhi, M., Gomez-Pilar, J., Zilio, F., Northoff, G., 2022. Intrinsic neural timescales: temporal integration and segregation. Trends Cogn. Sci. 26 (2), 159–173. https://doi.org/10.1016/j.tics.2021.11.007.

Wu, J., 2013. Hierarchy theory: an overview. In: Rozzi, R., Pickett, S.T.A., Palmer, C., Armesto, J.J., Callicott, J.B. (Eds.), Linking Ecology and Ethics for a Changing World: Values, Philosophy, and Action. Springer, Netherlands, pp. 281–301. https://doi.org/10.1007/978-94-007-7470-4_24.

Xia, M., Wang, J., He, Y., 2013. BrainNet Viewer: a network visualization tool for Human Brain connectomics. PLoS One 8 (7). https://doi.org/10.1371/journal.pone.0068910.

Yang, E., Milisav, F., Kopal, J., Holmes, A.J., Mitsis, G.D., Misic, B., Finn, E.S., Bzdok, D., 2023. The default network dominates neural responses to evolving movie stories. Nat. Commun. 14 (1). https://doi.org/10.1038/s41467-023-39862-y.

Yang, H., Susilo, T., Duchaine, B., 2016. The anterior temporal face area contains invariant representations of face identity that can persist despite the loss of right FFA and OFA. Cereb. Cortex 26 (3), 1096–1107. https://doi.org/10.1093/cercor/bhu289.

Yeh, M.C., Tsai, Y.W., Hsu, H.C., 2016. A content-based approach for detecting highlights in action movies. Multimed. Syst. 22 (3), 287–295. https://doi.org/10.1007/s00530-015-0457-6.

Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., Fisch, B., Liu, H., Buckner, R. L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106 (3), 1125–1165. https://doi.org/10.1152/jn.00338.2011.

Yeshurun, Y., Nguyen, M., Hasson, U., 2021. The default mode network: where the idiosyncratic self meets the shared social world. Nat. Rev. Neurosci. 22 (3), 181–192. https://doi.org/10.1038/s41583-020-00420-w.

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C.J., Hasson, U., 2017. Same story, different story: the neural representation of interpretive frameworks. Psychol. Sci. 28 (3), 307–319. https://doi.org/10.1177/0956797616682029.

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. Nat. Neurosci. 4 (6), 651–655. http://neurosci.nature.com.

Zacks, J.M., Speer, N.K., Reynolds, J.R., 2009. Segmentation in reading and film comprehension. J. Exp. Psychol.: Gen. 138 (2), 307–327. https://doi.org/10.1037/a0015305.

Zacks, J.M., Speer, N.K., Swallow, K.M., Maley, C.J., 2010. The brain's cutting-room floor: segmentation of narrative cinema. Front. Hum. Neurosci. 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. Cereb. Cortex 27 (10), 4988–5000. https://doi.org/10.1093/cercor/bhx202.

Zemeckis, R., 1994. *Forrest Gump* [Video recording]. Paramount Pictures.

Zhen, Z., Yang, Z., Huang, L., Kong, X., Wang, X., Dang, X., Huang, Y., Song, Y., Liu, J., 2015. Quantifying interindividual variability and asymmetry of face-selective regions: a probabilistic functional atlas. Neuroimage 113, 13–25. https://doi.org/10.1016/j.neuroimage.2015.03.010.

Zhu, W., Huang, Y., Xie, X., Liu, W., Deng, J., Zhang, D., Wang, Z., Liu, J., 2023. AutoShot: a short video dataset and State-of-the-art shot boundary detection. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2023-June*, pp. 2238–2247. https://doi.org/10.1109/CVPRW59228.2023.00218.