Semantic novelty modulates neural responses to visual change across the human brain

Maximilian Nentwich[1], Marcin Leszczynski[2,3,4], Brian E. Russ[3,5,6], Lukas Hirsch[1], Noah Markowitz[7], Kaustubh Sapru[1], Charles E. Schroeder[2,3], Ashesh Mehta[7,8], Stephan Bickel[3,7,8], Lucas C Parra[1]

[1] Department of Biomedical Engineering, The City College of New York, New York, USA;
[2] Departments of Psychiatry and Neurology, Columbia University College of Physicians and Surgeons, New York, USA;
[3] Translational Neuroscience Lab Division, Center for Biomedical Imaging and Neuromodulation, Nathan Kline Institute, Orangeburg, USA;
[4] Cognitive Science Department, Institute of Philosophy, Jagiellonian University, Kraków, Poland;
[5] Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine, New York, USA;
[6] Department of Psychiatry, New York University at Langone, New York, USA;
[7] The Feinstein Institutes for Medical Research, Northwell Health, Manhasset, USA;
[8] Departments of Neurosurgery and Neurology, Zucker School of Medicine at Hofstra/Northwell, Manhasset, USA

Abstract

Our continuous visual experience in daily life is dominated by change. Previous research has focused on visual change due to stimulus motion, eye movements or unfolding events, but not their combined impact across the brain, or their interactions with semantic novelty. We investigate the neural responses to these sources of novelty during film viewing. We analyzed intracranial recordings in humans across 6328 electrodes from 23 individuals. Responses associated with saccades and film cuts were dominant across the entire brain. Film cuts at semantic event boundaries were particularly effective in the temporal and medial temporal lobe. Saccades to visual targets with high visual novelty were also associated with strong neural responses. Specific locations in higher-order association areas showed selectivity to either high or low-novelty saccades. We conclude that neural activity associated with film cuts and eye movements is widespread across the brain and is modulated by semantic novelty.

Introduction

To study the neural processing of natural visual stimuli, recent work has focused on the experience of watching movies (*1–4*). Movies offer a balance between experimental control and the realisms of natural environments (*1*). Visual dynamics during movie watching are dominated by motion of objects in the scene, the viewer's own eye movements, and film cuts introduced by the film editor to change view angle, but also to transition between events.

Change to the visual input due to stimulus motion in movies is associated with strong neural responses, widespread across the occipital, parietal and temporal lobes (*3*, *4*). It is a more powerful driver of neural responses than low-level features such as luminance and contrast (*3*, *4*). Change to the visual input to the retina is also caused by saccades, the rapid eye movements between fixations, which many studies consider a confound (*5*, *6*). Neural Responses to saccades were thought to be confined to visual processing areas and largely suppressed in higher-order association areas (*7*). More recent work, however, indicates that saccades play an important role in the organization of many perceptual and cognitive processes (*8–11*). Saccades modulate neural responses across the visual system (*8*, *12*, *13*), medial temporal lobe (*14*, *15*), non-visual nuclei of the anterior thalamus (*9*) and even the auditory cortex (*16*). Complicating matters, stimulus motion in movies can also attract saccades (*17*), and a novel complex visual scene transiently increases saccade rate (*18*). In addition, responses to stimulus motion and to saccades are modulated by the semantics of the visual stimulus. For instance, responses to stimulus motion are particularly pronounced when they are associated with 'social' stimuli (*4*), and saccade-locked potentials in the medial temporal lobe (MTL) can be specific to the target of a saccade, such as faces or objects (*10*).

Besides stimulus motion and saccades, movies also allow us to investigate the processing of narratives, which relies on unique mechanisms in the human brain (*19*). The theory of event segmentation proposes that continuous narratives are segmented and remembered as discrete events (*20*). Event boundaries, the moments of change between events, are associated with shifts in brain states as well as transient neural responses (*6*, *21–24*). In movies, event boundaries typically coincide with film cuts. Film cuts between events contain semantic changes and are associated with neural activity in higher-order association areas (*6*, *21*). On the other hand, film cuts that maintain continuity (of space, time and action) are mainly associated with changes in low-level visual areas (*5*, *21*). Consistent with event segmentation theory, increased activation in MTL following a film cut is predictive of subsequent recall of the preceding event (*25*). Similarly, firing rate of a distinct population of cells in the MTL following film cuts between different clips are predictive of recognition memory (*6*).

In summary, the main drivers of visual change -- stimulus motion, film cuts and saccades -- elicit neural activity in various visual and higher-order association areas. However, they have been studied in isolation, so that the relative strength and extent of responses across the brain is not well established. Given that these sources of visual change are correlated it is important to analyze them in combination. Additionally, the effects of event boundaries (*20*, *22*, *23*), social context (*4*) and faces (*15*, *26*) during film suggest a role for novelty at the semantic level. Here,

we hypothesized that neural responses associated with stimulus motion, saccades and scene cuts are modulated by semantic novelty in the visual scene. We test this with intracranial EEG recordings that offer high signal-to-noise ratio (*27*, *28*) and capture fast responses (<1s), which may not be detectable with fMRI that is typically used in functional mapping studies. Using a multiple-regression approach we found that neural responses related to stimulus motion are mostly confined to the low-level visual brain areas, while neural responses related to saccades and film cuts are widespread across the whole brain. We refrain from interpreting these responses as being specific to stimulus motion, cuts or saccades, since a variety of features in the film stimuli could be associated with these regressors. To better control for unobserved factors, we then focused on contrasts between high and low semantic novelty while keeping other factors constant. Here we defined low-level semantic novelty in terms of visual features across saccades using deep-networks (*29*), and on a higher level, in terms of event boundaries judged by human observers across film cuts (*20*, *22*, *23*). We found that responses to saccades and film cuts are modulated by low and high-level semantic novelty, respectively. Importantly, we found that specific locations in higher-order association areas tend to be modulated by either high or low semantic novelty. Higher-order association areas are defined as those that lie higher on the sensory-association hierarchy, as defined by cortical microstructure (*30*). The underlying factors that drive these wide ranging and specific responses related to natural dynamic stimuli remain to be untangled. However, we conclude that semantic novelty is an important factor modulating neural responses to film.

Results

Patients (N=23, Table S1) were implanted with intracranial electrodes for seizure onset localization totaling 6328 contacts with a wide coverage across the whole brain (Figure 1A). Intracranial electroencephalography (iEEG) was recorded simultaneously with eye movements while patients watched video clips totaling 43.6 minutes (Figure 1B). We were interested in neural responses related to the main sources of visual change in movies, motion in the video stimulus, film cuts and eye movements (Figure 1C). Stimulus motion was quantified as the magnitude of optical flow averaged across the whole screen (Figure 1C), and captures motion of objects or the whole scene. It is subsequently referred to as motion. Saccades and film cuts were quantified as a series of impulses at the onset of saccades or film cuts, respectively (Figure 1C). We observed that film cuts are followed by a significant decrease of saccade frequency after 100 ms and a rebound at 250ms (Figure 1D). In addition, motion increases prior to a saccade (Figure 1E), suggesting that both film cuts and motion drive saccades. Motion also decreases before film cuts (Figure 1F), as a result of the video editing process (Figure S1). In total, motion, saccades and cuts clearly interact: saccade frequency increases after film cuts, an increase in motion attracts saccades, and motion slows down before film cuts.

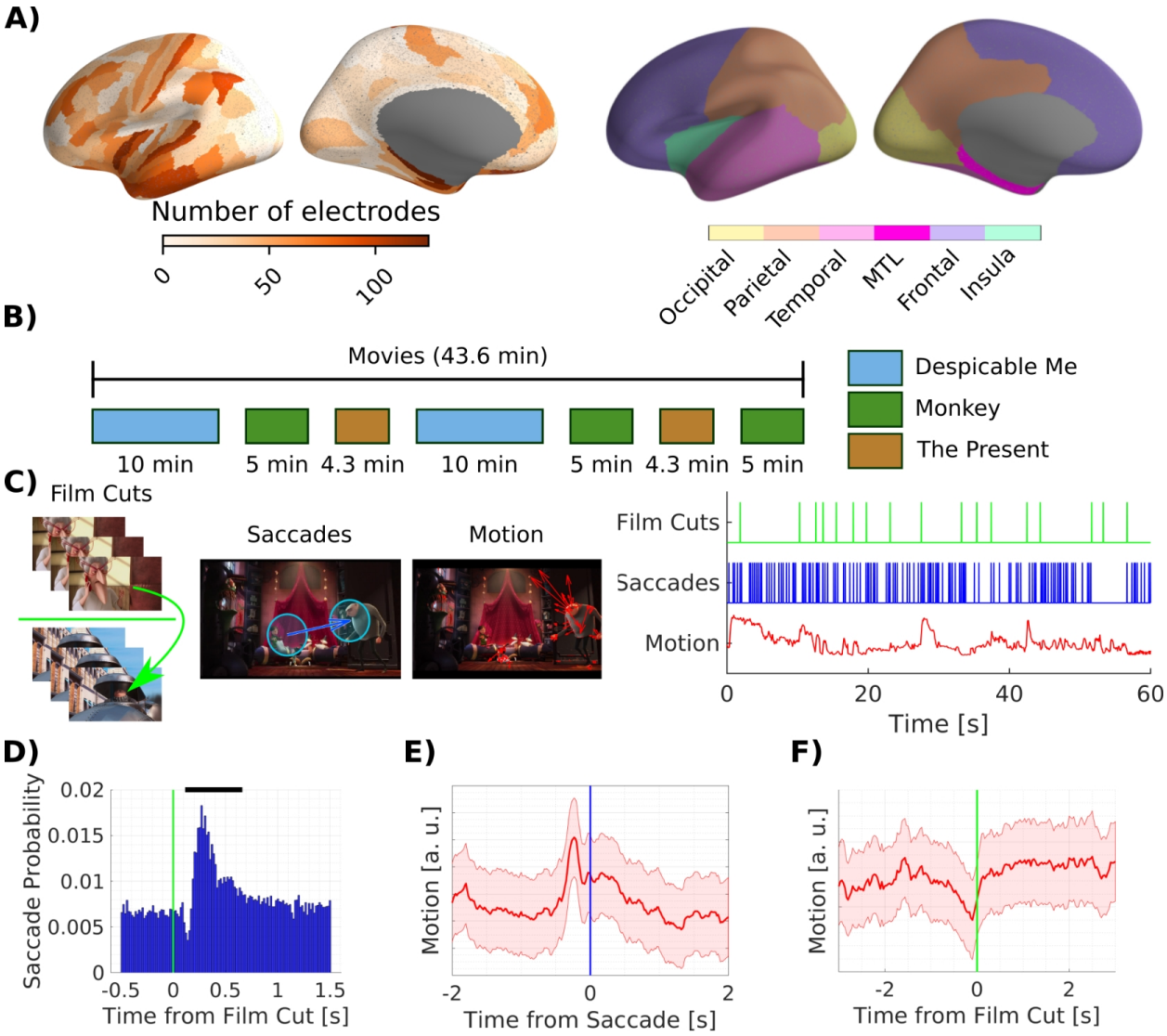

Figure 1: Electrodes coverage, film stimuli and regressors quantifying sources of visual change. A) We collected eye movement and simultaneous intracranial EEG data from 6328 contacts across 23 patients. We analyze neural data in 5378 bipolar channel pairs, referred to as channels. We localize channels to parcels of the Glasser cortical surface atlas (*30*, *31*). Electrodes cover the whole brain but are denser in the temporal and sparser in the occipital lobe. Right panel shows the location of cortical lobes with color coding used in subsequent analysis. The medial temporal lobe (MTL) includes the Amygdala, Hippocampus, entorhinal and parahippocampal cortex. Contacts in the entorhinal and parahippocampal cortex are not included in the temporal lobe. Image courtesy of Assoc Prof Frank Gaillard, Radiopaedia.org, rID: 46846 B) Patients watched up to 43.6 min of video clips. Two different 10 min clips of the animated comic 'Despicable Me' (*32*) were presented, one in English, the other Hungarian. 'The Present' is a short, 4.3min, animated movie presented twice. 'Monkey' videos are three distinct clips of short scenes from documentaries on macaques presented without sound (*4*, *33*). C) Sample film cut, with three frames before the cut on top and three frames after the cut on the bottom. Sample saccade on a frame. Blue circles denote the foveal visual field at 5 degree visual angle. Optical flow vectors (red) on a sample frame. The regressors specifying film cuts and saccade are a series of impulses at the time of the cuts and saccade onset, respectively. The regressor for motion captures the total optic flow in between video frames. D) Saccade probability as a function of time from film cuts. Black bar indicates time bins

with a saccade probability significantly different from the mean ($p < 0.001$). Significance has been tested against a surrogate distribution of film cuts at random time points. E) Average motion as a function of saccade onset time. F) Average motion as a function of time from film cuts across all clips. Cuts tend to follow periods of low motion -- an effect mostly driven by 'Despicable Me' (Figure S1). Shaded area depicts the standard error of the mean.

Neural responses associated with stimulus motion, saccades and film cuts

Based on previous literature we hypothesized that visual change relates to strong and widespread neural responses. To test this we analyzed broad-band high-frequency amplitude (BHA, 70-150 Hz), a signal of dendritic origin that is highly correlated with neuronal firing (*34*, *35*). To increase spatial specificity we performed bipolar re-referencing. All further analysis therefore considers 5378 bipolar channel pairs. We will refer to these channel pairs as channels. To capture neuronal responses, we used a conventional systems identification approach (Figure S2A) (*36*). Specifically, BHA in each channel is treated as the output of a linear system, with motion, saccades and cuts as its input (*37*). The resulting impulse responses are often referred to as "temporal response functions" (TRF) and are obtained for each channel separately.

We test statistical significance separately for responses associated with motion, film cuts and saccades so that each electrode is characterized as "responsive" to one, several, or none of these stimuli. In each brain area a subset of channels shows statistically significant responses. For example, of the 928 channels in the parietal lobe (Figure 2A), 238 were responsive to film cuts, 106 to saccades and 65 to motion (FDR corrected with $q<0.05$, see methods). We analyzed all three sources of change simultaneously to account for the correlation between them and disentangle their contributions to the neural responses (*36*, *37*). For saccade and film cuts this is demonstrated in Figure S2B. To map responsive channels to the cortical surface we compute a weighted average of the response amplitude in each parcel of the Glasser atlas (*31*) following a method described by Gao et al. (*30*) (Figure 2B).

TRFs vary widely in terms of amplitude, duration and onset across the brain (Figure 2, S3, S4 & S5). Motion is associated with increased BHA across the brain (Figure 2A&B & S3A), whereas film cuts and saccades are related to differential increases or decreases in activity, especially beyond sensory areas in the occipital lobe (47% of the responses related to film cuts and 18% of responses related to saccades show a decrease of activity, Figure 2A&B, S3A). Saccades are associated with a suppression, in particular, prior and during the saccade (in 18% of responsive channels, Figure 2A, S3A & S5C). The onset of responses associated with film cuts predominantly starts after the cuts, while in the case of motion the onset occurs before the stimulus (Figure S5). The onset of responses associated with saccades varies widely across the brain with onsets before saccades in some, and after in other areas (Figure S5). The long-lasting responses of up to 1s after the saccade are likely associated with processing of the visual stimulus after fixation (*38*), as opposed to correlated of the saccadic motor command, which takes effect before saccade onset (*39*). Consistent with this, we obtain similar responses when using fixation onset as regressor (Figure S6). However, we prefer to use saccades onset

as a marker of visual change, as it is conceptually more closely related to visual change across the retina.

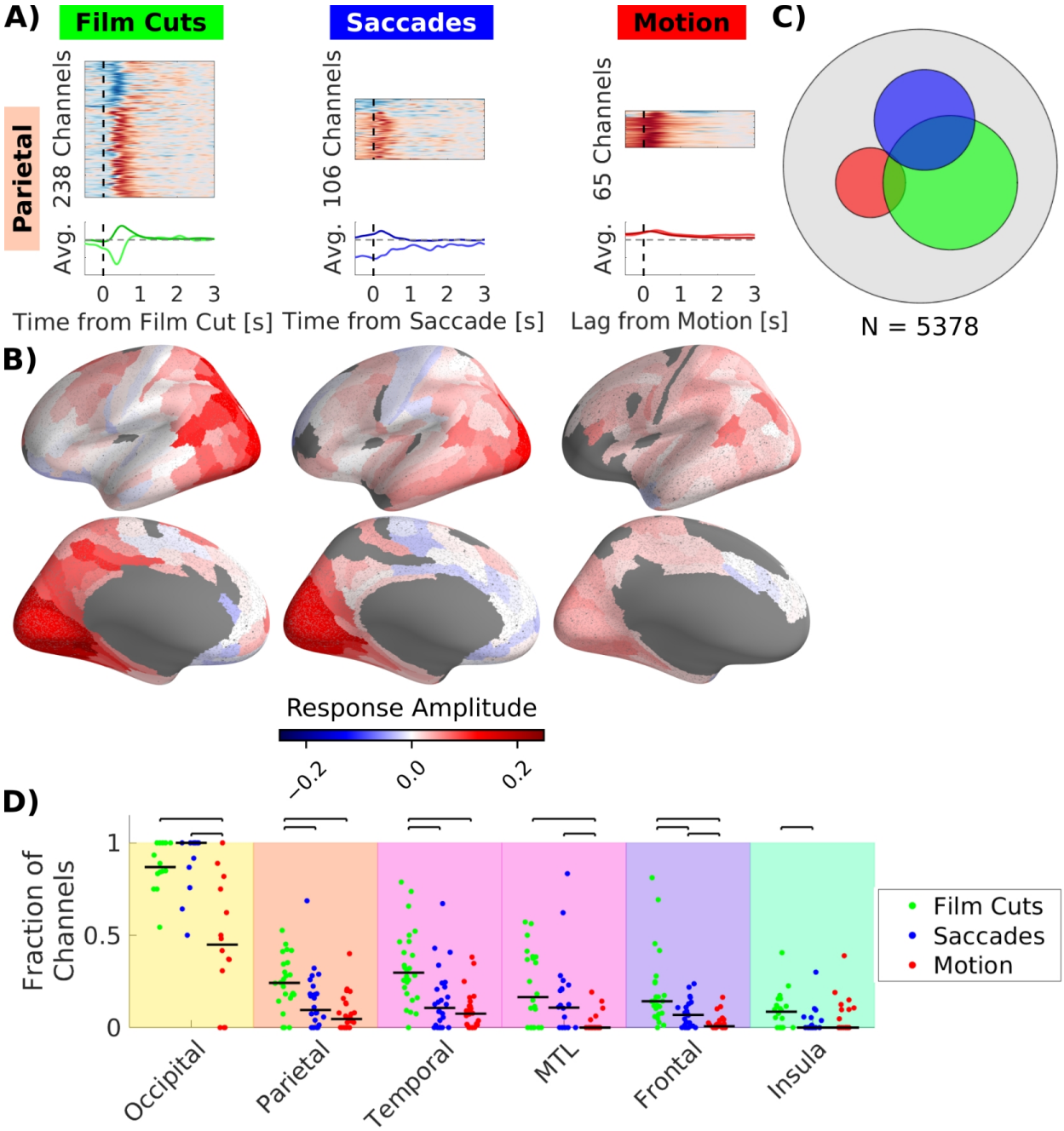

Figure 2. More channels respond to film cuts and saccades than to stimulus motion. A) Top row: Temporal response functions in the parietal lobe for channels with statistically significant response in BHA. Red indicates an increase, blue a decrease in BHA compared to the baseline. The sizes of each image reflect the number of significant channels. Similar TRFs were grouped by clustering TRFs that are highly correlated to each other (*40*, *41*). Bottom row: Average TRF for each cluster, scaled to reflect the relative strength of responses to each regressor within each channel. Time for motion indicates the lag of the neural response in relation to the optical flow signal. B) Weighted average response amplitude of

significant channels in parcels of the Glasser atlas (*31*). Red colors indicate an increase of BHA, blue color a decrease of BHA. The number of patients with significant electrodes in each parcel, is plotted in Figure S4B. C) Number of significant channels for each regressor indicated as area of each circle. The area of the gray circle indicates the total number of channels. D) Fraction of channels out of all channels within each brain area with significant response. Each dot shows the fraction of channels with a significant response for each patient. Straight black bars depict the median across patients. Black bars with whiskers above the plot indicate pairwise comparisons with significant differences between the medians ($p < 0.05$, Wilcoxon rank sum test, FDR corrected at $q = 0.05$). Background colors correspond to different lobes in Figure 1A. For results in a more detailed parcellation of the brain see Figure S3.

Contrary to our expectation we found the most widespread BHA responses are associated with film cuts followed by saccades and motion (Figure 2D & S4B). Across the whole brain, the sets of channels responsive to either film cuts, saccades or motion were largely distinct (Figure 2C). The overlap between responses related to saccades and film cuts (41% of saccade-responsive electrodes, Figure 2C) could be due to unobserved common factors. Surprisingly, the responses related to saccades and film cuts were also stronger than those to motion in areas such as the precuneus and middle temporal area (Figure S3). These areas have been shown to be important in motion processing (*33*, *42*). However, in the transverse temporal gyrus (Heschl's gyrus) the largest fraction of channels respond to motion (Figure S3). We refrain from interpreting these responses related to motion, cuts or saccades as specific to these individual factors. Various features in films could be associated with these regressors. For instance, sound is associated with motion ($r=0.24$, $p<10^{-4}$, permutation statistics). This partly explains the responses associated with motion in Heschl's gyrus, which is considered an auditory processing area. To better control for these unobserved factors, we analyzed how each of these factors is modulated by semantic novelty, keeping other confounds constant.

Cuts at event boundaries are associated with distinct neural responses across the brain

We observed that film cuts are associated with fluctuations in neural activity throughout the brain (Figure 2). Film cuts cause abrupt changes in various low-level visual features, such as luminance, contrast and color, but for some cuts there are additional changes in semantic information and the narrative. To determine the effect of semantic versus low-level visual changes we divide film cuts into two categories: "event cuts" and "continuous cuts". (Figure S7A). To this end, we collected event boundary annotations in a separate population of participants recruited online (N=200, Table S2). Participants were instructed to watch the videos and "indicate when […] a meaningful segment has ended by pressing the spacebar". These event boundaries annotations were consistent across participants and with data collected previously (*43*) (Figure S7B). For each video, we ranked all cuts by event salience – the number of participants that marked the end of a segment within one second after a cut. The film cuts with large event salience are termed 'event cuts' (using change point detection, see Methods). This resulted in 57 event cuts out of a total of 561 cuts (Table S2). Among the cuts with the lowest event salience we selected an equal number of 'continuous cuts' matched in low-level visual features, following (*6*) (see Methods). Continuous cuts lie between narrative event boundaries and are characterized by changes in camera angle or position (Figure 3C).

We first tested whether event cuts relate to stronger neural responses than continuous cuts. In each channel we fit the TRF identified previously (Figure 2A) to the neural signal after each individual film cut. The factor with which the TRF has to be multiplied to best fit the neural signal, quantifies the amplitude of the response (Figure S8). We perform a mixed-design ANOVA to test for a fixed effect of event versus continuous cuts and a fixed effect of brain region, while controlling for the random effect of patients (Table S3). We find a significant main effect of cut ($F(1,2270) = 33.8$, $p = 6.9*10^{-9}$) with larger amplitude for event cuts than continuous cuts. Amplitudes also differ significantly across brain regions ($F(5,2270) = 5.75$, $p = 2.1*10^{-5}$). A follow-up mixed-design ANOVA in each brain region with patient as random factor shows that amplitudes are significantly larger for event cuts in the temporal lobe and MTL (Figure 3A, temporal: $F(1,803) = 48.5$, $p = 7*10^{-12}$; MTL: $F(1,95) = 23.2$, $p = 5.4*10^{-6}$; FDR correction at $q = 0.05$, full results in Table S4).

We compute the weighted average of the amplitude difference of event versus continuous cuts across all channels in each parcel of the Glasser atlas to visualize this effect on the cortical surface (Figure 3B) (*30*, *31*). This analysis reveals a complex pattern of brain areas responding preferentially either to event or continuous cuts. We hypothesized that these distinct responses related to event cuts and continuous cuts may be particularly pronounced in higher-order association areas. To explore this possibility, we repeat the analysis but now identify separate TRFs for event cuts and continuous cuts in each channel (Figure 3C), i.e. using separate regressors indicating each type of cut. We included saccades as a regressor to remove correlated activity associated with saccades following cuts. We obtain similar results when also including motion as a regressor (Figure S9), or when performing the analysis on different types of videos (Figure S10). We found channels responsive only to event cuts, channels responsive only to continuous cuts, and channels responsive to both (Figure S11). We quantify this specificity of responses as the fraction of electrodes responsive to only one type of cut. In the occipital lobe specificity is low (Figure 3D), while in temporal and frontal lobe as well as the MTL it is higher than expected by chance (Table S5). Specificity is mostly driven by channels responsive to event cuts, but not continuous cuts (Figure S12). In most channels both event cuts and continuous cuts are associated with increased neural activity (Figure S11). To directly test whether specificity increases in higher-order association areas, we compute the spatial correlation of specificity with maps of cortical hierarchy (Figure 3E). In particular, gray matter myelination, measured by the T1w/T2w ratio, has been suggested as an indirect measure hierarchy from sensory to association processing (*30*, *44*). We find a significant negative correlation of specificity with the T1w/T2w ratio (Figure 3E, right) indicative of increased specificity along the sensory-to-association hierarchy (*30*, *44*). This correlation holds when the analysis is conducted on a more fine-grained atlas and when testing significance with permutation statistics accounting for the spatial autocorrelation of cortical myelination across adjacent brain areas (Figure S13). In addition, specificity is negatively correlated with the second principal gradient of functional connectivity, which separates somatomotor and auditory cortices (Figure S13C) (*45*). There was no consistent association of specificity to the first principal gradient of functional connectivity capturing sensorimotor-to-transmodal progression (Figure S13B) (*45*).

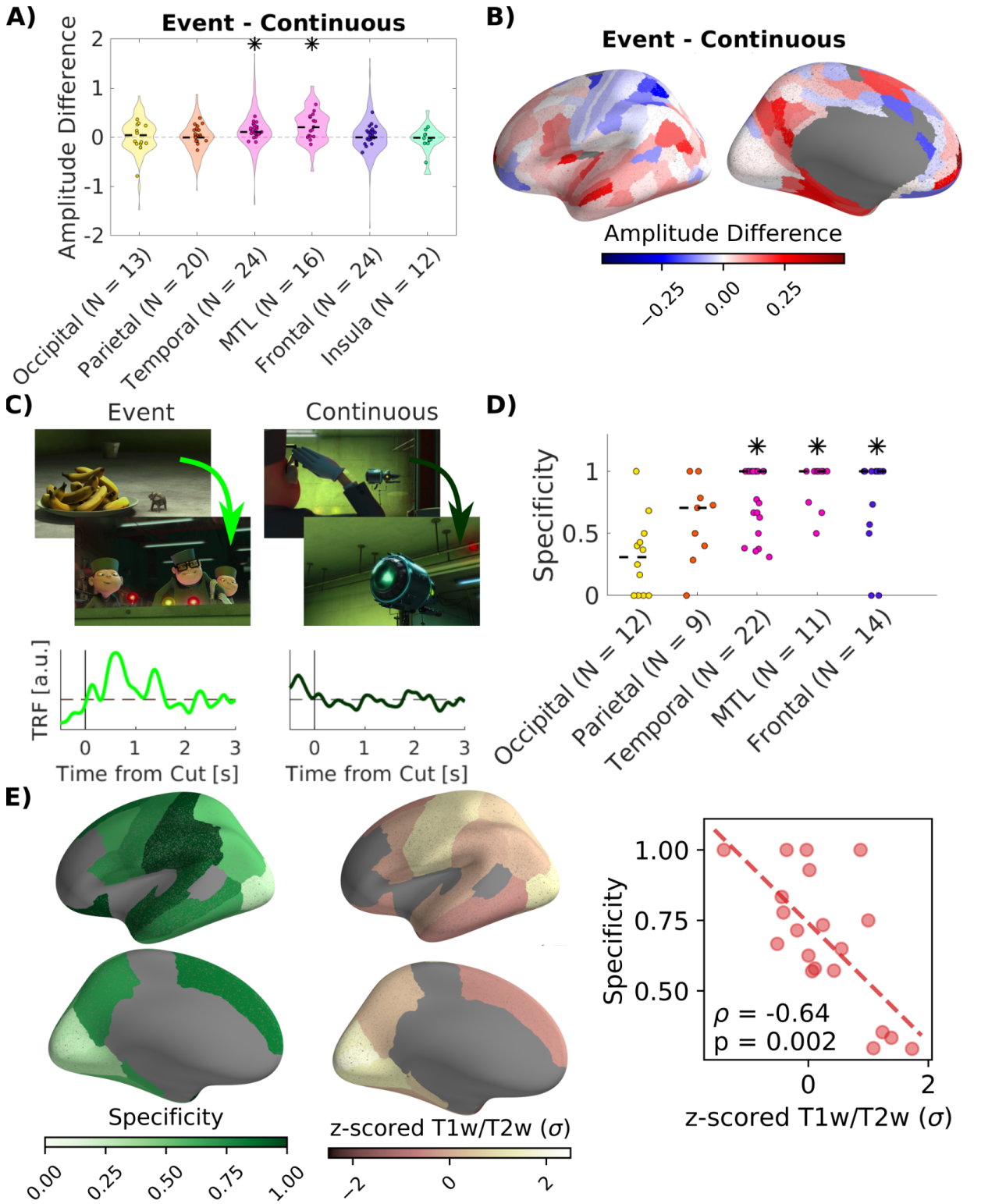

Figure 3: Neural responses associated with film cuts are stronger, and in higher-order association areas respond specifically to event cuts as opposed to continuous cuts. A) Difference in magnitude of the responses to individual event cuts and continuous cuts (Figure S7). Magnitude of the response is estimated by scaling the temporal response function in each channel to best fit the response to each cut (Figure S8). Only channels with significant responses to film cuts in Figure 2A are considered. Violin plots show the distribution of the difference in magnitude of responses across channels. Dots show the

average difference of magnitude within each patient. Stars indicate significant main effects of the type of film cut within regions with patients as a random factor in a mixed-design ANOVA (FDR corrected at q = 0.05, Table S4). B) Weighted average of the amplitude difference of responses to event versus continuous cuts across significant channels within parcels of the Glasser atlas (*31*). C) Temporal response functions are obtained for two separate regressors coding separately for event cuts and continuous cuts. In this example channel in the supramarginal gyrus there is a significant response after 0.5s to event cuts, but no response to continuous cuts. TRFs in all channels are shown in Figure S11. D) Specificity of responses in each region measured as a fraction of responsive channels responding to only one type of film cut. Dots show the specificity in each region for each patient. Stars indicate specificity that is significantly higher than expected by random assignment of channels to event or continuous cuts. Regions with only one significant channel have been removed from analysis. E) Left: Specificity computed in parcels of the Desikan-Killiany atlas (*46*). Specificity is computed in all parcels with at least five significant channels. Middle: Mean T1w/T2w ratio in each parcel of the Desikan-Killiany atlas. Right: Spearman correlation coefficient between T1w/T2w ratio and specificity. Spatial correlations of specificity with the T1w/T2w map in a finer grained parcellation, as well as correlations with gradients of functional connectivity are shown in Figure S13.

Novelty across saccades

As with film cuts, BHA responses associated with saccades are widespread across the brain, with long latencies suggestive of semantic visual processing (Figure 2 & S3). We hypothesized that responses associated with saccades are modulated by changes in low-level and semantic visual features between the foveal image before and after each saccade. If true, we would expect stronger saccade-related responses when the image features before and after a saccade are less similar to one another, i.e. when the upcoming target of a saccade is novel. To measure novelty we leverage a deep convolutional neural network trained with contrastive learning. Specifically, we use a ResNet pre-trained to extract features shared across different image patches (*29*). These features capture semantic properties of objects in the images (*47*). We compute the euclidean distance of features for image patches of 5x5 degree in visual angle around the gaze position before and after each saccade (Figure S14A&B). In doing so, each saccade is associated with a numerical value indicating the novelty of the upcoming fixation. Interestingly, on average the novelty associated with observed saccades was larger than random simulated saccades, matched in distance and direction (Figure S14C). This indicates that viewers tend to direct their gaze towards locations with higher novelty. This is particularly true shortly after film cuts, whereas saccades later during scenes tend to move towards low-novelty targets (Figure S15).

Next, we divided all 55,334 saccades from all 23 patients and videos into two equally sized groups with high and low novelty (Figure 4C), while controlling for saccade amplitude (Figure S14D). We excluded saccades across cuts and the first saccade after each cut. We predicted that saccades with high novelty will be related to stronger BHA responses. We estimate the magnitude of the response to each saccade as before (Figure S8). We find significant main effects for saccade novelty and brain region in a mixed-design ANOVA with patients as a random factor (Figure 4A, $F(1,1632) = 19.77$, $p = 9.3*10^{-6}$; Region: $F(5,1632) = 20.41$, $p = 8.2^{-20}$, Table S6). A follow-up mixed-design ANOVA for each brain region shows that novelty increases response amplitude in occipital and temporal lobes (Figure 4A, occipital: $F(1,264) = 9.06$, $p = 0.0043$, temporal $F(1,513) = 11.83$, $p = 0.0011$, FDR corrected at q = 0.05, Table S7). An analysis of the amplitude difference in smaller parcels shows that most of the cortex responds

stronger to high-novelty saccades. However, there are some exceptions, particularly in the parietal and temporal lobes, responding stronger to low-novelty saccades (Figure 4B). We also computed separate TRFs for saccades with high and low novelty, including film cuts and optical flow as regressors to remove correlated activity (Figure 4C). In the occipital lobe the majority of channels are associated with responses to either high or low novelty, i.e. respond with low specificity (Figure 4D). In contrast, in the frontal lobe most channels are responsive to either high or low novelty saccades (Figure 4D & S17), with a specificity above chance (median = 1, $p = 0.024$; 1000 permutations. FDR corrected across brain regions at $q = 0.05$, Table S8). This suggests a specialization for high novelty, but at the same time, low-novelty saccades. Interestingly, in the parietal and frontal lobes, some responses related to low-novelty saccades show an inhibition of neural activity (Figure S16). In contrast, channels responsive to high-novelty saccades show only increases in neural activity. The comparison of specificity with T1w/T2w ratio leads to strong negative correlations indicating that specificity increases along the sensory-to-association hierarchy (Figure 4E & S18A). In addition, specificity of responses to high- and low-novelty saccades is related to the second gradient of functional connectivity, separating somatomotor and auditory cortices (Figure S18C). There was no consistent association with the first gradient of functional connectivity (Figure S18B). As such, there results parallel those for specificity to type of cuts.

Behavioral data shows that saccade novelty does not significantly differ between repeats of “The Present” when controlling for saccade amplitude (difference of median novelty: $-7*10^{-4}$, $p=0.46$, $N_{repeat_1}=6{,}843$, $N_{repeat_2}=5{,}121$, Wilcoxon rank sum test). However, saccade amplitude increases significantly in the second repeat of “The Present” (difference of median amplitude 0.55 DVA, $p=1.5*10^{-31}$, $N_{repeat_1}=6{,}843$, $N_{repeat_2}=5{,}121$, Wilcoxon rank sum test).

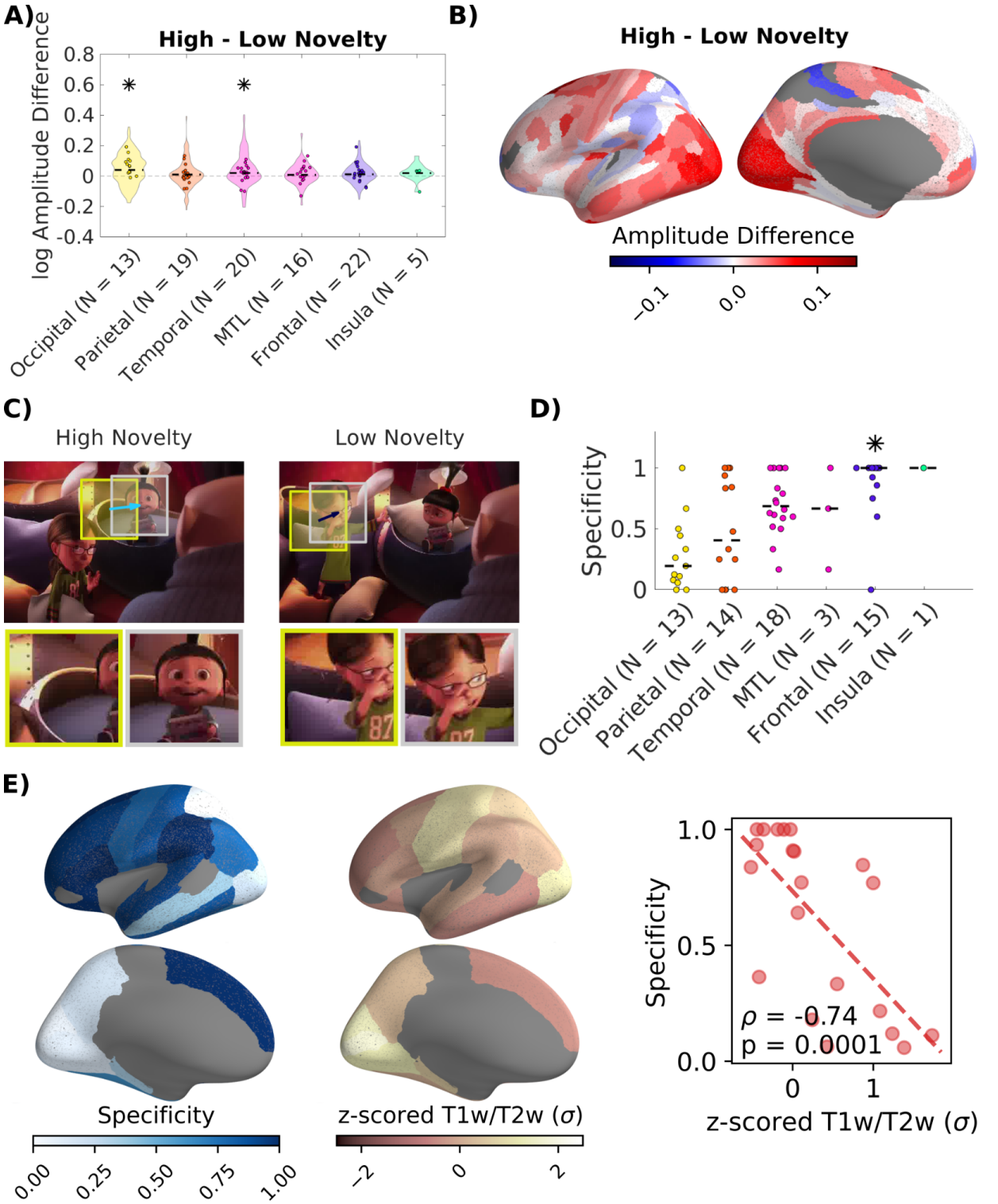

Figure 4: Differential responses associated with saccades to high and low novelty visual stimuli in higher-order association areas. A) Difference in magnitude of neural responses to individual saccades with high and low novelty targets. Magnitude was estimated using filters for all saccades as described in Figure 3A. Stars indicate significant main effects of the type of film cut within regions with patients as a random factor in a mixed-design ANOVA (FDR corrected at q = 0.05, Table S7). B) Weighted average of

the amplitude difference in parcels of the Glasser atlas (*31*). C) Temporal response functions are estimated for saccades with high and low novelty separately. Example of a high-novelty saccade (feature distance = 8.49) and a low-novelty saccade (feature distance = 4.85) with similar saccade amplitude. Novelty is computed as the distance between features from a convolutional neural network (Figure S14). TRFs in all channels are shown in Figure S16. D) Specificity of responses to saccades with high and low novelty defined as in Figure 3C. E) Left: Specificity computed for channels in each parcel of the Desikan-Killiany atlas (*46*). Parcels with less than five responsive channels are not included. Middle: Map of T1w/T2w ratio, a proxy of gray matter myelination (*30*, *44*). Right: Spearman correlation of specificity and the T1w/T2w gradient. Spatial correlations of specificity with the T1w/T2w map in a finer grained parcellation, as well as correlations with gradients of functional connectivity are shown in Figure S18.

Saccades to faces

A more direct way to quantify semantic changes across saccades is by the content of the saccade target. Specifically, saccades to faces have been reported to relate to stronger neural responses than saccades to other objects (*10*). To detect faces in the movies we finetuned a pre-trained object detection and segmentation network with a subset of labeled frames from our videos. This network was then used to detect faces in the unlabeled frames. We then divide saccades to faces and saccades to other objects (non-face saccades, 10,225 each) and analyze their neural response functions (Figure S19).

As above we calculate the difference in the magnitude of responses related to face and non-face saccades. We find a significant main effect of face vs non-face and brain region in a mixed-design ANOVA with patient as random effect (Face vs non-face: $F(1,1551) = 40.4$, $p = 2.72*10^{-10}$; Brain Region: $F(5,1551) = 11.46$, $p = 6.99^{-11}$, Table S9). Surprisingly, the responses related to non-face saccades have a significantly larger magnitude than those related to face-saccades in the occipital, parietal, mediotemporal (MTL) and frontal lobe (Figure 5A) (follow-up mixed-design ANOVA for each brain region with patient as random effect, Figure 5A: Occipital: $F(1,263) = 30.84$, $p = 1.65*10^{-7}$; Parietal: $F(1,200) = 19.01$, $p = 3.57*10^{-5}$; MTL: $F(1,100) = 9.26$, $p = 0.0036$; Frontal: $F(1,432) = 13.43$, $p = 3.72*10^{-4}$; FDR control, at a level of $q = 0.05$, Table S10). This is surprising because saccades to faces have a higher novelty ($p=1.6*10^{-60}$, $N_{face}=7636$, $N_{non-face}=7784$, Mann-Whitney U-test), so if anything, the opposite effect would have been expected. Consistent with this, computing the weighted average of the amplitude difference of responses confirms stronger responses to non-face saccades in most brain areas (Figure 5B). However, this fine grained analysis also shows some areas in the parietal and temporal lobes with larger responses to face saccades (Figure 5B) To determine the specificity of responses we compute separate TRFs for face and non-face saccades, including film cuts as a regressor to remove correlated activity. Most channels throughout the parietal, temporal, frontal lobe and insula are selectively responsive to either face or non-face saccades (Figure 5D, Parietal: median = 0.94, $p = 0.002$; Temporal: median = 0.97, $p = 0.002$; Frontal: median = 1, $p = 0.002$; Insula: median = 1, $p = 0.025$. 1000 permutations. FDR correction at $q = 0.05$, Table S11). Interestingly, in the superior temporal gyrus, which contains the auditory cortex, the largest fraction of responsive channels is to face-saccades (Figure S20). The spatial correlation of specificity with the T1w/T2w map is not significant for larger parcels in the Desikan-Killiany atlas (Figure 5E). This result is unsurprising given that responses across the brain are specific to non-face saccades

starting from sensory areas (Figure 5B&D). There is a significant correlation, however, of the specificity to face and non-face saccades and the second functional gradient, which separates somatomotor and auditory cortices (Figure S21). Overall, these results indicate that faces and other objects are associated with selective responses in brain areas extending far beyond traditional visual processing areas. However, again, we can not rule out that some of these responses are due to other unobserved factors associated with the presence or absence of faces.

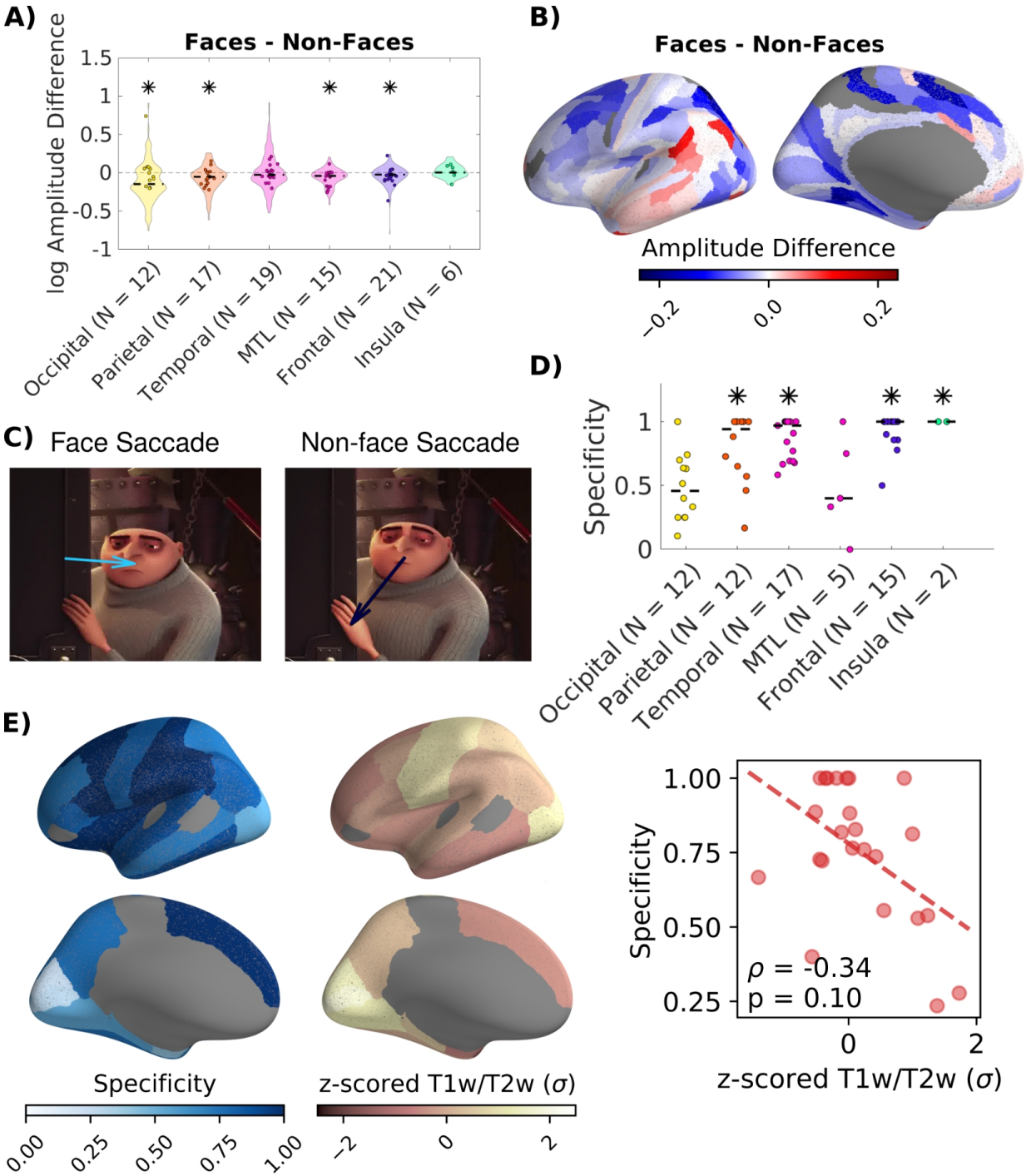

Figure 5: Responses that are specific for face and non-face saccades are found in higher-order association areas. A) Difference in magnitude of neural responses to individual saccades to faces and non-faces. Magnitude was estimated using filters for all saccades as described in Figure 3A. Stars indicate significant main effects of the type of film cut within regions with patients as a random factor in a mixed-design ANOVA (FDR corrected at q = 0.05, Table S10). B) Weighted average of the amplitude difference in parcels of the Glasser atlas. C) Separate TRFs are computed for face and non-face saccades (Figure S19). D) Specificity of responses to saccades to faces and non-faces defined as in Figure 3C. E) Left: Specificity computed for channels in each parcel of the Desikan-Killiany atlas (*46*). Parcels with less than five responsive channels are not included. Middle: Map of T1w/T2w ratio, a proxy of gray matter myelination (*30*, *44*). Right: Spearman correlation of specificity and the T1w/T2w gradient. Spatial correlations of specificity with the T1w/T2w map in a finer grained parcellation, as well as correlations with gradients of functional connectivity are shown in Figure S21.

Face Motion

Responses to motion are dominated by responses related to face motion compared to total motion (Figure S22). Specifically, face motion is associated with responses in distinct clusters of channels in the lateral occipital cortex, fusiform gyrus and superior temporal sulcus (Figure S22&S23). These areas include face-processing areas (*48–50*), known to respond stronger to moving faces (*4*, *26*, *48*)**.** Other areas, however, show specific responses related to total motion, notably the Hippocampus (Figure S23), which might reflect spatial remapping during camera movements (*51*).

Discussion

Much of the existing neuroscience literature related to motion perception, saccades or dynamic visual stimuli focuses on the effects of individual stimulus properties on neural activity in anatomically constrained brain areas. This approach links specific effects to existing reports for the same brain areas. However, implicitly, it also ascribes a narrow specialization to individual brain areas which may not be warranted in a real-life setting. Here we take a more inclusive approach to study the effect of multiple sources of visual change in a more ecologically valid setting of watching movies. The system identification approach controls for the observed correlation between motion, saccades, and film cuts. We found that low-level and semantic visual changes across film cuts and saccades result in fluctuations of neural activity in distinct and widespread neural populations across the whole brain. In contrast, motion appears to affect a more confined area of visual brain regions. In the following we will put these results in context of the specific literature in each of these areas.

Overall, responses related to motion are more confined to visual areas compared to responses related to film cuts and saccades (Figure 2D & S4). Therefore, motion in naturalistic stimuli is perhaps a less dominant driver of neural responses than previously thought (*3*, *4*). The responses related to motion we find in our data, however, are consistent with previous reports. Studies using simple motion stimuli as well as naturalistic stimuli have identified specific motion sensitive brain areas in the occipital, temporal and parietal lobes, such as the medial temporal area (MT), or the ventral intraparietal area (VIP) (*3*, *4*, *42*, *52*). However, motion in naturalistic

stimuli activates much wider areas across the temporal lobe, largely driven by motion of socially relevant stimuli, such as faces or body parts (*3*, *4*, *33*). Similarly, we find widespread responses associated with optical flow in occipital, temporal and parietal lobes (Figure 2).

We find that responses associated with film cuts, especially those at event boundaries, are widespread across the entire brain (Figure 2&3). These results consolidate previous findings of responses to event boundaries in individual brain areas, most prominently in the medial temporal lobe. This supports the proposed role of event boundaries in organizing memory of continuous experience (*6*, *20*, *22*, *23*, *43*, *53*). Our results show that event cuts elicit stronger responses (than continuous cuts) in association areas like the angular gyrus and posterior medial cortex (Figure 3B), consistent with previous fMRI results (*22*). We find the opposite effect in the superior parietal and premotor cortex (Figure 3B), similar to a previous fMRI study (*21*). This is consistent with the suggestion that the superior parietal cortex maintains internal sensory and motor representations (*54*). While film cuts are not usually considered in the context of event boundaries (*5*, *55*), recent work has identified distinct responses depending on the nature of visual change across film cuts. For example, different neural populations in the medial temporal lobe show specific responses to film cuts within or between different video clips (*6*). In addition, film cuts with action discontinuities activate parietal and frontal areas (*21*), while film cuts in general strongly activate visual areas (*5*, *21*). Here we propose that while all film cuts are associated with various changes in low-level visual content, some coincide with semantic change at event boundaries. We defined the latter as “event cuts" and cuts within events as “continuous cuts” based on standard event segmentation (*20*, *21*). Our data shows that film cuts are associated with widespread neural responses across the whole brain (Figure 2). Importantly, channels in the MTL and temporal lobe are related to stronger responses to event cuts than to continuous cuts (Figure 3A). By showing that event boundaries can be linked to film cuts we were able to investigate their fast neural correlates. We show that event boundaries are processed in a widespread ensemble with increased specificity in higher-order association areas (Figure 3E).

Saccades create shifts of the image on the retina that result in neural responses in early visual areas (*7*, *56*, *57*). These low-level responses are suppressed in higher-order visual areas, probably due to integration of visual responses with a copy of the motor signal (*7*, *56*, *57*). In contrast, previous work has found modulation of neural activity by saccades in the non-visual thalamus (*9*), medial temporal lobe (*14*, *15*), auditory, frontal and parietal cortices (*16*). These findings collectively suggest that saccades are linked to the neural processing of higher-order cognitive processes, such as attention and memory (*11*). Our findings support the view that neural activity across most, if not all the brain (Figure 2), is modulated by saccades and in particular by their association with semantic novelty.

There is competing evidence that saccades in static visual scenes sample either semantically similar (*58*), or dissimilar objects in static natural scenes (*59*). In movies, scenes change rapidly and we find that both types of saccades can be observed (Figure S15). In static natural scenes, eye movements target objects with similar semantics (*58*). We propose that this type of saccades corresponds to saccades with low semantic novelty in our video data. Saccades are

also attracted by visual features that are novel in the context of static scenes (*59*). We propose that this corresponds to saccades with high semantic novelty in our data. Exploitation of a visual scene by sampling similar objects competes with the necessity to explore novel objects in dynamic environments. We suggest that these objectives are concurrent and the balance may vary with task demands. For instance, we observed that saccade amplitude, a measure of exploratory behavior, increases during the second repeat of the video “The Present”. In movies, eye movements are synchronized between subjects, suggesting a stereotypical behavior (*60*, *61*). Increased saccade amplitude upon repeal could therefore indicate idiosyncratic exploration (*61*). Interestingly the opposite behavior is found for static images which are less synchronized across subjects (*60*) and results in shorter saccades upon repetition (*62*). Together this suggests that the balance between exploitation and exploitation depend on the task. Indeed, we find that after film cuts, saccades are directed primarily to novel targets (Figure S15). We suggest that neural responses to high-lovelty saccades support exploration, and responses to low-novelty saccades support exploitation. The concurrent demands are reflected in different sets of channels that are responsive to saccades with high and low semantic novelty (Figure 4). The suppression of neural processing in the parietal and frontal lobes that we find in particular for low-novelty saccades (Figure S16), may facilitate perceptual stability across eye movements (*63*).

Defining saccade novelty by face or non-face targets provides a clear link to semantic and social information related to faces. Faces contain a wide variety of features that are processed in distinct specialized areas (*48*). Low-level, view specific features are processed in the occipital lobe, while face specific areas in the fusiform gyrus and superior temporal gyrus process face identity, face motion and eye gaze. An extended face network, distributed in the auditory cortex, limbic system and prefrontal cortex, has been proposed to process higher level semantic information, such as speech, directed attention, emotion and biographical information (*10*, *16*, *64*, *65*). In support of this model, we find that saccades to faces and other objects are associated with activity fluctuations in distinct sets of channels across the whole brain (Figure 5). In particular, we find responses related to saccades to faces in auditory areas, such as the STS, and the frontal cortex, as well as, distinct sets of channels responsive to face and non-face saccades in the medial temporal lobe (Figure S20). Together these results show that the extended face network can be studied in real world scenarios, where multiple semantic features interact, by locking face processing to saccades.

Limitations

In most recording locations we found responses to only one of the main sources of visual change (Figure 2C). However, some locations respond to both film cuts and saccades, which may be due to unobserved common factors. For instance, after a film cut often a new face appears, but faces also attract saccades. Therefore, responses to faces may activate the same channels even after accounting for the correlation between film cuts and saccades. More generally, we note that in film there could be a variety of unobserved factors that are associated with cuts, saccades or movement. For instance, stimulus movement is associated with sound.

This could explain responses to motion we observed in the auditory cortex, including the transverse temporal (Heschl's) and superior temporal gyrus (FIgure S3 & S23) (*66*). Therefore, we caution against the interpretation that these widespread responses are *specific* to cuts, saccades or motion. We also performed a number of contrasts individually for different types of film cuts, types saccades and types of stimulus motion. These contrasts aim to capture the effect of semantic novelty, while keeping other factors constant (e.g. saccade amplitude). Nevertheless, there is a complex interaction between different features of the stimulus, as well as with saccade behavior so that confounds cannot be ruled out.

The relatively weak responses related to optical flow in our data could be due to several factors. First, our total motion feature is unspecific, different types of stimulus motion. The face motion regressor captures only a small part of socially relevant motion. Thus, the related responses could be restricted in space. Second, electrode coverage in our patient population is not chosen to specifically cover motion sensitive areas. We further did not run any motion localizers. The locations we recorded from might not cover motion sensitive areas. However, we would still have expected widespread responses to motion. It is therefore possible that responses to film cuts and saccades simply are related to stronger neural responses than motion because they are associated with more pronounced changes and novel information.

The system-identification approach used here captures time-delayed responses only to first order. It does not capture higher-order (non-linear) effects on BHA or multiplicative interactions between regressors (motion, saccades and cuts). Techniques for doing this within a system-identification approach are readily available (*36*, *37*), but would lead to significantly more complex exposition. The current approach does, however, account and compensate for correlation in the regressors, contrary to more conventional reverse-correlation (evoked response) analysis.

We have focused on BHA as the best available correlate of neuronal firing that is available in iEEG recordings. One could perform a more thorough analysis of local field potentials, for example modulation of power in other frequency bands and phase alignment, which are a rich source of information on neuronal dynamics (*67*, *68*). As shown by recordings in macaque V1, such analysis may be particularly important when analyzing modulatory top-down effects (*8*).

The analysis focused on saccade onset, rather than fixation onset. Our motivation was to look for change across a saccade, rather than focusing on the content of individual fixations. However, early on we determined that our analysis approach gives similar results if we used fixation onset instead of saccade onset (as the two are tightly coupled in time, 32±14ms, Figure S6). However, we note that previous work showed that locking visual responses to saccade versus fixation onset highlight either top-down or bottom-up influences at the level of V1 (*8*).

While we control for various low level features between event and continuous cuts, similar features could drive differences between responses to high- and low-novelty saccades. However, we control for saccade amplitude, which we expect to have the strongest influence on neural responses (*69*).

Including two repeats of “The Present” video in the same film cut TRF model might lead to temporally smeared responses, due to anticipatory responses (*70*) in the second repeat. While we present comics in Hungarian and English we have not investigated the effect of different languages. The narrative in comics is easy to follow regardless of language.

Conclusion

Taking a data-driven approach to analyze intracranial EEG data during movies we found widespread responses related to film cuts and saccades. We show that semantic changes across film cuts and saccades can be defined through event boundaries, visual novelty and the presence of faces. These semantic changes modulate neural activity in distributed locations across the whole brain, particularly in higher-order association areas. This extends previously known anatomical locations with functional specificity to these visual features. Further studies will be needed to fully characterize the factors underlying these widespread yet specific responses related to visual change we have found here. The richness of film makes this analysis challenging, but we do believe that this richness is necessary to capture the full extent of neural processing of natural stimuli.

Methods

Dataset

Intracranial electroencephalography (iEEG) along with eye movements were recorded from 23 patients (mean age 37.96 years, age range 19-58 years, 11 female; Table S1) with pharmacoresistant focal epilepsy at North Shore University Hospital (Manhasset, New York). Patients were chronically implanted with depth and/or grid electrodes to identify epileptogenic foci. Three patients were implanted twice at different times. We recorded the same session twice with these patients with different electrode coverage. The study was approved by the institutional review board at the Feinstein Institute for Medical Research and all patients gave written informed consent before implantation of electrodes. Across patients electrode locations cover most of the brain (Figure 1A). However, most dense coverage is available on the temporal lobe and coverage of the occipital lobe is more sparse (Figure 1A). iEEG data was recorded continuously at 3kHz (16-bit precision, range ± 8 mV, DC) on a Tucker-Davis Technologies data processor (TDT, Alachua, FL, USA). Gaze position was recorded simultaneously with iEEG data with a Tobii TX300 eye tracker (Tobii Technology, Stockholm, Sweden) at 300Hz. The eye tracker was calibrated before each video to prevent drift. We used parallel port triggers sent from the stimulus PC to the eye tracker and data processor to align the different data streams. A custom script using psychtoolbox (*71*) for movie presentation and the Tobii SDK for collecting eye tracking data was implemented in MATLAB. For additional accuracy in the alignment of the movie features to iEEG and eye tracking data we recorded timestamps at the onset of each frame with the clock of the eye tracker.

Patients watched up to 43.6 minutes of video clips (Figure 1B). Video clips included segments of an animated feature film ('Despicable Me', two different clips, 10 min each, in English and Hungarian language), a animated short film with a mostly-visual narrative shown twice ('The Present, 4.3' min), and three clips of documentaries of macaques ('Monkey', 5 min each, without sound).

Electrode localization

Each electrode shank/grid contains multiple recording contacts. Contact locations were identified using the iELVis MATLAB toolbox (*72*). All subjects received a preoperative T1-weighted 1mm isometric scan on a 3T scanner. Tissue segmentation and reconstruction of the pial surface was performed with the freesurfer package (*73*, *74*). Postoperative CT scans were acquired and coregistered to the freesurfer reconstruction. Contacts were then semi-manually localized using the bioimagesuite software (*75*). All contacts were then coregistered to the freesurfer fsaverage brain for visualization and assignment to anatomical atlases (*76*). Subdural contacts were shifted to the closest vertex of the lepto-meningeal surface to correct for brain shift while preserving the geometry of grid contacts. Freesurfer coordinates of subdural contacts are determined by finding the nearest vertex on freesurfer spherical pial surface. In contrast, stereotactic electrode shanks were coregistered to fsaverage space using a linear affine transformation. Stereotactic contacts close to the pial surface (< 4mm) are assigned to cortical atlases by finding the nearest vertex on freesurfer spherical pial surface. For further analyses stereotactic contacts close to the pial surface were shifted to the nearest vertex of the native pial surface and then moved to fsaverage space in the same manner as subdural contacts.

Data Preprocessing

iEEG data was minimally processed by removing line noise at 60Hz, 120Hz, and 180Hz, with a 5th order butterworth bandstop filter, and low frequency drift at 0.5Hz with 5th order butterworth high-pass filter. The data was then re-referenced to a bipolar montage. Artifacts with an absolute voltage 5 times of the interquartile range of voltage of each channel were removed. Further, after visual inspection, channels with spiking activity and channels outside the skull were identified manually and removed from analysis. The power of the signal in each frequency band is calculated by the absolute value of the Hilbert transformation of the bandpass filtered signal. The broadband high-frequency amplitude (BHA) power is defined in the range of 70-150Hz. The power envelope is then downsampled to 60Hz.

Stimulus motion

As a measure of motion we extract optical flow from each video using the Horn-Schunck method as implemented in MATLAB (*77*). The Horn-Schunck method computes the displacement vectors of pixels from one frame to the next, assuming smooth flow across neighboring pixels. We average the displacement vectors across all pixels within each frame. This results in a regressor of average motion throughout the video (Figure 1C). The motion

regressor contains artifacts from film cuts. To remove these artifacts, we replace the samples within a window of 166 ms around film cuts with a linear interpolation. For movies with sound ("Despicable Me English", "Despicable Me Hungarian", "The Present") we correlate the motion regressor with the sound envelope. The sound envelope is calculated as the absolute value of the Hilbert transform of the sound extracted from the movie files. The correlation is computed across the concatenated signal for all three movies. To test the significance of this correlation we compute a surrogate distribution of the motion regressor by 10,000 circular shuffles.

Film cuts

Film cuts in the movies were identified as peaks in the temporal contrast between consecutive video frames. Temporal contrast is the mean square difference of luminance between consecutive video frames (*3*). Film cuts with smooth transitions do not show up as sharp peaks in the temporal contrast and are missed. Moments of sudden motion on the other hand, might be mistaken for Film cuts. Film cuts detection is corrected by visual inspection to account for these errors.

Event cuts versus continuous cuts

To classify film cuts based on changes in semantic information we align film cuts to event boundaries. We record event boundary annotations for all videos in a separate study conducted online. 200 participants were recruited on Prolific (www.prolific.co). The task was implemented in PsychoJS scripts created from the psychopy builder (*78*). The task was hosted on Pavlovia (https://pavlovia.org/). Participants watched one of the videos each with the following instructions: *"The movie can be divided into meaningful segments. You will have to indicate when you feel like a meaningful segment has ended by pressing the 'spacebar. You will likely detect multiple events throughout the movie."*. For all videos, except Despicable Me, we included a task to check attention. Participants saw a black screen with 10 white boxes flashed at random times. Participants had to respond with a button press every time they saw a white box. We excluded the data from 20 out of 200 participants because either no event boundaries were annotated or the attention test failed (Table S2). The attention test was considered failed if participants responded to less than 8 of the white boxes in the task after the movies.

Event boundary annotations are delayed after film cuts due to reaction time and processing of visual information. To adjust for this, 1s was subtracted from every event boundary (similar to (*23*)). Event boundaries from all participants were aggregated in one vector consisting of impulses at the time of button presses per video. This vector was then smoothed with a Gaussian of 0.5s, to account for variance in the manual annotation of event boundaries (Figure S7A) (*23*, *24*). The resulting vector is a measure of event salience (Figure S7A) (*24*)**.** Event salience in our data is consistent with salience from data collected by Cohen et al. (Figure S7B&C) (*43*). This allowed us to compute event salience at the time of each film cut. Film cuts were sorted by event salience. 'Event cuts' are the film cuts with event salience above a threshold. For each movie this threshold is defined as the change point detected with the findchangepts() function in MATLAB. This method minimizes the residual error from the mean in

the segments before and after the change point. We select 57 event cuts out of a total of 561 cuts (Table S2). An equal number of cuts with lowest salience were selected, which we refer to as ‘Continuous cuts’. We test if event cuts and continuous cuts differ in changes of low-level visual features (*79*). These features are luminance, contrast, complexity, entropy, color, and features from layer fc7 of AlexNet (*6*). Complexity was quantified as the ratio of file size after JPEG compression (*80*). We compute a p-value for each feature across cuts with a Wilcoxon signed rank test and correct for multiple comparisons with the Benjamini-Hochberg procedure. If event and continuous cuts differ in low-level features we iteratively select a random set of continuous cuts from the film cuts in the lower half of event salience until there is no difference in low-level features.

Face Detection

We use an object detection algorithm made available through facebook's Detectron2 platform (*81*). We selected a ResNeXt-101-32x8d model backbone (*82*) trained in the mask R-CNN framework (*83*) on the COCO dataset (*84*) due to its high segmentation accuracy compared to other models on the Detectron2 platform. Neural networks for face detection exhibit high performance on natural movies, however, face detection in comics requires retraining of the networks. We therefore annotated faces in 4551 frames in ‘Despicable Me English’ and 1575 frames in ‘Despicable Me Hungarian’ using the ‘Labelme’ (https://github.com/wkentaro/labelme) and ‘Roboflow’ (Roboflow Inc, Des Moines, Iowa) tools. We applied flip and 90° rotations for data augmentation and created a training and validation set with a 80%-20% train-validation split ratio. For ‘Despicable Me English’ achieved a mean average precision of bounding box annotations mAP=0.61 and classification accuracy 74.5% on the validation dataset (mAP=0.74 and 80% classification accuracy on a subset of frames). For ‘Despicable Me Hungarian’ we achieved a mAP=0.58 and a classification accuracy of 78% on the validation dataset. Missing bounding boxes and wrong labels were corrected manually. Faces in the video ‘The Present’ were annotated manually with ‘Labelme’ in the whole video.

Saccade Detection

For saccade detection we apply a 20th order median filter to smooth the gaze position data and compute eye movement velocity. Samples of the eye velocity that were faster than 2 standard deviations from average eye velocity were labeled as saccades. Often we observe a short adjustment of the eye movement after a saccade until it fixates on the new target. We correct this overshoot by merging these samples to the saccade. To combine samples for the saccade and the overshoot, we perform a morphological closing operation with a kernel size of 5 samples (16.7 ms) on the samples belonging to the saccade and overshoot. We label the first sample in the saccade as the saccade onset. The fixation onset corresponds to the first sample after which eye velocity drops under the 70th percentile, computed from velocity values within 33ms before and 120ms after saccade onset. The eye tracker provides labels for data quality when the gaze was not detected, for example during eye blinks. Saccades within 83ms of samples with low data quality are removed. We also remove saccades within 110ms after a previous saccade.

Classification of saccades to faces

Saccades to faces could be determined simply using the location of the fixation onset. If the saccade lands on the bounding box of a face annotation the saccade could be classified as a face saccade. However, several saccades move towards faces but land just outside the face bounding box (Figure S19A). Other saccades land within a face bounding box, but move away from the center of the face (Figure S19B). Therefore, we generate several handcrafted features and classify face and non-face saccades using an SVM (Figure S19). The first feature is a binary variable indicating whether the saccade points towards or away from the centroid of the closest face annotations bounding box. The second feature measures the overlap of a circle with a radius of 5 degree visual angle with all face annotation bounding boxes. The third feature is the distance to the closest face annotation centroid. The fourth feature is the angle between the vector from saccade onset to the face annotation centroid and the vector from saccade to fixation onset. The fifth feature is the angle between the vector from saccade onset to the face annotation centroid and the vector from fixation onset to the face annotation centroid. We manually label a total of 1288 face and non-face saccades for saccades in one video from one patient to obtain training data. We fit an SVM with a Gaussian kernel in MATLAB using fitcsvm() and a kernel scale of 2.2. We achieve a cross-validation accuracy of 0.964 using 10 fold cross-validation. Saccades in all other videos and patients are classified in face and non-face saccades using this SVM model. The SVM classifies saccades and provides a score, indicating the signed distance to the decision boundary. Negative scores indicate saccades predicted as non-face saccades. Saccades with scores above 1 are classified as face saccades.

Saccade Novelty

To quantify the change of semantic novelty across saccades we use convolutional neural networks trained through contrastive learning (*29*). In contrastive learning neural networks are trained on subsets of transformation of images, in order to learn more generalizable representations. The networks are trained to minimize the distance of features from transformations of the same image. The most useful transformations to improve performance are random crops of images (*29*). These random crops are similar to saccades in images. In fact, crops based on simulated saccades improve performance of networks trained with contrastive learning compared to random crops (*85*). Here, we compute the feature distance between image patches extracted around gaze position at saccade and fixation onset (Figure S14A). Patches have a size of 200x200 pixels corresponding to the size of the foveal visual field of 5 degree visual angle. Features of a convolutional neural network of the pre- and post-saccadic patch are computed with a ResNet-50 trained with SimCLR (*29*). Saccade novelty is then defined as the euclidean distance between the features of the pre- and post-saccadic image patch (Figure S14B). A large distance between features corresponds to high saccade novelty. Novelty, thus defined, trivially correlates with saccade amplitude as patches further away will look less alike (Figure S14D). Thus, we divide all saccades into two groups of high and low saccade novelty, while controlling for saccade amplitude. We do this by fitting a linear regression model to describe the relationship between saccade novelty and saccade amplitude.

Saccades with higher novelty than predicted with this linear model comprise the group of saccades with high novelty (Figure S14D). The groups of saccades with high and low novelty are matched in number. To control for the possible confound of film cuts, saccades across film cuts and saccades within 1 seconds after film cuts are removed from the analysis.

System identification approach to establish temporal response functions.

We identify neural responses to features in the movies with a conventional linear system identification approach (*36*, *37*). Each channel is analyzed individually. The inputs to the system are the time courses for motion, film cuts, and saccade onset (Figure 1). These are the same for all channels (from a patient). The output is the time course of the BHA neural signal for every channel. All signals are (re)sampled at 60Hz - twice the frame rate. An impulse response, or temporal response function (TRF), is estimated (with ordinary leasts squares with ridge regression with 0.3 as the ridge parameter: Equation S1, (*36*)) that maps the stimulus to the BHA signal through a convolution (Figure S2A). For each channel TRFs are estimated simultaneously for all inputs to remove correlation. We fit the TRFs in with latencies from 0.5 seconds before to 3 seconds after the visual stimulus. After estimation, TRFs are smoothed with a Gaussian window with a standard deviation of 53ms to filter higher-frequency noise. Similar smoothing is common in cluster-based statistical analysis often performed in fMRI research and increases sensitivity of the analysis (*86*, *87*). To determine responses to semantic change associated with film cuts, saccades and motion we construct TRF models with different regressors. Different types of film cuts were modeled as event cuts and continuous cuts as inputs, also including film cuts to control for correlated activity (Figure 3). We found that including motion as an additional regressor did not change results (Figure S9). The model to test differences of novelty across saccades includes separate inputs for saccades with high and low novelty, as well as film cuts and motion (Figure 4). To test differences in saccades by saccade target we construct a model with separate inputs for face and non-face saccades, also including film cuts (Figure 5).

Statistically significant responses

To determine the statistical significance of responses, i.e. predictable fluctuations in BHA, we compute a surrogate distribution of TRFs with time-shuffled output signals. Surrogates output signals are constructed by random circular shifts in time of the BHA in each electrode (i.e. multiple channels). Input signals are left unchanged to preserve their correlation structure. For all analyses we construct 10,000 surrogate outputs. Surrogate TRFs are computed as above using the surrogate output signals. We then determine which channels have time points in the TRF that are significantly different from surrogate TRFs. We refer to it as a “significant response”. This is done separately for each input (regressor). Thus, for a given channel, we may find a significant response for saccades (i.e. the saccade TRF has a significant time point) or we may find a significant response for cuts, or the responses may be significant for both (i.e. both TRFs have a significant time point). Corrections for multiple comparisons across time points and channels are addressed through cluster-based statistics (*87*). Significant clusters are determined in two steps. First, we determine a test statistic α for each time point as the

proportion of surrogate TRFs that have a more extreme amplitude than the original TRF. Clusters are defined as connected time points and channels on a shaft/grid that satisfy the test statistic of $\alpha < 0.001$. For each cluster a weight is computed as the squared amplitude of the TRF summed over the cluster. Second, a distribution of surrogate weights is found for each electrode by taking the maximum weight in each electrode. For each cluster in the original data we compute p-values as the proportion of surrogate weights that is larger than the cluster weight. Finally, the p-values for all clusters in all electrodes and patients are corrected for multiple comparisons using the Benjamini-Hochberg procedure implemented in mafdr() in MATLAB (*88*) at a false discovery rate of $q < 0.05$. Clusters with corrected p-values above 0.05 are considered significant. For example, 1151 channels showed significant responses to film cuts (Figure 2 & S3). At a FDR of 0.05, this means that on average, 58 channels may be false discoveries.

To determine the extent of responses to cuts, saccades and motion we compute the ratio of channels with significant responses in each brain area and patient (Figure 2D). Within each brain area we test whether the differences between the ratio of responsive channels to different regressors are statistically significant across patients with a Wilcoxon rank sum test. We correct for multiple comparisons across all p-values with the Benjamini-Hochberg procedure implemented in mafdr() in MATLAB (*88*) at a false discovery rate of $q < 0.05$.

Response amplitude

To estimate the amplitude of the BHA responses in each channel to individual film cuts (or saccades) we fit the TRFs to the neural data in the same window around the specific film cut (or saccade) (Figure S8). The amplitude is estimated using ordinary least squares regression. The regression coefficient describes the factor the TRF is multiplied with to best fit the neural data.

To test whether response amplitudes differ between conditions, for example event and continuous cuts, we conduct a mixed-design ANOVA at the level of individual channels. We model the fixed effect of condition (e.g. event or continuous cuts), fixed effect of brain region and random effect of patients as independent variables. To test in which regions response amplitudes differ between conditions we perform a follow-up comparison in each brain region with a mixed-design ANOVA at the level of channels. Here, we model the fixed effect of condition with the patient as a random effect. We control p-values for the main effects of condition and patients for multiple comparisons using the Benjamini-Hochberg procedure at a false discovery rate of $q < 0.05$.

Specificity of responses

To measure whether responses are specific to condition (e.g. high vs low novelty in saccade target) we define a measure of specificity. Assume N_1 channels respond to condition 1, N_2 channels respond to condition 2, and among all those N_{12} are responsive to both. Then specificity is defined here as a the fraction of responsive channels that respond to only one of the two conditions:

$$Specificity = \frac{N_1 + N_2 - N_{12}}{N_1 + N_2} = 1 - \frac{N_{12}}{N_1 + N_2}$$

The second expression makes clear that this definition is one minus the Jaccard index, which is a standard measure of overlap. A specificity of 1 indicates that there is no overlap in responsive channels, while a specificity of 0 indicates that any channel responding to one condition also responds to the other.

To test whether specificity differs between brain regions we apply a Skillings-Mack test (*89*), which is a non-parametric alternative to a two-way ANOVA for unbalanced data. This test determines statistical significance of the fixed effect of brain regions on specificity, while controlling for the random effect of patients.

If responsive channels were selected at random to respond to either condition, we may find by chance some level of specificity. To test whether the observed level of specificity in a given brain region is larger than this chance value we construct a surrogate distribution of specificity by shuffling the assignment of responsive channels to a condition at random. For instance if 7 channels in a given brain region and patient respond to event cuts, 5 respond to continuous cuts and 2 to both, we create 1,000 permutations of 10 channels. In each permutation the number of channels responding to event cuts, continuous cuts or both is randomly assigned. This results in a surrogate distribution of median specificity in each brain area. We test the median across patients of the original data against this surrogate data to obtain a p-value. P-values are corrected for multiple comparisons across brain regions with the Benjamini-Hochberg procedure at a false discovery rate of $q < 0.05$. We remove regions with only 1 responsive channel in any given patient from analysis, because no distinct permutations can be constructed.

Removal of saccadic spike artifacts

Significant TRFs to saccades often consist of sharp spikes at the time of the saccade (Figure S24B). These channels are localized close to the orbit of the eyes and are likely artifacts of muscle movements Figure S24C) (*90*). To remove these channels from all further analysis we construct a correlation matrix between all significant TRFs (Figure S24A). We then cluster this correlation matrix to visually identify the group of channels with saccadic spikes (Figure S24A&B) (*40*). Saccade related artifacts are identified and removed from all analyses with this method.

Registration of responses to cortical atlases

Our analysis is based on bipolar channel pairs. In the spatial plots all contacts that are part of a significant bipolar channel pair are included. In order to localize effects to specific anatomically or functionally defined brain areas we adapt the pipeline developed by Gao et al. (*30*) (https://github.com/rdgao/field-echos). Electrode coordinates from the iElvis pipeline described above are transformed from MNI305 to MNI152 space with an affine transformation (*91*). In Figure 1A electrodes within 4 mm distance to a parcel of the Glasser atlas are counted. For maps of weighted average responses in Figures 2B, 3B, 4B, 5B and S4A we follow the approach described in Gao et al. (*30*). Briefly, responses of all electrodes close to a parcel are weighted by distance following a Gaussian with full-width-half-max of 4 mm. Parcels where the highest weighted count is below 0.5 are excluded from analysis (gray parcels in plots). The pipeline in Gao et al has been implemented with the Glasser atlas (*31*). When analyzing specificity, responsive electrodes are sparse. To address this we utilize the coarser Desikan-Killiany atlas (Figures 3E, 4E, 5E, S13, S18 and S21). Voxel resolution T1w/T2w ratio data is available through the neuromaps toolbox (*92*). We transform this data onto the freesurfer surface using the neuromaps toolbox. This enables us to average T1w/T2w data within parcels of the Desikan-Killiany atlas available through freesurfer (*46*). We follow the same approach to obtain functional gradients in the Glasser and Desikan-Killiany parcellation (Figures S13, S18, S21).

Visualization

For visualization of the time-course in Fig. 2A-top, each TRF is z-scored and baseline corrected individually for each channel and each regressor (also corresponding panel in S3, S11, S16). The baseline is computed between 2.1s and 3s after stimulus onset (film cuts, saccades or motion). Electrodes are grouped in clusters of similar waveforms (*40*, *41*). For Figure 2A-bottom we average similar TRF waveforms in the same cluster. Here, the scale is the same across all regressors for better comparison. For Figure 2B we compute the maximum amplitude across time (preserving the sign) and average over all channels within a parcel of the Glasser atlas (*31*). Positive values (red) indicate a dominant enhancement of TRF in that area. Negative values (blue) indicate a dominant suppression. Zero values (white) indicate that enhancement and suppression are similarly strong.

Correlation to gradients

Registering results to standardized brain maps allows correlation of our results to anatomical and functional gradients. For brain maps in the coarser Desikan-Killiany atlas we compute the Spearman correlation coefficient and the associated p-value assuming that these larger parcelations provide independent measures. For brain maps in the finer Glasser atlas we follow Gao et al. and compute significance using permutation statistics taking into consideration spatial autocorrelation (*30*).

Funding
National Institutes of Health grants R01MH111439, P50MH109429, and R01CA247910
National Science Foundation grant DRL-1660548

Author contributions
Conceptualization: MN, LCP, CES; Methodology: MN, ML, BER, LCP; Software: MN, LH, NM, KS; Formal Analysis: MN; Data Curation: MN, KS; Data Collection: MN, ML, SB; Visualization: MN, NM; Supervision: LCP, CES, AM; Writing – original draft: MN, LCP; Writing – review & editing: MN, ML, BER, LH, NM, CES, SB, LCP

Competing interests
Authors declare that they have no competing interests.

Data and materials availability
Result data is available on OSF (https://osf.io/n6vpc/) and code for analysis provided on github (https://github.com/MaxNentwich/semantic-novelty). There are also plans for a future release of the entire raw data including other experiments that were performed with these patients. This will be part of a separate data-release publication involving a larger group of authors in this consortium effort.

Acknowledgments
The authors thank Atanas Stankov and Jens Madsen for their technical and intellectual input concerning data analysis, interpretation and core scientific ideas.

We thank Elizabeth Espinal, Sabina Gherman and Gelana Tostaeva for their assistance in data collection.

Additionally, the authors thank Tejaswini Sudhakar, Mohigul Nasimova, Raydi Camilo Jimenez, Rasha Hussain, Nicole Zheng, Justin Lam and Samantha Lee for tirelessly annotating faces in all movies.

References

1. S. A. Nastase, A. Goldstein, U. Hasson, Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*. **222**, 117254 (2020).
2. S. S. Cohen, S. Henin, L. C. Parra, Engaging narratives evoke similar neural activity and lead to similar time perception. *Sci. Rep.* **7**, 4578 (2017).
3. J. P. Dmochowski, J. J. Ki, P. DeGuzman, P. Sajda, L. C. Parra, Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity. *NeuroImage*. **180**, 134–146 (2018).
4. B. E. Russ, D. A. Leopold, Functional MRI mapping of dynamic visual features during natural viewing in the macaque. *NeuroImage*. **109**, 84–94 (2015).
5. L. Isik, J. Singer, J. R. Madsen, N. Kanwisher, G. Kreiman, What is changing when: Decoding visual information in movies from human intracranial recordings. *NeuroImage*. **180**, 147–159 (2018).
6. J. Zheng, A. G. P. Schjetnan, M. Yebra, B. A. Gomes, C. P. Mosher, S. K. Kalia, T. A. Valiante, A. N. Mamelak, G. Kreiman, U. Rutishauser, Neurons detect cognitive boundaries to structure episodic memories in humans. *Nat. Neurosci.* **25**, 358–368 (2022).
7. T. Golan, I. Davidesco, M. Meshulam, D. M. Groppe, P. Mégevand, E. M. Yeagle, M. S. Goldfinger, M. Harel, L. Melloni, C. E. Schroeder, L. Y. Deouell, A. D. Mehta, R. Malach, Increasing suppression of saccade-related transients along the human visual hierarchy. *eLife*. **6**, e27819 (2017).
8. A. Barczak, S. Haegens, D. A. Ross, T. McGinnis, P. Lakatos, C. E. Schroeder, Dynamic Modulation of Cortical Excitability during Visual Active Sensing. *Cell Rep.* **27**, 3447-3459.e3 (2019).
9. M. Leszczynski, L. Chaieb, T. Staudigl, S. J. Enkirch, J. Fell, C. E. Schroeder, Neural activity in the human anterior thalamus during natural vision. *Sci. Rep.* **11**, 17480 (2021).
10. T. Staudigl, J. Minxha, A. N. Mamelak, K. M. Gothard, U. Rutishauser, Saccade-related neural communication in the human medial temporal lobe is modulated by the social relevance of stimuli. *Sci. Adv.* **8**, eabl6037.
11. M. Leszczynski, C. E. Schroeder, The Role of Neuronal Oscillations in Visual Active Sensing. *Front. Integr. Neurosci.* **13** (2019) (available at https://www.frontiersin.org/article/10.3389/fnint.2019.00032).
12. C. Rajkai, P. Lakatos, C.-M. Chen, Z. Pincze, G. Karmos, C. E. Schroeder, Transient Cortical Excitation at the Onset of Visual Fixation. *Cereb. Cortex*. **18**, 200–209 (2008).
13. A. M. Bartlett, S. Ovaysikia, N. K. Logothetis, K. L. Hoffman, Saccades during Object Viewing Modulate Oscillatory Phase in the Superior Temporal Sulcus. *J. Neurosci.* **31**, 18423–18432 (2011).
14. M. J. Jutras, P. Fries, E. A. Buffalo, Oscillatory activity in the monkey hippocampus during visual exploration and memory formation. *Proc. Natl. Acad. Sci.* **110**, 13144–13149 (2013).
15. T. Staudigl, E. Hartl, S. Noachtar, C. F. Doeller, O. Jensen, Saccades are phase-locked to alpha oscillations in the occipital and medial temporal lobe during successful memory encoding. *PLOS Biol.* **15**, e2003404 (2017).
16. M. Leszczynski, S. Bickel, M. Nentwich, B. E. Russ, L. Parra, P. Lakatos, A. Mehta, C. E. Schroeder, Saccadic modulation of neural excitability in auditory areas of the neocortex (2022), p. 2022.05.24.493336, , doi:10.1101/2022.05.24.493336.
17. T. J. Smith, P. K. Mital, Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *J. Vis.* **13**, 16 (2013).
18. P. J. A. Unema, S. Pannasch, M. Joos, B. M. Velichkovsky, Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Vis. Cogn.* **12**, 473–494 (2005).

19. S. Zuo, L. Wang, J. H. Shin, Y. Cai, B. Zhang, S. W. Lee, K. Appiah, Y. Zhou, S. C. Kwok, Behavioral evidence for memory replay of video episodes in the macaque. *eLife*. **9**, e54519 (2020).
20. J. M. Zacks, N. K. Speer, K. M. Swallow, T. S. Braver, J. R. Reynolds, Event perception: a mind-brain perspective. *Psychol. Bull.* **133**, 273 (2007).
21. J. P. Magliano, J. M. Zacks, The Impact of Continuity Editing in Narrative Film on Event Segmentation: Cognitive Science. *Cogn. Sci.* **35**, 1489–1517 (2011).
22. C. Baldassano, J. Chen, A. Zadbood, J. W. Pillow, U. Hasson, K. A. Norman, Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*. **95**, 709-721.e5 (2017).
23. A. Ben-Yakov, R. N. Henson, Eds., The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. *J. Neurosci.* **38**, 10057–10068 (2018).
24. A. Jafarpour, S. Griffin, J. J. Lin, R. T. Knight, Medial Orbitofrontal Cortex, Dorsolateral Prefrontal Cortex, and Hippocampus Differentially Represent the Event Saliency. *J. Cogn. Neurosci.* **31**, 874–884 (2019).
25. M. Silva, C. Baldassano, L. Fuentemilla, Rapid Memory Reactivation at Movie Event Boundaries Promotes Episodic Encoding. *J. Neurosci.* **39**, 8538–8548 (2019).
26. D. Pitcher, G. Ianni, L. G. Ungerleider, A functional dissociation of face-, body- and scene-selective brain areas based on their response to moving and static stimuli. *Sci. Rep.* **9**, 1–9 (2019).
27. T. Ball, M. Kern, I. Mutschler, A. Aertsen, A. Schulze-Bonhage, Signal quality of simultaneously recorded invasive and non-invasive EEG. *NeuroImage*. **46**, 708–716 (2009).
28. J. Parvizi, S. Kastner, Promises and limitations of human intracranial electroencephalography. *Nat. Neurosci.* **21**, 474–483 (2018).
29. T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A Simple Framework for Contrastive Learning of Visual Representations. *ArXiv200205709 Cs Stat* (2020) (available at http://arxiv.org/abs/2002.05709).
30. R. Gao, R. L. van den Brink, T. Pfeffer, B. Voytek, Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. *eLife*. **9**, e61277 (2020).
31. M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, S. M. Smith, D. C. Van Essen, A multi-modal parcellation of human cerebral cortex. *Nature*. **536**, 171–178 (2016).
32. A. Petroni, S. S. Cohen, L. Ai, N. Langer, S. Henin, T. Vanderwal, M. P. Milham, L. C. Parra, The Variability of Neural Responses to Naturalistic Videos Change with Age and Sex. *eNeuro*. **5** (2018), doi:10.1523/ENEURO.0244-17.2017.
33. B. E. Russ, T. Kaneko, K. S. Saleem, R. A. Berman, D. A. Leopold, Distinct fMRI Responses to Self-Induced versus Stimulus Motion during Free Viewing in the Macaque. *J. Neurosci.* **36**, 9580–9589 (2016).
34. R. Mukamel, H. Gelbard, A. Arieli, U. Hasson, I. Fried, R. Malach, Coupling Between Neuronal Firing, Field Potentials, and fMRI in Human Auditory Cortex. *Science*. **309**, 951–954 (2005).
35. M. Leszczyński, A. Barczak, Y. Kajikawa, I. Ulbert, A. Y. Falchier, I. Tal, S. Haegens, L. Melloni, R. T. Knight, C. E. Schroeder, Dissociation of broadband high-frequency activity and neuronal firing in the neocortex. *Sci. Adv.* **6**, eabb0977 (2020).
36. M. J. Crosse, G. M. Di Liberto, A. Bednar, E. C. Lalor, The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Front. Hum. Neurosci.* **10** (2016), doi:10.3389/fnhum.2016.00604.
37. O. Dimigen, B. V. Ehinger, Regression-based analysis of combined EEG and eye-tracking data: Theory and applications. *J. Vis.* **21**, 3–3 (2021).

38. A. D. Stankov, J. Touryan, S. Gordon, A. J. Ries, J. Ki, L. C. Parra, *bioRxiv*, in press, doi:10.1101/2021.02.11.430486.
39. M. A. Sommer, R. H. Wurtz, A Pathway in Primate Brain for Internal Monitoring of Movements. *Science*. **296**, 1480–1482 (2002).
40. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, B. Fischl, H. Liu, R. L. Buckner, The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
41. D. Lashkari, E. Vul, N. Kanwisher, P. Golland, Discovering structure in the space of fMRI selectivity profiles. *NeuroImage*. **50**, 1085–1098 (2010).
42. M. Nau, A. Schindler, A. Bartels, Real-motion signals in human early visual cortex. *NeuroImage*. **175**, 379–387 (2018).
43. S. S. Cohen, N. Tottenham, C. Baldassano, Developmental changes in story-evoked responses in the neocortex and hippocampus. *eLife*. **11**, e69430 (2022).
44. X.-J. Wang, Macroscopic gradients of synaptic excitation and inhibition in the neocortex. *Nat. Rev. Neurosci.* **21**, 169–178 (2020).
45. D. S. Margulies, S. S. Ghosh, A. Goulas, M. Falkiewicz, J. M. Huntenburg, G. Langs, G. Bezgin, S. B. Eickhoff, F. X. Castellanos, M. Petrides, E. Jefferies, J. Smallwood, Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci.* **113**, 12574–12579 (2016).
46. R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, R. J. Killiany, An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*. **31**, 968–980 (2006).
47. Z. Wen, Y. Li, Toward Understanding the Feature Learning Process of Self-supervised Contrastive Learning (2021), (available at http://arxiv.org/abs/2105.15134).
48. B. Duchaine, G. Yovel, A Revised Neural Framework for Face Processing. *Annu. Rev. Vis. Sci.* **1**, 393–416 (2015).
49. N. Kanwisher, J. McDermott, M. M. Chun, The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *J. Neurosci.* **17**, 4302–4311 (1997).
50. K. Nagy, M. Greenlee, G. Kovács, The Lateral Occipital Cortex in the Face Perception Network: An Effective Connectivity Study. *Front. Psychol.* **3** (2012) (available at https://www.frontiersin.org/article/10.3389/fpsyg.2012.00141).
51. M. Nau, J. B. Julian, C. F. Doeller, How the Brain's Navigation System Shapes Our Visual Experience. *Trends Cogn. Sci.* **22**, 810–825 (2018).
52. A. Bartels, S. Zeki, N. K. Logothetis, Natural Vision Reveals Regional Specialization to Local Motion and to Contrast-Invariant, Global Flow in the Human Brain. *Cereb. Cortex*. **18**, 705–717 (2008).
53. Y. Pu, X.-Z. Kong, C. Ranganath, L. Melloni, Event boundaries shape temporal organization of memory by resetting temporal context. *Nat. Commun.* **13**, 622 (2022).
54. D. M. Wolpert, S. J. Goodbody, M. Husain, Maintaining internal representations: the role of the human superior parietal lobe. *Nat. Neurosci.* **1**, 529–533 (1998).
55. K.-H. Lu, S.-C. Hung, H. Wen, L. Marussich, Z. Liu, Influences of High-Level Features, Gaze, and Scene Transitions on the Reliability of BOLD Responses to Natural Movie Stimuli. *PLOS ONE*. **11**, e0161797 (2016).
56. E. Podvalny, Erin Yeagle, P. Mégevand, N. Sarid, M. Harel, G. Chechik, A. D. Mehta, R. Malach, Invariant Temporal Dynamics Underlie Perceptual Stability in Human Visual Cortex. *Curr. Biol.* **27**, 155–165 (2017).
57. S. K. Miura, M. Scanziani, Distinguishing externally from saccade-induced motion in visual cortex. *Nature*. **610**, 135–142 (2022).
58. C.-C. Wu, H.-C. Wang, M. Pomplun, The roles of scene gist and spatial dependency

among objects in the semantic guidance of attention in real-world scenes. *Vision Res.* **105**, 10–20 (2014).
59. D. Ernst, S. Becker, G. Horstmann, Novelty competes with saliency for attention. *Vision Res.* **168**, 42–52 (2020).
60. M. Dorr, T. Martinetz, K. R. Gegenfurtner, E. Barth, Variability of eye movements when viewing dynamic natural scenes. *J. Vis.* **10**, 28–28 (2010).
61. J. Madsen, L. C. Parra, Cognitive processing of a common stimulus synchronizes brains, hearts, and eyes. *PNAS Nexus*. **1**, pgac020 (2022).
62. O. C. Lancry-Dayan, G. Kupershmidt, Y. Pertzov, Been there, seen that, done that: Modification of visual exploration across repeated exposures. *J. Vis.* **19**, 2 (2019).
63. M. Ibbotson, B. Krekelberg, Visual perception and saccadic eye movements. *Curr. Opin. Neurobiol.* **21**, 553–558 (2011).
64. J. V. Haxby, E. A. Hoffman, M. I. Gobbini, The distributed human neural system for face perception. *Trends Cogn. Sci.* **4**, 223–233 (2000).
65. M. Ozker, D. Yoshor, M. S. Beauchamp, Frontal cortex selects representations of the talker's mouth to aid in speech perception. *eLife*. **7**, e30387 (2018).
66. V. B. Penhune, R. J. Zatorre, J. D. MacDonald, A. C. Evans, Interhemispheric Anatomical Differences in Human Primary Auditory Cortex: Probabilistic Mapping and Volume Measurement from Magnetic Resonance Scans. *Cereb. Cortex*. **6**, 661–672 (1996).
67. C. E. Schroeder, P. Lakatos, Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* **32**, 9–18 (2009).
68. P. Lakatos, A. S. Shah, K. H. Knuth, I. Ulbert, G. Karmos, C. E. Schroeder, An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *J. Neurophysiol.* **94**, 1904–1911 (2005).
69. A. Yagi, Saccade size and lambda complex in man. *Physiol. Psychol.* **7**, 370–376 (1979).
70. C. S. Lee, M. Aly, C. Baldassano, Anticipation of temporally structured events in the brain. *eLife*. **10**, e64972 (2021).
71. D. H. Brainard, The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
72. D. M. Groppe, S. Bickel, A. R. Dykstra, X. Wang, P. Mégevand, M. R. Mercier, F. A. Lado, A. D. Mehta, C. J. Honey, iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. *J. Neurosci. Methods*. **281**, 40–48 (2017).
73. A. M. Dale, B. Fischl, M. I. Sereno, Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*. **9**, 179–194 (1999).
74. B. Fischl, M. I. Sereno, A. M. Dale, Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System. *NeuroImage*. **9**, 195–207 (1999).
75. X. Papademetris, M. P. Jackowski, N. Rajeevan, M. DiStasio, H. Okuda, R. T. Constable, L. H. Staib, BioImage Suite: An integrated medical image analysis suite: An update. *Insight J.* **2006**, 209 (2006).
76. B. Fischl, M. I. Sereno, R. B. H. Tootell, A. M. Dale, High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* **8**, 272–284 (1999).
77. B. K. P. Horn, B. G. Schunck, Determining optical flow. *Artif. Intell.* **17**, 185–203 (1981).
78. J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, J. K. Lindeløv, PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods*. **51**, 195–203 (2019).
79. J. Zheng, A. Gómez Palacio Schjetnan, M. Yebra, C. Mosher, S. Kalia, T. A. Valiante, A. N. Mamelak, G. Kreiman, U. Rutishauser, "Cognitive boundary signals in the human medial temporal lobe shape episodic memory representation" (preprint, Neuroscience, 2021), , doi:10.1101/2021.01.16.426538.
80. A. Townsend, JPEG: Image compression algorithm (2017), (available at http://pi.math.cornell.edu/~web6140/TopTenAlgorithms/JPEG.html#19).
81. Y. Wu, A. Kirillov, F. M. Massa, W.-Y. Lo, R. Girshick, Detectron2 (2019), (available at

https://github.com/facebookresearch/detectron2).
82. S. Xie, R. Girshick, P. Dollar, Z. Tu, K. He, "Aggregated Residual Transformations for Deep Neural Networks" in (2017; https://openaccess.thecvf.com/content_cvpr_2017/html/Xie_Aggregated_Residual_Transformations_CVPR_2017_paper.html), pp. 1492–1500.
83. K. He, G. Gkioxari, P. Dollar, R. Girshick, "Mask R-CNN" in (2017; https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html), pp. 2961–2969.
84. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, "Microsoft COCO: Common Objects in Context" in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars, Eds. (Springer International Publishing, Cham, 2014), *Lecture Notes in Computer Science*, pp. 740–755.
85. B. Wang, D. Mayo, A. Deza, A. Barbu, C. Conwell, On the use of Cortical Magnification and Saccades as Biological Proxies for Data Augmentation. *ArXiv211207173 Cs Q-Bio* (2021) (available at http://arxiv.org/abs/2112.07173).
86. K. J. Friston, A. Holmes, J.-B. Poline, C. J. Price, C. D. Frith, Detecting Activations in PET and fMRI: Levels of Inference and Power. *NeuroImage*. **4**, 223–235 (1996).
87. J. Chumbley, K. Friston, False discovery rate revisited: FDR and topological inference using Gaussian random fields. *NeuroImage*. **44**, 62–70 (2009).
88. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
89. J. H. Skillings, G. A. Mack, On the Use of a Friedman-Type Statistic in Balanced and Unbalanced Block Designs. *Technometrics*. **23**, 171–177 (1981).
90. K. Jerbi, S. Freyermuth, S. Dalal, P. Kahane, O. Bertrand, A. Berthoz, J.-P. Lachaux, Saccade Related Gamma-Band Activity in Intracerebral EEG: Dissociating Neural from Ocular Muscle Activity. *Brain Topogr.* **22**, 18–23 (2009).
91. CoordinateSystems - Free Surfer Wiki, (available at https://surfer.nmr.mgh.harvard.edu/fswiki/CoordinateSystems).
92. R. D. Markello, J. Y. Hansen, Z.-Q. Liu, V. Bazinet, G. Shafiei, L. E. Suárez, N. Blostein, J. Seidlitz, S. Baillet, T. D. Satterthwaite, M. M. Chakravarty, A. Raznahan, B. Misic, neuromaps: structural and functional interpretation of brain maps. *Nat. Methods*. **19**, 1472–1479 (2022).