FRONT MATTER

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

Authors

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

Affiliations

[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] Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine, New York, USA.
[5] Department of Psychiatry, New York University at Langone, New York, USA.
[6] The Feinstein Institutes for Medical Research, Northwell Health, Manhasset, USA.
[7] Departments of Neurosurgery and Neurology, Zucker School of Medicine at Hofstra/Northwell, Manhasset, USA.

Correspondence:
Maximilian Nentwich and Lucas C. Parra
Department of Biomedical Engineering, The City College of New York
Email: max.nentwich@gmail.com; parra@ccny.cuny.edu

Abstract

Our continuous visual experience in daily life is dominated by change. Previous research has focused on the effects of visual motion, eye movements or the transition between events, but has failed to capture their full impact across the brain. Using intracranial recordings in humans, we investigate the neural responses to these sources of novelty during the natural experience of watching film. Responses to saccades and film cuts were much stronger than those to visual motion, extending far beyond traditional visual processing areas. Film cuts associated with semantic event boundaries elicit strong and specific responses in higher-order brain areas. Saccades associated with high visual novelty also elicit strong neural responses. Specific locations in higher-order brain areas show selectivity to either high or low-novelty saccade, as well as face or non-face targets. In summary, visual and semantic novelty drive much of the human brain, while exhibiting specialization to specific forms of novelty.

Teaser

When watching movies, the entire brain responds to film cuts, eye movements and motion, with stronger responses when something new happens.

MAIN TEXT

Introduction

In order 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 scenes.

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). Visual motion is also caused by eye movements, which many studies consider to be a confound (5, 6). Complicating matters, motion in movies can also attract and guide eye movements (7). Regardless, it is clear that eye movements cause neural responses that are distinct from what is expected due to the associated visual change alone (8). Responses to saccades, the rapid movements of the eyes between fixations, were thought to be confined to visual processing areas and largely suppressed in higher-order perceptual areas (9). More recent work, however, indicates that saccades play an important role in the organization of many perceptual and cognitive processes (10–13). Saccades modulate neural responses across the visual system (10, 14, 15), medial temporal lobe (16, 17), non-visual nuclei of the anterior thalamus (11) and even the auditory cortex (18). It is also worth noting that responses to visual motion and saccades are modulated by the semantics of the visual stimulus. For instance, responses in fMRI to 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 (12).

Besides motion and saccades, movies also allow us to investigate the processing of narratives. The theory of event segmentation proposes that continuous narratives are segmented and remembered as discrete events (19). Event boundaries, the moments of change between events, are associated with shifts in brain states as well as transient neural responses (6, 20–23). 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 brain areas (6, 20). 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, 20). Consistent with event segmentation theory, increased activation in MTL following a film cut is predictive of subsequent recall of the preceding event (24). Specific neural responses to different types of cuts have also been observed in single cell data from the MTL (6). For instance, modulation in firing rate in cells that respond to 'hard boundaries' is predictive of later recognition of the subsequent scene.

In summary, the main drivers of visual change -- motion, film cuts and saccades -- elicit neural activity in various visual and higher-order brain areas. However, they have been studied in isolation, so that the relative strength of responses and extent of their responses across the brain is not well established. Given that these sources of visual change interact it is important to analyze them in combination to control for their correlations to one another. We hypothesized that visual change causes strong neural responses that are modulated by semantic novelty in the visual scene. We measured semantic novelty in terms of visual features across saccades using deep-networks (25), and on a higher level,

in terms of event boundaries judged by human observers across film cuts. We found that neural activity related to motion is mostly confined to the low-level visual brain areas, while neural activity related to saccades and film cuts is widespread across the whole brain. In addition, we found that responses to saccades and film cuts are enhanced by semantic novelty. Importantly, we find specific brain areas that respond to high semantic novelty and semantically meaniniful stimuli, in particular faces. However, a distinct set of brain areas responds exclusively to low semantic novelty and non-face saccades.

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 various video clips totaling up to 45 minutes (Figure 1B). We were interested in neural responses to the main sources of visual change in this task, namely visual motion in the videos, film cuts and eye movements (Figure S1). Visual motion was quantified as the magnitude of optical flow averaged across each frame (Figure 1C). Saccades and film cuts were quantified as a series of impulses at saccade onset and the time of film cuts, respectively (Figure 1C). We observed that film cuts are followed by a decrease of saccade frequency with the dip 100ms after cuts and a rebound at 250ms (Figure 1D). In addition, visual motion is larger prior to a saccade (Figure 1E), suggesting that both film cuts and motion drive saccades. We also note that visual motion decreases before film cuts (Figure 1F), as a result of the video editing process. This decrease in motion is particularly prominent in professionally edited movies (Figure S2). In total, the three sources of visual change (motion, saccades and cuts) are clearly correlated to one another: saccade frequency increases after film cuts, an increase in motion attracts saccades, and motion slows down before film cuts.

Responses to low level features in the movies

Based on previous literature we hypothesized that visual change leads 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 (27, 28). 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 for simplicity. To capture neuronal responses, we used a conventional systems identification approach (Figure S3A) (29). Specifically, BHA in each channel is treated as the output of a linear system, with motion, saccades and cuts as its input (30). 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 motion, film cuts and saccades so that each electrode is characterized as responsive to one, several, or none of these sources of visual change. In each brain area a subset of channels shows statistically significant responses. For example, of the 928 channels located 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 (29, 30). That the respective contributions to the neural responses can be successfully disentangled from one another is demonstrated for saccade and film cuts in Figure S3B.

Contrary to our expectation we found the most widespread BHA responses to saccades and film cuts (Figure 2). The visual motion led to responses mostly in the occipital lobe,

while film cuts and saccades also resulted in widespread responses in frontal, temporal and parietal lobes (Figure 2B&D). Across the whole brain, the sets of channels responding to either film cuts, saccades or motion are distinct, with little overlap (Figure 2C). Surprisingly, the responses to saccades and film cuts were also stronger than those to motion in areas such as the precuneus and middle temporal area (Figure S4). Both of these areas have been shown to be important in processing of motion (8, 33). However, in the transverse temporal gyrus (Heschl's gyrus) the largest fraction of responsive channels responded to motion (Figure S4).

Motion predominantly increases BHA across the brain (Figure 2A & S4A), whereas film cuts and saccades differentially increase or decrease activity, especially in higher order areas (Figure 2A & S4A). Duration of responses to saccades, and to a lesser extent, to film cuts increase from occipital towards frontal areas (Figure S4). Responses to film cuts are also higher in amplitude than responses to motion and saccades, especially in higher-order brain areas (Figure 2A & S4).

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

The main observation from Figure 2 is that film cuts drive neural activity throughout the brain. 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 S5A). 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 consistent with data collected in a previous study (34) (Fig. S5B). For each video, we ranked all cuts by agreement, i.e. the fraction of participants that marked the end of a segment within one second after a cut. The film cuts with large agreement are termed 'event cuts' (using change point detection, see Methods). As a result, we selected 57 event cuts out of a total of 561 cuts that were reliably marked as event cuts (Table S2). Among the cuts with low agreement we selected an equal number of cuts matching in low-level visual features, following (6) (see Methods). These cuts were labeled 'continuous cuts'. Continuous cuts lie between narrative event boundaries and are characterized mostly by changes in camera angle or position (Figure 3B).

We first tested whether event cuts lead 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, gives us a measure of the amplitude of the response (Figure S6). For each channel we compute the difference between the average amplitude of responses to event cuts and the average amplitude of responses to continuous cuts. In the temporal lobe and MTL this difference across all channels is significantly larger than zero indicating event cuts lead to stronger responses than continuous cuts (Figure 3A). In all other brain areas the magnitudes of responses to event cuts and continuous cuts are similar on average.

However, the distributions of these magnitude differences appear to be bimodal in the occipital lobe and MTL. Therefore, we suspected that there are distinct responses to event cuts and continuous cuts in these areas. To explore this possibility, we repeat the analysis but now identify separate TRFs for event cuts and continuous cuts in each channel (Figure 3B), i.e. using separate regressors indicating each type of cut. We found channels that

respond selectively to event cuts, channels that respond selectively to continuous cuts, and non-selective channels that respond to both (Figure 3C&D, for full list of TRFs see Figure S9). Most channels in the occipital lobe respond non-selectively to either event or continuous cuts (Figure 3C). In contrast, parietal, temporal and frontal lobes are more selective for event cuts. Selectivity for event cuts is most pronounced in MTL, and particularly for the hippocampus entorhinal and parahippocampal cortex (Figure S10), which is consistent with previous reports of strong fMRI responses to event boundaries in the hippocampus. In most channels both event cuts and continuous cuts increase neural activity (Figure S9). In this analysis 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 S7), or when performing the analysis on different types of videos (Figure S8).

Novelty across saccades

As with film cuts, BHA responses to saccades are widespread across the brain (Figure 2). We hypothesize that these responses are driven by changes in low-level and semantic visual features between the foveal image before and after each saccade. If that is true, we would expect stronger 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 that is pre-trained to extract features that are shared across different image patches (25). These features tend to capture semantic properties of objects in the images (35). 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 S11A&B). In doing so, each saccade is associated with a numerical value indicating the novelty of the upcoming fixation. Interestingly, on average the novelty of observed saccades was larger than emulated saccades (random fixation pairs matched in distance and direction to observed saccades) (Figure S11C, p=10-22 N=55,334). 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 S12).

Next, we divided all 55,334 saccades from all 23 patients and videos into two equally sized groups with high and low novelty, while controlling for saccade amplitude and excluding saccades across cuts (Figure 4B). We predicted that saccades with high novelty will result in stronger BHA responses. We estimate the magnitude of the response to each saccade as before (Figure S6). In the occipital, temporal and frontal lobes saccades with high novelty were associated with stronger neural responses (Figure 4A). We also computed separate TRFs for saccades with high and low novelty (Figure 4B). In the occipital lobe the majority of channels respond non-selectively to either high or low novelty. In contrast, for higher-order brain areas most channels respond selectively to either high or low novelty saccades (Figure 4C&D). This suggests that there is a specialization for high novelty, but also, a specialization for low-novelty saccades. These low-novelty saccades target areas of the scene that are semantically similar to the current gaze point. Interestingly, in the parietal and frontal lobes, some channels responding to low-novelty saccades show an inhibition of neural activity (Figure S13). In contrast, channels responding to high-novelty saccades show only increases in neural activity.

Saccades to faces

Another way to quantify semantic changes across saccades is by the content of the saccade target. Specifically, saccades to faces have been reported to have stronger neural responses than saccades to other objects (12). 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) and analyze their neural responses (Figure S15).

As above we calculate the difference in the magnitude of responses to face and non-face saccades. Surprisingly, the responses to non-face saccades have a larger magnitude than face-saccades in the occipital, parietal, mediotemporal (MTL) and frontal lobe (Figure 5A). This is in particular surprising because saccades to faces have a higher novelty (p=1.6*10-60, Nface=7636, Nnon-face=7784, Mann-Whitney U-test), so if anything, the opposite effect would have been expected. Most channels throughout the brain respond selectively to either face or non-face saccades (Figure 5C). Face saccades dominate responses primarily in the temporal lobe. Interestingly, in the superior temporal gyrus, which contains the auditory cortex, the largest fraction of responsive channels is to face-saccades (Figure 5D & S16). Overall, these results indicate that faces and other objects are processed selectively in brain areas extending far beyond traditional visual processing areas.

Face Motion

Optical flow captures various different types of visual motion in the videos. This includes global movement of the scene due to camera movements and the movement of characters (36). In primates, socially relevant stimuli accounts for the majority of neural activity related to motion, e.g. monkeys watching other monkeys (4). To capture socially relevant motion we used the same face annotations as before and computed the motion of faces throughout the movie (Figure 6A).

Using total motion and face motion as regressors we found that a larger fraction of channels respond to face motion over total motion (Figure 6B&C). This includes the lateral occipital cortex, superior parietal lobe and fusiform gyrus (Figure 6B & S17). The lateral occipital lobe and fusiform gyrus in particular are areas known to be involved in face processing (37–39). Therefore, the responses to face motion, likely capture known face processing areas. In contrast, channels in the MTL, particularly the hippocampus, respond predominantly to total motion and not face motion Figure (6B and S17). Since total motion also captures motion of the camera, these responses could be related to spatial remapping. Spatial remapping is necessary when neural representations of locations in the visual scene need to be reassigned after movement of the scene.

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 allows one to link 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 allows us to disentangle the observed correlation between visual motion, saccades, and film cuts. We found that low-level and semantic visual changes across film cuts and saccades are processed by distinct, widespread neural populations across the whole brain. In contrast, visual motion is processed in a more confined area of visual brain regions. In the following we will put our results in context of the specific literature on motion, film cuts, and saccades.

Motion

Motion processing has been studied with various simple visual stimuli, such as moving dots. Several specific brain areas within the occipital, temporal and parietal lobes have been identified to be important in processing motion in this context, for instance V3, the medial temporal area (MT), or the ventral intraparietal area (VIP) (33). Motion processing in these areas has been confirmed for naturalistic stimuli (3, 4, 36). However, motion in naturalistic stimuli activates much wider areas across the temporal lobe (3, 4, 8). Here, we confirm these results in intracranial EEG data recorded from human subjects, where we find widespread responses to optical flow in temporal and parietal lobes (Figure 2). The widespread responses to motion in naturalistic conditions have been shown to be largely driven by motion of socially relevant stimuli, such as faces or body parts (4). We confirm this observation in our data, where motion of faces dominated neural responses compared to total motion (Figure 6). Specifically, face motion activates distinct clusters of channels in the lateral occipital cortex, fusiform gyrus and superior temporal sulcus (Figure 6 & S17), which include known face-processing areas (37–39). These areas are also known to respond stronger to moving faces (4, 37, 40). Some areas, however, respond specifically to total motion, notably the Hippocampus (Figure S17). The total motion regressor also includes camera movement, which requires remapping of spatial representations, in which the hippocampus plays a major role (41). Importantly, responses to motion are more confined to visual areas compared to responses to film cuts and saccades, which are more widespread in higher order brain areas. Our direct comparison of different types of visual change in movies thus shows that motion in naturalistic stimuli is perhaps a less dominant driver of neural responses than previously thought (3, 4). An interesting exception is the transverse temporal gyrus (Heschl's gyrus), where a larger fraction of channels responds to optical flow than to film cuts and saccades (Figure S4). The channels in the Heschl's gyrus and the superior temporal gyrus, comprising the auditory cortex (42), specifically respond to total motion as opposed to face motion (Figure S17). These responses, therefore, may represent processing of the sounds related to moving objects or characters, highlighting the multimodal nature of the auditory cortex (43–47).

Event boundaries

The other obvious source of visual change in film are cuts. Film makers use cuts to change view angle, but also to change location and time between scenes. Such scene cuts not only change low-level visual content but also change in semantics, which may be perceived by viewers as a boundary between distinct events. The analysis of film cuts thus naturally links to the concept of "event boundaries", and we defined "continuous cuts" versus "event cuts" based on standard event segmentation (19, 20). Many previous studies

have focused on the medial temporal lobe because of the proposed role of event boundaries in organizing memory of continuous experience (6, 19, 21, 22, 34, 48). Most recently, Zheng et al., have analyzed different types of film cuts as a proxy of cognitive boundaries (6). They identified neurons in the medial temporal lobe with specific responses to film cuts within and between different video clips. Our data shows similar specificity of channels in the medial temporal lobe to event and continuous cuts. However, we also find a large number of channels with specific responses to event cuts widely distributed across the brain outside of the medial temporal lobe (Figure 3C&D). Other work with intracranial recordings in humans has been able to decode scene identity from recording locations widespread across the whole temporal lobe (5). Specific responses to event salience have also been identified in the orbitofrontal cortex (23). Our data, with wide coverage and rich naturalistic stimuli, shows that event boundaries are indeed processed in distributed areas across the whole brain (Figure 3C&D). This view is supported by work on event boundaries with fMRI data demonstrating that different features of events are processed in successive stages of the visual processing hierarchy (20, 21). Notably, Magliano and Zacks report that film cuts with action discontinuities activate parietal and frontal areas, while film cuts in general strongly activate visual areas (20). Additionally, in electrophysiological data, information flow from the hippocampus to cortical region has been shown to be timed to event boundaries (49). Together, this suggests that event boundaries have a broad impact on the brain, and that film cuts are a particularly effective tool to study the effect of event boundaries on brain and cognition.

Saccades

Saccades are often studied in the context of visual processing, for example saccadic suppression in early visual areas (50). In higher-order areas effects of saccades are thought to be suppressed as well, which is reflected in the observation that shifts in the retinal image across saccades are not consciously perceived (9). Our findings, however, support the view that saccades modulate neural activity across most, if not all the brain (Figure 2) (18). These results extend previous work that has found modulation of neural activity by saccades in the non-visual thalamus (11), medial temporal lobe (16, 17), auditory, frontal and parietal cortices (18). These findings collectively suggest that saccades are essential in the organization of perceptual and cognitive processes across the whole brain, which has been suggested for processes such as attention and memory (13). This claim is further supported by our findings that responses to saccades depend on the type of semantic change.

Saccade novelty

There is competing evidence that saccades sample either semantically similar (51), or semantically dissimilar objects in static natural scenes (52). We show that, in dynamic natural scenes, both types of saccades occur at different times.

Visual representations that are maintained across saccades have been proposed to consist of rough schemas (53). In natural scenes, eye movements have been shown to target objects with similar semantics (51). We propose that saccades with low semantic novelty mainly sample semantically similar locations in order to build congruent schemas across saccades. This claim is supported by the widespread responses specific to saccades with low semantic novelty we find in our data (Figure 4).

In contrast, saccades are also attracted by visual features that are novel in the context of the scene (52). Similarly, EEG saccade-evoked responses show differences for saccades to objects that are semantically congruent or incongruent with the context of a scene (54). We propose that saccades with high semantic novelty are associated with the shift of attention to novel and semantically incongruent objects in a visual scene. We find that different sets of channels are recruited to process information across saccades with high and low semantic novelty (Figure 4).

Previous work on saccade guidance by either semantically similar, or semantically dissimilar objects used a confined set of static visual scenes of objects or abstract visual stimuli (51–54). These paradigms constrain saccade behavior to either type. In movies, scenes change rapidly and we find that both types of saccades can be observed. Saccades targeting semantically similar or dissimilar objects show distinct patterns of neural responses across the brain (Figure 4). Our results, therefore, support the notion that multiple behaviors compete for eye movement control in natural environments (55). Construction of a coherent schema of a visual scene competes with the necessity to sample novel objects in dynamic environments. Consistent with this we find that after film cuts, saccades are directed primarily to novel targets.

In addition to the construction of schemas of visual representations, low-novelty saccades suppress neural processing in in the parietal and frontal lobes (Figure S13). These effects could be indicative of mechanisms to maintain perceptual stability across eye movements (56).

Saccades to faces

Faces contain a wide variety of features that are processed in distinct specialized areas (37). 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 (57). In support of this model, we find that saccades to faces and other objects are processed by distinct sets of channels across the whole brain (Figure 5). Notably we find specific sets of channels that respond to face and non-face saccades in the medial temporal lobe (Figure S16). Similarly, responses in distinct groups of neurons in the medial temporal lobe to saccades to faces and other objects were recently reportent by Staudigl et al. (12). They suggested that that these responses reflect the coordination of detecting socioemotional features and memory. We also find specific responses to saccades to faces in auditory areas, such as the STS, and the frontal cortex (Figure S16). Excitability in the auditory system was found to be enhanced late during a fixation and during saccades by Leszczynski et al. (18). In the frontal cortex, responses to the talker's mouth have been reported (58). Since faces in movies are associated with speech, we propose that the specific responses to saccades to faces in auditory and frontal areas are related to speech processing. 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

The relatively weak responses to optical flow in our data could be due to several factors. First, our motion feature is unspecific, capturing global and local motion. It has been shown that specifically the motion of socially relevant objects, such as faces is associated with strong responses (4, 40). However, face motion in our study only captures a small part of the socially relevant motion defined by Russ et al (4), which includes motion of the whole body. Second, electrode coverage in our patient population is not chosen to specifically cover motion sensitive areas. We further did not run any motion localizers to identify motion sensitive channels. 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. Higher-order (non-linear) effects on BHA are not captured. Similarly, we have not

included multiplicative interactions between regressors (motion, saccades and cuts). Techniques for doing this within a system-identification approach are readily available (29, 30), but would lead to significantly more complex exposition. The current approach does however account for, and compensates 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. The same intracranial EEG recordings, however, could be analyzed more extensively for modulation of power in other frequency bands and phase alignment. In other words, one could perform a more thorough analysis of local field potentials, which are a rich source of information on neuronal dynamics (59, 60). As shown by recordings in macaque V1, such analysis may be particularly important when analyzing modulatory top-down effects (10).

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). 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. (10).

Future work

The widespread responses to saccades and film cuts demonstrate the opportunity to study a wide range of sensory and cognitive processes in movies in a broad view. Film cuts are characterized by various changes in semantics. Further characterization of the specific types of change would allow investigation of relevant cognitive processes. For example, event boundaries are associated with changes in space, time or action. These changes are potentially processed by distinct neural areas. In the case of novelty across saccades, we identified that saccades with high and low novelty are processed in distinct brain areas. We propose that these different types of saccades represent the construction of schemas of visual scenes and exploration of new scenes, respectively. Future work can address whether switches between these processes in natural environments are related to switches in attentional states.

In conclusion, taking a data driven approach to analyze intracranial EEG data during movies we found widespread responses 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, extending previously known anatomical locations with functional specificity to these visual features. We argue that film cuts and saccades in movies offer the opportunity to study several cognitive processes, such as attention or memory in naturalistic conditions. Future studies developing relevant memory and attention tasks in combinations with movies can contribute to the understanding of perception and cognition in natural environments.

Materials and Methods

Experimental Design

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 psychotoolbox (61) 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', 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). Inkscapes does not contain film cuts, humans or animals and was therefore not analyzed.

Electrode localization

Each electrode shank/grid contains multiple recording contacts. Contact locations were identified using the iELVis MATLAB toolbox (62). 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 (63, 64). Postoperative CT scans were acquired and coregistered to the freesurfer reconstruction. Contacts were then semi-manually localized using the bioimagesuite software (65). All contacts were then coregistered to the freesurfer fsaverage brain for visualization and assignment to anatomical atlases (66). 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.

Features extraction from videos

Total motion

As a measure of motion we extract optical flow from each video using the Horn-Schunck method as implemented in MATLAB (67). 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 around film cuts with a linear interpolation.

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.

Definition of 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 (68). 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 'space' bar. 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.

To account for reaction time and processing of visual information we subtracted about 1s from the timing of event boundary annotations to match the timing of scene cuts (similar to (22)). Event boundaries from all participants were aggregated in one regressor consisting of impulses at the time of button presses per video. This regressor was then smoothed with a Gaussian of 0.5s. The resulting regressor is a measure of event salience (Figure S5A) (23). Event salience in our data is consistent with salience from data collected by Cohen et al. (Figure S5B&C) (34). 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 the highest event salience above a "change point". For each movie the change point is detected with the findchangepts() function in MATLAB, which 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 'Continuous cuts' is selected from cuts with low event salience. Event cuts and continuous cuts were matched in changes of low-level visual features (69). 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 (70).

Face Detection

We use an object detection algorithm made available through facebook's Detectron2 platform (71). We selected a ResNeXt-101-32x8d model backbone (72) trained in the mask R-CNN framework (73) on the COCO dataset (74) 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.

Face Motion

To estimate face motion we compute the velocity of the centroid of the face annotations from frame to frame. We sum the velocity of all bounding boxes within each frame to capture motion of all faces within a frame.

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 S15A). Other saccades land within a face bounding box, but move away from the center of the face (Figure S15B). Therefore, we generate several handcrafted features and classify face and non-face saccades using an SVM (Figure S15). 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 (25). 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 (25). 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 (75). Here, we compute the feature distance between image patches extracted around gaze position at saccade and fixation onset (Figure S11A). 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 (25). Saccade novelty is then defined as the euclidean distance between the features of the pre- and post-saccadic image patch (Figure S11B). A large distance between features corresponds to high saccade novelty. We divide all saccades into two groups of high and low saccade novelty, 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. 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 (29, 30). Each channel is analyzed individually. The inputs to the system are the time courses for the visual 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, (29)) that maps the stimulus to the BHA signal through a convolution (Figure S3A). 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.

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 S6). 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.

Statistical Analysis

To determine the statistical significance of responses, i.e. predictable fluctuations in BHA, we compute a surrogate distribution of TRFs with time-shuffled input signals (visual motion, saccades, cuts). 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 (76). 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 a 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 (77) 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 S4 & 2D). At a FDR of 0.05, this means that on average, 58 channels may be false discoveries.

Removal of saccadic spike artifacts

Significant TRFs to saccades often consist of sharp spikes at the time of the saccade (Figure S18B). These channels are localized close to the orbit of the eyes and are likely artifacts of muscle movements Figure S18C) (78). To remove these channels from all further analysis we construct a correlation matrix between all significant TRFs (Figure S18A). We then cluster this correlation matrix to visually identify the group of channels

with saccadic spikes (Figure S18) (31). Saccade related artifacts are identified and removed from all analyses with this method.

Visualization

We visualize channel locations with the iELVis MATLAB toolbox (62). All spatial plots are based on the freesurfer fsaverage brain (66). Our analysis is based on bipolar channel pairs. For clearer visualization, in spatial plots all contacts that are part of a significant bipolar channel pair are plotted. We group responsive channels in anatomical regions based on the Desikan-Killiany and Aseg atlases (79, 80). Bar plots depict the number of channel pairs in each anatomical area. When only one channel is in a given anatomical area the channel pair is counted as 0.5 channels in the given area.

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. 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).
8. 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).
9. 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).
10. 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).
11. 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).
12. 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.
13. 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).
14. 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).
15. 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).
16. 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).
17. 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).

18. 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.
19. 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).
20. 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).
21. 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).
22. 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).
23. 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).
24. M. Silva, C. Baldassano, L. Fuentemilla, Rapid Memory Reactivation at Movie Event Boundaries Promotes Episodic Encoding. J. Neurosci. 39, 8538–8548 (2019).
25. 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).
26. 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.
27. 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).
28. 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).
29. 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.
30. O. Dimigen, B. V. Ehinger, Regression-based analysis of combined EEG and eye-tracking data: Theory and applications. J. Vis. 21, 3–3 (2021).
31. 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).
32. D. Lashkari, E. Vul, N. Kanwisher, P. Golland, Discovering structure in the space of fMRI selectivity profiles. NeuroImage. 50, 1085–1098 (2010).
33. M. Nau, A. Schindler, A. Bartels, Real-motion signals in human early visual cortex. NeuroImage. 175, 379–387 (2018).
34. S. S. Cohen, C. Baldassano, bioRxiv, in press, doi:10.1101/2021.04.12.439526.
35. Z. Wen, Y. Li, Toward Understanding the Feature Learning Process of Self-supervised Contrastive Learning (2021), (available at http://arxiv.org/abs/2105.15134).
36. 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).
37. B. Duchaine, G. Yovel, A Revised Neural Framework for Face Processing. Annu. Rev. Vis. Sci. 1, 393–416 (2015).

38. 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).
39. 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).
40. 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).
41. 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).
42. 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).
43. P. Lakatos, C.-M. Chen, M. N. O'Connell, A. Mills, C. E. Schroeder, Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. Neuron. 53, 279–292 (2007).
44. P. Lakatos, M. N. O'Connell, A. Barczak, A. Mills, D. C. Javitt, C. E. Schroeder, The Leading Sense: Supramodal Control of Neurophysiological Context by Attention. Neuron. 64, 419–430 (2009).
45. K.-M. G. Fu, T. A. Johnston, A. S. Shah, L. Arnold, J. Smiley, T. A. Hackett, P. E. Garraghty, C. E. Schroeder, Auditory Cortical Neurons Respond to Somatosensory Stimulation. J. Neurosci. 23, 7510–7515 (2003).
46. G. A. Calvert, E. T. Bullmore, M. J. Brammer, R. Campbell, S. C. R. Williams, P. K. McGuire, P. W. R. Woodruff, S. D. Iversen, A. S. David, Activation of Auditory Cortex During Silent Lipreading. Science. 276, 593–596 (1997).
47. A. A. Ghazanfar, C. E. Schroeder, Is neocortex essentially multisensory? Trends Cogn. Sci. 10, 278–285 (2006).
48. 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).
49. S. Michelmann, A. R. Price, B. Aubrey, W. K. Doyle, D. Friedman, P. C. Dugan, O. Devinsky, S. Devore, A. Flinker, U. Hasson, K. A. Norman, bioRxiv, in press, doi:10.1101/2020.12.09.416438.
50. B. Krekelberg, Saccadic suppression. Curr. Biol. 20, R228-9 (2010).
51. 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).
52. D. Ernst, S. Becker, G. Horstmann, Novelty competes with saliency for attention. Vision Res. 168, 42–52 (2020).
53. M. M. Hayhoe, Integration of Visual Information Across Saccades, 6.
54. M. I. Coco, A. Nuthmann, O. Dimigen, Fixation-related Brain Potentials during Semantic Integration of Object–Scene Information. J. Cogn. Neurosci. 32, 571–589 (2020).
55. M. Hayhoe, D. Ballard, Eye movements in natural behavior. Trends Cogn. Sci. 9, 188–194 (2005).
56. M. Ibbotson, B. Krekelberg, Visual perception and saccadic eye movements. Curr. Opin. Neurobiol. 21, 553–558 (2011).
57. J. V. Haxby, E. A. Hoffman, M. I. Gobbini, The distributed human neural system for face perception. Trends Cogn. Sci. 4, 223–233 (2000).
58. 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).

59. C. E. Schroeder, P. Lakatos, Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. 32, 9–18 (2009).
60. 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).
61. D. H. Brainard, The Psychophysics Toolbox. Spat. Vis. 10, 433–436 (1997).
62. 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).
63. A. M. Dale, B. Fischl, M. I. Sereno, Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. NeuroImage. 9, 179–194 (1999).
64. 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).
65. 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).
66. 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).
67. B. K. P. Horn, B. G. Schunck, Determining optical flow. Artif. Intell. 17, 185–203 (1981).
68. 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).
69. 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.
70. A. Townsend, JPEG: Image compression algorithm (2017; http://pi.math.cornell.edu/~web6140/TopTenAlgorithms/JPEG.html#19).
71. Y. Wu, A. Kirillov, F. M. Massa, W.-Y. Lo, R. Girshick, Detectron2 (2019; https://github.com/facebookresearch/detectron2).
72. S. Xie, R. Girshick, P. Dollar, Z. Tu, K. He, (2017; https://openaccess.thecvf.com/content_cvpr_2017/html/Xie_Aggregated_Residual_Transformations_CVPR_2017_paper.html), pp. 1492–1500.
73. K. He, G. Gkioxari, P. Dollar, R. Girshick, (2017; https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html), pp. 2961–2969.
74. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, 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.
75. 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).
76. J. Chumbley, K. Friston, False discovery rate revisited: FDR and topological inference using Gaussian random fields. NeuroImage. 44, 62–70 (2009).
77. 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).

78. 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).
79. 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).
80. B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, A. M. Dale, Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. Neuron. 33, 341–355 (2002).

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.

Funding:
National Institutes of Health grant R01MH111439
National Institutes of Health grant P50MH109429
National Institutes of Health grant 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 will be made available on OSF (https://osf.io/) and code for analysis provided on github (https://github.com/) by the time of publication. 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.

Figures and Tables

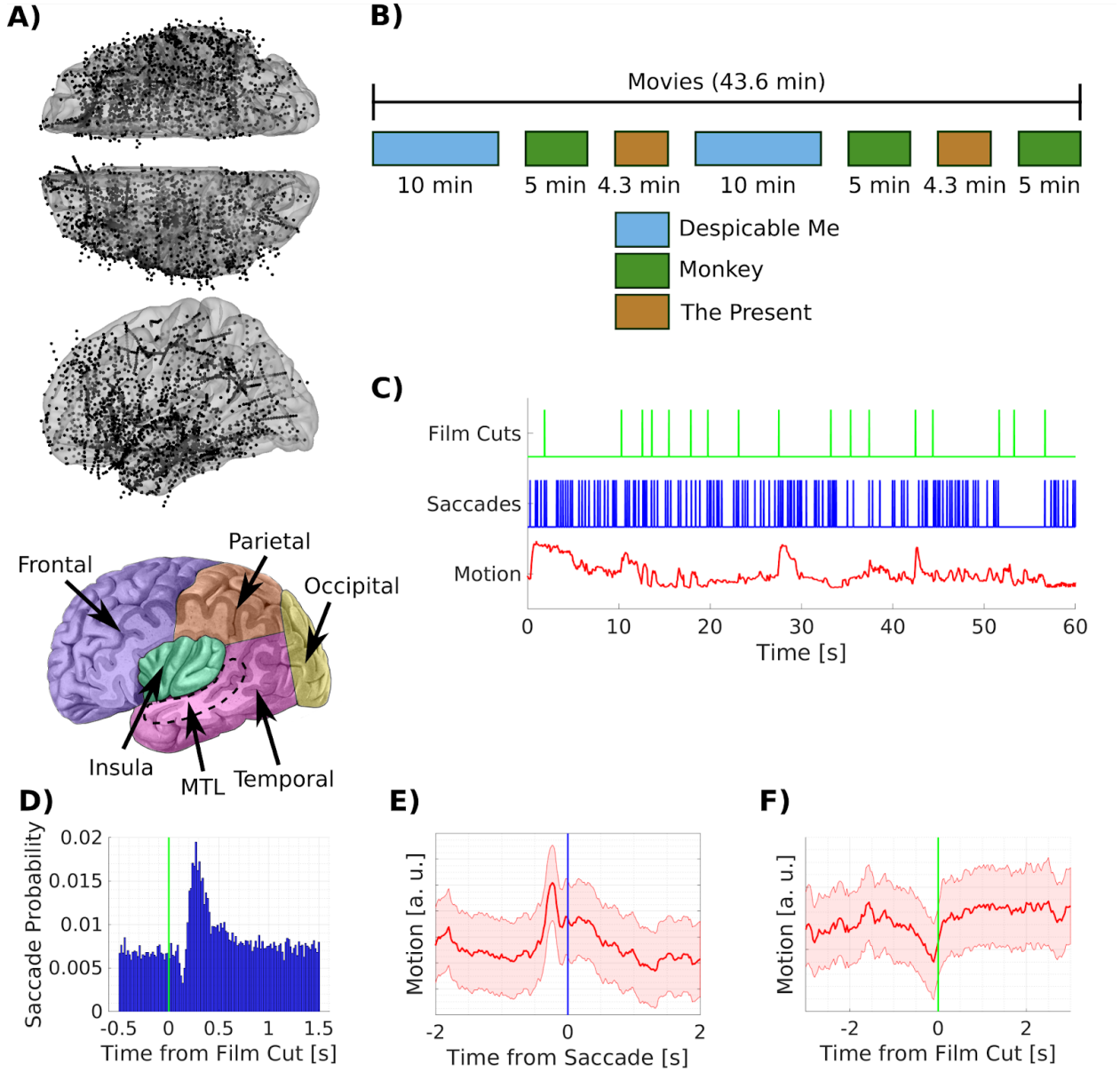

Figure 1: Electrodes coverage across the whole brain and movie features capturing novelty. A) We analyzed eye tracking and simultaneous intracranial EEG data from 6328 contacts across 23 patients. Electrodes cover the whole brain but are more sparse in the occipital lobe. Each electrode contains multiple recording contacts (points on the graph). We analyze neural data in 5378 bipolar channel pairs, referred to as channels. Left 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 10min clips of the animated comic 'Despicable Me' (26) 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, 8). C) 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 visual motion captures the total optic flow in each video frame. D)

Saccade probability as a function of time from film cuts. E) Average visual 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 S2). Shaded area depicts the standard error of the mean.

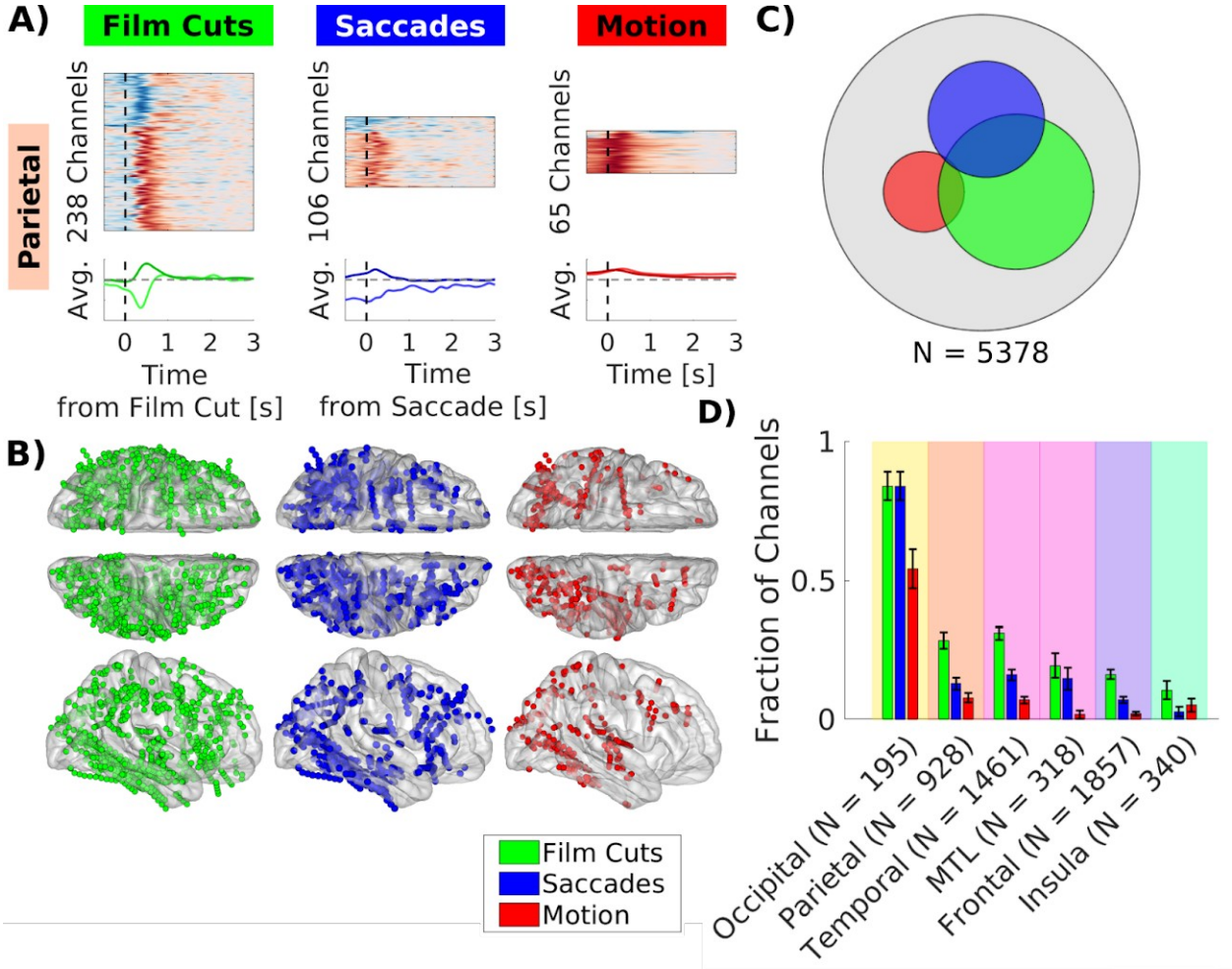

Figure 2. More channels respond to film cuts and saccades than to visual motion. A) Temporal response functions in the parietal lobe for channels with statistically significant response in BHA. The sizes of the blocks in each column reflect the relative number of significant channels. TRFs in each channel were normalized by z-scoring. Red indicated an increase, blue a decrease in BHA. Similar TRFs were grouped by clustering TRFs that are highly correlated to each other (31, 32). The bottom row shows the average TRF for each cluster. The magnitude of the responses was scaled to reflect the relative strength of responses to each stimulus within each channel. TRFs are smoothed with a Gaussian window with a standard deviation of 53ms. Time for motion indicated the delay of the neural response in relation to the optical flow signal. B) Location of all channels with significant response plotted on an average brain. C) Number of significant channels for each condition 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. Error bars correspond to the 95% confidence interval of the proportion of channels with significant responses. Background colors correspond to different brain areas in Figure 1A. For results in a more detailed parcellation of the brain see Figure S4.

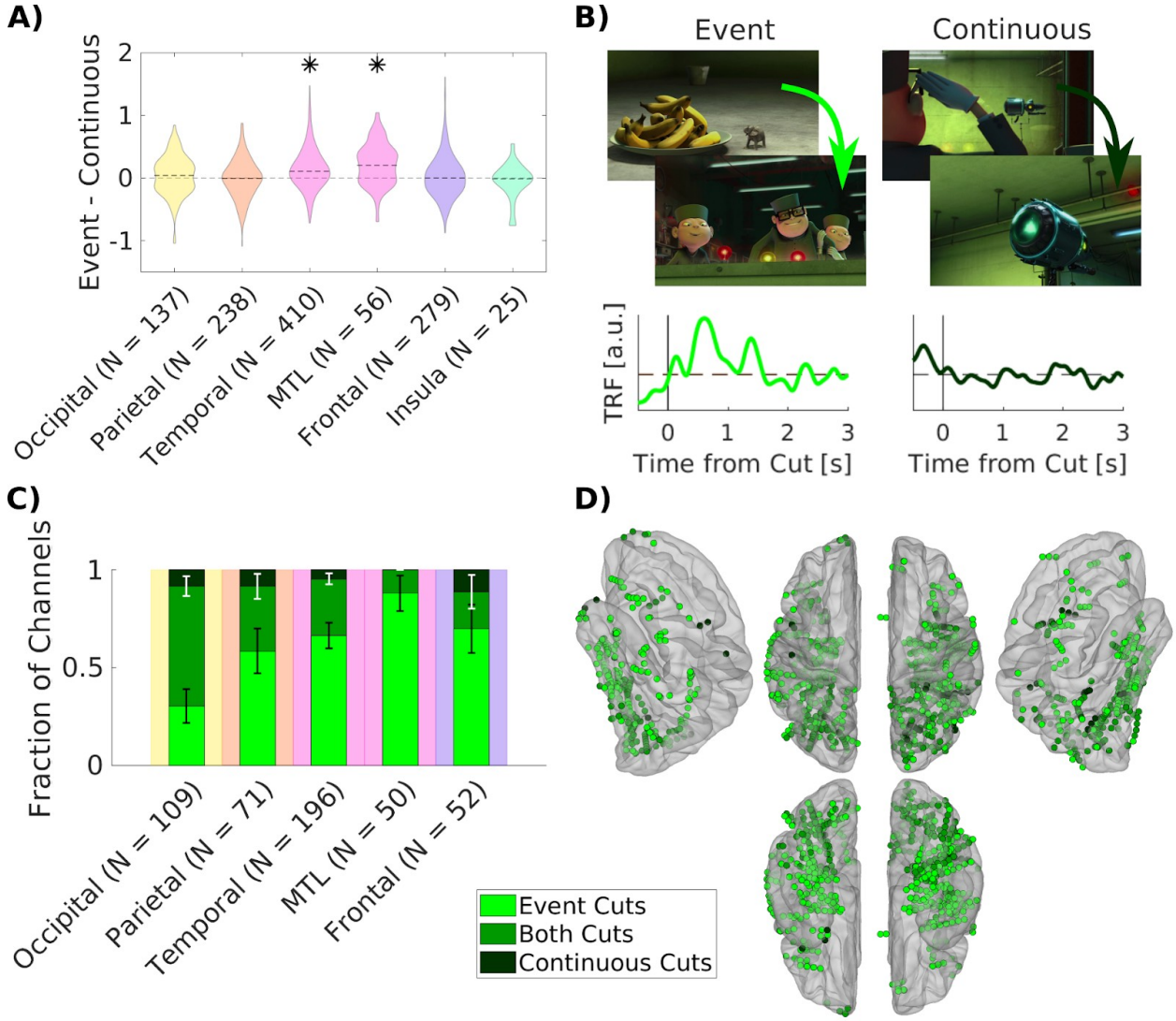

Figure 3: Responses to film cuts associated with event boundaries dominate neural responses in higher order brain areas. A) Difference in magnitude of the responses to individual event cuts and continuous cuts (Figure S5). Magnitude of the response is estimated by scaling the temporal response function in each channel to best fit the response to each cut (Figure S6). Only channels with significant responses to film cuts in Figure 2B are considered. Difference of medians Δ: Occipital: Δ=0.0425, p=0.2, N=137; Parietal: Δ=-0.005, p=0.67, N=238; Temporal: Δ=0.11, p = 1.2*10-16, N=410, MTL: Δ=0.2, p=8.2*10-5, N=56; Frontal: Δ=0.0009, p=0.67, N=279; Insula: Δ=-0.01, p=0.67, N=25; Wilcoxon signed-rank test, False discovery rate (FDR) control, at a level of q = 0.05. B) Temporal response functions are obtained for two separate regressors coding 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. This channel is therefore only responsive to event cuts. TRFs in all channels are shown in Figure S9. C) Fraction of responsive channels with selective response to event cuts or continuous cuts. White error bars correspond to the 95% confidence interval of the proportion of channels responsive to continuous cuts, black error bars correspond to the proportion of channels responsive to event cuts. Background colors correspond to different brain areas in Figure 1A. D) Location of channels selectively responding to event cuts only (bright green), continuous cuts only (dark green), or responding non-selectively to either types of cuts (medium green). For results in a more detailed parcellation of the brain see Figure S10.

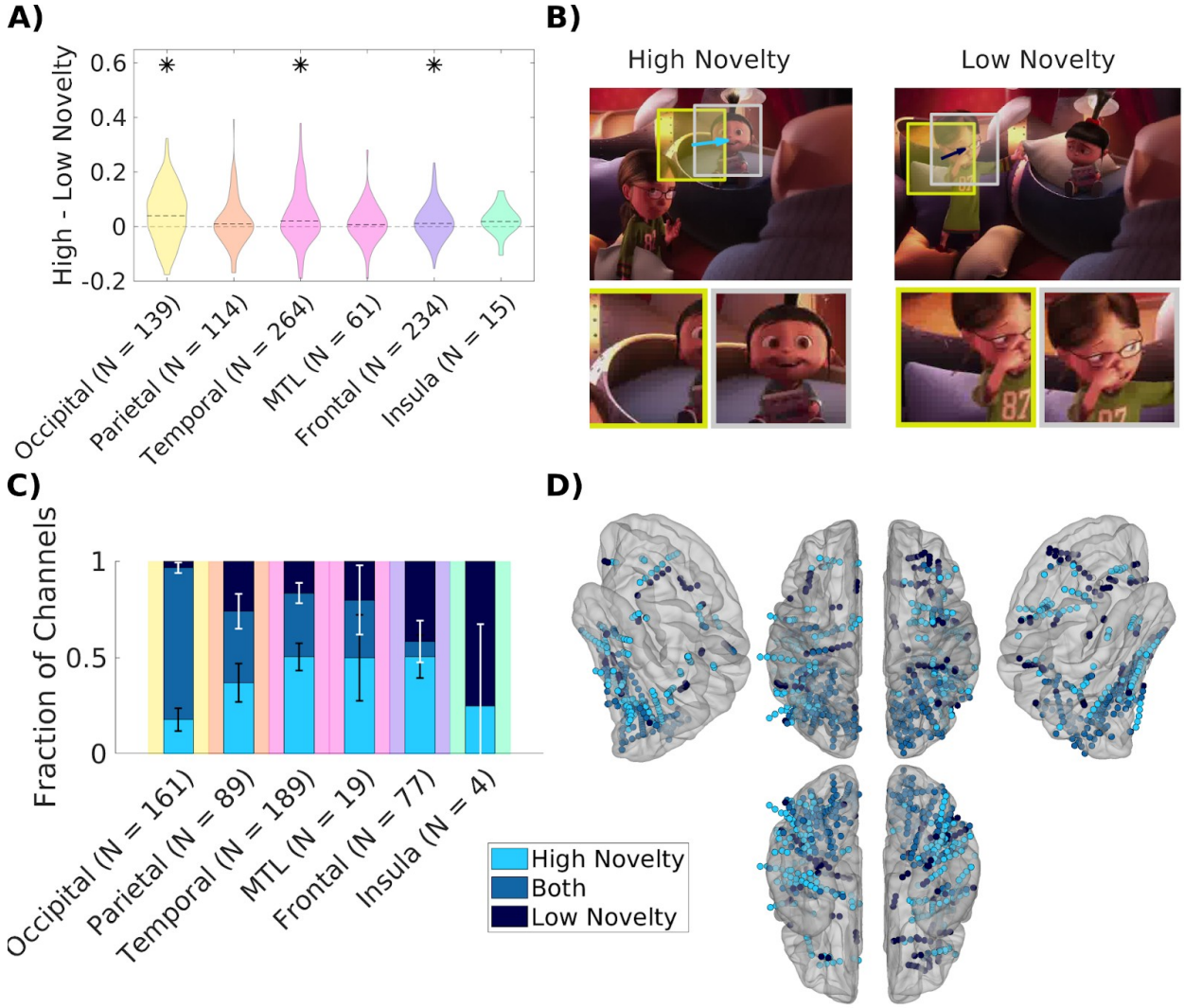

Figure 4: Differential responses to saccades with high and low novelty in higher order brain areas. A) Difference in magnitude of neural responses to individual saccades with high and low novelty targets. Magnitude was estimated using filters for all saccades as described in Figure 3. Difference of medians Δ: Occipital: Δ=0.04, p=3.5*10-6, N=139; Parietal: Δ=0.01, p=0.19, N=114; Temporal: Δ=0.021, p = 3.8*10-7, N=264, MTL: Δ=0.008, p=0.17, N=61; Frontal: Δ=0.012, p=2.8*10-4, N=234; Insula: Δ=0.02, p=0.062, N=15; Wilcoxon signed-rank test, FDR control, at a level of q = 0.05. B) 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 S11). TRFs in all channels are shown in Figure S13. C) Fraction of responsive channels with selective (low/high novelty) or non-selective response (both). Background colors correspond to different brain areas in Figure 1A. D) Locations of channels with significant TRFs selectively responding to saccades with high novelty only (light blue), saccades with low novelty only (dark blue) and non-selective responses to both high and low novelty saccades (medium blue). For results in a more detailed parcellation of the brain see Figure S14.

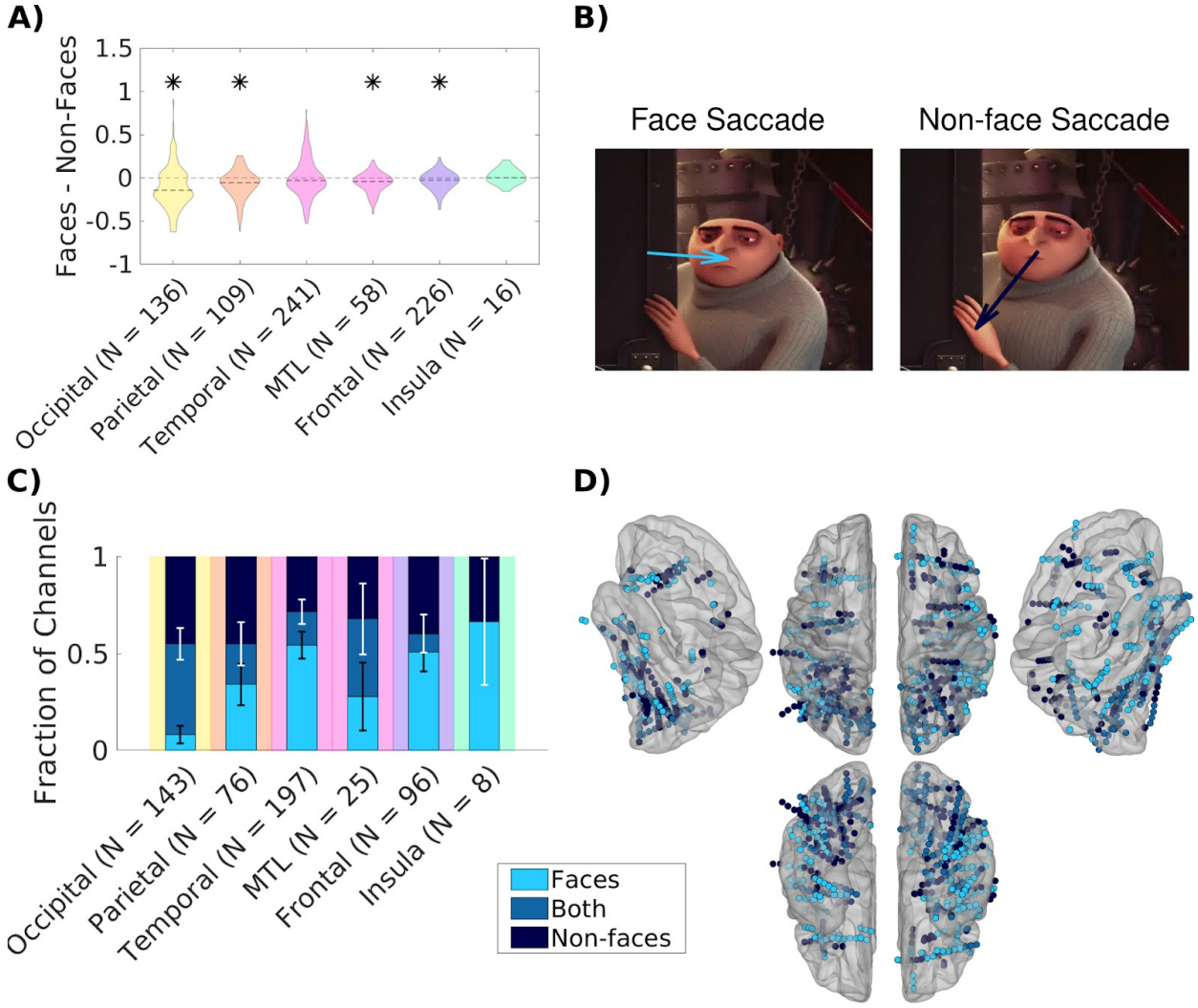

Figure 5: Responses that are selective for face and non-face saccades are found in higher-order brain areas. A) Difference in magnitude of neural responses to individual saccades to faces and non-faces. Magnitude was estimated using filters for all saccades as described in Figure 3. Difference of medians Δ: Occipital: Δ=-0.14, p=3.9*10-7, N=139; Parietal: Δ=-0.053, p=5.5*10-5, N=109; Temporal: Δ=-0.029, p=0.37, N=241, MTL: Δ=-0.041, p=5*10-4, N=58; Frontal: Δ=-0.024, p=2.5*10-5, N=226; Insula: Δ=0.0028, p=0.5, N=16; Wilcoxon signed-rank test, FDR control, at a level of q = 0.05. B) Separate TRFs are computed for face and non-face saccades (Figure S15). C) Fraction of responsive channels with selective (face/non-face) or non-selective response (both). Background colors correspond to different brain areas in Figure 1A. D) Location of channels with significant TRFs to face saccades only (light blue), non-face saccades only (dark blue), and face and non-face saccades (medium blue). For results in a more detailed parcellation of the brain see Figure S16.

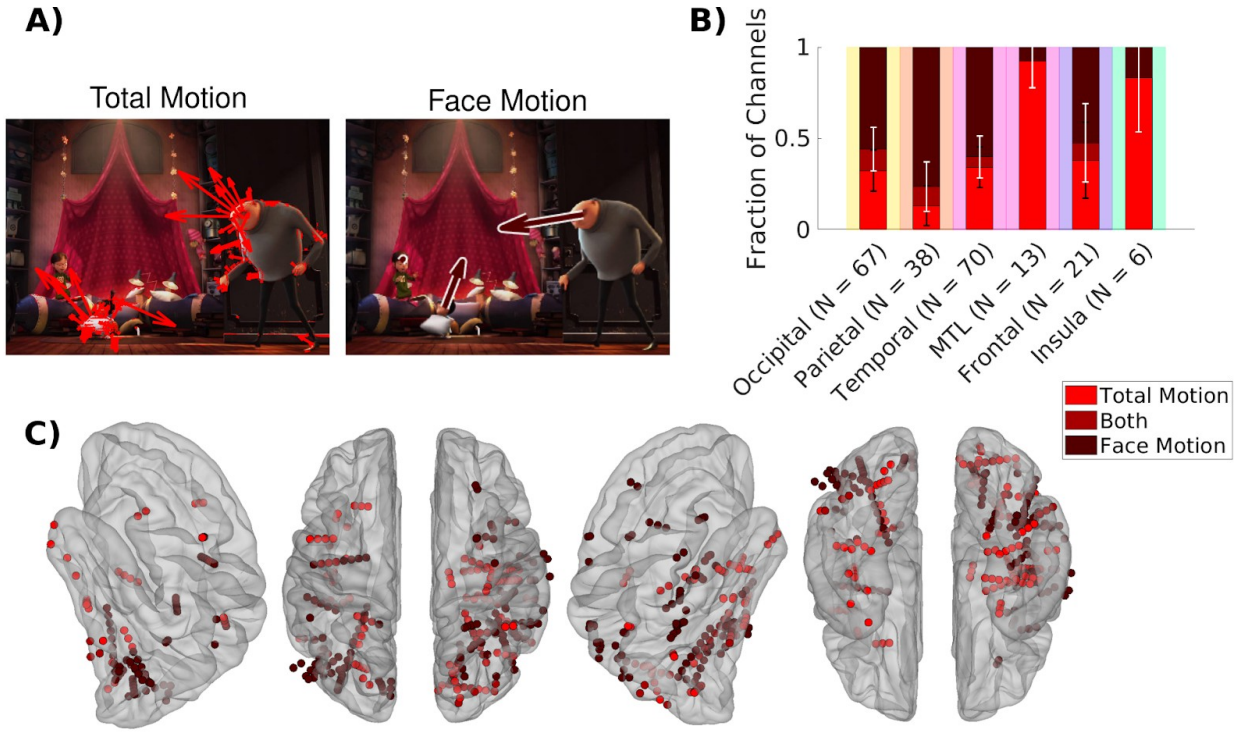

Figure 6: Total motion and face motion are processed in distinct visual areas. A) Separate TRFs are computed for total motion (optical flow) and face motion. B) Fraction of channels with selective responses (face/total motion) and non-selective responses (both). Background colors correspond to different brain areas in Figure 1A. C) Channels with significant response to total motion only (bright red), face motion only (dark red), and total and face motion (medium red) on the fsaverage brain. For results in a more detailed parcellation of the brain see Figure S17.