NeuroImage 216 (2020) 116491

Contents lists available at ScienceDirect

NeuroImage

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

Following Forrest Gump: Smooth pursuit related brain activation during free movie viewing

Ioannis Agtzidis [a,*], Inga Meyhöfer [b], Michael Dorr [a], Rebekka Lencer [b]

[a] Department of Electrical and Computer Engineering, Technical University of Munich, Germany
[b] Department of Psychiatry and Psychotherapy & Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Germany

ARTICLE INFO

Keywords:
Eye movements
fMRI
Smooth pursuit eye movement
Dynamic naturalistic scenes

ABSTRACT

Most fMRI studies investigating smooth pursuit (SP) related brain activity have used simple synthetic stimuli such as a sinusoidally moving dot. However, real-life situations are much more complex and SP does not occur in isolation but within sequences of saccades and fixations. This raises the question whether the same brain networks for SP that have been identified under laboratory conditions are activated when following moving objects in a movie.

Here, we used the publicly available studyforrest data set that provides eye movement recordings along with 3 T fMRI recordings from 15 subjects while watching the Hollywood movie "Forrest Gump". All three major eye movement events, namely fixations, saccades, and smooth pursuit, were detected with a state-of-the-art algorithm. In our analysis, smooth pursuit (SP) was the eye movement of interest, while saccades were acting as the steady state of viewing behaviour due to their lower variability. For the fMRI analysis we used an event-related design modelling saccades and SP as regressors initially. Because of the interdependency of SP and content motion, we then added a new low-level content motion regressor to separate brain activations from these two sources.

We identified higher BOLD-responses during SP than saccades bilaterally in MT+/V5, in middle cingulate extending to precuneus, and in the right temporoparietal junction. When the motion regressor was added, SP showed higher BOLD-response relative to saccades bilaterally in the cortex lining the superior temporal sulcus, precuneus, and supplementary eye field, presumably due to a confounding effect of background motion. Only parts of V2 showed higher activation during saccades in comparison to SP.

Taken together, our approach should be regarded as proof of principle for deciphering brain activity related to SP, which is one of the most prominent eye movements besides saccades, in complex dynamic naturalistic situations.

1. Introduction

Because of the dramatically space-variant resolution of their visual system, humans make several eye movements per second to sample their surroundings with the high-resolution fovea. Consequently, the neural implementation of gaze behaviour as one of the fundamental aspects of visual information processing is an active research topic. In particular, functional magnetic resonance imaging (fMRI) has been previously used along with eye tracking in order to identify brain areas (localized BOLD-responses) and networks related to specific eye movements such as fixations and saccades (Luna et al., 1998; Beauchamp et al., 2001; Sestieri et al., 2007; McDowell et al., 2008; Ettinger et al., 2008; Kleiser et al., 2009; Lukasova et al., 2018). However, brain areas subserving smooth pursuit (SP) eye movements have been studied to a lesser extent only (Petit and Haxby, 1999; Lencer et al., 2004; Nagel et al., 2006; Kimmig et al., 2008; Kellar et al., 2018), possibly due to technical challenges in the analysis of dynamic setups. Therefore, there is a need to further investigate possible solutions and their confounds in situations closer to real-world visual tracking.

When segmented eye tracking data are directly related to brain activation, the majority of experiments use specifically designed synthetic stimuli (Lencer et al., 2004; Nagel et al., 2006; Kimmig et al.,

* Corresponding author.
E-mail addresses: ioannis.agtzidis@tum.de (I. Agtzidis), inga.meyhoefer@ukmuenster.de (I. Meyhöfer), michael.dorr@tum.de (M. Dorr), rebekka.lencer@ukmuenster.de (R. Lencer).

https://doi.org/10.1016/j.neuroimage.2019.116491
Received 8 August 2019; Received in revised form 13 December 2019; Accepted 22 December 2019
Available online 7 January 2020

2008). Such stimuli can take the form of fixation crosses that change position when saccades are studied, or of linearly or sinusoidally moving dots when smooth pursuit is investigated. The biggest advantage of synthetic stimuli is that their properties are well defined and can explicitly represent specific features, which simplifies the analysis of both the eye tracking and BOLD signals. By ideally modulating the stimulus along one isolated feature dimension only, the link to specific brain activations can be established more precisely. These advantages come at the cost of using paradigms that are not representative of normal human vision, because ecologically valid visual input is much more complex and real-world SP does not occur in isolation but within sequences of saccades and fixations. Thus, the use of synthetic stimuli moving on a uniform background ignores the possible influence of background information (Brenner and Smeets, 2015), crowding effects (Sanocki et al., 2015), and overall eye movement planning processes (Gold and Shadlen, 2007; Tatler et al., 2017). Another important limitation is that following a uniform synthetic stimulus over a longer time interval can result in reduced maintenance of attention (Tagliazucchi and Laufs, 2014; Vanderwal et al., 2015).

Because of the increased complexity of naturalistic stimuli, some studies have restricted themselves to the presentation of static naturalistic scenes, i.e. images (Kay et al., 2011; Mannion, 2015). Going beyond static scenes, significant improvements in both vigilance and head motion were achieved by (Vanderwal et al., 2015), who used an abstract dynamic pattern to enhance the participant's attention, together with fMRI resting state analysis. An even better approximation to unconstrained human vision then are fully naturalistic dynamic stimuli. Consequently, both the neuroimaging and eye tracking communities have recently started to explore the possibilities of more immersive experiments (Hasson et al., 2004, 2008; Lahnakoski et al., 2012; Nardo et al., 2014; Andric et al., 2016; Marsman et al., 2016). Some recorded data sets of naturalistic fMRI (Hanke et al., 2016) and eye tracking studies (Dorr et al., 2010; Mathe and Sminchisescu, 2012; Linnea Larsson et al., 2013) have even become publicly available.

Under conditions that resemble daily life more, parallel processing of conflicting information is often required during smooth pursuit on background textures, which by themselves may contain multiple dynamic objects moving in different directions. This is of particular interest to the present study because it has been concluded from monkey studies that neurons in V5 play an important role not only during smooth pursuit, but also for the interaction of different, even conflicting retinal stimuli (for a review see (Ilg and peter, 2008)). However, in such naturalistic settings, care has to be taken to disentangle neural activations arising from dynamic visual input and those that arise from smooth pursuit eye movements. Fortuitously, recent advances in computer vision are beginning to enable an automated understanding of dynamic complex visual scenes even for large and diverse data sets.

Compared to synthetic stimuli and experiments with typically explicit instructions on how to move the eyes (e.g. "follow the dot", "make a saccade when the target appears", etc.), segmentation of a gaze trace into its constituent eye movements is also more challenging when unconstrained dynamic naturalistic stimuli are used. As reported by (Hooge et al., 2018), hand-labelling of "ground-truth data", which is considered the gold standard in eye movement segmentation, of only fixations and saccades can take anywhere between 4 and 15 s for each second of gaze signal. This process becomes infeasible when data sets increase in size. For example, the studyforrest project provided by (Hanke et al., 2016) contains roughly 30 h of simultaneous fMRI and gaze recordings. Therefore, the authors of the studyforrest data set reported brain activations related to visual versus non-visual cues, but did not differentiate between different eye movement subtypes.

In recent years, however, progress has been made on the automatic analysis of eye movement data and smooth pursuit in particular. Several eye movement classification algorithms have been developed based on publicly available large-scale data sets along with partial or full ground-truth annotations of eye movements (Dorr et al., 2010; Mathe and Sminchisescu, 2012; Hooge et al., 2018; Startsev et al., 2019). Most of the established classification algorithms only label fixations and saccades because they were developed with static stimuli in mind. As a consequence, these algorithms misclassify SP as fixations or saccades, thereby preventing the identification of neural correlates of smooth pursuit. To solve this problem, the automatic multi-observer smooth pursuit algorithms (MOSP) have been developed (Agtzidis et al., 2016) that are able to differentiate all three major eye movements, i.e. fixations, saccades, and SP, with high classification quality (for an evaluation, see (Startsev et al., 2019)).

In the present study, we make use of several of these recent developments. Using state-of-the-art computer vision and eye movement classification algorithms, we analyse gaze and fMRI recordings from the large-scale studyforrest data set and are able to identify specific brain areas related to smooth pursuit and saccadic eye movements evoked by complex naturalistic scenes, i.e. a Hollywood movie.

2. Methods

2.1. Data set

For our analysis, we used the publicly available studyforrest data set as an approximation to a complex natural environment; for full experimental details, we refer to the paper presenting the original data set (Hanke et al., 2016). Briefly, this data set includes 15 participants who watched the Hollywood movie "Forrest Gump" while their gaze was tracked in an fMRI scanner, and another 15 participants with in-lab gaze only recordings (which we used here only to improve the automatic detection of smooth pursuit events, see below). The stimulus was presented to the in-scanner participants through an LCD projector in combination with a front-reflective mirror and to the in-lab participants through an LCD monitor. The gaze data were recorded with a high-frequency eye tracker (EyeLink 1000, set to 1000 Hz sampling rate with telephoto lens for the fMRI recordings) and a 13-points calibration was performed at the beginning of each session. The fMRI recordings were acquired with a 3 T scanner (Philips Achieva dStream MRI scanner) with repetition time (TR) of 2 s and $3 \times 3 \times 3$ mm^3 voxel size.

2.2. Motion estimation in the stimulus

Because smooth pursuit behaviour is tightly linked to moving targets, we estimated the overall motion per video frame with computer vision techniques. Despite all recent advances, such algorithms can still yield noisy outputs, so we used two different algorithms for additional robustness. The first algorithm computed motion based on the minors of the structure tensor as described by (Barth, 2000) with the aim to provide a sparse optical flow field by estimating motion only at points that are not susceptible to the aperture problem, i.e. corners. Initially, the input video was spatially subsampled by a factor of two, and then a spatio-temporal Gaussian pyramid with five spatial and two temporal levels was created. For each level of this multiscale representation, velocity per pixel was computed. These velocity estimates were normalized relative to the original video resolution and combined in a procedure similar to pyramid synthesis described by (Adelson and Burt, 1981); higher speed values were clipped to the 90th percentile speed. The second algorithm used edge-preserving interpolation of correspondences for optical flow (EpicFlow) computation as described by (Revaud et al., 2015). The algorithm first used dense matching with edge-preserving interpolation, followed by an energy minimization step. An example content motion computation of the EpicFlow algorithm is provided in Fig. 1b. For both algorithms, finally, the mean length of pixel displacements was computed per video frame.

2.3. Eye movement classification

From the provided data we created a quadruplet of values for each

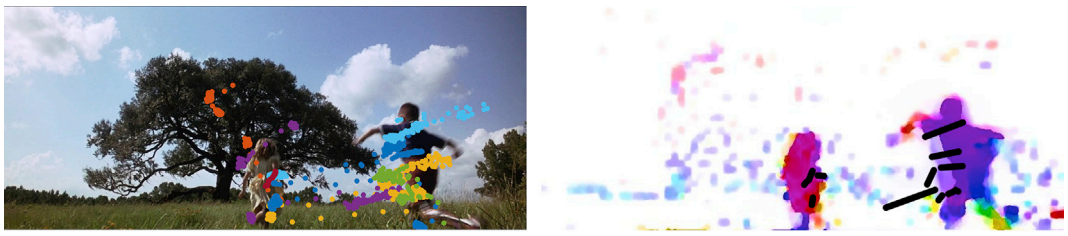

Fig. 1. a) Example frame from the studyforrest data set with superimposed gaze traces (over a 400 ms period; one colour per subject) from the in-scanner participants. Smooth pursuit is evidenced by the elongated point clouds. **b)** Optical flow computed by the EpicFlow algorithm. The estimated motion corresponds well with the actual motion in the video. Black lines indicate the Multiple Observer Smooth Pursuit (MOSP) algorithm output, i.e. automatically detected smooth pursuit segments (in the 400 ms window).

gaze sample that comprised time, x and y coordinates on the monitor coordinate system, and a confidence estimation of the eye tracking quality. Since the data set used monocular eye tracking, a confidence value of 1 meant good tracking of the eye and a value of 0 meant tracking loss. After inspection of the data, lost tracking varied from 1.2% to 16.7% among subjects with the notable exception of subjects 05 and 20. For these two subjects the lost tracking was 86.7% and 39.0%, respectively, and they were excluded from all subsequent analyses.

The remaining gaze traces were segmented into eye movements by the MOSP algorithms from (Agtzidis et al., 2016) as implemented by (Startsev et al., 2019). This algorithm achieved state-of-the-art performance in an online benchmark against a manually labelled data set (Startsev et al., 2016) compared to several recent eye movement detection algorithms (Linnéa Larsson et al., 2015; Dar et al., 2019). Another aspect of the MOSP algorithm that is advantageous for our application is the fact that it uses simple-to-understand thresholds that can be easily tweaked (unlike deep learning approaches such as (Startsev et al., 2019; Zemblys et al., 2019)). In particular, MOSP achieves high SP detection precision (i.e. low number of false positives), which is central to our analysis.

The MOSP algorithm has two distinct components, with the first one being responsible for saccade detection and the second for differentiation of fixations and SP. The saccade detector is based on the algorithm described by (Dorr et al., 2010) and uses a high and a low speed threshold. The high speed threshold is used for initiation of saccade detection, which is then extended on both sides until the speed falls below the low threshold. This high speed threshold, which is higher than the speed of sample-to-sample noise of the input, makes the algorithm specifically robust to noise. The second component of MOSP further processes the intersaccadic intervals by initially assigning the fixation label to the samples that almost certainly belong to a fixation. The remaining samples are then marked as SP candidates and pooled among all the participants. Then they are clustered with an algorithm (DBSCAN) that creates a cluster only if the gaze sample density is above a certain threshold. This density-based clustering reliably detects SP because of two significant SP properties. Firstly, SP can occur only if motion is present in the stimulus at a given time, and the number of moving targets per video frame is typically low. Secondly, moving targets attract attention (especially in a Hollywood movie) and they are usually pursued by more than one participant. An example of the clustering property along with the output of the MOSP algorithm is presented in Fig. 1. The combination of these two properties allows the algorithm to robustly distinguish drift and SP-like motion from actual SP, which is particularly important for gaze recordings in fMRI experiments because they tend to have higher noise levels than lab recordings. For example, if some gaze samples are erroneously marked as SP candidates, they will not be labelled as SP as long as no other subject has a similar pattern in the same area in space and time. While this has the possible drawback of missing some SP episodes when not enough subjects pursue a target (reduced sensitivity), the increased specificity is more important in the context of this study to identify brain areas related to SP processing.

Since the original MOSP algorithms were designed and optimized for the GazeCom data set (Dorr et al., 2010), some parameters were adjusted (see online data for full details). We further improved the SP detection by using both in-lab and in-scanner recordings together because the SP detection algorithm improves with increasing number of gaze traces. Despite the different stimulus sizes, we used the same pixel-space for both sets of recordings by scaling the pixel-per-degree values for the (less noisy) in-lab recordings; the agreement in detected SP episodes for the two sets was high (r^2 of 0.84 for share of SP in 2-s intervals).

2.4. fMRI analysis

The fMRI data analysis was performed with SPM12 using Matlab 9.2. We initially followed a standard preprocessing pipeline for each recording (Poldrack et al., 2011). The process comprised realigning the functional data to the mean image of each session (without slice timing correction), coregistering them to the anatomical T1 scan, normalizing them to the MNI template, and resampling them into $3 \times 3 \times 3$ mm^3 voxels. Finally, we applied smoothing with a Gaussian kernel of 8 mm at full width half maximum (FWHM).

During the recording of the studyforrest data set its authors split the movie stimulus into 8 different segments of approximately 15 min each with each one displayed separately in the scanner. In the first level analysis we combined all 8 recording sessions into one design matrix in order to model the full Forrest Gump movie. For each session in the design matrix we fitted an SP, a saccade, and a movie motion regressor when needed. In order to account for variations in the onset and width of the hemodynamic response among subjects, we used the canonical hemodynamic response function (HRF) along with its time and dispersion derivatives. Apart from the previous regressors, we also used the six head movement components that were returned from the realignment step during preprocessing as nuisance regressors.

The eye movement and motion regressors were modelled as event time series with events placed 2 s apart, which by design coincides with the scanner's TR and therefore each event was representing the regressor variance between scans. The amplitude of each event was modulated by the prevalence of the corresponding eye movement or the amount of motion in the 2-s window: It had a value of 0 when it was the same as the overall mean and was linearly increasing up to a maximum value of 1. A detailed description of the regressor modelling procedure is given in the next section. As it becomes evident from how the regressors were modelled, it would have been impossible to model both fixations and SPs with this process without creating strong (negative) correlations between the two. To make this interdependence more clear, let us consider that a subject starts pursuing a target. Then consequently the amplitude of the SP regressor would increase with the fixation amplitude decreasing proportionally.

After fitting the GLM to the data of each subject independently, we used the amplitude component of the HRF of each regressor that spanned 8 recording sessions in order to compute the contrasts of interest. These contrasts included the main effect of the eye movements and motion, the

comparison between SP and saccades, and the comparison of the eye movements to motion. Finally, at the second level of the fMRI analysis we performed a one-sample *t*-test for each of the previous contrasts for the 13 valid subjects. The resulting clusters ($p < 0.05$ Family Wise Error [FWE] corrected with an initial threshold of $p < 0.001$) were overlaid on a three-dimensional brain model and are presented in the results section.

2.4.1. Regressor modelling

As outlined above, our regressors were not modelling each eye movement event independently, but were placed in 2-s intervals, which were modulated by the amount of the respective eye movement in that window. For the experiments that were taking movie motion into account, this was modelled through the mean movie motion and as before in consecutive 2-s windows.

More specifically, the computation of the magnitude of the eye movement modulation parameters was taking into account three main factors: **i)** The first factor was capturing the changes in eye movements between different naturalistic stimuli and was represented by the mean percentage of each eye movement of each subject; this is equivalent to the mean viewing behaviour. **ii)** The second factor was capturing the differences in prevalence and variance between different eye movements and it was a constant value with the modulation parameter being inversely proportional to it. The value of this factor was chosen from the data in order to bring approximately 95% of the modulated values below 1 (for a visualization, see Fig. 2). Therefore, it was set to $\text{modulation}_{\text{sacc}} = 1.5$ for saccades due to their small variance in relation to different input stimuli. For SP it was set to $\text{modulation}_{\text{SP}} = 5$ in order to reflect the large variance of smooth pursuit eye movements, which cannot occur in the absence of a moving target but can be continuously performed for long periods of time when a salient moving object exists. **iii)** The third factor was capturing the variance among subjects. This subject-specific factor was based on the observation that the prevalence of each eye movement varies among subjects and it may directly or indirectly relate to the differences in brain connectivity (Mueller et al., 2013; Vanderwal et al., 2017). In the case of the studyforrest data set, saccades varied from 5.8% to 12.4% and SPs from 11.5% to 19.3% among the subjects; thus, if the overall mean were used, the relevant activations in some subjects would be suppressed and in some would be amplified.

As an illustrative example, consider a hypothetical subject that has an overall mean SP percentage of $\text{overall}_{\text{SP}} = 15\%$ and performed SP $\text{clip}_{\text{SP}} = 10\%$ of the time in a given clip. Now in a particular 2-s window $\text{window}_{\text{SP}} = 85\%$ of its duration was labelled as SP. The modulation magnitude will be $(\text{window}_{\text{SP}}\text{-clip}_{\text{SP}})/(\text{modulation}_{\text{SP}}*\text{overall}_{\text{SP}}) = (85-10)/(5*15) = 1$. After computing the modulation parameters across all 2-s intervals of the data set according to the previous formula, we found that the SP and saccade regressors were uncorrelated (Pearson correlation $r = 0.02$), which is a good indication of no shared variability between the two.

For the magnitude of the motion estimation modulation parameters, we followed a similar process as with the modelling of eye movement parameters. Again, here the steady state was captured through the mean content motion for each stimulus independently. The resulting value was normalized with the 90th percentile of the motion values across all clips and was bound to a maximum value of 1 in order to limit the influence of outliers.

2.5. Additional validation regressors

Apart from the eye movement regressors of SP and saccade we also used a motion regressor, which in the nominal case was modelling the global motion in the video. We further explored two variations of it. In the first variation of motion modelling, we used a window around the gaze position to get a local estimation of the motion and in the second variation we subtracted the smooth pursuit velocity from the mean content velocity in the same window with the aim of approximating the retinal motion. Since the results for local motion were subpar in comparison to global motion, we do not present them in the results section but discuss them at the end of the manuscript.

To further understand what drives eye movements, we also ran models which included scene complexity and edge density estimation as additional regressors, with their values being modelled identically to motion regressor as explained in the previous section. The scene complexity was computed as the entropy of the saliency of each frame using a standard saliency model (Itti et al., 1998). Similar to the entropy of image saliency, we calculated edge density as the per-frame entropy of the absolute pixel values on the third level of a Laplacian pyramid (which represents edges in the spatial frequency range of approximately 3–6 cycles per degree, i.e. close to the peak of the human contrast sensitivity function). Again these results are discussed at the end of the manuscript.

2.6. Data and code availability

The full source code and resulting labels for the eye movement analysis are available at https://web.gin.g-node.org/ioannis.agtzidis/studyforrest_analysis. The fMRI and eye tracking data was taken from the publicly available data set of (Hanke et al., 2016).

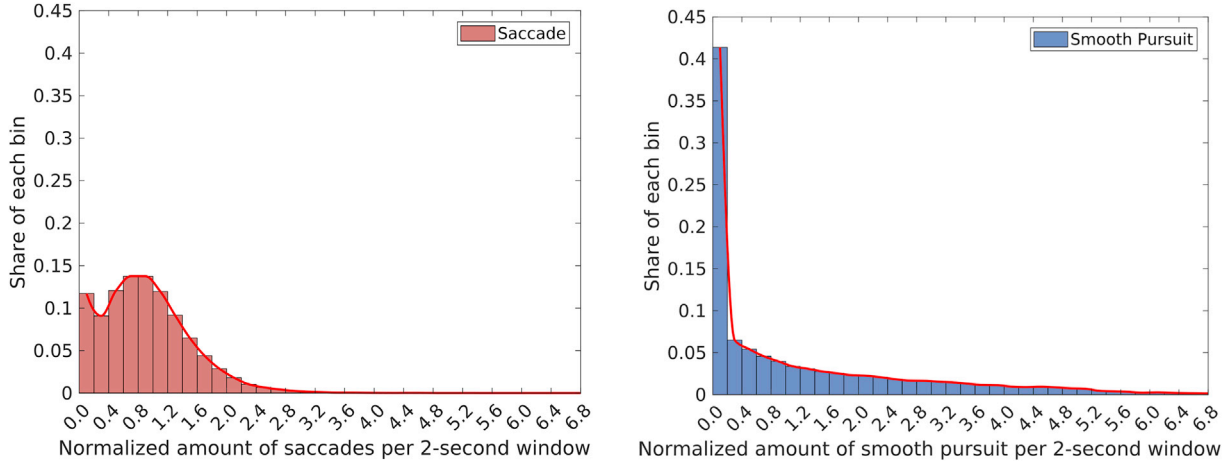

Fig. 2. Probability distribution of saccade (left) and smooth pursuit (right) ratios as detected in the 2-s windows that were used during the event-related 1st level analysis, normalized so that 1 corresponds to each subject's mean. A wide distribution indicates high variability across subjects and time. Saccades (red) have lower variability and are centred around 1. SP ratios (blue) are more variable and the peak close to 0 represents the absence of SP (e.g. no SP target is moving in the scene).

3. Results

The presented functional group results of this section were mapped to the three-dimensional cortical template of the "Population-Average, Landmark- and Surface-based" Atlas (PALS) (Van Essen, 2005) with the metric-enclosing-voxel algorithm in Caret (version 5.65) (Van Essen et al., 2001). When needed, the provided coordinates are reported in the Montreal Neurological Institute (MNI) coordinate system.

3.1. Eye movement statistics

Overall, for the valid in-scanner subjects the algorithm classified 53% of gaze samples as fixations, 8.4% as saccades, and 14.8% as SP with the rest being labelled as noise (tracking loss, blinks, cluster noise, etc.). Because we here were interested in separating e.g. saccades and SP as cleanly as possible, this relatively high noise level was acceptable. Fixations showed the highest absolute variation among participants (std: 10.1%), which is to be expected since the fixation detection is very sensitive to eye tracking noise and our objective was not to model this type of eye movement. Saccades (std: 2.5%) and SP (std: 3%) had lower absolute variance but very high relative variations among participants. This relatively high between-subject variability was captured by the subject-specific modulation factor during the first level analysis.

Apart from the between-subject variability there exists within-subject variability, which varies for different eye movements. In Fig. 2 we visualize the probability distributions of the ratios in 2-s windows of saccades and SP per subject in relation to the same subject's overall mean. Because the range of the distributions differed between eye movement types, we chose the eye movement specific modulation factors of Section 2.4.1 with the aim of normalizing them into comparable ranges. Here, a value of 1 indicates that the share of each eye movement type in a given interval is equal to the overall subject mean. A value of 0 denotes that the respective eye movement does not occur in that interval and values above 1 mean that we have above-average occurrence. As can be seen from Fig. 2, saccades show lower within-subject variability and are centred around the mean ratio of 1. On the other hand, the occurrence of SP shows higher variability with a peak close to 0, which represents the absence of SP when no moving target is present in the stimulus, i.e. movie.

3.2. SP- and saccade-related activations

The mean effects of SP- and saccade-related BOLD-responses are given in Fig. 3, where we present clusters at $p_{FWE} < 0.05$ using an initial threshold of $p < 0.001$. This procedure yielded three clusters related to SP (SP1-SP3) and two clusters related to saccades (Sac1-Sac2), see Table 1. The notable difference between the SP1 and Sac1 clusters is the strong activation within the middle temporal gyrus, presumably visual motion area MT+/V5 in SP1 but not Sac1, which is to be expected since this area is associated both with SP and motion processing. The second large cluster marked as SP2 mainly covers parts of the middle cingulate cortex and the precuneus. Also there exists a much smaller saccade-related cluster that covers part of the precuneus and is marked as Sac2 and a small SP-specific cluster related to the right temporoparietal junction (rTPJ) is marked as SP3. In order to check whether the presented clusters robustly represent brain areas that are involved in each eye movement, we ran a simple cross-validation procedure (for more details, see Supplementary Material). The resulting high correlation ($r^2 = 0.81$) between the activations of the two independent models (Fig. S1) shows that our regressors are fitting a pattern instead of only the provided data.

Table 1
List of clusters with peak activation T-value and location along with cluster level FWE-corrected p-values that are related to SP and saccadic eye movements.

Cluster Name	Peak activation			Cluster Size	$P_{FWE-corr}$	Peak T-value
	X	Y	Z			
SP1	−3	−91	14	7647	<0.001	15.11
SP2	6	−43	56	2048	<0.001	8.48
SP3	57	−40	17	109	0.011	6.94
Sac1	−9	−82	17	6437	<0.001	16.84
Sac2	−6	−52	56	245	<0.001	6.25

Table 2 lists a more detailed description of anatomical and functional areas included in the identified clusters SP1-SP3 and Sac1-Sac2, respectively. Anatomical areas were parcellated with the automated anatomical atlas (Tzourio-Mazoyer et al., 2002; Rolls et al., 2015). In order to avoid cluttering the table with anatomical areas that are represented by relatively few voxels, we applied a cutoff threshold as a percentage of the total voxel count in each cluster. For the two biggest clusters of Table 2 the threshold was set at ~2% and for the rest at 5%.

3.3. SP-saccade related activations

In the way that we structured our analysis, saccades were used as a proxy to represent the steady-state condition of our visual system because of their lower variability, with smooth pursuits being the eye movement of interest. Hence, we were interested in the specific differences between SP- and saccade-related brain activations during natural viewing. These contrasts with $p_{FWE} < 0.05$ and initial threshold of $p < 0.001$ are visualized in Fig. 4.

This procedure identified three areas with stronger activation during

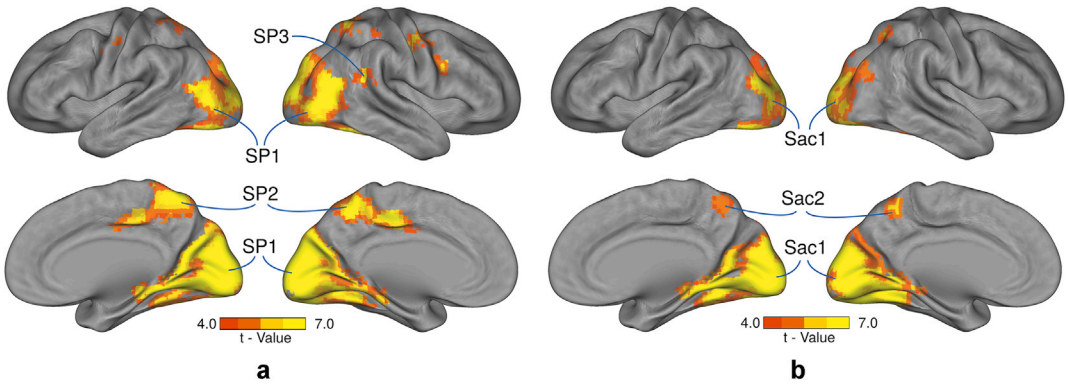

Fig. 3. a) SP-related activity with $p_{FWE} < 0.05$ with initial threshold of $p < 0.001$. Activations span bilaterally the visual areas of the brain (SP1: $k_E = 7647$, including the SP-related MT+/V5), bilaterally the middle cingulate expanding to the precuneus (SP2: $k_E = 2048$), and the right temporoparietal junction (SP3: $k_E = 109$). **b)** Saccade-related activity with $p_{FWE} < 0.05$ with initial threshold of $p < 0.001$. Activations span bilaterally the visual areas of the brain (Sac1: $k_E = 6437$) and the precuneus (Sac2: $k_E = 245$). For a detailed list of the subareas refer to Table 1.

Table 2
List of brain areas involved in both SP- and saccade-related clusters (coloured gray) and areas that are unique to SP (coloured white). The threshold for visualization was chosen at ~2% for the big clusters and 5% for the smaller clusters of Table 1. Therefore, the values do not sum up to the total number of voxels in each cluster.

Anatomical Area	Brodmann Area (Functional Region)	SP Activations						Saccade Activations					
		Peak activation			Part of	Num. of Voxels	Peak T-value	Peak activation			Part of	Num. of Voxels	Peak T-value
		X	Y	Z				X	Y	Z			
L Lingual	17, 18	-15	-76	2	SP1	528	12.04	-18	-79	2	Sac1	522	14.00
R Lingual	17, 18	12	-85	13	SP1	578	8.59	9	-85	-13	Sac1	599	10.73
L Calcarine	17, 18, 30	-3	-91	14	SP1	505	15.11	-9	-79	14	Sac1	503	15.94
R Calcarine	17, 18, 30	12	-85	8	SP1	447	11.32	12	-79	14	Sac1	309	15.01
L Cuneus	18, 19	-3	-91	17	SP1	368	13.33	-9	-82	17	Sac1	329	16.84
R Cuneus	18, 19	12	-94	14	SP1	405	10.72	12	-79	17	Sac1	309	11.75
L Occipital Sup	18, 19	-15	-97	23	SP1	273	12.51	-15	-82	11	Sac1	255	12.49
R Occipital Sup	18, 19	15	-97	17	SP1	219	10.42	18	-94	5	Sac1	164	8.04
L Occipital Mid	19, 37 (V5)	-48	-73	5	SP1	532	10.47	-15	-10	8	Sac1	555	7.18
R Occipital Mid	19, 37	27	-88	14	SP1	303	8.09	36	-67	29	Sac1	291	6.45
L Occipital Inf	18	-27	-82	-10	SP1	142	8.30	-27	-70	-10	Sac1	131	7.36
L Fusiform	18, 19	-24	-79	-10	SP1	347	8.44	-33	-49	-10	Sac1	352	11.90
R Fusiform	18, 19	33	-79	-16	SP1	415	11.09	27	-82	-16	Sac1	365	12.03
L Cerebellum 6	-	-6	-73	-16	SP1	181	8.14	-18	-73	-16	Sac1	185	9.41
R Cerebellum 6	-	15	-85	-16	SP1	214	9.76	24	-82	-19	Sac1	187	11.86
L Precuneus	5, 7	-9	-49	47	SP2	325	8.20	-6	-52	56	Sac2	100	6.25
R Precuneus	5, 7	6	-43	56	SP2	278	8.48	6	-52	53	Sac2	69	5.30
L Temporal Mid	19, 39 (V5)	-48	-70	8	SP1	122	8.61						
R Temporal Mid	19, 39 (V5)	41	-64	8	SP1	223	11.16						
R Temporal Inf	37 (V5)	48	-46	-25	SP1	101	8.02						
L Cingulate Mid	23, 24, 31	-9	-22	44	SP2	184	7.55						
R Cingulate Mid	23, 24, 31	12	-25	44	SP2	177	6.32						
R Paracentral lobule	5	12	-40	56	SP2	110	6.63						
R Temporal Sup	40	57	-40	17	SP3	55	6.94						
R Supramarginal	40	51	-40	23	SP3	45	6.57						

SP compared to saccades. In Fig. 4a the first area has bilateral activations of the motion processing and SP-related area MT+/V5 (right: $k_E = 169$, left: $k_E = 89$), with the second area containing the middle cingulate and extending to precuneus ($k_E = 655$). Lastly, the third area comprises of an activation in the right temporo-parietal junction (rTPJ; $k_E = 158$). Fig. 4b shows that the saccade > SP contrast has significant activations in V2 (right: $k_E = 91$). The full list of anatomical areas that are part of these clusters is provided in Table 3.

3.4. Accounting for movie motion

To differentiate SP from content motion related brain activations, we added an additional motion regressor during the first level analysis, which was again modelled as time-series with its values computed in a process similar to eye movement modulation of Section 2.4.1. Here, we present the results using the EpicFlow algorithm and whole frame mean motion modelling (results for the algorithm based on the minors of the structure tensor were qualitatively similar; data not shown). To better understand the relation between the EpicFlow motion estimation and our motion regressor, refer to the Supplementary Material and Fig. S2. The resulting motion regressor was uncorrelated with the saccade regressor (Pearson $r = -0.11$) and the same held true for the SP regressor (Pearson $r = 0.18$). The mean effects of SP-, saccade-, and motion-related BOLD-responses are visualized in Fig. 5. As can be seen from Fig. 5a and b, the activations for SP and saccades are qualitatively very close to the activations of Fig. 3 but with reduced size and intensity for SP when motion was included in the model (Fig. 5a). This reduction in SP-related activations followed from strong positive motion-related (Fig. 5c) activations

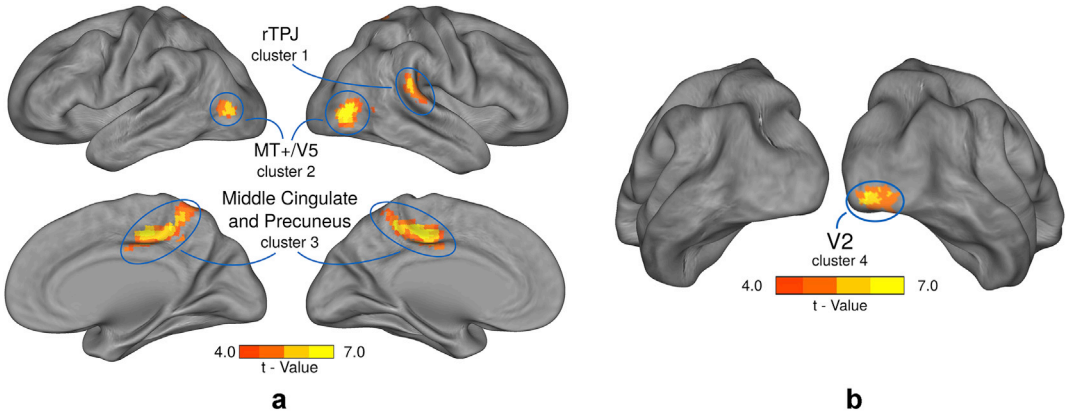

Fig. 4. a) Activations for SP > saccade at $p_{FWE} < 0.05$ with an initial threshold of 0.001. Activations in bilateral MT+/V5 (right: $k_E = 169$, left: $k_E = 89$), in the middle cingulate extending to precuneus ($k_E = 665$), and in the right temporo-parietal junction (rTPJ) ($k_E = 158$). **b)** Activations for saccade > SP contrast with $p_{FWE} < 0.05$ with initial threshold of 0.001. Activation in V2 (right: $k_E = 91$).

in roughly the same areas as the SP-related activations shown in Fig. 3a. Moreover, the activity in the cortex lining the superior temporal sulcus (STS) and in the supplementary motor area including the supplementary eye field (SEF) was negatively correlated with our motion regressor.

Following this model, the contrast SP > saccade yielded significant activations only in the middle cingulate ($k_E = 106$) and rTPJ ($k_E = 104$) areas (not shown graphically). The MT+/V5 and precuneus activations seen in Fig. 4a did not reach the significance threshold of 0.05 FWE.

The SP > motion contrast (Fig. 6) revealed bilateral activations in the cortex lining the superior temporal sulcus (STS; right: $k_E = 536$, left: $k_E = 194$), the precuneus ($k_E = 102$), and the supplementary motor area including the supplementary eye field (SEF; $k_E = 177$). To the contrary, the saccade > motion contrast did not reveal any significantly activated areas.

Table 3
List of areas involved in the SP > saccade and saccade > SP contrasts. The threshold for visualization was set to 15 voxels for all clusters and they do not sum up to the total voxel number for each cluster.

Anatomical Area	Peak Activation			Part of	Number of Voxels	Peak T-value
	X	Y	Z			
R Temporal Sup	60	−31	20	Cluster 1	73	5.95
R Supramarginal	60	−34	23	Cluster 1	36	5.83
L Temporal Mid	−51	−73	8	Cluster 2	23	5.79
R Temporal Mid	51	−70	−1	Cluster 2	103	10.11
L Occipital Mid	−48	−76	5	Cluster 2	89	7.46
L Cingulate Mid	−9	−31	44	Cluster 3	157	10.18
R Cingulate Mid	9	−22	44	Cluster 3	145	8.01
L Precuneus	−12	−44	44	Cluster 3	30	5.63
R Precuneus	12	−52	58	Cluster 3	92	6.72
R Paracentral Lobule	12	−37	47	Cluster 3	32	6.36
R Postcentral	15	−49	68	Cluster 3	32	6.77
R Parietal Sup	18	−49	68	Cluster 3	26	6.06
R Occipital Inf	24	−97	−7	Cluster 4	42	7.32
R Occipital Mid	39	−88	−1	Cluster 4	18	7.11
R Lingual	24	−91	−4	Cluster 4	15	5.64

4. Discussion

The aim of this study was to analyse brain activations related to SP and saccades as the most prominent eye movements in complex dynamic naturalistic scenes. To this end, we here presented methods based on off-the-shelf algorithms and modelling techniques that can handle the noisy and unstructured nature of motion and eye tracking data coming from scanner recordings, when dynamic natural scenes are used as stimuli. Our main results are in line with previous studies showing activations in the MT+/V5 area during SP, when SP and saccades were modelled separately. When an additional regressor representing motion content of the stimulus was included in the model, specific attention-related areas were identified, while some other brain areas (including MT+/V5) fell below the significance threshold due to the similar SP and motion evoked BOLD mean effects. These results were shown to be robust in a simple cross-validation procedure (see Supplementary Material and Fig. S1).

4.1. Validity of eye movement classification

The MOSP algorithm that we used for automatic eye movement detection has previously achieved state-of-the-art performance in a manually annotated data set (Startsev et al., 2016). To ensure that the MOSP algorithm returned high quality output in the studyforrest data set, we manually tuned its parameters based on visual inspection of a small portion of the results. A full manual annotation of a data set as big as the studyforrest (ca. 30 h) was not feasible given the fact that it takes approximately 15 s to label 1 s of gaze, and multiple annotators are needed for best results (Startsev et al., 2019).

4.2. Validity of algorithms defining motion content

A potential weak point in using motion estimation algorithms to define motion content of a stimulus is the fact, that they tend to give noisy results. For that reason, we validated the presented results by using two different motion estimation algorithms (Barth, 2000; Revaud et al., 2015). In both cases the identified brain activations were comparable, underlining the validity of our approach. In the second analysis of our study, we were interested in specifically identifying what drives SP in humans in the presence of motion. To this end, we used the mean frame motion as an approximation to background motion. However, there exist many other ways of modelling motion and we investigated two of them in more detail. In the first approach we modelled motion in a five degree

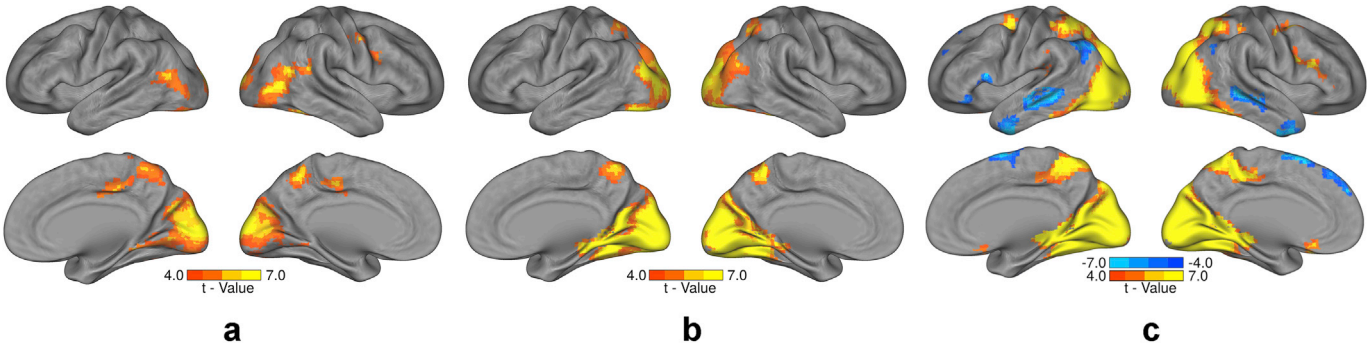

Fig. 5. Mean effects of SP and saccade when motion is included in the model, and the mean effect of motion itself, with $p_{FWE} < 0.05$ with initial threshold of $p < 0.001$. **a)** SP-related activations span bilaterally the visual areas of the brain ($k_E = 3706$, including the MT+/V5) and bilaterally the middle cingulate expanding to the precuneus ($k_E = 605$) **b)** Saccade-related activations span bilaterally the visual areas of the brain up to precuneus ($k_E = 7715$) **c)** Motion-related activations span bilaterally the visual areas of the brain (including the MT+/V5) and extend up to the middle cingulate and precuneus ($k_E = 10876$). Also negative activations exist bilaterally in the cortex lining the superior temporal sulcus (right: $k_E = 456$, left: $k_E = 450$), the temporo-parietal junction (right: $k_E = 118$, left: $k_E = 242$), and the supplementary motor area ($k_E = 304$).

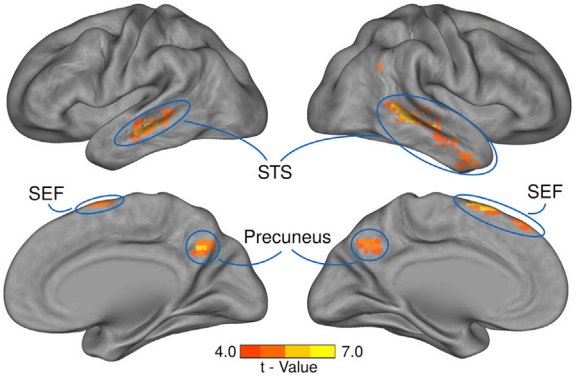

Fig. 6. Activations for SP > motion contrast with the motion regressor computed with EpicFlow and $p_{FWE} < 0.05$ with initial threshold of 0.001. Activations appear in the right superior and middle temporal gyri (posterior and anterior STS) ($k_E = 536$, peak xyz: 60, −55, 26) and in the left middle temporal gyrus (posterior STS) ($k_E = 194$, peak xyz: −60, −28, 11), bilaterally in the precuneus ($k_E = 102$, peak xyz: 6, −58, 35), and bilaterally in the supplementary eye field ($k_E = 177$, peak xyz: −6, 11, 62).

window around each gaze position. In the second approach we aimed at decorrelating the two regressors by modelling retinal motion. For this purpose, we subtracted the SP velocity (speed and direction) from the motion velocity in the same window and then used the magnitude of the resulting vector in our model. The resulting activations, while qualitatively similar, were weaker for both approaches in their extent and intensity. This can be partially attributed to the fact that in both of these approaches the correlation between the motion and SP regressors was higher than when only mean frame motion was used (window $r = 0.21$, window - SP velocity $r = 0.51$ vs. mean frame $r = 0.18$). The changes in the correlation values can be attributed to many factors. Generally the noisy results of motion estimation algorithms may become even noisier as we use the mean of a smaller window instead of the full frame. Also the reported gaze can be noisy and oftentimes has spatial offsets, which can result in missing completely or partially the moving target in the motion computation. As a result, SP velocity disproportionately influences the result of its subtraction from the window motion and thus returns higher correlation values. A similar effect appears with targets of very small size. It should be noted that the reported gaze position from the eye tracker was much noisier in the scanner than in the lab: the median dispersion of 25 ms windows of gaze data was 31 pixels in the scanner vs. 10 pixels in the lab for the studyforrest data set.

4.3. Brain areas related to variance in smooth pursuit

The contrast of SP > saccade with only the SP and saccade regressors included in the first level design matrix revealed activations in middle cingulate and precuneus, which have been previously associated with SP eye movement control (Tanabe et al., 2002; Kimmig et al., 2008) and visuo-spatial processing (Berman et al., 1999; Cavanna and Trimble, 2006). Additionally, this contrast yielded higher activation during SP than saccades related to the rTPJ, an area that is involved in guidance towards unattended areas (Corbetta et al., 2000; Wu et al., 2015; Marsman et al., 2016). Most importantly, this contrast revealed bilateral activations related to area MT+/V5, which is regarded a core motion processing area and has been associated with SP eye movements in previous studies (Kimmig et al., 2008; Petit and Haxby, 1999; Lencer and Peter, 2008; Nagel et al., 2006; Ohlendorf et al., 2010; Marsman et al., 2016). Notably, the MT+/V5 area became non-significant in the same contrast when a third regressor modelling the overall stimulus motion was added. This may be best explained by the fact that the variance of the BOLD response in this area was now shared between two regressors (SP and motion) instead of one (Ohlendorf et al., 2010), as can be seen from the mean effect of SP and motion in Fig. 5a and c. This demonstrates the difficulty in finding a single source of activation in natural scenes where many different factors may provoke activation of a specific area, and a complete disentanglement of such confounds may prove elusive.

4.4. Benefits of considering motion content in the model

Adding motion as a regressor to the model allowed us to identify SP-related activations that were not per se driven by the overall motion of the stimulus (Fig. 5a). Interestingly, motion itself additionally resulted in negative effects related to STS and SEF areas. Thus, when directly contrasting SP > motion, these two areas together with the precuneus occurred as being significantly stronger activated during SP than by motion content alone (Fig. 6). STS is considered a hub for information processing including the processing of biological motion (Saygin, 2007; Jastorff and Orban, 2009; Grossman et al., 2010) as well as processing of faces in situations requiring social cognition (Allison et al., 2000; Hoffman and Haxby, 2000; Lahnakoski et al., 2012). In line with this model, inhibiting STS activity by transcranial magnetic stimulation (TMS) resulted in difficulties perceiving biological motion (Grossman et al., 2005). Also, reduced activity in the STS (Freitag et al., 2008; Alaerts et al., 2014) has been associated with difficulties in understanding biological motion and emotional content in autism spectrum disorder patients (Hubert et al., 2007; Nackaerts et al., 2012; Alaerts et al., 2014). SEF activations have been associated with anticipatory eye movements, even in situations with invisible targets, reflecting cognitive input to

smooth pursuit planning independent from visual input (Lencer et al., 2004; Missal and Heinen, 2004; Ohlendorf et al., 2010).

When interpreting our finding related to motion content, it should be considered that our motion regressor was based on a low-level account of pixel-wise motion energy which might have failed to capture semantic properties of natural scenes. Thus, high values of motion content from our analyses were related to background and camera motion, which are both extensively used in professionally shot cinematic videos (Cutting et al., 2011), see Supplementary Material and Fig. S2. In contrast, moving mid-sized objects, i.e. socially meaningful targets, were linked to low motion content values. Thus, irrelevant motion modelled by our motion content regressor may have led to the observed negative activations bilaterally in the STS and the SEF unless SP to a meaningful target was performed.

Given the current rapid pace of progress in computer vision algorithms for high-level scene segmentation and understanding (Sevilla-Lara et al., 2016; Zhang et al., 2018), more complex modelling of the semantics of different types of motion information might enable a more fine-grained analysis of such effects in the future. Moreover, a correlation analysis as the one performed by (Hasson et al., 2004) could offer further insights into how each brain area relates to the scene flow and its semantic content.

4.5. Considering additional possible confounds

To at least partially alleviate the potential confounds of the motion energy analysis, we included additional regressors modelling basic video characteristics. In two control experiments, we modelled scene complexity based on saliency and edge density as attention-grabbing parameters in order to test whether these parameters interfere with the activations related to SP and motion content. In both cases, the mean effect of the validation regressor showed significant activations in some very small clusters (approx. 150–300 voxels overall in the posterior part of the brain and mostly in the visual cortex), but did not influence the activations regarding the main contrasts of interest. From these observations, we conclude that the eye movement planning processes are predominantly driven by the underlying motion based on the way we modelled each characteristic. However, a more exhaustive search of all the potential parameters and modelling techniques may be required in future studies of SP in dynamic natural scenes.

4.6. Lack of associations with frontal eye fields under natural viewing conditions

We did not identify any activations related to the frontal eye fields (FEF) which have been described to be involved in planning and execution of both SP and saccades (MacAvoy et al., 1991; Berman et al., 1999; Gagnon et al., 2006; Kimmig et al., 2008). One possible explanation might be that in typical experiments, participants switch between baseline periods of prolonged fixation and e.g. dot following or scene viewing. Instead, in the data set used here, participants were likely to constantly engage in some form of eye movement planning during continuous movie viewing, which is more representative of real-world viewing behaviour. Therefore, the variance of e.g. saccades in consecutive 2-s windows may not have been sufficient to identify all saccade-related activations, including FEF. Another limiting factor may be the small size of the FEF regions and the big variance in their reported location (Vernet et al., 2014) along with their activation being dependent on specific experimental conditions and instructions (Lencer et al., 2004).

5. Conclusions

In this study, we demonstrate brain networks specifically related to the often-overlooked smooth pursuit eye movements in complex dynamic naturalistic scenes. Our findings underline the notion that special care has to be taken to model variance across subjects, within subjects, and for different eye movement types. We also identified some of the confounds which arise from the semantic variation in movie content and which cannot be captured by a low-level image-based analysis alone. Nevertheless, our results show that findings from previous research with impoverished synthetic scenes can be qualitatively confirmed for highly complex, ecologically valid naturalistic stimuli.

Acknowledgments

This research was supported by the Elite Network Bavaria, funded by the Bavarian State Ministry of Science and the Arts.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.116491.

References

Adelson, E., Burt, P., 1981. Image data compression with the Laplacian pyramid. In: Proceeding of the Conference on Pattern Recognition and Image Processing, 218–23. IEEE Computer Society Press.

Agtzidis, I., Startsev, M., Dorr, M., 2016. Smooth pursuit detection based on multiple observers. In: Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications - ETRA '16. ACM Press. https://doi.org/10.1145/2857491.2857521.

Alaerts, K., Woolley, D.G., Jean, S., Di Martino, A., Swinnen, S.P., Wenderoth, N., 2014. Underconnectivity of the superior temporal sulcus predicts emotion recognition deficits in autism. Soc. Cogn. Affect. Neurosci. 9 (10), 1589–1600. https://doi.org/10.1093/scan/nst156.

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

Andric, M., Goldin-Meadow, S., Small, S.L., Hasson, U., 2016. Repeated movie viewings produce similar local activity patterns but different network configurations. Neuroimage 142 (November), 613–627. https://doi.org/10.1016/j.neuroimage.2016.07.061.

Barth, Erhardt, 2000. The minors of the structure tensor. Informatik Aktuell. https://doi.org/10.1007/978-3-642-59802-9_28.

Beauchamp, M.S., Petit, L., Ellmore, T.M., Ingeholm, J., Haxby, J.V., 2001. A parametric fMRI study of overt and covert shifts of visuospatial attention. Neuroimage 14 (2), 310–321. https://doi.org/10.1006/nimg.2001.0788.

Berman, R.A., Colby, C.L., Genovese, C.R., Voyvodic, J.T., Luna, B., Thulborn, K.R., Sweeney, J.A., 1999. Cortical networks subserving pursuit and saccadic eye movements in humans: an FMRI study. Hum. Brain Mapp. 8 (4), 209–225.

Brenner, E., Smeets, J.B.J., 2015. How moving backgrounds influence interception. PLoS One 10 (3), 1–21. https://doi.org/10.1371/journal.pone.0119903.

Cavanna, A.E., Trimble, M.R., 2006. The precuneus: a review of its functional anatomy and behavioural correlates. Brain: J. Neurol. 129 (3), 564–583. https://doi.org/10.1093/brain/awl004.

Corbetta, M., Kincade, J.M., Ollinger, J.M., McAvoy, M.P., Shulman, G.L., 2000. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat. Neurosci. 3 (3), 292–297. https://doi.org/10.1038/73009.

Cutting, James E., Brunick, Kaitlin L., Delong, Jordan E., Iricinschi, Catalina, Candan, Ayse, 2011. Quicker, faster, darker: changes in Hollywood film over 75 years. I-Perception 2 (6), 569–576.

Dar, A.H., Wagner, A.S., Hanke, M., 2019. REMoDNaV: robust eye movement detection for natural viewing. bioRxiv 619254. https://doi.org/10.1101/619254.

Dorr, M., Martinetz, T., Gegenfurtner, K.R., Barth, E., 2010. Variability of eye movements when viewing dynamic natural scenes. J. Vis. 10 (10), 28. https://doi.org/10.1167/10.10.28.

Ettinger, Ulrich, Ffytche, Dominic H., Kumari, Veena, Kathmann, Norbert, Reuter, Benedikt, Zelaya, Fernando, Steven, C., Williams, R., 2008. Decomposing the neural correlates of antisaccade eye movements using event-related FMRI. Cerebr. Cortex 18 (5), 1148–1159.

Freitag, C.M., Konrad, C., Häberlen, M., Kleser, C., Gontard, A. von, Reith, W., Troje, N.F., Krick, C., 2008. Perception of biological motion in autism spectrum disorders. Neuropsychologia 46 (5), 1480–1494. https://doi.org/10.1016/j.neuropsychologia.2007.12.025.

Gagnon, D., Paus, T., Grosbras, M.-H., Bruce Pike, G., O'Driscoll, G.A., 2006. Transcranial magnetic stimulation of frontal oculomotor regions during smooth pursuit. J. Neurosci. 26 (2), 458–466. https://doi.org/10.1523/JNEUROSCI.2789-05.2006. The Official Journal of the Society for Neuroscience.

Gold, J.I., Shadlen, M.N., 2007. The neural basis of decision making. Annu. Rev. Neurosci. 30 (1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038.

Grossman, Emily D., Battelli, Lorella, Pascual-Leone, Alvaro, 2005. Repetitive TMS over posterior STS disrupts perception of biological motion. Vis. Res. 45 (22), 2847–2853.

Grossman, E.D., Jardine, N.L., Pyles, J.A., 2010. fMR-adaptation reveals invariant coding of biological motion on the human STS. Front. Hum. Neurosci. 4 (March), 15. https://doi.org/10.3389/neuro.09.015.2010.

Hanke, M., Adelhöfer, N., Kottke, D., Iacovella, V., Sengupta, A., Kaule, F.R., Roland, N., Waite, A.Q., Baumgartner, F., Stadler, J., 2016. A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. Scientific Data 3 (October), 160092. https://doi.org/10.1038/sdata.2016.92.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., David, J., Heeger, 2008. Neurocinematics: the neuroscience of film. Projections 2 (1), 1–26. https://doi.org/10.3167/proj.2008.020102.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303 (5664), 1634–1640. https://doi.org/10.1126/science.1089506.

Hoffman, E.A., Haxby, J.V., 2000. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat. Neurosci. 3, 80–84. https://doi.org/10.1038/71152.

Hooge, I.T.C., Niehorster, D.C., Nyström, M., Andersson, R., Hessels, R.S., 2018. Is human classification by experienced untrained observers a gold standard in fixation detection? Behav. Res. Methods 50 (5), 1864–1881. https://doi.org/10.3758/s13428-017-0955-x.

Hubert, B., Wicker, B., Moore, D.G., Monfardini, E., Duverger, H., Da Fonséca, D., Deruelle, C., 2007. Brief report: recognition of emotional and non-emotional biological motion in individuals with autistic spectrum disorders. J. Autism Dev. Disord. 37 (7), 1386–1392. https://doi.org/10.1007/s10803-006-0275-y.

Ilg, U.J., Peter, T., 2008. The neural basis of smooth pursuit eye movements in the rhesus monkey brain. Brain Cogn. 68 (3), 229–240. https://doi.org/10.1016/j.bandc.2008.08.014.

Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (11), 1254–1259. https://doi.org/10.1109/34.730558.

Jastorff, J., Orban, G.A., 2009. Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. J. Neurosci. 29 (22), 7315–7329. https://doi.org/10.1523/JNEUROSCI.4870-08.2009.

Kay, K.N., Naselaris, T., Gallant, J.L., 2011. fMRI of human visual areas in response to natural images. https://doi.org/10.6080/K0QN64NG.

Kellar, D., Newman, S., Franco, P., Hu, C., Port, N.L., 2018. Comparing fMRI activation during smooth pursuit eye movements among contact sport athletes, non-contact sport athletes, and non-athletes. NeuroImage: Clinical 18 (January), 413–424. https://doi.org/10.1016/j.nicl.2018.01.025.

Kimmig, H., Ohlendorf, S., Speck, O., Sprenger, A., Rutschmann, R.M., Haller, S., Greenlee, M.W., 2008. fMRI evidence for sensorimotor transformations in human cortex during smooth pursuit eye movements. Neuropsychologia 46 (8), 2203–2213. https://doi.org/10.1016/j.neuropsychologia.2008.02.021.

Kleiser, R., Konen, C.S., Seitz, R.J., Frank, B., 2009. I know where you'll look: an fMRI study of oculomotor intention and a change of motor plan. Behav. Brain Funct.: BBF 5 (July), 27. https://doi.org/10.1186/1744-9081-5-27.

Lahnakoski, Juha M., Glerean, Enrico, Juha Salmi, Jääskeläinen, Iiro P., Sams, Mikko, Hari, Riitta, Nummenmaa, Lauri, 2012. Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. In: Frontiers in Human Neuroscience. https://doi.org/10.3389/fnhum.2012.00233.

Larsson, L., Nyström, M., Andersson, R., Martin, S., 2015. Detection of fixations and smooth pursuit movements in high-speed eye-tracking data. Biomed. Signal Process. Control 18, 145–152. https://doi.org/10.1016/j.bspc.2014.12.008.

Larsson, L., Nyström, M., Martin, S., 2013. Detection of saccades and postsaccadic oscillations in the presence of smooth pursuit. IEEE Trans. Biomed. Eng. 60 (9), 2484–2493. https://doi.org/10.1109/tbme.2013.2258918.

Lencer, R., Nagel, M., Sprenger, A., Zapf, S., Erdmann, C., Heide, W., Binkofski, F., 2004. Cortical mechanisms of smooth pursuit eye movements with target blanking. An fMRI study. Eur. J. Neurosci. 19 (5), 1430–1436. https://doi.org/10.1111/j.1460-9568.2004.03229.x.

Lencer, R., Peter, T., 2008. Neurophysiology and neuroanatomy of smooth pursuit in humans. Brain Cogn. 68 (3), 219–228. https://doi.org/10.1016/j.bandc.2008.08.013.

Lukasova, K., Nucci, M.P., Machado de Azevedo Neto, R., Vieira, G., Sato, J.R., Amaro Jr., E., 2018. Predictive saccades in children and adults: a combined fMRI and eye tracking study. PLoS One 13 (5), 1–17. https://doi.org/10.1371/journal.pone.0196000.

Luna, B., Thulborn, K.R., Strojwas, M.H., McCurtain, B.J., Berman, R.A., Genovese, C.R., Sweeney, J.A., 1998. Dorsal cortical regions subserving visually guided saccades in humans: an fMRI study. Cerebr. Cortex 8 (1), 40–47. https://doi.org/10.1093/cercor/8.1.40.

MacAvoy, M.G., Gottlieb, J.P., Bruce, C.J., 1991. Smooth-pursuit eye movement representation in the primate frontal eye field. Cerebr. Cortex 1 (1), 95–102. https://doi.org/10.1093/cercor/1.1.95.

Mannion, D., 2015. fMRI responses of human visual cortex (v1, v2, v3) to natural image patches obtained from above and below the centre of gaze of an observer freely-navigating an outdoor environment. https://doi.org/10.6080/K0JS9NC2.

Marsman, J.-B.C., Cornelissen, F.W., Dorr, M., Vig, E., Barth, E., Remco, R.J., 2016. A novel measure to determine viewing priority and its neural correlates in the human brain. J. Vis. 16 (6), 3. https://doi.org/10.1167/16.6.3.

Mathe, S., Sminchisescu, C., 2012. Dynamic eye movement datasets and learnt saliency models for visual action recognition. Computer Vision – ECCV 2012 842–856. https://doi.org/10.1007/978-3-642-33709-3_60.

McDowell, J.E., Dyckman, K.A., Austin, B.P., Clementz, B.A., 2008. Neurophysiology and neuroanatomy of reflexive and volitional saccades: evidence from studies of humans. Brain Cogn. 68 (3), 255–270. https://doi.org/10.1016/j.bandc.2008.08.016.

Missal, M., Heinen, S.J., 2004. Supplementary eye fields stimulation facilitates anticipatory pursuit. J. Neurophysiol. 92 (2), 1257–1262. https://doi.org/10.1152/jn.01255.2003.

Mueller, S., Wang, D., Fox, M.D., Yeo, B.T.T., Jorge, S., Sabuncu, M.R., Shafee, R., Lu, J., Liu, H., 2013. Individual variability in functional connectivity architecture of the human brain. Neuron 92 (2), 586–595. https://doi.org/10.1016/j.neuron.2012.12.028.

Nackaerts, E., Wagemans, J., Werner, H., Swinnen, S.P., Wenderoth, N., Alaerts, K., 2012. Recognizing biological motion and emotions from point-light displays in autism spectrum disorders. PLoS One 7 (9), 1–12. https://doi.org/10.1371/journal.pone.0044473.

Nagel, M., Sprenger, A., Zapf, S., Erdmann, C., Kömpf, D., Heide, W., Binkofski, F., Lencer, R., 2006. Parametric modulation of cortical activation during smooth pursuit with and without target blanking. An fMRI study. Neuroimage 29 (4), 1319–1325. https://doi.org/10.1016/j.neuroimage.2005.08.050.

Nardo, D., Santangelo, V., Macaluso, E., 2014. Spatial orienting in complex audiovisual environments. Hum. Brain Mapp. 35 (4), 1597–1614. https://doi.org/10.1002/hbm.22276.

Ohlendorf, S., Sprenger, A., Speck, O., Glauche, V., Haller, S., Hubert, K., 2010. Visual motion, eye motion, and relative motion: a parametric fMRI study of functional specializations of smooth pursuit eye movement network areas. J. Vis. 10 (14), 21. https://doi.org/10.1167/10.14.21.

Petit, L., Haxby, J.V., 1999. Functional anatomy of pursuit eye movements in humans as revealed by fMRI. J. Neurophysiol. 82 (1), 463–471. https://doi.org/10.1152/jn.1999.82.1.463.

Poldrack, Russell A., Mumford, Jeanette A., Nichols, Thomas E., 2011. Handbook of Functional MRI Data Analysis. Cambridge University Press.

Revaud, Jerome, Weinzaepfel, Philippe, Harchaoui, Zaid, Schmid, Cordelia, 2015. EpicFlow: edge-preserving interpolation of correspondences for optical flow. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr.2015.7298720.

Rolls, E.T., Joliot, M., Tzourio-Mazoyer, N., 2015. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. Neuroimage 122 (November), 1–5. https://doi.org/10.1016/j.neuroimage.2015.07.075.

Sanocki, T., Islam, M., Doyon, J.K., Lee, C., 2015. Rapid scene perception with tragic consequences: observers miss perceiving vulnerable road users, especially in crowded traffic scenes. Atten. Percept. Psychophys. 77 (4), 1252–1262. https://doi.org/10.3758/s13414-015-0850-4.

Saygin, A.P., 2007. Superior temporal and premotor brain areas necessary for biological motion perception. Brain 130 (9), 2452–2461. https://doi.org/10.1093/brain/awm162.

Sestieri, C., Pizzella, V., Cianflone, F., Luca Romani, G., Corbetta, M., 2007. Sequential activation of human oculomotor centers during planning of visually-guided eye movements: a combined fMRI-MEG study. Front. Hum. Neurosci. 2, 1. https://doi.org/10.3389/neuro.09.001.2007.

Sevilla-Lara, L., Sun, D., Jampani, V., Black, M.J., 2016. Optical flow with semantic segmentation and localized layers. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. https://doi.org/10.1109/cvpr.2016.422.

Startsev, M., Agtzidis, I., Dorr, M., 2016. Smooth pursuit. http://michaeldorr.de/smoothpursuit/.

Startsev, M., Agtzidis, I., Dorr, M., 2019. 1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits. Behav. Res. Methods 51 (2), 556–572. https://doi.org/10.3758/s13428-018-1144-2.

Startsev, M., Agtzidis, I., Dorr, M., 2019. Characterising and automatically detecting smooth pursuit in a large-scale ground-truth data set of dynamic natural scenes. J. Vis. 19 (14), 10. https://doi.org/10.1167/19.14.10.

Tagliazucchi, E., Laufs, H., 2014. Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. Neuron 82 (3), 695–708. https://doi.org/10.1016/j.neuron.2014.03.020.

Tanabe, J., Tregellas, J., Miller, D., Ross, R.G., Freedman, R., 2002. Brain activation during smooth-pursuit eye movements. Neuroimage 17 (3), 1315–1324. https://doi.org/10.1006/nimg.2002.1263.

Tatler, B.W., Brockmole, J.R., Carpenter, R.H.S., 2017. LATEST: a model of saccadic decisions in space and time. Psychol. Rev. 124 (3), 267–300. https://doi.org/10.1037/rev0000054.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15 (1), 273–289. https://doi.org/10.1006/nimg.2001.0978.

Vanderwal, T., Eilbott, J., Finn, E.S., Craddock, R.C., Adam, T., Castellanos, F.X., 2017. Individual differences in functional connectivity during naturalistic viewing conditions. Neuroimage 157 (August), 521–530. https://doi.org/10.1016/j.neuroimage.2017.06.027.

Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L.C., Castellanos, F.X., 2015. Inscapes: a movie paradigm to improve compliance in functional magnetic resonance imaging. Neuroimage 122 (November), 222–232. https://doi.org/10.1016/j.neuroimage.2015.07.069.

Van Essen, D.C., 2005. A population-average, Landmark- and surface-based (PALS) atlas of human cerebral cortex. Neuroimage 28 (3), 635–662. https://doi.org/10.1016/j.neuroimage.2005.06.058.

Van Essen, D.C., Drury, H.A., Dickson, J., Harwell, J., Hanlon, D., Anderson, C.H., 2001. An integrated software suite for surface-based analyses of cerebral cortex. J. Am. Med. Inform. Assoc. 8 (5), 443–459. https://doi.org/10.1136/jamia.2001.0080443.

Vernet, M., Quentin, R., Chanes, L., Mitsumasu, A., Valero-Cabré, A., 2014. Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal

regions involved in eye movements and associated cognitive operations. Front. Integr. Neurosci. 8 (August), 66. https://doi.org/10.3389/fnint.2014.00066.

Wu, Q., Chang, C.-F., Xi, S., Huang, I.-W., Liu, Z., Juan, C.-H., Wu, Y., Fan, J., 2015. A critical role of temporoparietal junction in the integration of top-down and bottom-up attentional control. Hum. Brain Mapp. 36 (11), 4317–4333. https://doi.org/10.1002/hbm.22919.

Zemblys, R., Niehorster, D.C., Holmqvist, K., 2019. gazeNet: end-to-end eye-movement event detection with deep neural networks. Behav. Res. Methods 51 (2), 840–864. https://doi.org/10.3758/s13428-018-1133-5.

Zhang, Hang, Dana, Kristin, Shi, Jianping, Zhang, Zhongyue, Wang, Xiaogang, Tyagi, Ambrish, Agrawal, Amit, 2018. Context encoding for semantic segmentation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/cvpr.2018.00747.