Neural Representations of Self-Motion During Natural Scenes in the Human Brain

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von

Didem Korkmaz Hacıalihafız
aus İzmir, Turkey

September – 2016

Tag der mündlichen Prüfung: 14.12.2016

Dekan der Math.-Nat. Fakultät: Prof. Dr. W. Rosenstiel
Dekan der Medizinischen Fakultät: Prof. Dr. I. B. Autenrieth

1. Berichterstatter: Dr Andreas Bartels
2. Berichterstatter: Prof. Dr. Uwe Ilg

Prüfungskommission: Dr. Andreas Bartels
Prof. Dr. Uwe Ilg
Prof. Dr. Hanspeter Mallot
Prof Dr. Steffen Gais

Declaration:

I hereby declare that I have produced the work entitled „ Neural representations of self-motion during natural scenes in the human brain", submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Tübingen, ...

Date Signature

Abstract

Navigating through the environment is one of the important everyday tasks of the visual system. This task relies on processing of at least two visual cues: visual motion, and scene content. Our sense of motion heavily relies on understanding and separating visual cues resulting from object motion and self-motion. Processing and understanding of visual scenes is an equally abundant task we are exposed to in our everyday environment. Together, motion and scene processing allow us to fulfill navigation tasks such as way finding and spatial updating.

In terms of neural processing, both, regions involved in motion processing, and regions involved in scene processing have been studied in great detail. However, how motion regions are influenced by scene content and how scene regions are involved in motion processing has barely been addressed.

In order to understand how self-motion and scene processing interact in the human brain, I completed a series of studies as part of this thesis. First of all, using planar horizontal motion and visual scenes, the first study of this thesis investigates motion responses of scene regions. The next study investigates whether eye-centered or world-centered reference frames are used during visual motion processing in scene regions, using objective 'real' motion and retinal motion during pursuit eye movements and natural scene stimuli. The third study investigates the effect of natural scene content during objective and retinal motion processing in motion regions. The last study investigates how motion speed is represented in motion regions during objective and retinal motion. Since many visual areas are optimized for natural visual stimuli, the speed responses were tested on Fourier scrambles of natural scene images in order to provide natural scene statistics as visual input.

I found evidence that scene processing regions parahippocampal place area (PPA) and occipital place area (OPA) are motion responsive while retrosplenial cortex (RSC) is not. In addition, PPA's motion responses are modulated by scene content. With respect to reference frames, I found that PPA prefers a world-centered reference frame while viewing dynamic scenes.

The results from motion regions (MT/V5+, V3A, V6 and cingulate sulcus visual area (CSv)) revealed that motion responses of all of them are enhanced during exposure

to scenes compared to Fourier-scramble, whereas only V3A responded also to static scenes. The last study showed that all motion responsive regions tested (MT/V5, MST, V3A, V6 and CSv) are modulated by motion speed but only V3A has a distinctly stronger speed tuning for objective compared to retinal motion.

These results reveal that using natural scene stimuli is important while investigating self-motion responses in human brain: many scene regions are modulated by motion and one of them (PPA) even differentiates object motion from retinal motion. Conversely, many motion regions are modulated by scene content and one of them (V3A) is even responsive to still scenes. Moreover, the objective motion preference of V3A is even stronger during higher speeds. These results question a strong separation of ‘where’ and ‘what’ pathways and show that scene region PPA and motion region V3A have similar objective motion and scene preferences.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Andreas Bartels for his supervision and support during my PhD. His positive comments, challenging questions and willingness to share his knowledge helped me enormously during my PhD. Besides my supervisor, I would also like to thank the other members of my thesis committee, Professor Dr. Uwe Ilg and Professor Dr. Hanspeter Mallot for their insightful comments and suggestions as well as their support. I thank all members of Vision and Cognition lab for their help. Further, I would like to thank all participants who took part in my experiments.

I would also like to thank my family and friends for their continuous support and encouragement despite being hundreds of kilometres away. I must also thank my beloved cats, Guru, Mofu and Boyoz. A special thanks goes to my husband Bleda for being in my life and supporting me.

Table of Contents

1. Synopsis

The visual system performs many important roles that we are not necessarily aware of. Keeping a stable perception of the world despite continuous head-, body- and eye-movement is one of the important roles of the visual system. In order to distinguish whether the perceived motion is self-induced motion or motion induced by external movement, i.e. real motion, the visual system needs to utilize retinal inputs and non-retinal motion cues. As a result, visual system generates an implicit map of self-generated motion and uses this map on different tiers of visual processing such as motion processing, scene processing and navigation. Yet, the mechanisms behind stable perception during eye movements are still not well known.

Self-motion processing is important for spatial processing. While moving through space, the visual system helps mapping the space for further use. The visual system needs to collect information about the environment it interacts with, and the visual basis of this environment consists of visual scenes. However, most studies investigating self-motion used simple dot-field stimuli. High-level visual motion processing involving complex stimuli remains unexplored. Scenes represent one such class of complex stimuli and possess additional features of interest.

Given that the content that is perceived as stable during eye movements typically consists of real-world scenes, it is also crucial to investigate how these mechanisms are affected by scene content. For instance, naturalistic stimuli have been shown to be more optimal for studying properties of the visual system (Kayser, Kording, & Konig, 2004; Parraga, Troscianko, & Tolhurst, 2000; Vinje & Gallant, 2000). In order to perceive the spatial layout correctly, the visual system integrates the scene from different viewpoints. Visual scenes we see and interact with in daily life are mostly moving. Yet, our brain recognizes the scenes and most of the time processes them without us even noticing. The visual system needs to keep both scene specific representations and scene invariant representations at the same time. This could be done either by representing details of scene snapshots or by an extended representation including multiple successive views of the scene. Most studies on scene processing were done using static scenes or snapshots of scenes. Similarly, most studies on motion processing were done using abstract stimuli such as gratings or random dot displays. In real life, we encounter dynamic scenes. Little is known on how the content such as real world scenes affects motion responses and how motion affects scene processing in the brain.

During self-motion, the visual system needs to use both world-centered and retinotopic representations of visual data for spatial updating, such as keeping track of the location of objects and self separately. As we move through space, objects' locations on our retina constantly change. Yet, we do know and can keep track of the location information of objects. This is called spatial updating. In order to be able to do spatial updating, the visual system needs to use both world-centered and retinotopic representations of visual data, such as object location. Indeed, many studies established which brain regions are associated with world-centered representation. Higher-level motion processing regions such as MST, V3A and V6, as well as many parietal regions, are modulated by world-centered representation of the motion (Erickson & Thier, 1991; Galletti, Battaglini, & Fattori, 1990; Ilg, Schumann, & Thier, 2004). World-centered and eye-centered representations of scenes might exist in different visual regions.

The current PhD thesis focuses on how the naturalistic scene stimuli affect self-motion processing in human brain. In more detail, this thesis investigates how the visual system unravel self-generated motion from object motion during smooth pursuit eye movements on naturalistic stimuli such as visual scenes, by utilizing self-induced retinal signals. Here, I address these questions using continuous horizontal motion and natural scene stimuli. Horizontal motion is very common in nature and it resembles lateral head movements, eye movements or looking out of a moving train. It is also found in movies and visual system prefers horizontal motion (Bartels, Zeki, & Logothetis, 2008).

This thesis presents four studies that use functional magnetic resonance imaging (fMRI) technique. In the experiments explained in chapters 3, 4 & 5, I used a well-established paradigm with 2 factors; objective motion and pursuit eye movements (Fischer, Bulthoff, Logothetis, & Bartels, 2012a). This design helped me to separate world-centered ('objective') and eye-centered ('retinal') motion responses during smooth pursuit eye movements. Before starting to investigate world- centered and eye-centered motion responses in scene regions, I first conducted the study explained in chapter 2, to characterize motion responses in scene regions, since this was not done before. The paradigm was background moving or static when eyes were fixated and the background content was also varied i.e. either natural outdoor scenes or Fourier scramble images of those scenes. That way, the motion responses in scene regions and whether/how those responses are modulated by scene content was nicely investigated in one experiment. Next, in chapter 3, I investigated whether scene regions use world-centered or eye-centered reference frames. The stimuli consisted of the aforementioned paradigm with objective motion and pursuit, as well as an additional factor for scenes.

The study in chapter 4 investigates how the motion processing regions are modulated by scene content and how their scene modulation depends on objective and retinal motion, respectively. Finally, Chapter 5 examines the speed tuning of motion regions during objective and retinal motion. The speed tuning properties are investigated using scrambled scene images in order to provide naturalistic image statistics for the visual system.

In this chapter, I give an overview of scene processing and motion processing, a definition of objective and retinal motion, an overview of literature on eye movements, speed processing and reference frames, as well as an overview of the studies in this thesis and their results.

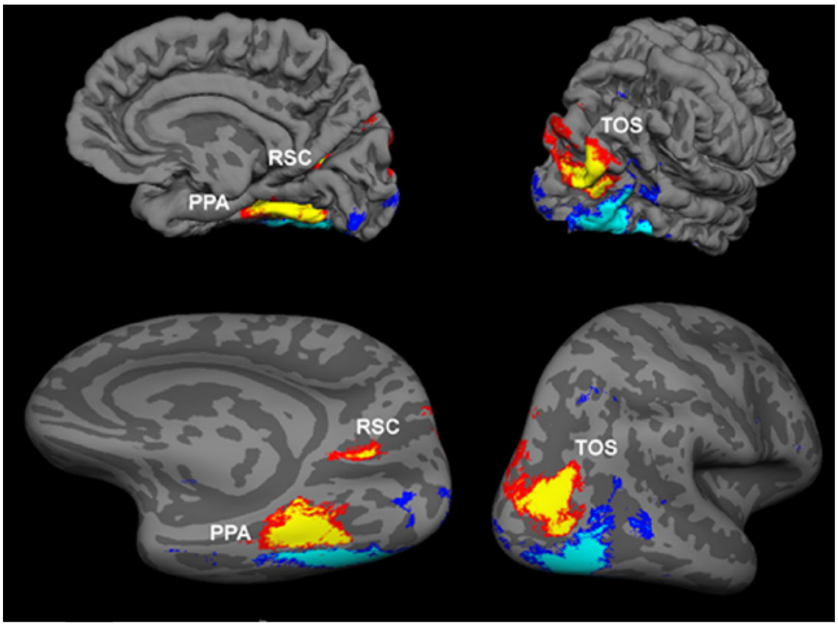
1.1. Neural processing of scenes

Understanding and perceiving the environment around us is one of everyday tasks of the visual system. Scene processing is crucial for higher-level functions such as navigation or self-motion perception. Visual system can recognize scenes and extract information regarding scene category or scene properties such as scene layout easily and in a very short time. Previous studies on scene processing have shown that the parahippocampal place area (PPA) (Epstein & Kanwisher, 1998) and the retrosplenial cortex (RSC) (Maguire, 2001), as well as a less studied area occipital place area (OPA) (also known as transverse occipital sulcus (TOS)) (Dilks, Julian, Paunov, & Kanwisher, 2013; Epstein, Higgins, & Thompson-Schill, 2005; Hasson, Harel, Levy, & Malach, 2003), have a key role in understanding and representing scenes. The perception of scenes around us seems to be very well integrated and continuous. To enable both specificity and invariance during scene processing, multiple levels of scene representation might coexist in human brain. During natural scene viewing, the retinal image constantly changes and scene regions may be playing a role in stable perception of scenes, for instance by using large-scale features of scenes to keep the representation of the scenes stable. Henderson et al. found that PPA responds in a gradual way as snapshots of scenes become more like a complete scene while RSC responds just to the presence of a complete scene (Henderson, Larson, & Zhu, 2008). These results suggest scene-processing areas are particularly sensitive to the 3D geometric structure that distinguishes scenes from other types of complex and meaningful visual stimuli (Henderson et al., 2008). However, the exact role of scene-processing regions in scene processing is still under debate.

PPA is the most studied region amongst these scene responsive regions. It prefers scenes or houses to faces, different objects and/or scrambled scenes (Epstein, Harris, Stanley, & Kanwisher, 1999). Lesions in the PPA can lead to a total inability to identify scenes and landmarks (Mendez & Cherrier, 2003). It is thought that PPA is selectively processing spatial layout of the scenes (Epstein, 2008) and it mainly encodes spatial aspects of scenes (Kravitz, Peng, & Baker, 2011). Yet, its actual role in scene processing is still not entirely understood. Parahippocampal region has two retinotopic regions, PHC-1 and PHC-2 ((Arcaro, McMains, Singer, & Kastner, 2009), but also see (MacEvoy & Epstein, 2007), which concludes that PPA and RSC are insensitive to retinal position of the stimuli). Many studies suggest that PPA encodes spatial layout in a viewpoint dependent way (Epstein, Graham, & Downing, 2003; Epstein, Higgins, Jablonski, & Feiler, 2007; Epstein et al., 2005; Park & Chun, 2009; Park, Chun, & Johnson, 2010). Another view on PPA focuses on its responses to physical features of a scene, such as high spatial frequencies (Rajimehr, Devaney, Bilenko, Young, & Tootell, 2011) or a preference for cardinal over oblique orientation (Nasr & Tootell, 2012). Whether PPA is only interested in low-level scene features or has a role in higher-level processing of scenes, such as scene category, is still under debate.

Figure 1 shows scene responsive regions in human brain. As can be seen in Figure 1B, the highest responses were induced by a landscape and cityscape scene images. In order to localize scene responsive regions, a contrast of scenes (landscape + cityscape) versus face images was used in this thesis.

A. Scene responsive areas



B. Response properties of PPA and RSC

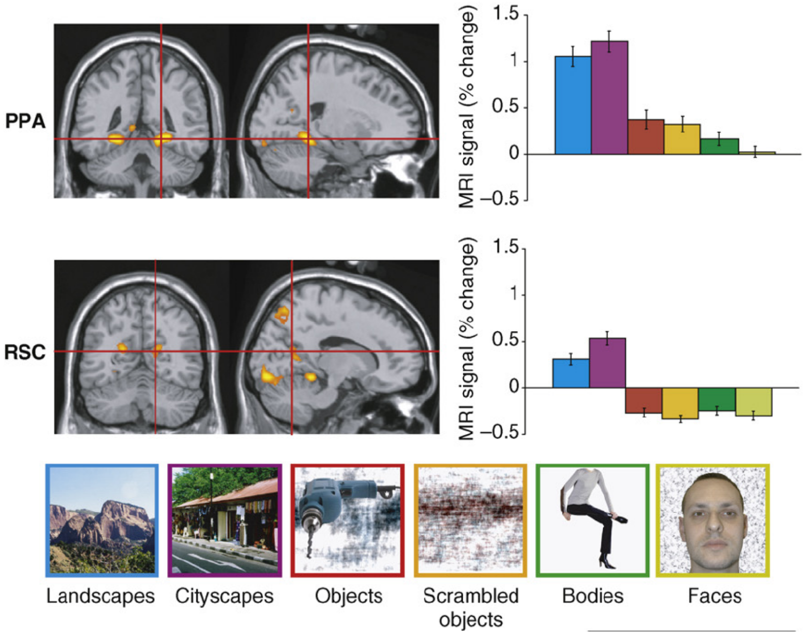


Figure 1. Scene responsive regions in human brain. (A) An overview of regions in human brain that prefers scenes versus faces. Red/yellow versus blue/cyan shows higher activity for scenes versus faces. RSC, PPA and TOS (OPA) are shown with white labels. (Nasr et al., 2011). (B) Response profiles of PPA and RSC for various object/stimuli categories (Epstein, 2008).

PPA consists of two subregions; while one part is processing low-level visual features and object shape and connected to OPA, the other subregion is more involved in memory and scene context and is connected to RSC (Baldassano, Beck, & Fei-Fei, 2013). PPA shows similarities to both RSC and OPA, yet there are also studies showing distinctions between PPA and OPA and between PPA and RSC. For instance, both PPA and RSC represent scenes in an extended way compared to the initial view (Park, Intraub, Yi, Widders, & Chun, 2007). Some evidence suggests that PPA and OPA might have distinct roles in scene processing; PPA is tolerant to mirror image reversals whereas OPA and RSC showed mirror image reversal sensitivity (Dilks, Julian, Kubilius, Spelke, & Kanwisher, 2011). PPA is not sensitive to egocentric distance information while this sensitivity was present in OPA and RSC (Persichetti & Dilks, 2016). In addition, scene representations in PPA were also dependent on how good navigators were the participants, whereas responses in OPA was not sensitive to the ability to navigate (Epstein et al., 2005). These studies reveal similarities and differences of scene regions, but still not much is known regarding how exactly the scene processing network works.

Another scene responsive region, RSC, is proposed to have a role in navigation and landmark representation. RSC is primarily involved in spatial memory (Epstein & Vass, 2014). It is also thought that RSC connects internal spatial representations to local topographical features (Marchette, Vass, Ryan, & Epstein, 2014). Lesion studies have shown that damage to the RSC can lead to an inability to use landmarks for self-orientation in space and deficits in navigational abilities (Maguire, 2001). During navigation, Sherrill et al found that optic flow responsive regions V6, V3A and MT+ shows a strong connectivity to regions involved in navigation, such as RSC and hippocampus (Sherrill et al., 2015). One view on RSC is that it is a 'higher-tier' region in scene-processing network. Supporting this view, RSC did not show any topographical organization (Epstein, Parker, & Feiler, 2007; Ward, MacEvoy, & Epstein, 2010) and it responds more to familiar scenes compared to unfamiliar ones (Epstein, Higgins, et al., 2007). Although it showed mirror image reversal sensitivity like OPA (Dilks et al., 2011), which still could be explained by the importance of this kind of sensitivity for navigation, previous studies showed that RSC is viewpoint invariant (Park & Chun, 2009; Park et al., 2010; Vass & Epstein, 2013). Moreover, A recent study showed that the most permanent landmarks activated RSC and that the strength of the responses in

RSC was dependent on how good navigators were the participants (Auger, Mullally, & Maguire, 2012). The integration of viewpoints in RSC is dependent on whether the location of the viewer in scene changed, and whether the viewpoints are continuous or not (Park & Chun, 2009). Viewpoint invariance is important in order to represent the scene in a broader space, to be able to navigate successfully. While many of these studies support the view that RSC has a role in spatial navigation, whether it uses different reference frames for this purpose is not known.

OPA is the least known among scene-responsive areas. While OPA is causally involved in scene recognition (Dilks et al., 2013; Ganaden, Mullin, & Steeves, 2013), its exact role in scene representation is yet unknown. It is less successfully localized compared to other scene responsive regions (Konkle & Oliva, 2012; Korkmaz Hacialihafiz & Bartels, 2015; Mullin & Steeves, 2013). OPA has retinotopic organisation (Grill-Spector, 2003; Levy, Hasson, Harel, & Malach, 2004; Nasr et al., 2011). Together with PPA, OPA is sensitive to peripheral stimuli (Levy, Hasson, Avidan, Hendler, & Malach, 2001; Levy et al., 2004). While the receptive field size of OPA is smaller than that of RSC or PPA (MacEvoy & Epstein, 2007), it has large enough receptive fields to cover across both hemifields (Ward et al., 2010). Distinct roles are suggested for OPA in scene processing; such as a role in encoding the higher order spatial relationships within a scene (Bettencourt & Xu, 2013) or encoding local scene elements, thus being involved in visually guided navigation and obstacle avoidance (Kamps, Julian, Kubilius, Kanwisher, & Dilks, 2016). It is thought to be a transitional region between lower and higher cortical levels, former being retinotopic and latter being scene selective regions (Hasson et al., 2003). Its role in early stages of scene processing is further supported by findings that show OPA is encoding the spatial relationship between objects within a scene (Nasr, Devaney, & Tootell, 2013). OPA preferred big objects in comparison to small objects (Konkle & Oliva, 2012) and it responded to any kind of scene, regardless of the number of objects within the scene: on the contrary, it showed a lower response to non-scene stimuli even in the presence of multiple objects (Bettencourt & Xu, 2013). OPA is also thought to have a role in navigation due to its sensitivity to egocentric distance information and was involved in processing of first- person motion information while viewing scenes (Kamps, Lall, & Dilks, 2016). As can be seen from the controversial results presented here, there is no consensus on the exact role of OPA in scene processing.

Motion responsive neurons are shown to be not limited to dorsal visual pathway, but also exist in the ventral visual pathway. A recent study showed that motion perception was dependent on ventral visual cortex (Gilaie-Dotan et al., 2013). Furthermore, motion responses were shown in object processing region LOC (lateral occipital cortex) (Self & Zeki, 2005), in face processing regions (Pitcher, Dilks, Saxe, Triantafyllou, & Kanwisher, 2011) and in monkey parahippocampal cortex (Sato & Nakamura, 2003). Prior studies on scene processing are mostly based on stationary scenes or snapshots of different views. Therefore, little is known how or whether their responses are modulated by visual motion, and to which degree their motion-response depends on scene content.

Although there are some studies on scene processing done using saccadic eye movements (Golomb, Albrecht, Park, & Chun, 2011; Ward et al., 2010) or snapshots of different viewpoints (Epstein et al., 2003; Epstein et al., 2005; Park & Chun, 2009; V. Sulpizio, Committeri, & Galati, 2014; Valentina Sulpizio, Committeri, Lambrey, Berthoz, & Galati, 2013), in real life, we perceive scenes frequently in motion, due to object-, eye- or self-motion. Prior evidence, based on saccadic changes in eye position during viewing of scenes, suggests that PPA and OPA encode scenes in eye-centered reference frame while RSC doesn't show any preference (Ward et al., 2010). Another study found that PPA adapted to views of the same scene during eye movements (Golomb et al., 2011). Yet the effect of continuous retinal motion and objective (real-world) motion in scene responsive regions is still poorly understood and further studies are indeed needed in order to fully understand the different roles of these regions in processing information arising from different types of motion and eye movements.

1.2. Visual motion processing responses in cortical regions

In visual neuroscience, motion processing is one of the most studied topics. Visual motion processing is one of the most important tasks of the visual system in daily life. The visual information that enters retina travels through LGN and thalamus to primary visual cortex. After primary visual cortex, the visual system is classically divided in two streams dorsal 'where' pathway and ventral 'what' pathway (Goodale & Milner, 1992). Traditionally, dorsal pathway is associated with action, object location and motion perception, whereas ventral pathway is associated with perception, object structure and form perception. Motion processing regions in human brain include

MT/V5, MST, VIP, VPS, V3A, V6 as well as CSv (Figure 2). V5/MT and MST are the most studied motion responsive regions. They are lower-tier regions in the hierarchy of motion processing network. Thus, they are highly responsive to most types of motion, while higher-tier motion regions such as V3A, V6 and CSv are more selective to complex motion patterns such as self-induced motion. However, higher-level motion responsive regions are less studied and more research is needed in order to understand how different motion types are represented in the human brain.

Previous studies in single neurons of monkey brain investigated various properties of V5/MT cells. Among these, motion and direction sensitivity are the most studied ones. Direction selective cells are first seen in V1 in motion processing hierarchy (Hubel & Wiesel, 1968) and most cells in V5/MT are direction selective (Dubner & Zeki, 1971). V5/MT neurons' receptive fields are approximately 10 times bigger than those of V1 neurons (Albright & Desimone, 1987) and V5/MT neurons show columnar organization (Albright, Desimone, & Gross, 1984). Lesions in V5/MT would result in akinotopsia, in which the patient is unable to perceive motion while visual perception works normally for other type of visual stimuli (Zeki, 1991). Since majority of studies investigating the response profile of V5/MT are done using abstract stimuli, little is known about its responses during natural visual stimuli such as scenes. Interestingly, V5/MT shows object responses regardless of motion (Kourtzi, Bulthoff, Erb, & Grodd, 2002; Kourtzi & Kanwisher, 2000). However, the effect of scene content in V5/MT responses is not investigated.

Located adjacent to V5/MT, area MST receives most of its inputs directly from V5/MT (Maunsell & Van Essen, 1983a; Ungerleider & Desimone, 1986). MST has very large receptive field size, extending to 60 degrees (Duffy & Wurtz, 1991). MST neurons have a role in executing smooth pursuit eye movements and they presumably receive extraretinal signals (Ilg & Thier, 2003; Newsome, Wurtz, & Komatsu, 1988; Thier & Erickson, 1992) and they are sensitive to optic flow stimuli (Bradley, Maxwell, Andersen, Banks, & Shenoy, 1996; Duffy & Wurtz, 1991). MST contains real motion cells and responds to object motion regardless of retinal motion (Erickson & Thier, 1991). In addition to visual responses, MST also has multisensory integration responses, since it combines vestibular signals and visual self-motion signals (Gu, DeAngelis, & Angelaki, 2007; Komatsu & Wurtz, 1988). Most of the neurons in dorsal subregion of MST (MSTd) prefer complex motion pattern that has planar motion as a component (Duffy & Wurtz, 1991). Whereas, lateral subregion of MST (MSTl) encodes object motion in world-centered coordinates and has visual tracking neurons (Ilg, 2008; Ilg et al., 2004). In addition, human fMRI studies defined homologues of MT/V5 and MST (e.g. see (Huk, Dougherty, & Heeger, 2002; Tootell et al., 1995)).

V3A is the second most motion responsive region in human brain, while in monkey it contains fewer motion responsive cells (Tootell et al., 1997). In monkey, V3A has been shown to include gaze dependent visual neurons and 'real' motion responsive cells (Galletti & Battaglini, 1989; Galletti et al., 1990). In addition, V3A responds to real-world motion while compensating for self-induced retinal changes (Fischer et al., 2012a). V3A has many connections both to ventral and dorsal regions (Tootell et al., 1997). Additionally, a recent study showed that V3A is connected to ventral regions via a white matter tract (Takemura et al., 2015). While previously object related responses were found in V3A, scene content responses were not investigated.

Another higher level motion region in human brain is V6, which is located in parieto-occipital sulcus, is another motion responsive region (Galletti, Fattori, Battaglini, Shipp, & Zeki, 1996; Galletti, Fattori, Gamberini, & Kutz, 1999; Pitzalis et al., 2006; Pitzalis et al., 2010). V6 is retinotopically defined and has receptive field size bigger than that of V3A (Galletti et al., 1999; Pitzalis et al., 2006). Similar to V3A, V6 is also involved in representing real-world motion (Fischer et al., 2012a). Being connected to many occipital and visual areas as well as parietal regions, V6 is thought to be a connecting hub between visual areas and sensorimotor areas (Galletti et al., 2001). V6 prefers egomotion compatible motion (Cardin & Smith, 2010; Galletti, Battaglini, & Fattori, 1991; Pitzalis et al., 2006). Additionally, V6 prefers translational motion, and is thought have a role in extracting information about near and far objects (Pitzalis et al., 2013).Thus, V6 is another candidate for investigating self-motion related responses during scene content.

Among the motion responsive regions identified in human brain, CSv has no clear homologue in monkey. Previous studies have shown that CSv prefers coherent motion (Antal, Baudewig, Paulus, & Dechent, 2008; Wall & Smith, 2008) and 2D motion (Fischer, Bulthoff, Logothetis, & Bartels, 2012b). CSv is thought to have a role in processing self-induced visual motion signals (Fischer, Bulthoff, Logothetis, & Bartels, 2011; Wall & Smith, 2008), vestibular signal responses (Smith, Wall, & Thilo, 2012) and heading direction responses (Furlan, Wann, & Smith, 2014). Not much is known about the responses in CSv.

Nevertheless, there is still not much knowledge about the exact roles of these motion regions on self-motion and object motion processing and how their responses change when encountered by natural scenes, which is the natural stimuli visual cortex handles during visual information processing in the real world.

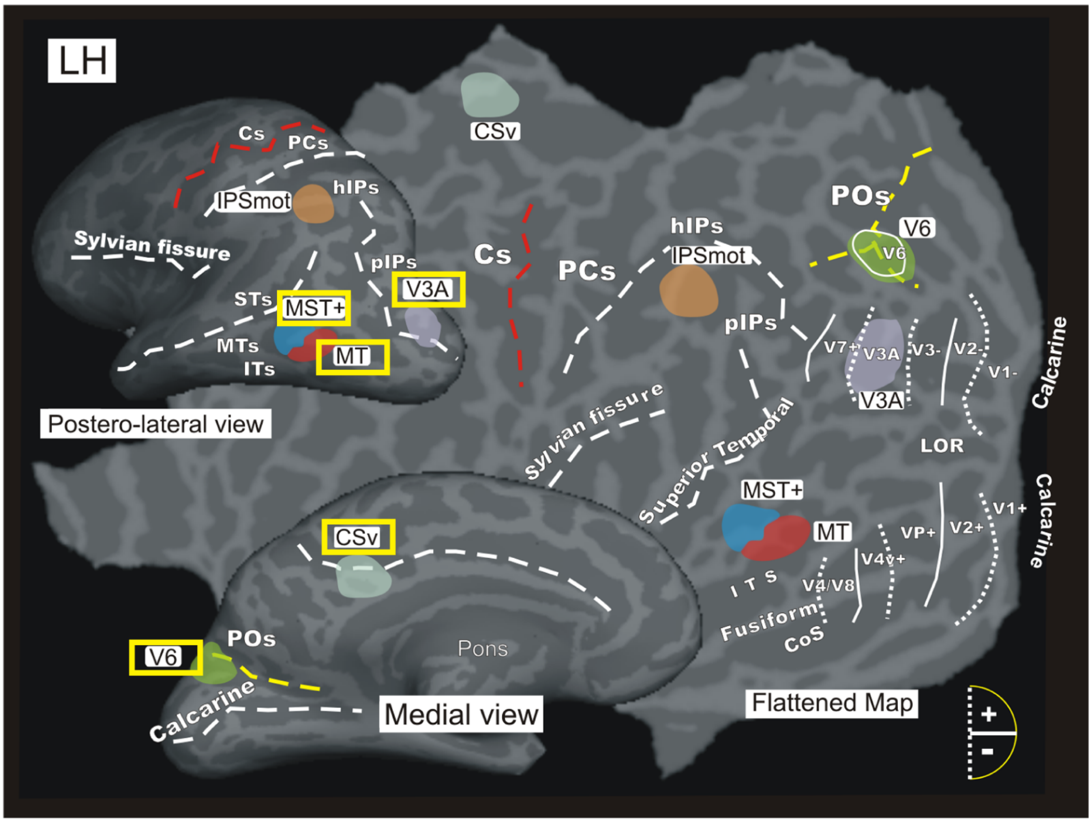


Figure 2. Motion responsive regions in the human brain. Here, motion regions and early visual areas are shown on flattened left hemisphere of a representative human brain. Regions of interest that are investigated in this thesis are shown in a yellow box (MT, MST, V3A, V6 and CSv). The figure is adapted from (Pitzalis et al., 2013).

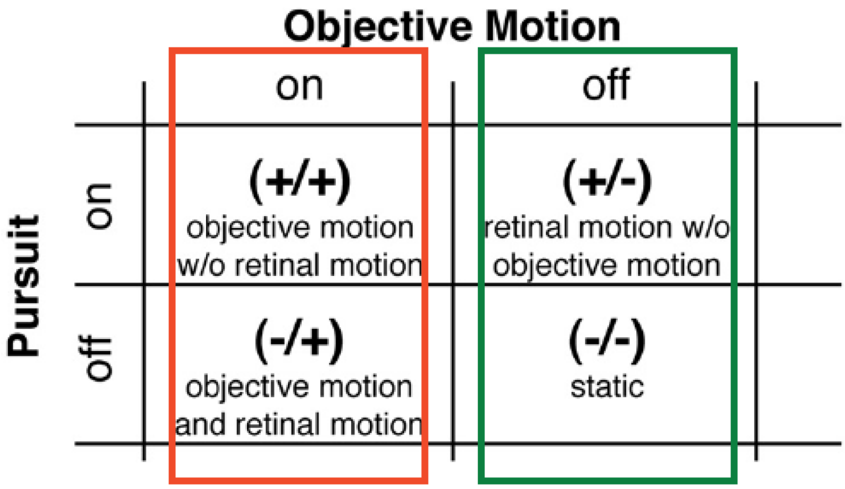
1.3. Objective and retinal motion

Retinal motion simply means that the location of an object changes physically on retina. This can be caused either (a) an object moving in the real world or (b) the observer moving its eyes, head or body. So how does visual system differentiate the two causes from each other? Or when smooth pursuit eye movements are used to track an object, the image of the object on retina is stable while the background image on the retina moves to the opposite direction of the eye movement. Yet, we perceive a stable background and the moving object, not vice versa. How does the visual system keep track of object motion and maintain visual stability at the same time? Many researchers tried to answer these questions. Objective motion is 'real' motion or object motion, representing the real movement of objects in external world.

Visual system uses non-retinal cues such as efference copies and vestibular signals to infer the origin of retinal motion. Multiple brain regions employ efference copies in order to differentiate real motion from self-generated retinal motion. For instance, 'real' motion cells are shown to exist in regions MST, VIP, parietal regions, V3A and V6 (Erickson & Thier, 1991; Galletti et al., 1990; Ilg et al., 2004).

Figure 3 shows an illustration of objective and retinal motion paradigm used in this thesis (Fischer et al., 2012a). Here, using a 2 x 2 factorial design with factors objective motion and smooth pursuit eye movements, objective and retinal motion during smooth pursuit eye movements. Since eye movements were present on both sides of the contrasts (as well as objective versus retinal contrast), any effect of eye movements were cancelled.

A. Objective motion



B. Retinal motion

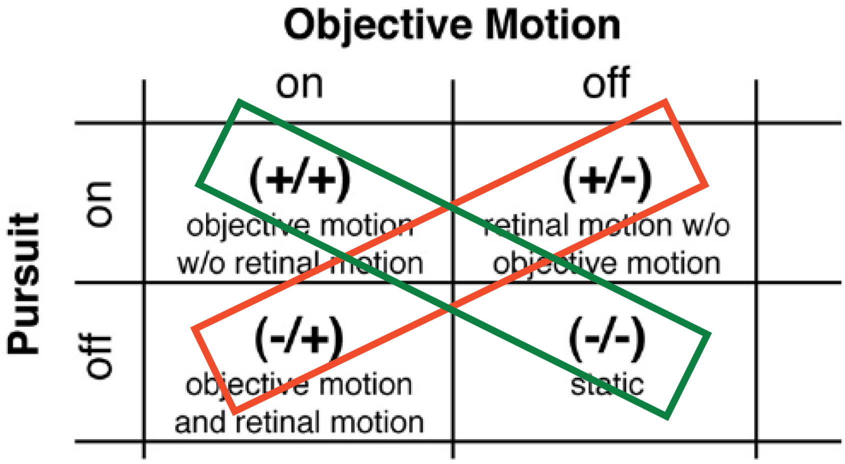


Figure 3. Objective and retinal motion in this thesis are defined using a 2 x 2 factorial design with factors objective motion and pursuit. (A) Objective motion. Here, the calculation of objective motion is visualized. Objective motion contrast is calculated as conditions in the red box versus conditions in the green box. Objective motion contrast is calculated as ((+/+) + (-/+)) vs. ((+/-) + (-/-)). (B) Similar to A, here the calculation of retinal motion is visualized as all conditions in the red box versus all conditions in the green box. Retinal motion is calculated as ((-/+) + (+/-)) vs. ((+/+) + (-/-)). Figure adapted from (Fischer et al., 2012a)

1.4. Eye movements

The fovea region of the eye provides clearer vision while it contains more photoreceptors. Thus, we move our eyes constantly in order to put the object of interest on the central part of retina, our fovea (Krauzlis, 2004). Two types of eye movements are used to achieve this goal; saccades and smooth pursuit eye movement. Saccades are very fast eye movements, usually employed during static visual input while smooth pursuit eye movement is used for tracking moving objects of interest.

Smooth pursuit eye movements are employed to keep a stable vision of the object of interest on fovea. Humans can pursue a target successfully up to speeds of 100 deg/s (Meyer, Lasker, & Robinson, 1985). Many regions take part in initiating and keeping smooth pursuit eye movements. V5/MT processes retinal motion of the target to pursue and MST keeps track of eye position and provides world-centered coordinates of the target (Ilg & Thier, 2003; Thier & Ilg, 2005). Additionally, neurons whose responses are modulated by smooth pursuit eye movements are also found in VIP and LIP regions (Bremmer, Ilg, Thiele, Distler, & Hoffmann, 1997; Colby, Duhamel, & Goldberg, 1993; Schlack, Hoffmann, & Bremmer, 2003). A subregion of frontal eye fields (FEF) is also involved in processing pursuit related signals (Gottlieb, MacAvoy, & Bruce, 1994) as well as cerebellum (Thier & Ilg, 2005).

1.5. Neural basis of speed processing in human brain

Understanding the speed of a moving object is an important role of the motion processing system. Previously, many studies showed that most neurons in V5/MT are tuned by motion speed (Maunsell & Van Essen, 1983c; Perrone & Thiele, 2001), while others found contradictory results, saying that only a quarter of neurons in V5/MT are speed sensitive (Priebe, Cassanello, & Lisberger, 2003). In addition, lesion studies show that V5/MT plays an important role in speed processing (Dursteler & Wurtz, 1988; Newsome, Wurtz, Dürsteler, & Mikami, 1985; Orban, Saunders, & Vandenbussche, 1995; Pasternak & Merigan, 1994; Yamasaki & Wurtz, 1991). The preferred speed of majority of V5/MT neurons is shown to be around 32 deg/s (Maunsell & Van Essen, 1983b; Mikami, Newsome, & Wurtz, 1986; Rodman & Albright, 1987). Although the speed responses of V5/MT neurons in monkey are widely investigated, speed processing in human brain is somewhat less studied.

Another region that is involved in speed processing is V3A. Human neuroimaging studies found that human V5/MT and V3A show speed tuning (Chawla et al., 1999; Chawla, Phillips, Buechel, Edwards, & Friston, 1998; Lingnau, Ashida, Wall, & Smith, 2009). Majority of the V3A neurons in monkey show speed related responses and these responses are present even at high speeds (Galletti et al., 1990). Moreover, using transcranial magnetic stimulation, McKeefry et al showed V3A's involvement in speed perception in human participants (McKeefry, Burton, Vakrou, Barrett, & Morland, 2008). V3A and V6 have been shown to be speed sensitive during head-centered motion (Arnoldussen, Goossens, & van den Berg, 2011, 2015). Nevertheless, the exact roles of these different regions in speed processing are still not well investigated.

Most of the previously mentioned studies on speed processing are done when eyes are fixed and the speed of the moving object on the background is varied. However, in real world, we would do pursuit eye movements to track the moving object. Additionally, the speed tuning properties of higher-tier motion processing regions during smooth pursuit eye movements are mostly not known.

1.6. Spatial reference frames: Eye-centered or world-centered?

During spatial navigation and interaction with objects around, visual system uses different spatial reference frames to represent the spatial layout of the surrounding environment. It is important for visual system to quickly switch between spatial reference frames in order to complete an action or movement.

Both eye-centered and world-centered representations of object locations are found in various brain regions. For instance, world-centered response to object location exists in V6 and VIP (Duhamel, Bremmer, BenHamed, & Graf, 1997; Galletti, Battaglini, & Fattori, 1993). Additionally, MST represents object locations in a world-centered manner (Ilg et al., 2004). World-centered responses are also found in V5/MT (d'Avossa et al., 2007). Further, human V3A and V6 prefer world-centered coding (Fischer et al., 2012a). Spatial reference frames are used during spatial navigation and are present in navigation related regions of the brain. Related to this, certain groups of cells in hippocampus such as place cells are shown to use world-centered representation (Moser, Rowland, & Moser, 2015; O'Keefe & Dostrovsky, 1971). Scene responsive area RSC has viewpoint independent scene responses. While some previous studies showed that another scene region, PPA, represents viewpoint dependent layout of the spatial surroundings and utilizes eye-centered representation of scenes, others found that PPA was involved in making judgements about world-centered object locations in a scene (Committeri et al., 2004; Galati, Pelle, Berthoz, & Committeri, 2010). Similar controversial results were also found for object processing region LOC; while a study found spatiotopic representations of object locations in LOC (McKyton & Zohary, 2007), another study found retinotopic organisation in LOC, as well as other ventral regions (Arcaro et al., 2009). According to one view on this topic, many brain regions use retinotopic representation of locations (Gardner, Merriam, Movshon, & Heeger, 2008; Golomb & Kanwisher, 2012)

Although these findings seem controversial, it is possible that eye-centered and world-centered representations exist together and are employed during different tasks of visual system. One such theory says that spatial attention affects whether eye-centered or world-centered reference frame is used (Crespi et al., 2011). Certainly, more studies are needed to clarify how eye-centered and world-centered reference frames coexist and explain their interchangeable use for a continuous flawless perception during everyday interactions with the world surrounding us.

1.7. Thesis overview

This thesis consists of three parts. The overall aim of this thesis is to advance our understanding of how visual scene and motion responses interact in the human brain during viewing naturalistic, dynamic stimuli.

After characterizing motion responses in scene regions in the first study, in the remaining studies I then consistently used a motion paradigm that helped to distinguish objective motion from retinal motion (Fischer et al., 2012a).

The first part consists of Chapter 2 and Chapter 3 and investigates the responses of scene responsive regions during moving scenes. Part 2, which consists of Chapter 4, investigates the responses of motion regions to scene content during motion. Lastly, Part 3, which includes Chapter 5, investigates speed tuning in motion regions during objective and retinal motion.

Part 1

Chapter 2: Motion responses in scene selective regions

Keeping a stable visual perception is one of the crucial roles of visual system. Although we move our eyes, head and body constantly, our visual system continuously updates our visual perception in order to maintain a flawless experience. Motion processing is a well-studied topic in both human and primate brain. However, most of these studies were interested only in well-known dorsal motion responsive regions.

Scene processing is important to understand our environment and has a role in interacting with this environment, such as during navigation.

In this study, I investigated whether scene regions are modulated by motion and if they are, how much their scene responses are affected by visual motion. In order to investigate this, I did an fMRI study on healthy human participants. We used a 2 x 2 factorial design with factors being visual motion (on/off) and scene (scene/scramble). Scramble images were obtained by phase scrambling of scene images, which resulted in images that do not have the scene content but all the low-level image properties such as luminance, contrast and spatial frequency would be preserved and same across image types. Results of this study showed that PPA and OPA were indeed responsive to motion. Further, PPA also showed an interaction between motion and scene content, meaning that it was more responsive to motion during scenes compared to during scrambled images. RSC was not modulated by motion at all, but it showed similar trend as PPA for the scene-motion interaction. These results showed a differentiation of scene regions based on their responsiveness to visual motion. I conclude that PPA has a special role in integrating visual scenes and visual motion.

Chapter 3: Objective (Real-world) Motion Responses in Scene Responsive Regions

In order to understand the visual scene surrounding us, we constantly make eye movements. Smooth pursuit eye movements are the type of eye movements used especially when tracking an object of interest. Thus, it is very important for the visual system to distinguish the real motion of the object from retinal changes resulting from our own eye movements. Although previous studies investigated whether scene-processing regions encode visual information in eye-centered or world-centered reference frames, they used saccadic changes and static snapshots of scenes. How the scene responsive regions utilize pursuit eye movements during the visual analysis of scenes is still not well known.

In the second study, I investigated how different types of motion are processed in scene responsive areas and which reference frames are used by different scene responsive regions during scene motion responses. In this fMRI study in healthy human participants, I used a 2 x 2 x 2 factorial design paradigm with factors objective motion (on/off), pursuit (on/off) and scene content (natural scenes versus Fourier scrambled version of these scenes). PPA was responsive to objective motion while cancelling out retinal motion responses completely, so that it can differentiate world-centered (objective) motion from eye-centered (retinal) motion and thus use a world-centered reference frame. PPA was the only region significantly preferring objective motion, even during scrambled images, and thus can be differentiated from OPA and RSC. Interestingly, RSC did not show any motion responses. OPA, on the hand, was responsive to both objective and retinal motion but did not differentiate between them; which fits to the view on the role of OPA that it is a lower-tier scene-processing region. Put together, these results indicate a distinct role for PPA in differentiating real-world motion while compensating for pursuit related eye movement signals, in order to create stable perception. Moreover, all three scene-processing regions could be distinguished from each other based on their motion response profiles.

Part 2

Chapter 4: Motion regions are modulated by scene content

The visual system is optimized for processing naturalistic stimuli. However, most research in visual system has been done using abstract stimuli, such as gratings or moving dots (Born & Bradley, 2005; Boussaoud, Ungerleider, & Desimone, 1990; Erickson & Thier, 1991; Galletti & Fattori, 2003; Goossens, Dukelow, Menon, Vilis, & van den Berg, 2006; Gu et al., 2007; Huk et al., 2002; Maciokas & Britten, 2010; Smith, Wall, Williams, & Singh, 2006), except for one that investigated motion responses during movie viewing (Bartels et al., 2008). Additionally, previous studies found responses related to object content in motion regions (Kourtzi et al., 2002).

In this study, I investigated how the motion responses in motion processing regions V5+/MT+, V3A, V6 and CSV are modulated by scene content. I used a 2 x 2 x 2 factorial design with the factors being objective motion (on/ off), pursuit (on/off) (Fischer et al., 2012a) as well as a factor for scene content (scene/ scramble). I found that all motion regions showed scene responses. However, only in V3A, the scene response persisted after removing motion. V3A could be differentiated from other motion regions by its responsiveness for still scenes. V5+/MT+ showed a motion and scene content interaction both during objective and retinal motion, meaning higher motion responses during scene than scrambled images, whereas V3A and V6 were only responsive to scene and retinal motion interaction and CSv was not responsive to any scene-motion interaction. These results show that motion regions are indeed modulated by scene content, thus demonstrate the importance of using naturalistic scene stimuli while studying the responses of visual system. In addition, V3A has a possible role in scene representation in the brain.

Part 3

Chapter 5: Speed tuning to real-world- and retinal motion in cortical motion regions

Speed is one of two factors defining motion, the other being direction. Thus, perceiving and processing speed is an important task during motion processing. Many studies examined speed-tuning properties of neurons in monkey MT/V5. It has been shown that many neurons in V5/MT are speed selective, and tuned for an optimal speed of about 30 deg/s (Maunsell & Van Essen, 1983c; Newsome, Mikami, & Wurtz, 1986). Other electrophysiology studies investigated speed tuning of neurons in other motion responsive regions, such as MST, V3 and V6. To this date, speed-tuning profiling for CSv has not been done.

There are not many studies investigating speed tuning in human brain. The only ones that are investigating speed responses in human brain used a combined retinal and objective motion; they investigated speed tuning during fixation while the speed of background changed compared to still background (Chawla et al., 1999; Chawla et al., 1998).

In this study, I investigated the speed tuning of motion regions V5/MT, MST, V3A, V6 and CSv during objective and retinal motion separately for speed levels of 1, 2, 4, 8, 16 and 24 deg/s respectively. I used Fourier scrambles of natural scene images in order to provide naturalistic stimuli. I found that all regions were modulated by the speed of both objective and retinal motion. Only V3A showed a differentiation in its modulation by objective compared to by retinal motion speeds. All regions showed a higher mean response to objective motion compared to retinal motion across all speeds. In accord with previous studies, these results suggest that V3A is mainly encoding objective motion and this response profile is consistent across different speeds.

1.8. General discussion

The studies described in this thesis investigate how dynamic scenes are processed in human brain during retinal and objective motion. Despite constant feed of visual motion cues due to our own movement or objects moving in our visual field, the visual system keeps a stable perception of the world. Using natural scenes and phase-scrambled images of scenes matching in luminance, contrast and frequency spectra, the studies in this thesis were able to distinguish content effect of scenes.

Part 1

There are various opinions on the exact roles of scene processing regions PPA, RSC and OPA and their similarities or differentiation regarding these roles.

Previous studies showed motion responses in ventral stream (Gilaie-Dotan et al., 2013; Pitcher et al., 2011; Self & Zeki, 2005). The results in both chapter 2 & 3 add PPA to previous ventral stream regions with motion responses. Moreover, while PPA and OPA were motion responsive, RSC was not sensitive to visual motion. Motion response profiles of these regions can be used to differentiate them from each other. Motion responses in PPA (and OPA) and lack of motion responses in RSC is also compatible with PPA's viewpoint dependency and RSC's viewpoint independency (Epstein et al., 2003; Epstein & Higgins, 2007; Epstein et al., 2005; Epstein, Parker, et al., 2007; MacEvoy & Epstein, 2007; Park & Chun, 2009). The lack of motion responses in RSC is also in line with the view on RSC that it is 'higher-level' scene processing region and it is more involved in functions such as navigation.

OPA having the highest motion responses while lacking responses to scene-motion interaction fits the previous view suggesting that it is a "lower" region in scene processing.

Part 2

In daily life, our brain receives visual input from our natural surroundings. Visual system is optimized for processing of natural stimuli statistics, which are features such as spatial frequency, luminance and contrast of natural scenes (Parraga et al., 2000). Previous studies show that motion regions' responses are modulated by object content (Kourtzi et al., 2002). We found that motion regions have scene content responses. More importantly, V3A can be differentiated from other motion regions by its responsiveness to still scenes. Scene preference in V3A fits well with the aforementioned functional properties of the region, such as taking part in contour processing, or object and shape processing.

Given the fact that V5/MT and V3A were shown to have viewpoint specific object responses (Konen & Kastner, 2008), the scene and motion interaction seen in V5/MT+, V3A and V6 could be driven by viewpoint changes in the given scene.

Part 3

Previously, speed tuning of motion regions during objective or retinal speed has not been systematically studied. In this study, using stimuli with natural image statistics, we found that all regions we investigated (V3A, V6, V5/MT, MST and CSv) showed speed tuning responses for both objective and retinal motion. Our results seem to contradict previous studies that found an inverted 'U' shape type of response profile for speed responses in V5/MT and V3A (Chawla et al., 1999; Chawla et al., 1998). However, these studies were more than 15 years old and made observations only with 3 participants. Hence, their results are quite noisy and with more participants, hence, it is possible that the present study achieved a better understanding of speed tuning in these regions.

The monotonic increase in both objective and retinal motion responses during speed tuning is plausible when previous single cell studies are considered. Optimal speed for the majority of V5/MT neurons is reported around 30 deg/s in various studies (Maunsell & Van Essen, 1983c; Newsome et al., 1986). Similarly, V3A neurons are still responsive even for high speeds (higher than 50 deg/s) (Galletti et al., 1990). Since the highest speed used in this study, which was only 24 deg/s, was smaller than the optimal values mentioned in the literature, it is possible that the results here are showing only the lower end of the range.

Human motion regions V3A (and V6) prefers object motion while compensating for retinal motion. This study explored whether their objective motion preference is specific to a particular motion speed or was it independent from motion speed. Here, V3A showed a difference in the speed tuning for objective and retinal motion. Its objective motion responses get higher with the speed whereas the change in the retinal motion responses is not as big.

1.9. Outlook and Conclusion

The findings of the studies in this thesis extend our understanding of self-motion processing in the brain and its communication with scene processing network, which is especially important for tasks such as spatial navigation. The results of this PhD thesis show the importance of using naturalistic stimuli while investigating functional properties of visual system. In order to tackle the underpinnings of visual functions, future studies investigating the effect of scene content on more complex motion patterns, such as 3D motion responses or optic flow responses are needed. Additionally, future studies could further explore the exact role of V3A in spatial processing, especially in processing of visual scenes.

In order to maintain a stable vision, many regions in the brain work together. The results of the present studies point out that world-centered motion responses are present both in motion and scene processing regions of human brain. These results agree with previous literature about V3A that it encodes primarily objective motion. Additionally, present results add to the existing literature that V3A even shows different speed tuning for objective and retinal motion. Similarly, PPA is also involved in world-centered encoding of visual information, even in the absence of visual scenes. Interestingly, V3A is sensitive to visual scenes, even during absence of motion. Although located in different branches of traditionally defined dorsal ‘where’ and ventral ‘what’ streams, the similarity of the responses of scene region PPA and motion region V3A is striking. Perhaps the separation between ‘where’ and ‘what’ pathways is not as strong as thought initially and these two regions exchange information in order to provide a stable perception.

1.10. Declaration of contributions

The PhD thesis comprises of four manuscripts that are published or are in preparation to be submitted for publication. The detailed explanation of contribution of authors is given for each manuscript:

1. Korkmaz Hacialihafiz, D. and Bartels, A. (2015). Motion responses in scene-selective regions. *Neuroimage* 118, 438–444. D.K.H. programmed and prepared the experiment, collected the data, and analyzed the data. D.K.H. and A.B. planned the experiment and data analysis and wrote the manuscript.

2. Korkmaz Hacialihafiz, D. and Bartels, A. (2016) Objective (Real-world) motion responses in scene responsive regions (manuscript in preparation): D.K.H. programmed and prepared the experiment, collected the data, and analyzed the data. D.K.H. and A.B. planned the experiment and data analysis and wrote the manuscript.

3. Korkmaz Hacialihafiz, D. and Bartels, A. (2016) Motion regions are modulated by scene content (manuscript in preparation): D.K.H. programmed and prepared the experiment, collected the data, and analyzed the data. D.K.H. and A.B. planned the experiment and data analysis and wrote the manuscript.

4. Korkmaz Hacialihafiz, D. and Bartels, A. (2016) Speed tuning to real-world- and retinal motion in cortical motion regions (manuscript in preparation): D.K.H. programmed and prepared the experiment, collected the data, and analyzed the data. D.K.H. and A.B. planned the experiment and data analysis and wrote the manuscript.

1.11. References

Albright, T. D., & Desimone, R. (1987). Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. *Experimental Brain Research, 65*, 582-592.

Albright, T. D., Desimone, R., & Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology, 51*(1), 16-31.

Antal, A., Baudewig, J., Paulus, W., & Dechent, P. (2008). The posterior cingulate cortex and planum temporale/parietal operculum are activated by coherent visual motion. *Visual neuroscience, 25*, 17-26. doi: 10.1017/S0952523808080024

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 29*, 10638-10652. doi: 10.1523/JNEUROSCI.2807-09.2009

Arnoldussen, D. M., Goossens, J., & van den Berg, A. V. (2011). Adjacent visual representations of self-motion in different reference frames. *Proceedings of the National Academy of Sciences of the United States of America, 108*(28), 11668-11673. doi: 10.1073/pnas.1102984108

Arnoldussen, D. M., Goossens, J., & van den Berg, A. V. (2015). Dissociation of retinal and headcentric disparity signals in dorsal human cortex. *Frontiers in Systems Neuroscience, 9*, 16. doi: 10.3389/fnsys.2015.00016

Auger, S. D., Mullally, S. L., & Maguire, E. A. (2012). Retrosplenial Cortex Codes for Permanent Landmarks. *PLoS ONE, 7*(8), e43620. doi: http://dx.doi.org/10.1371/journal.pone.0043620

Baldassano, C., Beck, D. M., & Fei-Fei, Li. (2013). Differential connectivity within the Parahippocampal Place Area. *NeuroImage, 75*, 228-237. doi: 10.1016/j.neuroimage.2013.02.073

Bartels, A., Zeki, S., & Logothetis, N. K. (2008). Natural Vision Reveals Regional Specialization to Local Motion and to Contrast-Invariant, Global Flow in the Human Brain. *Cerebral Cortex, 18*(3), 705-717. doi: 10.1093/cercor/bhm107

Bettencourt, K. C., & Xu, Y. (2013). The role of transverse occipital sulcus in scene perception and its relationship to object individuation in inferior intraparietal sulcus. *J Cogn Neurosci, 25*(10), 1711-1722. doi: 10.1162/jocn_a_00422

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT *Annual Review of Neuroscience* (Vol. 28, pp. 157-189). Palo Alto: Annual Reviews.

Boussaoud, D., Ungerleider, L. G., & Desimone, R. (1990). Pathways for Motion Analysis - Cortical Connections of the Medial Superior Temporal and Fundus of the Superior Temporal Visual Areas in the Macaque. *Journal of Comparative Neurology, 296*(3), 462-495.

Bradley, D. C., Maxwell, M., Andersen, R. A., Banks, M. S., & Shenoy, K. V. (1996). Mechanisms of heading perception in primate visual cortex. *Science, 273*(5281), 1544-1547.

Bremmer, F., Ilg, U. J., Thiele, A., Distler, C., & Hoffmann, K. P. (1997). Eye position effects in monkey cortex. I. Visual and pursuit-related activity in extrastriate areas MT and MST. *Journal of neurophysiology, 77*(2), 944-961.

Cardin, V., & Smith, A. T. (2010). Sensitivity of human visual and vestibular cortical regions to egomotion-compatible visual stimulation. *Cerebral cortex, 20*(8), 1964-1973. doi: 10.1093/cercor/bhp268

Chawla, D., Buechel, C., Edwards, R., Howseman, A., Josephs, O., Ashburner, J., & Friston, K. J. (1999). Speed-dependent responses in V5: A replication study. *Neuroimage, 9*(5), 508-515.

Chawla, D., Phillips, J., Buechel, C., Edwards, R., & Friston, K. J. (1998). Speed-dependent motion-sensitive responses in V5: an fMRI study. *Neuroimage, 7*(2), 86-96.

Colby, C. L., Duhamel, J. R., & Goldberg, M. E. (1993). Ventral intraparietal area of the macaque: anatomic location and visual response properties. *Journal of Neurophysiology, 69*(3), 902-914.

Committeri, G., Galati, G., Paradis, A. L., Pizzamiglio, L., Berthoz, A., & LeBihan, D. (2004). Reference frames for spatial cognition: Different brain areas are involved in viewer-, object-, and landmark-centered judgments about object location. *Journal of Cognitive Neuroscience, 16*(9), 1517-1535. doi: 10.1162/0898929042568550

Crespi, S., Biagi, L., d'Avossa, G., Burr, D. C., Tosetti, M., & Morrone, M. C. (2011). Spatiotopic Coding of BOLD Signal in Human Visual Cortex Depends on Spatial Attention. *PloS one, 6*(7), e21661. doi: 10.1371/journal.pone.0021661

d'Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nature neuroscience, 10*(2), 249-255. doi: 10.1038/nn1824

Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 31*, 11305-11312. doi: http://dx.doi.org/10.1523/JNEUROSCI.1935-11.2011

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 33*, 1331-1336a. doi: http://dx.doi.org/10.1523/JNEUROSCI.4081-12.2013

Dubner, R., & Zeki, S. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research, 35*, 528-532.

Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology, 65*(6), 1329-1345.

Duhamel, J. R., Bremmer, F., BenHamed, S., & Graf, W. (1997). Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature, 389*(6653), 845-848.

Dursteler, M. R., & Wurtz, R. H. (1988). Pursuit and optokinetic deficits following chemical lesions of cortical areas MT and MST. *Journal of Neurophysiology, 60*(3), 940-965.

Epstein, R. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences, 12*(10), 388-396. doi: 10.1016/j.tics.2008.07.004

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron, 37*(5), 865-876. doi: http://dx.doi.org/10.1016/s0896-6273(03)00117-x

Epstein, R., Harris, A, Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron, 23*(1), 115-125. doi: http://dx.doi.org/10.1016/S0896-6273(00)80758-8

Epstein, R., & Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb Cortex, 17*(7), 1680-1693. doi: http://dx.doi.org/10.1093/cercor/bhl079

Epstein, R., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *Journal of neurophysiology, 97*, 3670-3683. doi: http://dx.doi.org/10.1152/jn.00003.2007

Epstein, R., Higgins, J. S., & Thompson-Schill, S. L. (2005). Learning Places from Views: Variation in Scene Processing as a Function of Experience and Navigational Ability. *Journal of Cognitive Neuroscience, 17*(1), 73-83. doi: http://dx.doi.org/10.1162/0898929052879987

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598-601. doi: http://dx.doi.org/10.1038/33402

Epstein, R., Parker, W., & Feiler, A. (2007). Where Am I Now? Distinct Roles for Parahippocampal and Retrosplenial Cortices in Place Recognition. *The Journal of Neuroscience, 27*(23), 6141-6149. doi: http://dx.doi.org/10.1523/jneurosci.0799-07.2007

Epstein, R., & Vass, L. K. (2014). Neural systems for landmark-based wayfinding in humans. *Philosophical Transactions of ...*.

Erickson, R. G., & Thier, P. (1991). A Neuronal Correlate of Spatial Stability during Periods of Self-Induced Visual-Motion. *Experimental Brain Research, 86*(3), 608-616.

Fischer, E., Bulthoff, H. H., Logothetis, N. K., & Bartels, A. (2011). Visual Motion Responses in the Posterior Cingulate Sulcus: A Comparison to V5/MT and MST. *Cerebral cortex*. doi: 10.1093/cercor/bhr154

Fischer, E., Bulthoff, H. H., Logothetis, N. K., & Bartels, A. (2012a). Human areas V3A and V6 compensate for self-induced planar visual motion. *Neuron, 73*(6), 1228-1240. doi: 10.1016/j.neuron.2012.01.022

Fischer, E., Bulthoff, H. H., Logothetis, N. K., & Bartels, A. (2012b). Visual motion responses in the posterior cingulate sulcus: a comparison to V5/MT and MST. *Cerebral cortex (New York, N.Y. : 1991), 22*, 865-876. doi: 10.1093/cercor/bhr154

Furlan, M., Wann, J. P., & Smith, A. T. (2014). A Representation of Changing Heading Direction in Human Cortical Areas pVIP and CSv. *Cerebral Cortex, 24*(11), 2848-2858. doi: 10.1093/cercor/bht132

Galati, G., Pelle, G., Berthoz, A., & Committeri, G. (2010). Multiple reference frames used by the human brain for spatial perception and memory. *Experimental brain research, 206*, 109-120. doi: 10.1007/s00221-010-2168-8

Galletti, C., & Battaglini, P. P. (1989). Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *J Neurosci, 9*(4), 1112-1125.

Galletti, C., Battaglini, P. P., & Fattori, P. (1990). 'Real-motion' cells in area V3A of macaque visual cortex. *Exp Brain Res, 82*(1), 67-76.

Galletti, C., Battaglini, P. P., & Fattori, P. (1991). Functional Properties of Neurons in the Anterior Bank of the Parieto-occipital Sulcus of the Macaque Monkey. *Eur J Neurosci, 3*(5), 452-461.

Galletti, C., Battaglini, P. P., & Fattori, P. (1993). Parietal neurons encoding spatial locations in craniotopic coordinates. *Exp Brain Res, 96*(2), 221-229.

Galletti, C., & Fattori, P. (2003). Neuronal mechanisms for detection of motion in the field of view. *Neuropsychologia, 41*(13), 1717-1727.

Galletti, C., Fattori, P., Battaglini, P. P., Shipp, S., & Zeki, S. (1996). Functional demarcation of a border between areas V6 and V6A in the superior parietal gyrus of the macaque monkey. *Eur J Neurosci, 8*(1), 30-52.

Galletti, C., Fattori, P., Gamberini, M., & Kutz, D. F. (1999). The cortical visual area V6: brain location and visual topography. *Eur J Neurosci, 11*(11), 3922-3936.

Galletti, C., Gamberini, M., Kutz, D. F., Fattori, P., Luppino, G., & Matelli, M. (2001). The cortical connections of area V6: an occipito-parietal network processing visual information. *Eur J Neurosci, 13*(8), 1572-1588.

Ganaden, R. E., Mullin, C. R., & Steeves, J. K. (2013). Transcranial Magnetic Stimulation to the Transverse Occipital Sulcus Affects Scene but Not Object Processing. *J Cogn Neurosci, 25*(6), 961-968. doi: http://dx.doi.org/10.1162/jocn_a_00372

Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *The Journal of Neuroscience, 28*(15), 3988-3999.

Gilaie-Dotan, S., Saygin, A. P., Lorenzi, L. J., Egan, R., Rees, G., & Behrmann, M. (2013). The role of human ventral visual cortex in motion perception. *Brain, 136*(Pt 9), 2784-2798. doi: http://dx.doi.org/10.1093/brain/awt214

Golomb, J. D., Albrecht, A. R., Park, S., & Chun, M. M. (2011). Eye Movements Help Link Different Views in Scene-Selective Cortex. *Cerebral Cortex, 21*(9), 2094-2102. doi: http://dx.doi.org/10.1093/cercor/bhq292

Golomb, J. D., & Kanwisher, N. (2012). Higher Level Visual Cortex Represents Retinotopic, Not Spatiotopic, Object Location. *Cerebral Cortex, 22*(12), 2794-2810. doi: 10.1093/cercor/bhr357

Goodale, M.A., & Milner, A.D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15*(1), 20-25.

Goossens, J., Dukelow, S. P., Menon, R. S., Vilis, T., & van den Berg, A. V. (2006). Representation of Head-Centric Flow in the Human Motion Complex. *The Journal of Neuroscience, 26*(21), 5616-5627. doi: 10.1523/jneurosci.0730-06.2006

Gottlieb, J. P., MacAvoy, M. G., & Bruce, C. J. (1994). Neural responses related to smooth-pursuit eye movements and their correspondence with electrically elicited smooth eye movements in the primate frontal eye field. *Journal of Neurophysiology, 72*(4), 1634-1653.

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology, 13*(2), 159-166. doi: http://dx.doi.org/10.1016/s0959-4388(03)00040-0

Gu, Y., DeAngelis, G. C., & Angelaki, D. E. (2007). A functional link between area MSTd and heading perception based on vestibular signals. *Nature Neuroscience, 10*(8), 1038-1047.

Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-Scale Mirror-Symmetry Organization of Human Occipito-Temporal Object Areas. *Neuron, 37*(6), 1027-1041. doi: http://dx.doi.org/10.1016/S0896-6273(03)00144-2

Henderson, J. M., Larson, C. L., & Zhu, D. C. (2008). Full scenes produce more activation than close-up scenes and scene-diagnostic objects in parahippocampal and retrosplenial cortex: an fMRI study. *Brain Cogn, 66*(1), 40-49. doi: 10.1016/j.bandc.2007.05.001

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J Physiol, 195*(1), 215-243.

Huk, A. C., Dougherty, R. F., & Heeger, D. J. (2002). Retinotopy and Functional Subdivision of Human Areas MT and MST. *The Journal of Neuroscience, 22*(16), 7195-7205.

Ilg, U. J. (2008). The role of areas MT and MST in coding of visual motion underlying the execution of smooth pursuit. *Vision research, 48*, 2062-2069. doi: 10.1016/j.visres.2008.04.015

Ilg, U. J., Schumann, S., & Thier, P. (2004). Posterior parietal cortex neurons encode target motion in world-centered coordinates. *Neuron, 43*(1), 145-151. doi: 10.1016/j.neuron.2004.06.006

Ilg, U. J., & Thier, P. (2003). Visual tracking neurons in primate area MST are activated by smooth-pursuit eye movements of an "imaginary" target. *J Neurophysiol, 90*(3), 1489-1502.

Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N., & Dilks, D. D. (2016). The occipital place area represents the local elements of scenes. *NeuroImage, 132*, 417-424.

Kamps, F. S., Lall, V., & Dilks, D. D. (2016). The occipital place area represents first-person perspective motion information through scenes. *Cortex, 83*, 17-26.

Kayser, C., Kording, K. P., & Konig, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Curr Opin Neurobiol, 14*(4), 468-473.

Komatsu, H., & Wurtz, R. H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *Journal of Neurophysiology, 60*(2), 580-603.

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature neuroscience, 11*, 224-231. doi: 10.1038/nn2036

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron, 74*, 1114-1124. doi: http://dx.doi.org/10.1016/j.neuron.2012.04.036

Korkmaz Hacialihafiz, D., & Bartels, A. (2015). Motion responses in scene-selective regions. *NeuroImage, 118*, 438-444. doi: http://dx.doi.org/10.1016/j.neuroimage.2015.06.031

Kourtzi, Z., Bulthoff, H. H., Erb, M., & Grodd, W. (2002). Object-selective responses in the human motion area MT/MST. *Nat Neurosci, 5*(1), 17-18. doi: http://www.nature.com/neuro/journal/v5/n1/suppinfo/nn780_S1.html

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience, 12*(1), 48-55.

Krauzlis, R. J. (2004). Recasting the smooth pursuit eye movement system. *Journal of neurophysiology, 91*(2), 591-603. doi: 10.1152/jn.00801.2003

Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-world scene representations in high-level visual cortex: it's the spaces more than the places. *J Neurosci, 31*(20), 7322-7333. doi: 10.1523/jneurosci.4588-10.2011

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nat Neurosci, 4*(5), 533-539.

Levy, I., Hasson, U., Harel, M., & Malach, R. (2004). Functional analysis of the periphery effect in human building related areas. *Hum Brain Mapp, 22*(1), 15-26.

Lingnau, A., Ashida, H., Wall, M. B., & Smith, A. T. (2009). Speed encoding in human visual cortex revealed by fMRI adaptation. *Journal of Vision, 9*(13), 3-3. doi: 10.1167/9.13.3

MacEvoy, S., & Epstein, R. (2007). Position selectivity in scene- and object-responsive occipitotemporal regions. *Journal of Neurophysiology, 98*, 2089-2098. doi: http://dx.doi.org/10.1152/jn.00438.2007

Maciokas, J. B., & Britten, K. H. (2010). Extrastriate Area MST and Parietal Area VIP Similarly Represent Forward Headings. *Journal of Neurophysiology, 104*(1), 239-247. doi: 10.1152/jn.01083.2009

Maguire, E. A. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavian Journal of Psychology, 42*(3), 225-238. doi: http://dx.doi.org/10.1111/1467-9450.00233

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. (2014). Anchoring the neural compass: coding of local spatial reference frames in human medial parietal lobe. *Nature neuroscience, 17*(11), 1598-1606.

Maunsell, J. H., & Van Essen, D. C. (1983a). The connections of the middle temporal area and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience, 3*(12), 2563-2586.

Maunsell, J. H., & Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology, 49*, 1127-1147.

Maunsell, J. H., & Van Essen, D. C. (1983c). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J Neurophysiol, 49*(5), 1148-1167.

McKeefry, D. J., Burton, M. P., Vakrou, C., Barrett, B. T., & Morland, A. B. (2008). Induced deficits in speed perception by transcranial magnetic stimulation of human cortical areas V5/MT+ and V3A. *J Neurosci, 28*(27), 6848-6857. doi: 28/27/6848 [pii]

10.1523/JNEUROSCI.1287-08.2008

McKyton, A., & Zohary, E. (2007). Beyond retinotopic mapping: the spatial representation of objects in the human lateral occipital complex. *Cerebral Cortex, 17*(5), 1164-1172.

Mendez, M. F., & Cherrier, M. M. (2003). Agnosia for scenes in topographagnosia. *Neuropsychologia, 41*, 1387-1395. doi: 10.1016/S0028-3932(03)00041-1

Meyer, C. H, Lasker, A. G., & Robinson, D. A. (1985). The upper limit of human smooth pursuit velocity. *Vision research, 25*(4), 561-563.

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex I. Mechanisms of speed and directional selectivity in extrastriate area MT. *Journal of Neurophysiology, 55*, 1308-1327.

Moser, M.-B., Rowland, D. C., & Moser, E. I. (2015). Place cells, grid cells, and memory. *Cold Spring Harbor perspectives in biology, 7*(2), a021808.

Mullin, C. R., & Steeves, J. K. E. (2013). Consecutive TMS-fMRI Reveals an Inverse Relationship in BOLD Signal between Object and Scene Processing. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 33*, 19243-19249. doi: http://dx.doi.org/10.1523/JNEUROSCI.2537-13.2013

Nasr, S., Devaney, K. J., & Tootell, R. B. H. (2013). Spatial encoding and underlying circuitry in scene-selective cortex. *NeuroImage, 83*, 892-900. doi: http://dx.doi.org/10.1016/j.neuroimage.2013.07.030

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. H. (2011). Scene-selective cortical regions in human and nonhuman primates. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 31*, 13771-13785. doi: 10.1523/JNEUROSCI.2792-11.2011

Nasr, S., & Tootell, R. B. H. (2012). A cardinal orientation bias in scene-selective visual cortex. *The Journal of neuroscience : the official journal of the Society for*

Neuroscience, 32, 14921-14926. doi: http://dx.doi.org/10.1523/JNEUROSCI.2036-12.2012

Newsome, W. T., Mikami, A., & Wurtz, R. H. (1986). Motion Selectivity In Macaque Visual-Cortex .3. Psychophysics and Physiology Of Apparent Motion. *Journal of Neurophysiology, 55*(6), 1340-1351.

Newsome, W. T., Wurtz, R. H., Dürsteler, M. R. , & Mikami, A. (1985). Deficits in visual motion processing following ibotenic acid lesions of the middle temporal visual area of the macaque monkey. *Journal of Neuroscience, 5*(3), 825-840.

Newsome, W. T., Wurtz, R. H., & Komatsu, H. . (1988). Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal inputs. *Journal of Neurophysiology, 60*(2), 604-620.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain research, 34*(1), 171-175.

Orban, G. A., Saunders, R. C., & Vandenbussche, E. (1995). Lesions Of the Superior Temporal Cortical Motion Areas Impair Speed Discrimination In the Macaque Monkey. *European Journal Of Neuroscience, 7*(11), 2261-2276.

Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage, 47*(4), 1747-1756. doi: http://dx.doi.org/10.1016/j.neuroimage.2009.04.058

Park, S., Chun, M. M., & Johnson, M. K. (2010). Refreshing and integrating visual scenes in scene-selective cortex. *Journal of cognitive neuroscience, 22*, 2813-2822. doi: http://dx.doi.org/10.1162/jocn.2009.21406

Park, S., Intraub, H., Yi, D.-J., Widders, D., & Chun, M. M. (2007). Beyond the edges of a view: boundary extension in human scene-selective visual cortex. *Neuron, 54*, 335-342. doi: 10.1016/j.neuron.2007.04.006

Parraga, C. A., Troscianko, T., & Tolhurst, D. J. (2000). The human visual system is optimised for processing the spatial information in natural visual images. *Curr Biol, 10*(1), 35-38.

Pasternak, T., & Merigan, W. H. (1994). Motion perception following lesions of the superior tempora lsulcus in the monkey. *Cerebral Cortex, 4*, 247-259.

Perrone, J. A., & Thiele, A. . (2001). Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nat Neurosci, 4*(5), 526-532.

Persichetti, A. S., & Dilks, D. D. (2016). Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex, 77*, 155-163.

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage, 56*(4), 2356-2363. doi: http://dx.doi.org/10.1016/j.neuroimage.2011.03.067

Pitzalis, S., Galletti, C., Huang, R. S., Patria, F., Committeri, G., Galati, G., . . . Sereno, M. I. (2006). Wide-field retinotopy defines human cortical visual area v6. *J Neurosci, 26*(30), 7962-7973.

Pitzalis, S., Sdoia, S., Bultrini, A., Committeri, G., Di Russo, F., Fattori, P., . . . Galati, G. (2013). Selectivity to translational egomotion in human brain motion areas. *PloS one, 8*, e60241. doi: 10.1371/journal.pone.0060241

Pitzalis, S., Sereno, M. I., Committeri, G., Fattori, P., Galati, G., Patria, F., & Galletti, C. (2010). Human v6: the medial motion area. *Cerebral cortex (New York, N.Y. : 1991), 20*, 411-424. doi: 10.1093/cercor/bhp112

Priebe, N. J., Cassanello, C. R., & Lisberger, S. G. (2003). The Neural Representation of Speed in Macaque Area MT/V5. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 23*(13), 5650-5661.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. H. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS biology, 9*, e1000608. doi: 10.1371/journal.pbio.1000608

Rodman, H. R., & Albright, T. D. (1987). Coding of visual stimulus velocity in area MT of the macaque. *Vision Research, 27*(12), 2035-2048.

Sato, N., & Nakamura, K. (2003). Visual response properties of neurons in the parahippocampal cortex of monkeys. *J Neurophysiol, 90*(2), 876-886. doi: http://dx.doi.org/10.1152/jn.01089.2002

Schlack, A., Hoffmann, K. P., & Bremmer, F. (2003). Selectivity of macaque ventral intraparietal area (area VIP) for smooth pursuit eye movements. *J Physiol, 551*(Pt 2), 551-561.

Self, M. W., & Zeki, S. (2005). The integration of colour and motion by the human visual brain. *Cerebral cortex, 15*, 1270-1279. doi: http://dx.doi.org/10.1093/cercor/bhi010

Sherrill, K. R., Chrastil, E. R., Ross, R. S., Erdem, U. M., Hasselmo, M. E., & Stern, C. E. . (2015). Functional connections between optic flow areas and navigationally responsive brain regions during goal-directed navigation. *Neuroimage, 118*, 386-396.

Smith, A. T., Wall, M. B., & Thilo, K. V. (2012). Vestibular inputs to human motion-sensitive visual cortex Journal:. *Cerebral Cortex*.

Smith, A. T., Wall, M. B., Williams, A. L., & Singh, K. D. (2006). Sensitivity to optic flow in human cortical areas MT and MST. *European Journal of Neuroscience, 23*(2), 561-569.

Sulpizio, V., Committeri, G., & Galati, G. (2014). Distributed cognitive maps reflecting real distances between places and views in the human brain. *Frontiers in Human Neuroscience, 8*, 716. doi: 10.3389/fnhum.2014.00716

Sulpizio, Valentina, Committeri, Giorgia, Lambrey, Simon, Berthoz, Alain, & Galati, Gaspare. (2013). Selective role of lingual/parahippocampal gyrus and retrosplenial complex in spatial memory across viewpoint changes relative to the environmental reference frame. *Behavioural Brain Research, 242*(0), 62-75. doi: http://dx.doi.org/10.1016/j.bbr.2012.12.031

Takemura, H., Rokem, A., Winawer, J., Yeatman, J. D., Wandell, B. A., & Pestilli, F. (2015). A Major Human White Matter Pathway Between Dorsal and Ventral Visual Cortex. *Cerebral Cortex*. doi: 10.1093/cercor/bhv064

Thier, P., & Erickson, R.G. (1992). Responses of visual-tracking neurons from cortical area MST-l to visual, eye and head motion. *European Journal of Neuroscience, 4,* 539-553.

Thier, P., & Ilg, U. J. (2005). The neural basis of smooth-pursuit eye movements. *Curr Opin Neurobiol, 15*(6), 645-652.

Tootell, R. B. H., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., . . . Dale, A. M. (1997). Functional analysis of V3A and related areas in human visual cortex. *Journal of Neuroscience, 17*(18), 7060-7078.

Tootell, R. B. H., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., . . . Belliveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 15*, 3215-3230.

Ungerleider, L. G., & Desimone, R. (1986). Cortical connections of visual area MT in the macaque. *Journal of Comparative Neurology, 248*, 190-222.

Vass, L. K., & Epstein, R. (2013). Abstract representations of location and facing direction in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 33*, 6133-6142. doi: http://dx.doi.org/10.1523/JNEUROSCI.3873-12.2013

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science, 287*(5456), 1273-1276.

Wall, M. B., & Smith, A. T. (2008). The representation of egomotion in the human brain. *Curr Biol, 18*(3), 191-194.

Ward, E. J., MacEvoy, S., & Epstein, R. (2010). Eye-centered encoding of visual space in scene-selective regions. *Journal of Vision, 10*(14). doi: http://dx.doi.org/10.1167/10.14.6

Yamasaki, D. S ., & Wurtz, R. W. (1991). Recovery of function after lesions in the superior temporal sulcus in the monkey. *Journal of Neurophysiology, 66*(3), 651-673.

Zeki, S. (1991). Cerebral akinetopsia (visual motion blindness) - a review. *Brain, 114*(Pt 2), 811-824.

2. Motion responses in scene selective regions

This chapter was reproduced from an article published in Neuroimage:

Korkmaz Hacialihafiz, D. and Bartels, A. (2015). Motion responses in scene-selective regions. Neuroimage 118, 438–444.

Authors: Didem Korkmaz Hacialihafiz and Andreas Bartels*

Vision and Cognition Lab, Centre of Integrative Neuroscience, University of Tübingen, Germany

*Corresponding author:

Andreas Bartels

Vision and Cognition Lab

Centre for Integrative Neuroscience,

University of Tübingen,

Otfried-Müller-Str. 25

72076 Tübingen

Germany

Email: andreas.bartels@tuebingen.mpg.de

Phone: +49 7071 2989168

2.1. Abstract

The vast majority of studies on scene processing were conducted using stationary scenes. However, during natural vision, scene views change dynamically due to self-induced eye-, head- and body-motion, and these dynamic changes are crucial for other higher-level functions such as navigation, self-motion perception and spatial updating. Yet, we do not know whether or how scene selective regions are modulated by visual motion, and to which degree their motion response depends on scene content. In this study we used fMRI to examine both questions using a 2x2 factorial design with the factors 2D planar motion (motion versus static), and scene content (natural scenes versus their fourier scrambles). We found that among independently localized scene responsive regions, parahippocampal place area (PPA) and transverse occipital sulcus (TOS), also referred to as occipital place area, (OPA) and were significantly motion responsive, whereas retrosplenial cortex (RSC) was not. Additionally, PPA, but not TOS/OPA or RSC, showed an interaction between motion and scene in that it responded more to motion in context of scenes than scramble. These results suggest that motion stimuli evoke different responses in motion responsive TOS/OPA and PPA versus motion un-responsive RSC.

Highlights

- Motion responses of scene regions were examined using 2-factorial design using fMRI
- Factors were scenes (natural scenes vs. fourier scramble) and motion (on vs. off)
- TOS/OPA responded strongest, PPA less, RSC not at all to visual motion
- TOS could be dissociated from PPA and RSC in their motion responses
- PPA showed an interaction between scenes and motion, although there was no significant difference between regions in their scene and motion interaction.

Keywords:

fMRI, PPA, TOS, OPA, retrosplenial cortex, scenes

2.2. Introduction

Scene perception and processing is one of the everyday functions of the human visual system. Its neural processing has therefore been studied in great detail, identifying scene selective responses in the parahippocampal place area (PPA) (Aguirre, Zarahn, & D'Esposito, 1998; Epstein & Kanwisher, 1998) as well as in comparably less studied regions such as the retrosplenial cortex (RSC) (Maguire, 2001) and the transverse occipital sulcus (TOS) (Grill-Spector, 2003; Hasson, Harel, Levy, & Malach, 2003; Nakamura et al., 2000) that has also been referred to as "occipital place area" (OPA) due to its causal contribution to scene perception (Dilks, Julian, Paunov, & Kanwisher, 2013). In the past, scene processing regions have almost exclusively been studied using static snapshots of scenes. However, scene perception is not only important on its own, but also crucial for other higher-level functions such as navigation, self-motion perception and spatial updating. During natural vision, scene views change dynamically due to self-induced eye-, head- and body-motion. A subset of this dynamics, that of instant view changes such as induced by saccades, has been investigated by several prior previous studies, using static snapshots created by dividing larger panoramic scenes into partially overlapping subsections (Epstein, Higgins, & Thompson-Schill, 2005; Golomb, Albrecht, Park, & Chun, 2011; Park & Chun, 2009; Park, Chun, & Johnson, 2010). These studies found viewpoint specific responses in PPA (Epstein, Graham, & Downing, 2003; Epstein, Higgins, Jablonski, & Feiler, 2007; Epstein et al., 2005; Epstein, Parker, & Feiler, 2007, 2008; Park & Chun, 2009; Park et al., 2010; Vass & Epstein, 2013), and in TOS/OPA (Epstein, Higgins, et al., 2007; Epstein et al., 2005). RSC was reported to be viewpoint-independent for the same scene, and to distinguish between different scenes (Park & Chun, 2009; Vass & Epstein, 2013), or to be partly viewpoint independent while viewpoint dependent under some conditions (Park et al., 2010). One more recent study found viewpoint invariance in PPA but viewpoint sensitivity in TOS/OPA (Dilks, Julian, Kubilius, Spelke, & Kanwisher, 2011). These discrepancies illustrate that these regions exhibit distinct functional properties that have not been fully understood yet. In particular, the response to motion or to scene motion of scene-selective regions has not been addressed at all until now. Given the importance and abundance of scene motion in real life conditions, this seems an important question worth addressing in a systematic way.

Hence, in this fMRI study we examined two main questions: are scene responsive regions in the human brain modulated by motion during viewing of natural and scrambled scenes, and if they are, do their responses show any interaction between scene content and motion? We investigated these questions by using scene and non-scene stimuli that were either shown statically or in horizontal linear motion. This led to a two-by-two factorial design allowing for full factorial control over main effects and their interaction. The factorial design also ensured that every contrast was fully balanced in terms of low-level stimulus properties. The scene images were gray-scale photographs of land- and city-scapes. For non-scene stimuli we used phase-scrambled images of the scenes matched in luminance, contrast, and frequency spectra to the gray-scale scene images. We chose horizontal motion as it is among the most abundant motion types in natural scenes and as our visual system has a tuning bias to cardinal motion directions (see e.g. (Bartels, Zeki, & Logothetis, 2008; Gros, Blake, & Hiris, 1998). We found that scene responsive regions showed differential responses to motion, with TOS/OPA showing highest, PPA intermediate, and RSC lacking motion responses. TOS/OPA could be differentiated from PPA and RSC in their motion responses. Moreover, PPA showed a significant interaction between scene content and motion, whereas this effect was not significant in RSC or TOS/OPA, although there was no significant difference between regions in their responses to scene and motion interaction.

2.3. Materials and Methods

Subjects

17 healthy subjects (9 female, 1 left handed, age between 20 and 36 (mean = 27.8)) participated in this study. All subjects had normal or corrected-to normal vision and gave written informed consent before the experiments. The study was approved by the local ethics committee of the University Hospital of Tübingen.

Experimental Paradigm and Setup

Two functional experiments were carried out: a functional scene region localizer and the main experiment, plus a structural scan. The functional localizer was used to localize scene- responsive regions PPA, RSC, and TOS/OPA.

Visual stimuli were gamma corrected and projected via a projector outside the scanner room onto a screen behind the participants' head yielding a visual field of view subtending 19 x 15 visual degrees. The experiment was programmed using Psychtoolbox-3 (Brainard 1997; Kleiner, Brainard et al. 2007) on MATLAB 7.10.0 (The Mathworks, Natick, MA, 2010) and presented using a windows PC.

Main Experiment

Paradigm

The main experiment was designed as a 2x2 factorial design with the factors scene (on/off) and motion (on/off), resulting in four conditions: moving scenes, moving scramble, still scenes and still scramble (figure 1).

Each condition was presented 4 times per run in a block-design, with a total of 16 blocks per run. The condition sequence was pseudorandomized such that each condition preceded all conditions equally frequently. Each block lasted 12 seconds. In order to additionally counterbalance initial conditions, one additional block was added to the beginning of each run.

Each run started with 6.9 seconds of gray screen (luminance: 144 cd/m2) with fixation and ended with 10 seconds of gray screen with fixation. Throughout the full duration of the experiment, participants fixated on a gray fixation disk (width: 0.74 degrees, luminance: 282 cd/m2) and performed the fixation task described below to ensure matched attentional demands across conditions. There were a total of 4 runs per participant.

Stimuli

Stimuli consisted of 32 gray scale images of outdoor scenes, namely landscapes and cityscapes, and of their phase-scrambled versions. Half of these images were left-right flipped duplicates to balance potential horizontal differences in spatial frequency. We equated luminance and contrast across images (luminance: 144 cd/m2, contrast: 32.4 cd/m2 root-mean-square (RMS) contrast, leading to an average Michelson contrast of 0.9004 ± 0.0925). Image selection was randomized for each condition separately and the images chosen from different image categories (landscape and cityscape) were balanced across conditions, in order to prevent bias resulting in different spatial properties due to different scene categories (Oliva & Torralba, 2001; Walther, Caddigan E Fau - Fei-Fei, Fei-Fei L Fau - Beck, & Beck, 2009). Phase scrambled images were obtained by applying a Fourier transformation followed by a reconstruction with random phases. As a result, scene content was removed while preserving low-level image attributes like luminance and contrast and spatial frequencies. During each block, the same, randomly chosen, image was presented. Images were larger than display width, allowing for lateral left-right panning motion across the screen in motion conditions. Motion of the stimuli was horizontal, following a sine velocity trajectory that ensured smooth motion (velocity range: 0 - 3.08 deg/sec, mean: 2.53 deg/sec). The trajectory period spanned 4 cycles per block, and its spatial extent was limited to 1.98 visual degrees in each direction. The starting direction of motion was pseudorandom and counterbalanced across runs.

Fixation Task

During the main experiment, subjects performed a 1-back character-matching task in order to ensure fixation and to balance attention across conditions. It consisted of a sequence of randomly presented alphabetical characters (a-z) displayed one at a time on the fixation disk. Characters were shown for 1 second with 83 ms blank intervals, with repetitions occurring between 3 and 8 presentations. Subjects were instructed to report character repetitions via button press. Button responses were recorded and included in the GLM as regressor of no interest.

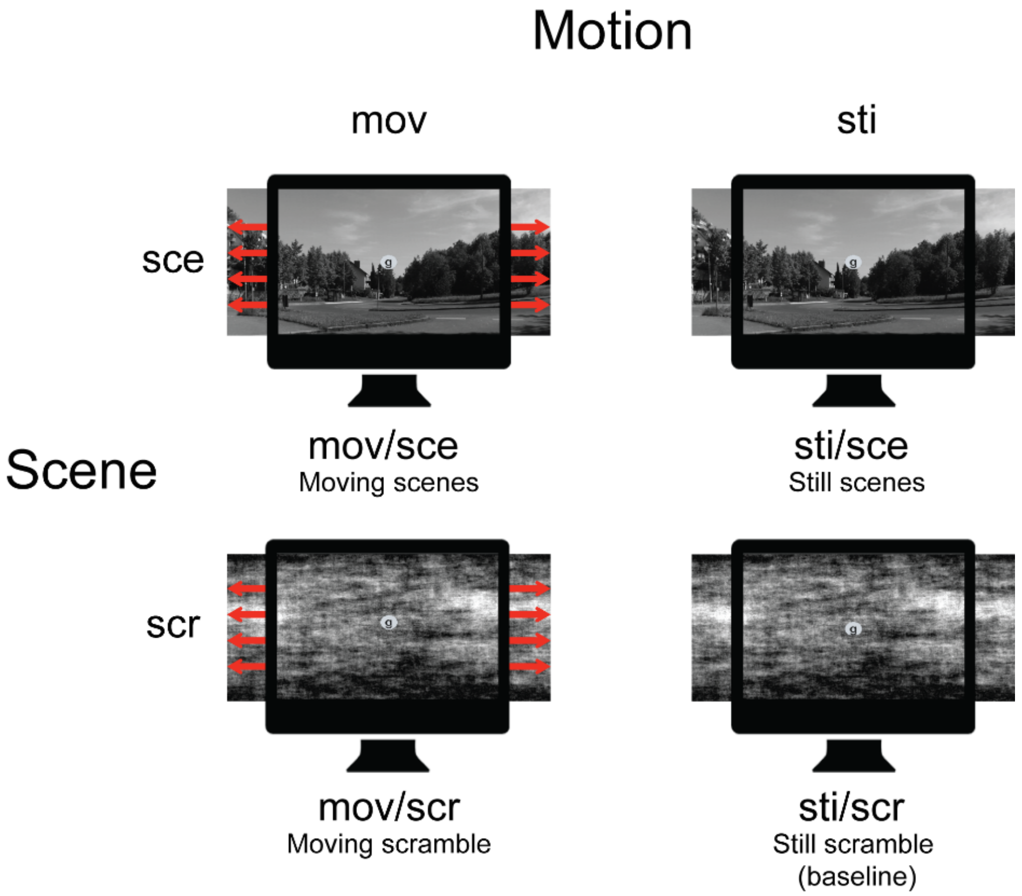


Figure 1. Illustration of the four stimulus conditions used in the main experiment. The conditions formed a 2 x 2 factorial design with the factors “scene” (scene/scramble) and “motion” (moving/still). There was a gray fixation disk present at all times, with a one-back matching character task. Motion was horizontal with a sinusoidal left-right velocity profile. The four conditions were as follows: mov/sce: moving scene, sti/sce: still scene, mov/scr: moving scramble, sti/scr: still scramble.

Functional Localizer

PPA, RSC and TOS/OPA were localized using an independent localizer experiment consisting of 3 conditions. Subjects viewed grayscale images of scenes, faces and phase scrambled versions of these in a block-design. Each block lasted 15 seconds. In each block, 5 different images from the same category were shown to subjects, each for 3 s. Blocks were separated by 1 second of blank screen. Each condition was shown 9 times during the run, which started with 6.9 seconds of gray screen with fixation and ended with 10 seconds of blank screen with fixation. PPA, RSC and TOS/OPA were localized using the contrast scenes versus faces (Epstein & Kanwisher, 1998). The regions were defined using an individual p-value for each subject (between $p < 0.05$ uncorrected and $p < 0.0001$ FWE corrected) in order to keep the ROIs approximately similar in size across participants and to account for individual variance in BOLD responsiveness, also referred to as variable thresholding technique (Fox, Iaria, & Barton, 2009; Murray & Wojciulik, 2004). Mean MNI coordinates and mean volumes of the ROIs are shown in table 1. There was a central fixation cross at all times.

Table 1. Mean MNI coordinates and volume of regions of interest locations across participants, for PPA, RSC and TOS/OPA. N indicates number of hemispheres in which ROIs could be identified.

ROI	X	Y	Z	Volume (mm^3)	N
Left PPA	-24,2 ± 2,5	-44,8 ± 4,8	-8,6 ± 5,5	1258,8 ± 399,4	17
Right PPA	24,7 ± 4,3	-43,3 ± 5,1	-9,5 ± 3,2	1354,9 ± 372,3	17
Left RSC	-15,0 ± 3,8	-57,6 ± 2,9	14,3 ± 5,7	728,5 ± 359,8	16
Right RSC	17,4 ± 4,2	-54,9 ± 5,2	16,2 ± 6,0	960,0 ± 482,2	17
Left TOS/OPA	-35.6 ± 6.7	-78,1 ± 4,5	16,4 ± 5,5	663,4 ± 347,3	14
Right TOS/OPA	34,1 ± 4,3	-77,9 ± 4.9	21.4 ± 5,8	568,0 ± 336,7	14

Image Acquisition

FMRI data were recorded with a Siemens Magnetom PRISMA 3 Tesla scanner using a 64-channel phased-array head coil (Siemens, Erlangen, Germany). T2* weighted functional images were recorded using a gradient-echo sequence to optimize blood-oxygen-level dependent (BOLD) contrast, with a TR of 2.3 s, TE of 35 ms, and a flip angle of 79°**.** Each brain volume consisted of 32 slices with a voxel size of 3 x 3 x 3 mm^3. The first 3 volumes of each run were discarded to allow for T1 equilibration.

In addition, T1-weighted high-resolution anatomical images were obtained with a resolution of 1 x 1 x 1 mm^3.

FMRI Data Preprocessing

Functional images were preprocessed using SPM5 (Friston et al. 1995, www.fil.ion.ucl.ac.uk/spm/) and MATLAB 7.10.0. Functional images were resliced and realigned. The structural image was coregistered to the mean functional image, and all images were normalized to the standard SPM template approximating the Montreal neurological institute (MNI) space. Images were spatially smoothed with a 6 mm kernel for single subject analyses, and with 12 mm for group analyses. For the extraction of ROIs, 6 mm smoothing was used. Time-series were high-pass filtered with a cut off of 128 s to remove low-frequency drifts in the signal.

Statistical Analysis

The general linear model (GLM) approach was used to analyze the responses from each subject separately. In addition, a second-level random effects group analysis was performed using contrast images from the first level analyses. The first level GLM included one regressor for each of the four conditions as well as one regressor for button responses. As regressors of no interest six motion realignment parameter time series were included, plus one additional regressor for global signal variance that was orthogonalized with respect to the conditions of interest (Desjardins, Kiehl, & Liddle, 2001; Van Dijk et al., 2010).

ROIs were defined based on the independent localizer for each individual subject. Mean beta values (averaged over all voxels) for each ROI of each subject were extracted from data of the main experiment.

In order to make response magnitudes comparable across the different ROIs, we normalized the ROI responses as follows. For a given ROI (e.g. PPA), we first subtracted the group mean response to the condition *still scramble*, then divided it by the group mean response to *still scenes*. This way, for each of the three ROIs, the group mean response to *still scramble* was set to zero, and the group mean response to *still scenes* to one. Note that this normalization preserved the between-subject variances for each condition in every ROI, therefore not affecting within-ROI analyses. Repeated measures ANOVAs were performed on the normalized ROI data to investigate the differential effects across the four conditions and their interactions within each ROI, and also across ROIs, using SPSS version 22 (IBM SPSS Statistics for Macintosh, Version 22.0). Greenhouse-Geisser correction was used in case of violation of sphericity, according to Mauchly's sphericity test. Since classical post-hoc tests are not permitted for repeated measures ANOVAs, paired t-tests were performed where significant interactions between effects and ROIs were found in the ANOVA in order to investigate which responses drove between-ROI differences. T-tests were then Bonferroni-Holm corrected for the number of tests performed.

Finally, whole-brain random effects group analyses were carried out by conducting t-tests on contrast images obtained from every subject for the scene-motion interaction.

Eye Tracking

Eye positions were recorded during the experiment at a sampling rate of 60 Hz using an infrared camera based eye tracker (Eye-Trac 6; Applied Science Laboratories). Eye position data were preprocessed, which included blink removal and smoothing of x and y gaze points using a 200 millisecond running average window. Fixation accuracy was then calculated using two measures. The first measure was the average distance of actual eye position relative to the fixation disk. The second measure was the standard deviation of the distances. These values were calculated for each condition separately across runs, yielding one value per subject and condition. Separate one-way ANOVAs were then conducted for each measure to examine differences across conditions.

2.4. Results

ROI Analyses

We investigated the main effects of visual horizontal motion (on vs. off), of scene content (outdoor scenes vs. phase-scramble), and of their interaction on independently localized scene processing regions PPA, TOS/OPA, and RSC. PPA was localized in 34 of 34 hemispheres, TOS/OPA in 28 of 34 hemispheres and RSC in 33 of 34 hemispheres (see table 1 for mean coordinates and volumes). Figure 2A shows raw mean beta responses to all conditions for each ROI.

To examine within-ROI effects, we analyzed each ROI separately using a two-way repeated measures ANOVA with the factors motion and scene. The main effects and interaction effects for each ROI are shown in Figure 3.

In PPA, there were significant effects of scene (F (1,33) = 227.4, p = 2.3 * 10^{-16}), motion (F (1,33) = 24.9. p = 1.9 * 10^{-5}) and a significant interaction between scene and motion (F (1,33) = 9.4, p = 0.004).

In RSC, there was a significant effect of scene (F (1,32) = 66.4, p = 2.6 * 10^{-9}) but there was neither a significant effect of motion (F (1,32) = 0.33, p = 0.57) nor a significant interaction between scene and motion (F (1,32) = 0.96, p = 0.33).

In TOS/OPA, there were significant effects of scene (F (1,27) = 71.9, p = 4.3 * 10^{-9}) and motion (F (1,27) = 34.4, p = 3.0 * 10^{-6}) but there was no significant interaction between scene and motion (F (1,27) = 1.2, p = 0.29).

In sum, all regions responded as expected to scenes, but only PPA and TOS/OPA had responses to visual motion, which were absent in RSC (Figure 3A). Only PPA showed a significant interaction between motion and scene-content, in that its motion response was higher in scene compared to scramble conditions, although the BOLD responses of RSC was also in a similar pattern (Figure 3B).

Next, we compared the different ROIs with each other. Since overall response magnitudes differed substantially between ROIs, presumably due to different distances to the MR-coil, we normalized them such that their mean-response to still scenes were equalized. We did this by first subtracting the group mean response to still scramble (i.e. this was now 0), then dividing all responses by the group mean response to still scenes (this now equaled one in each ROI). This normalization preserved individual variance across participants in each condition. The normalized beta estimates are shown in Figure 2B.

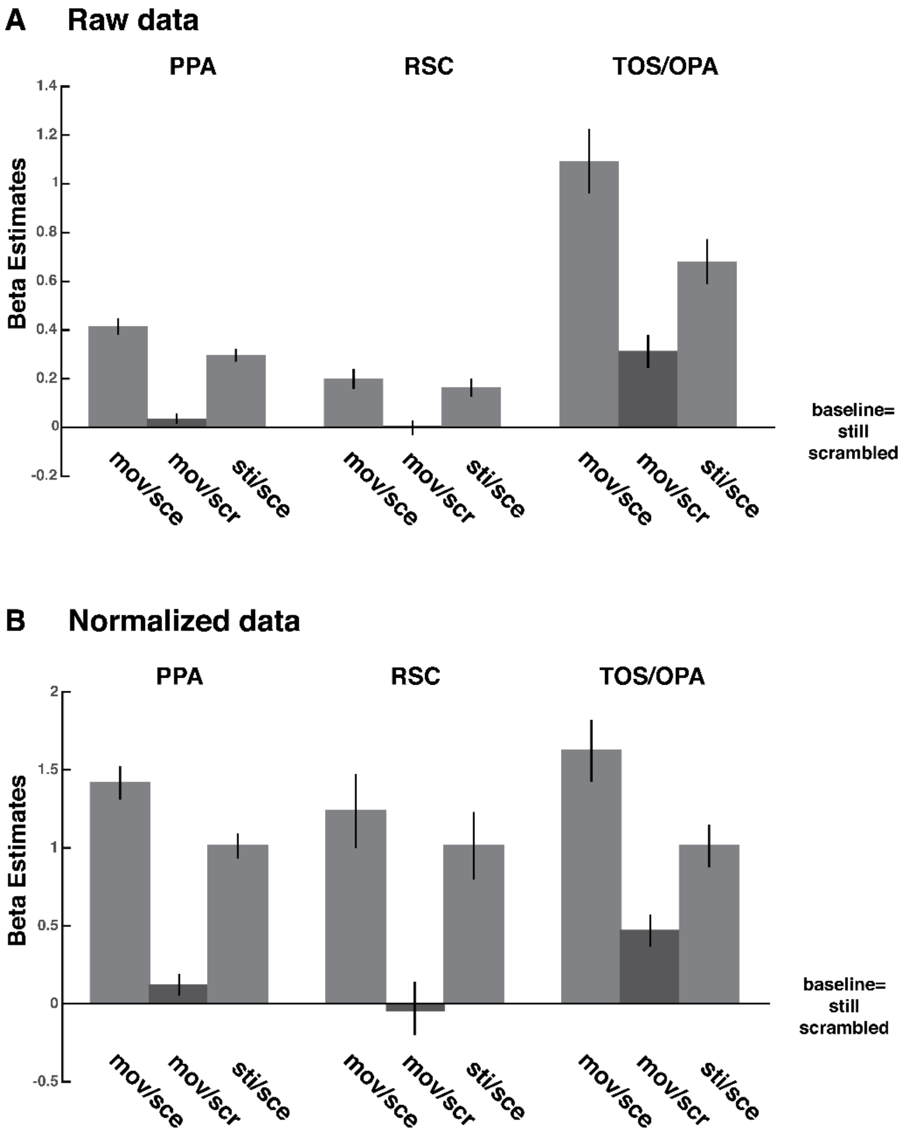


Figure 2. ROI responses to the experimental conditions. (A) Raw beta estimates of ROIs to the conditions moving scene, still scene and moving scramble. To provide a better estimate of variance for the remaining conditions, the condition still scramble was subtracted from all conditions for each subject (note that repeated-measures statistics were carried out without this step). (B) Normalized beta estimates. Same data as in (A), but all responses were divided by the mean response of still scene. Again, for statistics, the group mean of still scramble had been subtracted instead of each individual response. Error bars show standard error of the mean (SEM).

The normalized ROI data were submitted to a 3 x 2 x 2 x 2 repeated measures ANOVA, with the factors ROI (PPA/RSC/TOS (OPA)), hemisphere (left/right), motion (on/off) and scene (on/off). Since there was no significant effect of hemisphere or any significant interaction between hemisphere and any of the factors, we combined data from left and right hemispheres and conducted a more tractable three factor ANOVA with ROI, motion and scene as factors. The results revealed significant main effects for motion (F (1,27) = 17.0, p = 3.2 * 10^{-4}) and scene (F (1,27) = 142.5, p = 2.8 * 10^{-12}), and significant interactions between ROI and motion (F (1.27,35.8) = 4.3, p = 0.034) and between motion and scene (F (1,27) = 4.1, p = 0.05) but not between ROI, motion and scene (F (1.33,36.1) = 0.375, p = 0.6), ROI and scene (F (2,54) = 0.16, p < 0.86) and no significant main effects for ROI (F (1.30,35.1) = 1.9, p = 0.17).

Since post-hoc tests are invalid for repeated measures ANOVAs, we followed up the significant ANOVA results using Bonferroni-Holm corrected t-tests.

In order to examine which differences between ROIs drove the interaction between ROI and motion in the ANOVA, we compared ROI responses for motion using paired t-tests across ROIs, Bonferroni-Holm corrected for three comparisons (see Figure 3A). We found significant differences between PPA and TOS/OPA (t (27) = 2.678, p = 0.036 corrected) and RSC and TOS/OPA (t (27) = 2.402, p = 0.046 corrected) but not between PPA and RSC (t (32) = 1.071, p = 0.292).

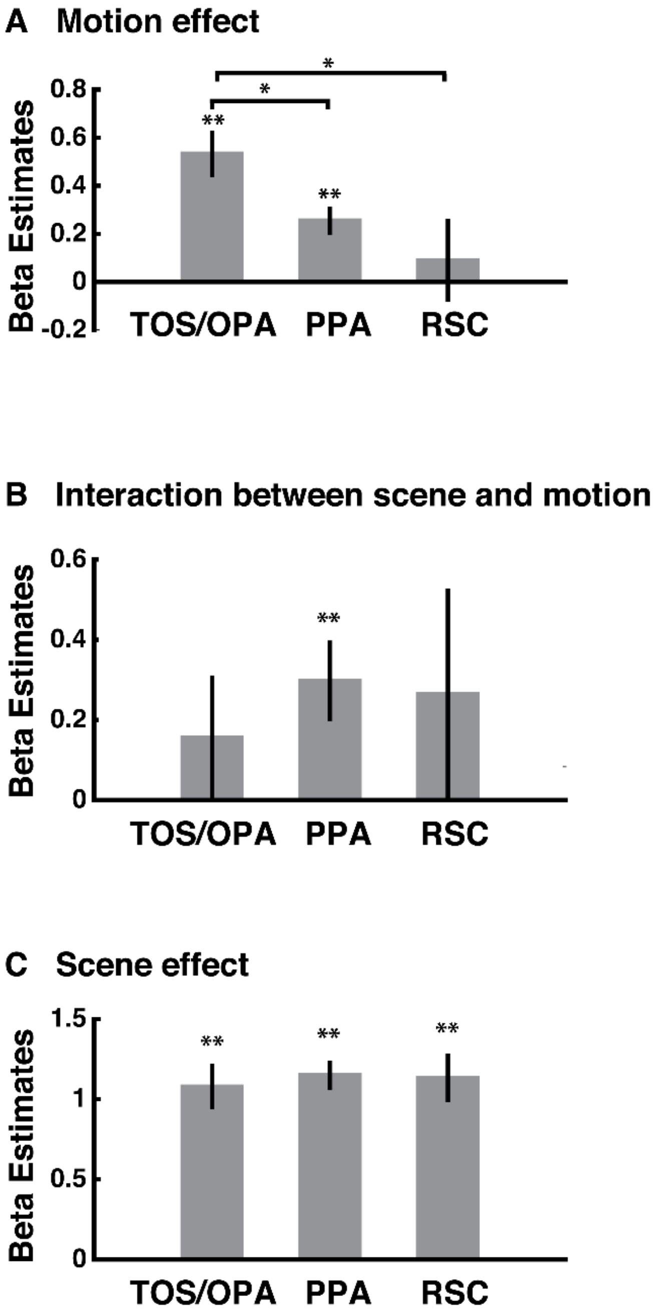


Figure 3. Main effects of motion, scene and their interaction for each ROI, shown for data normalized to static scene effects. (A) Main effect of motion, i.e. moving vs. still across scenes and scramble. A three-way ANOVA with factors ROI, scene and motion showed significant ROI-by-motion interaction, and targeted t-tests between ROIs showed significant motion preference of TOS/OPA compared to both, PPA and RSC. (B) Interaction between scene and motion effects, i.e. (moving scene vs moving scramble) vs. (still scene vs still scramble). Only PPA showed significant effects. (C) Main effect of scene, i.e. scene vs. scrambled across moving and still. Note that data were normalized such that static scene vs static scramble equaled one in each ROI (see Figure 2B). **: $p < 0.005$, *: $p < 0.05$, Bonferroni-Holm corrected. Error bars show standard error of mean (SEM).

Whole Brain Analyses

Whereas scene and motion responses per se have been reported in numerous prior studies, their interaction has not been examined brain-wide before. We therefore performed additional whole-brain random-effects analyses across all voxels of the brain in order to test whether regions beyond the scene selective ROIs responded to the interaction between motion and scene. The whole-brain RFX analysis showed no activation for the interaction of scenes and motion surviving FDR or FWE correction. However, Figure 4 shows that the interaction was evident in PPA at uncorrected levels, with peak T-statistic values of t = 2.54 in left PPA and t = 2.13 in right PPA.

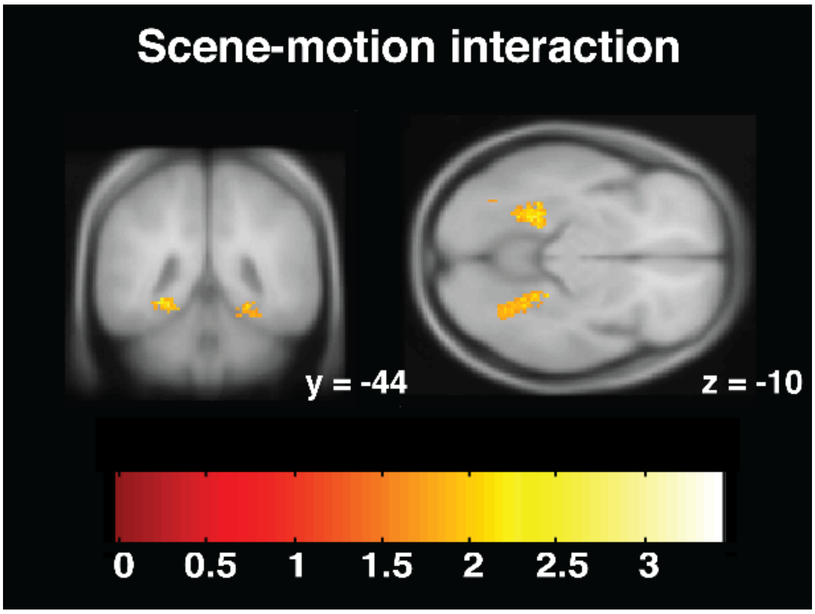


Figure 4. Group data for interaction of scenes and motion. A random effects whole-brain analysis for the interaction between scene and motion yielded no significant voxels surviving FDR or FWE correction. However, voxels in PPA were evident at uncorrected thresholds, shown here in coronal and sagittal slices with peak T-statistic values of t = 2.54 in left PPA and t = 2.13 in right PPA, shown thresholded at $p < 0.05$ uncorrected for illustration.

Behavioral Data

Throughout the experiment subjects had performed a fixation character back-matching task intended to maintain vigilance and to balance attention across conditions. The average rate of correct responses was near ceiling with 0.87 ± 0.06 (mean ± std). The mean response time was 0.56 s ± 0.14 s (mean ± std). As a measure of attentional engagement we performed an ANOVA over reaction times across conditions and found no difference (F (1,3) = 0.43, p = 0.73) across conditions.

Eye Tracking Data

Eye position data had been collected throughout the experiment, and we extracted two measures for each of the four conditions: average eye-position distance from the central fixation, and the standard deviation of the fixation-error. For each of these two measures, separate ANOVAs were calculated across the four stimulus conditions. Neither ANOVA showed differences across conditions (for the average distances, F (3,268) = 0.31, p = 0.8167, for the standard deviation of average distances, F (3,268) = 0.19, p = 0.9025).

2.5. Discussion

In this study we addressed two simple but fundamental questions: first, whether scene processing regions are modulated by visual planar motion and second, to which extent their motion responses are modulated as a function of scene content. To answer these questions we conducted fMRI experiments using scene and scramble stimuli combined with horizontal linear motion as visual stimuli. Horizontal panning motion was chosen as it is one of the most abundant motion types natural scenes (see e.g. (Bartels et al., 2008; Gros et al., 1998), occurring e.g. in the background when we track a laterally moving object in the foreground. We focused on independently identified scene responsive regions, namely PPA (Epstein & Kanwisher, 1998), RSC (Maguire, 2001) and TOS/OPA (Dilks et al., 2013; Grill-Spector, 2003; Hasson et al., 2003) as their motion responses have not been examined previously. The results showed that motion responses of TOS/OPA was significantly different from motion responses of PPA and RSC, and that only PPA showed a significant dependence of its motion-response on natural scene content, although there was no significant difference between PPA, RSC and TOS in their responses to scene and motion interaction.

Content-dependent motion responses in Scene Selective Regions

PPA responded significantly to motion, and showed a scene-motion interaction in that its motion response (vs. still) was higher in context of scenes than in context of scrambled scenes. This interaction was not found in RSC or TOS/OPA even though all three regions responded significantly to scenes compared to scramble and the three regions' responses did not differ from each other significantly for this interaction. This pattern of results is compatible with prior studies that used static stimuli and suggested that PPA responses are viewpoint dependent (Epstein et al., 2003; Epstein & Higgins, 2007; Epstein et al., 2005; Epstein, Parker, et al., 2007; MacEvoy & Epstein, 2007; Park & Chun, 2009). These studies showed that PPA responses were heightened upon presentation of statically presented but horizontally shifted scenes in comparison to non-shifted scenes. There are however other studies suggesting that PPA becomes viewpoint invariant when the difference between viewpoints is small (Ewbank, Schluppeck, & Andrews, 2005) or when subjects become familiar with the scenes (Epstein et al., 2005). For this reason, we intentionally limited the absolute extent of motion to a few visual degrees to minimize appearance of new scene content through motion. The maximal range of scene shift covered ±1.98 visual degrees (preserving 73.6 % of preserved overlap between extreme views), whereas some prior studies used 16 degrees (33 % of preserved overlap) (Park & Chun, 2009). Our results therefore suggest that even small continuous viewpoint changes of less than ±2.5 degrees drive PPA, specifically in context of scenes, indicating a high sensitivity of PPA for dynamic changes in viewpoint.

PPA and RSC are thought to have distinct roles in scene processing (Epstein, Parker, et al., 2007). RSC is thought to be more involved in navigation and route learning, integrating spatial representation from different sources (Wolbers & Buchel, 2005), mediating also the recognition of known scenes, whereas PPA has been proposed to have a more visual role supporting the immediate perception scenes, regardless of familiarity or navigational relevance (Epstein, Higgins, et al., 2007). For example, while RSC responses are viewpoint invariant, it is sensitive to mirror image reversal that implies distinct scenes, whereas the reverse was found in PPA (Dilks et al., 2011). Moreover, a recent study showed that RSC signaled permanent landmarks (Auger, Mullally, & Maguire, 2012).

Seen this way, the lack of scene-motion interactions found here in RSC is in line with previous work showing that while PPA is viewpoint sensitive, RSC is viewpoint invariant (Park & Chun, 2009; Park et al., 2010). The viewpoint invariance of RSC has been shown to depend on the continuity of the static snapshots presented (Park & Chun, 2009), consistent with the continuous viewpoint change conferred in our moving stimuli.

Another explanation for our results could be that the difference in the content information of dynamic scenes compared to still scenes drives the difference in PPA and RSC. Since the scenes that were shown for different conditions were balanced, the amount of information provided by the scenes differed between conditions because of moving scenes spanning across a wider view. Previously, PPA and RSC were shown to have a larger representation of the scene beyond the shown physical view, which is known as ‘boundary extension‘ and while PPA showed attenuation to same views, RSC did not (Park, Intraub, Yi, Widders, & Chun, 2007), which indicates that while RSC integrates the present view of the scene into a more general representation of the space, which is needed for navigation, PPA is more sensitive to physical attributes of the scene.

This could explain our content related findings in PPA and RSC, while PPA being more sensitive to physical properties of the viewpoint would result in scene motion interaction, this interaction would not be as significant in RSC since RSC places a given view of the scene into a larger view in the first encounter. Our results are compatible RSC being higher in the scene processing hierarchy than PPA, in the sense that it integrates the viewpoint of the scene to a broader scene perception, presumably for the purpose of spatial navigation, whereas PPA is in lower than RSC and is presumably responsible for encoding physical properties of the given viewpoint of the scene. However, in our stimuli, the order of conditions was balanced (moving scenes similar to wide view whereas still scenes are similar to close view) and their results showed the attenuation effect only happens when a close view was followed by wide view, but not the other way around. Moreover, our design can also be reviewed in the light of viewpoint specificity of PPA, where the moving scene condition could be seen as a continuous set of different viewpoints of an image, compared to a given viewpoint of the image in still scenes condition(Epstein et al., 2003; Schmidt et al., 2007). Yet, future studies are needed to establish the underlying mechanisms of the content dependent motion responses in scene responsive regions.

Motion responses in Scene Selective Regions

In recent years, there has been increasing interest in motion responses of regions in the ventral visual pathway that is traditionally not considered to be motion selective. Motion responses and even direction specificity have been shown to also exist in the ventral visual pathway, even in relatively early regions such as V4 (Tolias, Keliris, Smirnakis, & Logothetis, 2005; Tolias, Smirnakis, Augath, Trinath, & Logothetis, 2001). Motion responses have also been found in the object processing pathway, specifically in the lateral occipital cortex (LOC) (Self & Zeki, 2005). Even the perception of motion can depend on ventral visual cortex integrity (Gilaie-Dotan et al., 2013). A growing body of literature showed that face selective areas show increased responses to dynamic face stimuli (Pitcher, Dilks, Saxe, Triantafyllou, & Kanwisher, 2011; Schultz & Pilz, 2009), and even directional selectivity in both ventral and dorsal face processing regions (Reinl & Bartels, 2014).

Findings showing that some neurons in monkey parahippocampal cortex responded to motion stimuli are directly relevant for the current study (Sato &

Nakamura, 2003). Previous human studies failed to find motion responses in PPA (Epstein, Harris, Stanley, & Kanwisher, 1999; Hasson et al., 2003) or in TOS/OPA (Hasson et al., 2003), presumably because they did not achieve the sensitivity afforded by our dedicated factorial design and they did not use planar motion. One of the aforementioned studies used expanding and contradicting moving rings (Hasson et al., 2003), whereas the other one used a sequence of photographs of a scene taken during forward motion of the camera (Epstein et al., 1999). Relevant in this context may be the recently discovered specific preference of PPA for cardinal orientations (Nasr & Tootell, 2012) that would render horizontal motion as employed here a particularly salient stimulus for this region.

TOS/OPA is the least studied scene selective region, its functional role compared to that of the other regions is comparably poorly understood (Dilks et al., 2013; Ganaden, Mullin, & Steeves, 2013), and it is less easily localizable compared to other scene responsive regions (Konkle & Oliva, 2012; Mullin & Steeves, 2013) (and our own observations). While TOS/OPA plays a role in encoding the higher order spatial relationships of objects within a scene (Bettencourt & Xu, 2013; Nasr, Devaney, & Tootell, 2013), it has been suggested to contribute to early steps of scene processing (Dilks et al., 2011; Dilks et al., 2013; MacEvoy & Epstein, 2007), since it has smaller receptive fields than those of PPA or RSC (MacEvoy & Epstein, 2007), even though they cover both hemifields (Ward, MacEvoy, & Epstein, 2010).

Our results showing that TOS/OPA has the highest motion responses among scene processing regions is compatible with its dorsal location.

In contrast to PPA and TOS, RSC did not have any motion responses or interactions between motion and content. This is unlikely due to lack of signal in this region, as RSC showed highly significant scene responses. The reason is thus more likely to be found in its above-discussed high-level function and invariance to (continuous) changes of scene views.

2.6. Conclusion

We conclude that scene responsive regions can be differentiated from each other in the motion sensitivity of TOS/OPA and the lack thereof in RSC. Moreover, PPA's differed from both other regions in showing a significant interaction between scene and motion, indicating its particular sensitivity to continuous changes in scene-views rather than a low-level sensitivity to motion. To our knowledge, these results are first to systematically examine motion responses and motion-content relationships in scene processing regions.

Our results support the view of viewpoint dependent representation of scenes in PPA, content-driven, motion invariant responses in RSC, and comparably more low-level responses in TOS/OPA driven by both content and motion. The present results provide insight into modulation and invariance of TOS/OPA, PPA and RSC to the abundant low-level feature of planar motion that dominates visual input in natural conditions. Further studies are needed to investigate responses of scene responsive regions to different types of motion in interaction with scene content, and to understand their exact role in integrating scenes into a continuous, navigatable view despite constant motion.

Acknowledgements

This work was funded by the Centre for Integrative Neuroscience Tübingen through the German Excellence Initiative (EXC307) and by the Max Planck Society, Germany.

Conflict of Interest: The authors declare no competing financial interests.

2.7. References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An Area within Human Ventral Cortex Sensitive to "Building" Stimuli: Evidence and Implications. *Neuron, 21*(2), 373-383. doi: http://dx.doi.org/10.1016/S0896-6273(00)80546-2

Auger, S. D., Mullally, S. L., & Maguire, E. A. (2012). Retrosplenial Cortex Codes for Permanent Landmarks. *PLoS ONE, 7*(8), e43620. doi: http://dx.doi.org/10.1371/journal.pone.0043620

Bartels, A., Zeki, S., & Logothetis, N. K. (2008). Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. *Cerebral cortex, 18*, 705-717. doi: http://dx.doi.org/10.1093/cercor/bhm107

Bettencourt, K. C., & Xu, Y. (2013). The Role of Transverse Occipital Sulcus in Scene Perception and Its Relationship to Object Individuation in Inferior Intraparietal Sulcus. *Journal of Cognitive Neuroscience, 25*(10), 1711-1722. doi: http://dx.doi.org/10.1162/jocn_a_00422

Desjardins, A. E., Kiehl, K. A., & Liddle, P. F. (2001). Removal of confounding effects of global signal in functional MRI analyses. *Neuroimage, 13*(4), 751-758. doi: http://dx.doi.org/10.1006/nimg.2000.0719

Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 31*, 11305-11312. doi: http://dx.doi.org/10.1523/JNEUROSCI.1935-11.2011

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 33*, 1331-1336a. doi: http://dx.doi.org/10.1523/JNEUROSCI.4081-12.2013

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron, 37*(5), 865-876. doi: http://dx.doi.org/10.1016/s0896-6273(03)00117-x

Epstein, R., Harris, A, Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron, 23*(1), 115-125. doi: http://dx.doi.org/10.1016/S0896-6273(00)80758-8

Epstein, R., & Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb Cortex, 17*(7), 1680-1693. doi: http://dx.doi.org/10.1093/cercor/bhl079

Epstein, R., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *Journal of neurophysiology, 97*, 3670-3683. doi: http://dx.doi.org/10.1152/jn.00003.2007

Epstein, R., Higgins, J. S., & Thompson-Schill, S. L. (2005). Learning Places from Views: Variation in Scene Processing as a Function of Experience and Navigational Ability. *Journal of Cognitive Neuroscience, 17*(1), 73-83. doi: http://dx.doi.org/10.1162/0898929052879987

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598-601. doi: http://dx.doi.org/10.1038/33402

Epstein, R., Parker, W., & Feiler, A. (2007). Where Am I Now? Distinct Roles for Parahippocampal and Retrosplenial Cortices in Place Recognition. *The Journal of*

Neuroscience, 27(23), 6141-6149. doi: http://dx.doi.org/10.1523/jneurosci.0799-07.2007

Epstein, R., Parker, W., & Feiler, A. (2008). Two Kinds of fMRI Repetition Suppression? Evidence for Dissociable Neural Mechanisms. *Journal of Neurophysiology, 99*(6), 2877-2886. doi: http://dx.doi.org/10.1152/jn.90376.2008

Ewbank, M. P., Schluppeck, D., & Andrews, T. J. (2005). fMR-adaptation reveals a distributed representation of inanimate objects and places in human visual cortex. *NeuroImage, 28*, 268-279. doi: http://dx.doi.org/10.1016/j.neuroimage.2005.06.036

Fox, C. J., Iaria, G., & Barton, J. J. (2009). Defining the face processing network: optimization of the functional localizer in fMRI. *Hum Brain Mapp, 30*(5), 1637-1651. doi: http://dx.doi.org/10.1002/hbm.20630

Ganaden, R. E., Mullin, C. R., & Steeves, J. K. (2013). Transcranial Magnetic Stimulation to the Transverse Occipital Sulcus Affects Scene but Not Object Processing. *J Cogn Neurosci, 25*(6), 961-968. doi: http://dx.doi.org/10.1162/jocn_a_00372

Gilaie-Dotan, S., Saygin, A. P., Lorenzi, L. J., Egan, R., Rees, G., & Behrmann, M. (2013). The role of human ventral visual cortex in motion perception. *Brain, 136*(Pt 9), 2784-2798. doi: http://dx.doi.org/10.1093/brain/awt214

Golomb, J. D., Albrecht, A. R., Park, S., & Chun, M. M. (2011). Eye Movements Help Link Different Views in Scene-Selective Cortex. *Cerebral Cortex, 21*(9), 2094-2102. doi: http://dx.doi.org/10.1093/cercor/bhq292

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology, 13*(2), 159-166. doi: http://dx.doi.org/10.1016/s0959-4388(03)00040-0

Gros, B. L., Blake, R., & Hiris, E. (1998). Anisotropies in visual motion perception: a fresh look. *Journal of the Optical Society of America A, 15*(8), 2003-2011. doi: http://dx.doi.org/10.1364/JOSAA.15.002003

Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-Scale Mirror-Symmetry Organization of Human Occipito-Temporal Object Areas. *Neuron, 37*(6), 1027-1041. doi: http://dx.doi.org/10.1016/S0896-6273(03)00144-2

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron, 74*, 1114-1124. doi: http://dx.doi.org/10.1016/j.neuron.2012.04.036

MacEvoy, S., & Epstein, R. (2007). Position selectivity in scene- and object-responsive occipitotemporal regions. *Journal of Neurophysiology, 98*, 2089-2098. doi: http://dx.doi.org/10.1152/jn.00438.2007

Maguire, E. A. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavian Journal of Psychology, 42*(3), 225-238. doi: http://dx.doi.org/10.1111/1467-9450.00233

Mullin, C. R., & Steeves, J. K. E. (2013). Consecutive TMS-fMRI Reveals an Inverse Relationship in BOLD Signal between Object and Scene Processing. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 33*, 19243-19249. doi: http://dx.doi.org/10.1523/JNEUROSCI.2537-13.2013

Murray, S. O., & Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral occipital complex. *Nat Neurosci, 7*(1), 70-74. doi: http://dx.doi.org/10.1038/nn1161

Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., . . . Zilles, K. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing. *Brain, 123*(9), 1903-1912. doi: http://dx.doi.org/10.1093/brain/123.9.1903

Nasr, S., Devaney, K. J., & Tootell, R. B. H. (2013). Spatial encoding and underlying circuitry in scene-selective cortex. *NeuroImage, 83*, 892-900. doi: http://dx.doi.org/10.1016/j.neuroimage.2013.07.030

Nasr, S., & Tootell, R. B. H. (2012). A cardinal orientation bias in scene-selective visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 32*, 14921-14926. doi: http://dx.doi.org/10.1523/JNEUROSCI.2036-12.2012

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision, 42*(3), 145-175. doi: 10.1023/A:1011139631724

Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage, 47*(4), 1747-1756. doi: http://dx.doi.org/10.1016/j.neuroimage.2009.04.058

Park, S., Chun, M. M., & Johnson, M. K. (2010). Refreshing and integrating visual scenes in scene-selective cortex. *Journal of cognitive neuroscience, 22*, 2813-2822. doi: http://dx.doi.org/10.1162/jocn.2009.21406

Park, S., Intraub, H., Yi, D.-J., Widders, D., & Chun, M. M. (2007). Beyond the edges of a view: boundary extension in human scene-selective visual cortex. *Neuron, 54*, 335-342. doi: 10.1016/j.neuron.2007.04.006

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage, 56*(4), 2356-2363. doi: http://dx.doi.org/10.1016/j.neuroimage.2011.03.067

Reinl, M., & Bartels, A. (2014). Face processing regions are sensitive to distinct aspects of temporal sequence in facial dynamics. *Neuroimage, 102P2*, 407-415. doi: http://dx.doi.org/10.1016/j.neuroimage.2014.08.011

Sato, N., & Nakamura, K. (2003). Visual response properties of neurons in the parahippocampal cortex of monkeys. *J Neurophysiol, 90*(2), 876-886. doi: http://dx.doi.org/10.1152/jn.01089.2002

Schmidt, D, Krause, B J, Weiss, P H, Fink, G R, Shah, N J, Amorim, M-a, . . . Berthoz, a. (2007). Visuospatial working memory and changes of the point of view in 3D space. *NeuroImage, 36*, 955-968. doi: 10.1016/j.neuroimage.2007.03.050

Schultz, J., & Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Experimental brain research, 194*(3), 465-475. doi: http://dx.doi.org/10.1007/s00221-009-1721-9

Self, M. W., & Zeki, S. (2005). The integration of colour and motion by the human visual brain. *Cerebral cortex, 15*, 1270-1279. doi: http://dx.doi.org/10.1093/cercor/bhi010

Tolias, A. S., Keliris, G. A., Smirnakis, S. M., & Logothetis, N. K. (2005). Neurons in macaque area V4 acquire directional tuning after adaptation to motion stimuli. *Nat Neurosci, 8*(5), 591-593. doi: http://dx.doi.org/10.1038/nn1446

Tolias, A. S., Smirnakis, S. M., Augath, M. A., Trinath, T., & Logothetis, N. K. (2001). Motion processing in the macaque: revisited with functional magnetic resonance imaging. *J Neurosci, 21*(21), 8594-8601.

Van Dijk, K. R., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., & Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics:

theory, properties, and optimization. *J Neurophysiol, 103*(1), 297-321. doi: http://dx.doi.org/10.1152/jn.00783.2009

Vass, L. K., & Epstein, R. (2013). Abstract representations of location and facing direction in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 33*, 6133-6142. doi: http://dx.doi.org/10.1523/JNEUROSCI.3873-12.2013

Walther, D. B., Caddigan E Fau - Fei-Fei, Li, Fei-Fei L Fau - Beck, Diane M., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. (1529-2401 (Electronic)). doi: D - NLM: NIHMS143895

D - NLM: PMC2774133 EDAT- 2009/08/28 09:00 MHDA- 2009/09/11 06:00 CRDT- 2009/08/28 09:00 AID - 29/34/10573 [pii] AID - 10.1523/JNEUROSCI.0559-09.2009 [doi] PST - ppublish

Ward, E. J., MacEvoy, S., & Epstein, R. (2010). Eye-centered encoding of visual space in scene-selective regions. *Journal of Vision, 10*(14). doi: http://dx.doi.org/10.1167/10.14.6

Wolbers, T., & Buchel, C. (2005). Dissociable retrosplenial and hippocampal contributions to successful formation of survey representations. *J Neurosci, 25*(13), 3333-3340. doi: http://dx.doi.org/10.1523/JNEUROSCI.4705-04.2005

3. Objective (Real-world) Motion Responses in Scene Responsive Regions

Authors: Didem Korkmaz Hacialihafiz, Andreas Bartels*

Vision and Cognition Lab, Centre of Integrative Neuroscience, University of Tübingen, Germany

*Corresponding author:

Andreas Bartels
Vision and Cognition Lab
Centre for Integrative Neuroscience,
University of Tübingen,
Otfried-Müller-Str. 25
72076 Tübingen
Germany

Email: andreas.bartels@tuebingen.mpg.de
Phone: +49 7071 2989168

3.1. Abstract

We perceive scenes as stable even when eye movements induce retinal motion, for example during pursuit of a moving object. Mechanisms mediating perceptual stability have primarily been examined in motion regions of the dorsal visual pathway. Here we examined whether motion responses in human scene regions are encoded in eye- or world centered reference frames. We recorded brain responses in human participants using fMRI while they performed a well-controlled visual pursuit paradigm previously used to examine dorsal motion regions. In addition, we examined effects of content by using either natural scenes or their Fourier scrambles. We found that parahippocampal place area (PPA) responded to motion only in world- but not in eye-centered coordinates, regardless of scene content. The occipital place area (OPA) responded to both, objective and retinal motion equally, and retrosplenial cortex (RSC) had no motion responses but responded to pursuit. Only PPA's objective motion responses were higher during scenes than scrambled images, although there was a similar trend in OPA. These results indicate a special role of PPA in representing its content in real-world coordinates. Our results question a strict subdivision of dorsal "what" and ventral "where" streams, and suggest a role of PPA in contributing to perceptual stability.

3.2. Introduction

Keeping a stable visual perception is one of the crucial roles of the visual system. Our visual system continuously integrates retinal inputs with eye-, head- and body movements in order to keep our subjective visual percept stable and in register with the external world. The mechanisms behind this constant updating are only partially known. Prior studies have primarily focused on high-level parietal regions, identifying several that engage in so-called remapping, potentially allowing for stable perception (Fischer, Bulthoff, Logothetis, & Bartels, 2012a, 2012b; Galletti, Battaglini, & Fattori, 1990; Ilg, Schumann, & Thier, 2004; Zhang, Heuer, & Britten, 2004). Given that the content that is perceived as stable during eye movements typically consists of real-world scenes, it seems reasonable to examine this question also in scene-selective regions.

Scene processing has been shown to take place in PPA (Aguirre, Zarahn, & D'Esposito, 1998; Epstein & Kanwisher, 1998), RSC (Maguire, 2001) and OPA (also known as transverse occipital sulcus (TOS)) (Dilks, Julian, Paunov, & Kanwisher, 2013; Grill-Spector, 2003; Hasson, Harel, Levy, & Malach, 2003; Nakamura et al., 2000). Several studies have examined spatial updating in these regions using saccadic eye movements (Golomb, Albrecht, Park, & Chun, 2011; Ward, MacEvoy, & Epstein, 2010) or snapshots of different viewpoints (Epstein, Graham, & Downing, 2003; Epstein, Higgins, & Thompson-Schill, 2005; Park & Chun, 2009; V. Sulpizio, Committeri, & Galati, 2014; Valentina Sulpizio, Committeri, Lambrey, Berthoz, & Galati, 2013). These studies have however not led to a clear agreement with regard to the reference frames encoded in scene regions. One study suggested that PPA and OPA utilize eye-centered coding while RSC does not show any preference (Ward et al., 2010). Another study found that PPA partially adapted to views of the same scene during saccadic eye movements, but this adaptation did not differ when the scene-snapshot moved with the saccade or not (Golomb et al., 2011). This study hence suggested that scene encoding is primarily eye-centered, yet with limited world-centered contribution. A problem with these saccade studies was that due to methodological considerations, saccades were executed on a blank screen, hence preventing true spatial updating of the scene during the saccade.

Here, we re-examine the important question whether scene regions encode visual input in eye-centered or world-centered coordinates using continuous motion and visual pursuit instead of using snapshots and saccades. This has multiple advantages. Pursuit can be carried out on the scenes rather than on intermittent blank screens, and updating occurs continuously rather than only a few times per stimulus block or between blocks. Updating-related signal can hence be expected to be considerably higher as it is generated continuously throughout each block. PPA and OPA have previously been shown to be motion responsive (Korkmaz Hacialihafiz & Bartels, 2015), as well as neurons of the parahippocampal gyrus in monkey (Sato & Nakamura, 2003). However, no study differentiated between retinal and world centered reference frames of these motion responses. Even among dorsal motion-selective regions only a subset encodes motion primarily in world-centered reference frame, such as V3A and V6 (Fischer et al., 2012a).

We designed stimuli according to a well-controlled 2 x 2 factorial design with the factors objective motion (on/off), pursuit (on/off) that allows separating eye- from world-centered motion encoding (Fischer et al., 2012a). A third factor of scene content (gray scale landscape and cityscape scenes or Fourier their scrambles) (Korkmaz Hacialihafiz & Bartels, 2015) was added as a third factor to examine content-dependence of reference frame preference. To balance attention across all conditions, participants performed a central character-matching task at all times. Importantly, effects of eye movements cancelled out for the important contrasts, since conditions including pursuit eye movements were present in both sides of the equation. Scene responsive regions were identified using an independent localizer scan. We performed GLM whole-brain analyses as well as region of interest (ROI) analyses. Results show that the key scene responsive regions, PPA, OPA and RSC, can be completely dissociated on the basis of responsiveness to pursuit or their preference for the distinct reference frames.

3.3. Materials and Methods

Participants

17 healthy participants with normal or corrected-to normal vision (9 female, 1 left handed, age between 20 and 36, mean = 27.8 years) took part in this study after giving written informed consent. The study was approved by the ethics committee of the University Hospital of Tübingen.

Experimental Setup

This study consisted of one main experiment, one functional localizer and one structural scan. The functional localizer aimed to identify scene- responsive regions: PPA, RSC, and OPA.

Visual stimuli were gamma corrected and back-projected onto a screen via a projector outside the scanner room. The screen was viewed via an angled mirror and subtended a visual field of 19 x 15 visual degrees.

The experiment was programmed using Psychtoolbox-3 (Brainard 1997, Kleiner, Brainard et al. 2007) on MATLAB 7.10.0 (The Mathworks, Natick, MA, 2010) and was presented using a windows PC.

Main Experiment

Figure 1 illustrates the eight conditions of the main experiment. It was a 2 x 2 x 2 factorial design with the factors objective motion (on/off), pursuit (on/off) and scene (on/off). The design of the first two factors (on-screen motion and pursuit) was identical to that described in a prior study (Fischer et al., 2012a). Here, a background image (described by the third factor: either a natural scene or Fourier-scramble thereof) was either stationary or moved on a horizontal trajectory left- and right-wards. A horizontal trajectory was chosen as this corresponds to common eye- and head-rotations in natural situations and correspondingly predominates in feature movies (Bartels, Zeki, & Logothetis, 2008). The velocity followed a sine function with a cycle of 3 s, and each block contained 4 cycles, hence lasting 12 s. The velocity varied between 0 and 3.08 deg/s, yielding a mean velocity of 2.53 deg/s. The motion extended 1.98 visual degrees in each direction. The starting direction of motion of each block was pseudorandomized and counterbalanced across runs.

The same parameters applied to the fixation disc, that could also either be stationary or moving. In conditions when both, the fixation disc and the background moved, the motion of both was locked, such that there was no relative motion.

Background images

We picked 32 images of outdoor scenes (both landscapes and cityscapes) and converted them to gray scale. These gray scale images and their phase-scrambled versions composed the stimuli. In order to prevent unbalanced stimuli due to horizontal inequalities in the images, half of the images were left right flipped duplicates of the other half. All images were adjusted so that they had equal contrast and luminance (luminance: 144 cd/m2, contrast: 32.4 cd/m2 root-mean-square (RMS) contrast, resulting in an average Michelson contrast of 0.9004 ± 0.0925). Images were larger than the screen to allow their displacement while filling the screen at all times.

Phase-scrambled versions of the images were created using Fourier transformation and reconstruction with random phases. This resulted in preservation of low-level features of the image such as luminance, contrast and spatial frequencies while removing scene content. The same images were used in one of our prior studies (Korkmaz Hacialihafiz & Bartels, 2015).

Paradigm

The stimuli were presented in a block design. Each consisted of 33 stimulus blocks. Each block lasted 12 seconds. The eight conditions were pseudorandomized and back matched so that each condition was preceded by all conditions with equal frequency across two runs. Each participant took part in 4 runs in total. Moreover, one additional block was added to the beginning of each run in order to ensure full counterbalancing for the first block. Each condition was presented 4 times in total in each run. Background images were randomly chosen for each block and only one image was used for an entire block.

Each run started with 6.9 seconds of gray screen with fixation and ended with 10 seconds of gray screen with fixation (luminance of gray screens: 144 cd/m2) leading to a total duration of 412.9 seconds. During the experiment, there was a gray fixation disk (width: 0.74 deg, luminance: 282 cd/m2) present at all times on the center of the screen, with the fixation task described below.

Fixation Task

In order to ensure fixation and balanced attention, participants were required to perform a 1-back character-matching task. The task was as follows: on the fixation disk a randomly chosen alphabetical character (a-z) was presented for 1 second each with 83 ms blank intervals in between. At random intervals, every 3 to 8 presentations, a repetition of the presented character occurred, which participants were required to report via button press. The timings of button presses were recorded and included in the GLM analyses as a regressor of no interest.

Functional Localizer and ROI definition

We used a separate localizer experiment in order to localize scene selective regions PPA (Aguirre et al., 1998; Epstein & Kanwisher, 1998), RSC (Maguire, 2001) and OPA/TOS (Dilks et al., 2013; Grill-Spector, 2003; Hasson et al., 2003; Nakamura et al., 2000) for every subject in each hemisphere. The localizer consisted of 3 conditions; gray scale images of scenes (which were different than the ones used in the main experiment), faces and phase-scrambled versions of these scenes and faces. The localizer consisted of one run, and the stimuli were shown in a block design. In each block 5 different images from the same category were shown for 3 s each, yielding a block length of 15 s and 1 s gray screen following each block. The experiment started with 6.9 s of gray screen with fixation, had 27 blocks in total (9 times x 3 conditions) and ended with 10 s of gray screen with fixation. There was a fixation cross and participants were asked to fixate at this central fixation cross at all times. All three ROIs were identified using the contrast (scenes > faces), using the MarsBaR toolbox (Brett, Anton, Valabregue, & Poline, 2002). In order to keep the ROIs similar in size across participants, we used an individual p-value for each participant and ROI when defining the ROIs (Fox, Iaria, & Barton, 2009; Murray & Wojciulik, 2004). Out of a total of 34 hemispheres, PPA was defined in 34 hemispheres, OPA in 28, and RSC in 33 hemispheres.

Data Acquisition

T2* weighted functional images were acquired using a 64-channel phased-array head coil in a Siemens Magnetom PRISMA 3T scanner (Siemens, Erlangen, Germany) with the following parameters: voxel size 3 x 3 x 3 mm^3, TR: 2.3 seconds, TE: 35

milliseconds, flip angle was 79°, 32 slices acquired in ascending order. In order to allow T1 equilibration, the first 3 volumes of data were discarded. Anatomical images were collected for each participant using T1-weighted images (1 x 1 x 1 mm^3 resolution).

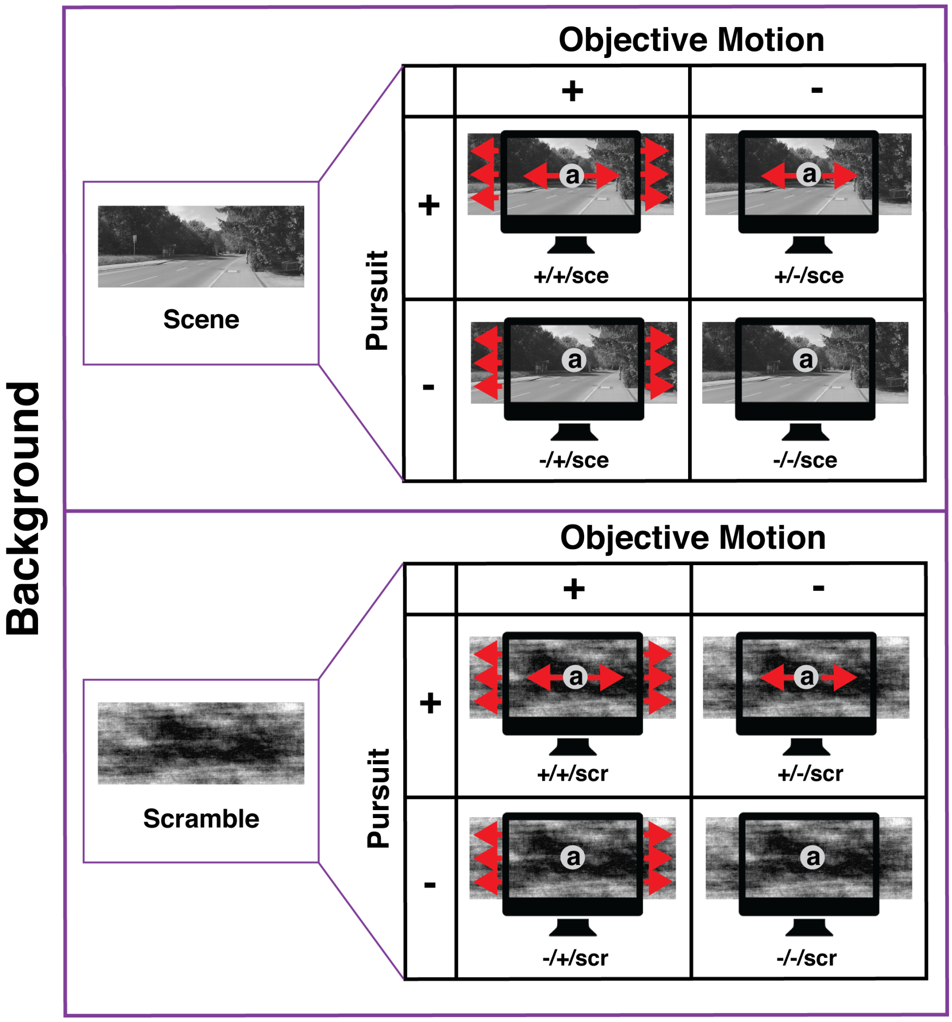


Figure 1. Conditions presented in the main experiment. A 2 x 2 x 2 factorial design with factors objective motion (on, off), pursuit (on, off) and scene (scene, scrambled) resulted in eight conditions shown above. In the "±/±" notation, the first position refers to pursuit, the second to objective motion. "+" refers to presence and "-" to absence. Objective motion was horizontal motion of the background image (scenes or scrambled images) and pursuit was horizontal motion of the fixation disk. There was a one-back character-matching task inside the fixation disk (shown larger for illustration).

FMRI Data Preprocessing and Statistical Analysis

Preprocessing was performed using SPM5 (www.fil.ion.ucl.ac.uk/spm) with the following steps: slice-time correction, realignment for motion correction, coregistration of the structural image to the mean functional image, normalization of the data to the

SPM template in Montreal neurological institute (MNI) space, and spatial smoothing with 6 mm full-width at half maximum Gaussian kernel for single participants and 12 mm for group level analyses, respectively.

Data of each participant were analyzed separately using the GLM (general linear model) in SPM5. We modeled each of the eight conditions as boxcars convolved by the canonical hemodynamic response function (hrf). Button presses were modelled as events. A total of seven regressors of no interest were included, consisting of six motion realignment regressors and one additional regressor for global signal variance (Desjardins, Kiehl, & Liddle, 2001; Van Dijk et al., 2010). The global signal variance regressor was orthogonalized to the conditions of interest. The data were high pass filtered using a cut-off value of 128 s. Beta images from the first level GLMs of each participant were used for group level analyses.

The ROI analyses were done by first extracting mean beta values for each ROI for each condition of each participant. Beta values were normalized in the range between 0 and 1 for each ROI and participant separately as follows: for each ROI of a given participant there were 32 mean beta values resulting from 4 runs and 8 conditions. The minimum of all 32 beta values was subtracted, and then all 32 beta values were divided by their maximal value. After this, an average beta value was calculated for each condition. This normalization ensured that all ROIs were comparable in mean and range of beta values. Repeated measures ANOVAs, as well as paired t-tests were conducted in order to analyze the effects of conditions using statistical analysis software IBM SPSS Statistics version 22.0. Greenhouse-Geisser correction was utilized in case of violation of sphericity as determined by Mauchly's sphericity test.

The contrasts used in the analysis were defined as follows: "objective motion" was defined using all conditions with moving background versus all conditions with still background, i.e. left vs. right column in Figure 1. "Retinal motion" was defined by all conditions where retinal input changed versus all conditions where retinal input remained the same, i.e. in Figure 1 (+/-) + (-/+) versus (+/+) + (-/-). Objective versus retinal motion was hence equivalent to (+/+) versus (+/-) (Fischer et al., 2012a). The contrast "Scene" was defined using all conditions with scene vs. all with scramble. "Pursuit" was defined as all conditions with pursuit versus all conditions with eyes fixed, i.e. top row vs. bottom row in figure 1. Note that all contrasts except for "pursuit" were balanced in terms of eye movements. The contrast “pursuit” contained both, eye-movement-related effects as well as effects related to peripheral visual stimulation induced by pursuit, but can nevertheless serve to functionally distinguish ROI properties.

Eye Tracking

Eye tracking of participants during the main experiment was done using an infrared camera based eye tracker system (Eye-Trac 6; Applied Science Laboratories). Preprocessing included blink removal, and smoothing of x and y positions using a running average window of 200 milliseconds. We calculated the fixation accuracy by the root mean square error of actual eye position relative to the fixation disk for each condition across participants and runs. We then used repeated measures ANOVAs and t-tests to examine condition-related effects.

3.4. Results

We investigated responses of independently localized scene-selective regions PPA, RSC and OPA to visual motion in world-centered (objective motion) and eye-centered (retinal motion) reference frames. The experiment was a 2 x 2 x 2 design with the factors pursuit (on, off) and visual motion (on, off), carried out using two types of content (natural scenes, their Fourier scrambles) as a third factor (see Figure 1). Figures 2A and 2B show raw and normalized mean beta responses to all conditions and each ROI, respectively. The contrasts of interest, such as for objective motion, retinal motion, or pursuit, were calculated using normalized responses that allowed for inter-region comparisons.

ROI Analyses

We examined the effects of objective motion, pursuit, scene and their interactions using repeated measures ANOVAs in PPA, RSC and OPA.

In order to test for lateralization effects, we performed a 3 x 2 x 8 repeated measures ANOVA with the factors ROI, hemisphere and condition. There was no main effect of hemisphere (F (1,12) = 0.027, p = 0.873), nor any interaction of hemisphere with any factor (hemisphere and ROI: F (2,24) = 0.002, p = 1.00, hemisphere and condition: F (3.488,41.855) = 0.25, p = 0.89 and hemisphere, ROI, condition: F (14,168) = 1.618, p = 0.079). For the remaining analyses we hence pooled data from both hemispheres for each ROI.

First, we conducted within-ROI analyses utilizing separate repeated measures ANOVAs with the factors scene, objective motion and pursuit for each ROI. We primarily focused on objective motion and retinal motion (i.e. the interaction between objective motion and pursuit), and on scene, and their interactions. These contrasts were completely balanced in terms of pursuit-related effects. We also report pursuit, but we note that it includes combined effects related to control of eye-movements, and to peripheral motion beyond the controlled visual screen induced by eye-movement, respectively.

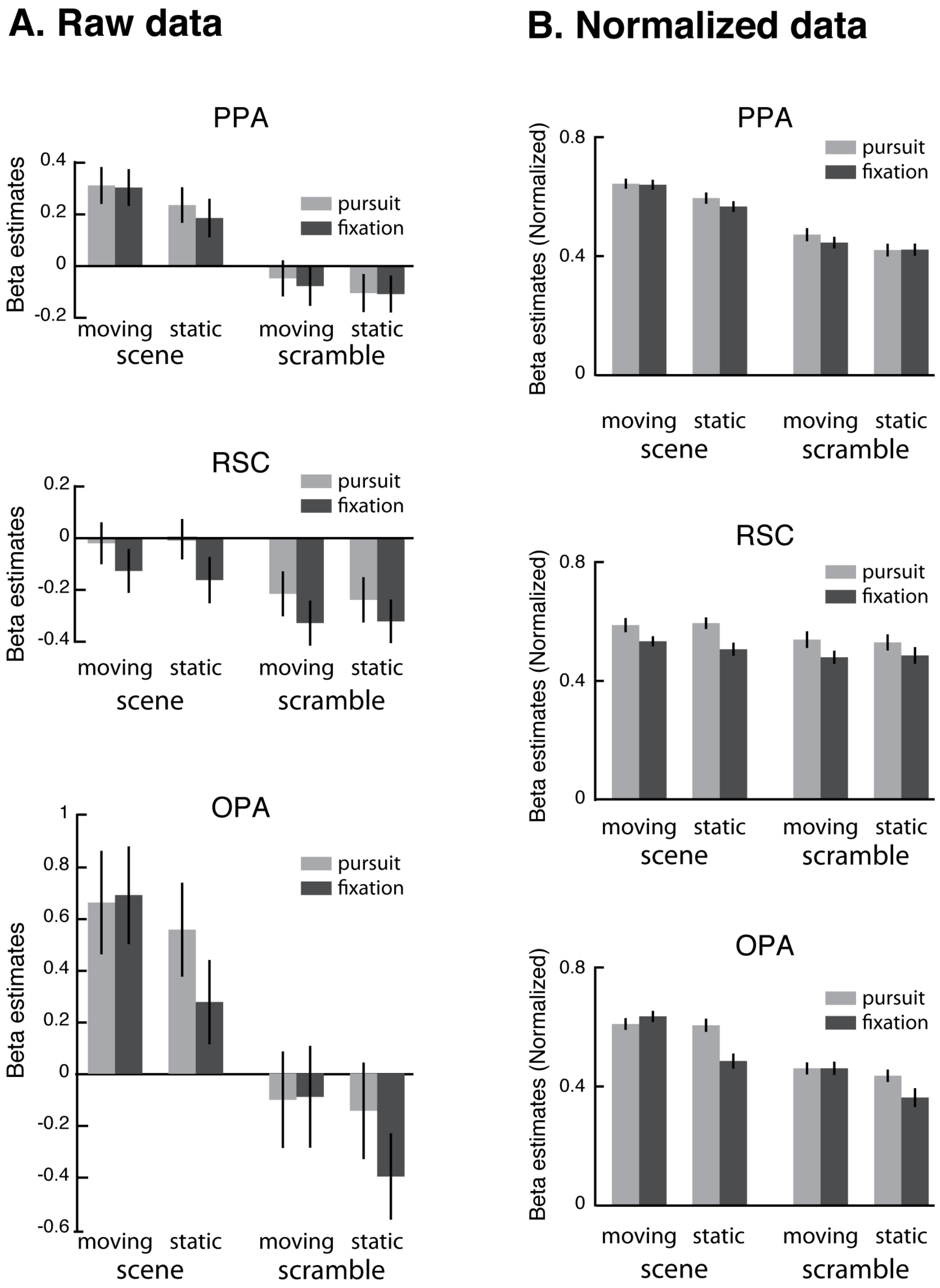


Figure 2. Responses to all eight conditions across ROIs. (A) Raw beta estimates in PPA, RSC and OPA. (B) Normalized beta estimates (see methods). 'Moving/static' refers to background motion whereas 'pursuit/fixation' notation refers to eye movement. Plots show mean ± standard error of mean (SEM).

Motion in eye- and world centered reference frames

Figure 3 shows main effects related to motion in eye- and world-reference frames (i.e. retinal and objective motion) and to pursuit for each ROI. Each scene region had a distinctly different signature in terms of its motion response.

PPA responded robustly to objective motion (F (1,33) = 34.56, p = 1 * 10^{-6}), but not at all to retinal motion (F (1,33) = 0.00047, p = 0.98), nor to pursuit (F (1,33) = 3.19, p = 0.083). RSC did not respond to objective motion (F (1,32) = 0.32, p = 0.57), or to retinal motion (F (1,32) = 0.41, p = 0.53), but it responded robustly to pursuit (F (1,32) = 17.58, p = 0.0002). OPA responded to both, objective motion (F (1,27) = 24.78, p = 0.00003), and retinal motion (F (1,27) = 18.61, p = 0.00019), as well as to pursuit (F (1,27) = 7.51, p = 0.011).

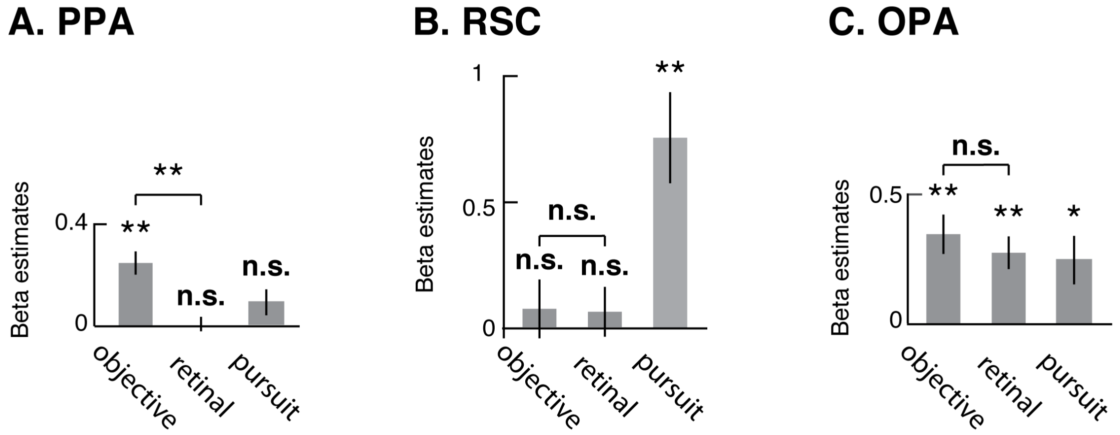


Figure 3. Responses of PPA, RSC, and OPA to objective motion, retinal motion, scene and pursuit. (A) PPA. (B) RSC (C) OPA. Note that data of each ROI was normalized. ** : $p < 0.001$, * : $p < 0.05$. Plots show mean ± standard error of mean (SEM).

Interactions between scene content and motion

Next we tested whether motion responses were differentially driven by scenes and scramble content. This is important, as in principle some of the scene region's motion responses could be accounted for by the increase in exposed scene-content by lateral motion, even though this was kept deliberately small in our stimuli, with 1.98 visual degrees in each direction out of 19 degrees stimulus width. The most important test was however whether objective motion responses persisted also for Fourier-scramble stimuli that lacked any scene content.

Figure 4A shows that all three ROIs responded robustly to scenes. Figure 4B shows that PPA responses were significant for objective motion, both during scenes (t (33) = 6.33, p = 1.1 * 10^{-6}) and scrambled (t (33) = 3.51, p = 0.003), and that it had a significant interaction between objective motion and scene (t (33) = 2.57, p = 0.015). In RSC, neither scenes (t (32) = 0.96, p = 0.34) nor scrambled images (t (32) = 0.06, p = 0.95) resulted in significant responses during objective motion and there were no significant interactions (t (32) = 0.66, p = 0.51). In OPA, objective motion responses were significant both during scenes (t (27) = 4.15, p = 0.0009) and scrambled images (t (27) = 4.14, p = 0.0006). Figure 4C shows that PPA did not have retinal motion responses during scenes (t (33) = 1.3, p = 0.2) or scrambled images (t (33) = -1.47, p = 0.15) and there was no significant interaction (t (33) = 1.61, p = 0.12). In RSC, retinal motion responses were similar to PPA's (during scenes: t (32) = 1.54, p = 0.13, during scrambled images: t (32) = -0.496, p = 0.62). In OPA, retinal motion responses were significant both during scenes (t (27) =3.39, p = 0.006) and scrambled images (t (27) = 3.46, p = 0.006), but there was no significant interaction (t (27) = 0.864, p = 0.395). P-values are Bonferroni-Holm corrected for 3 comparisons for each ROI.

Next, we tested which regions differed in the contrast ("objective" versus "retinal" motion). We analysed this contrast for scene and scrambled images separately. Figure 4D shows that PPA was the only region with a significant preference for objective motion both during scenes (t (33) = 3.79, p = 0.0025, Bonferroni corrected) and scramble (t (33) = 3.52, p = 0.0052, Bonferroni corrected). Neither RSC nor OPA had a preference for objective versus retinal motion. No region differed between scenes and scramble (PPA: t (33) = 0.11, p = 3.66, RSC: t (32) = -0.67, p = 2.03, OPA: t (27) = -0.31, p = 3.05).

Comparisons between ROIs

Next, we tested for differences between ROIs. We performed a 3 x 2 x 2 x 2 repeated measures ANOVA with the factors ROI (PPA, RSC, OPA), objective motion (on/off), pursuit (on/off) and scene (scene/ scrambled). The factor ROI had significant interactions with each of the remaining factors (objective motion: F (1.51, 40.71 = 13.53, p = 0.00013; retinal motion (i.e. objective motion and pursuit): F (1.62,43.79) =8.60, p = 0.001; scene: F (2,54) = 24.64, p = 2.5 * 10^{-8}; pursuit: F (1.54,41.47) = 4.37, p = 0.027). There were no triple interactions. Direct comparisons between ROIs using

Bonferroni corrected t-tests showed highly significant dissociations between ROIs. Figure 5 shows that PPA and OPA differed in their objective motion response from RSC (PPA vs. RSC (t (32) = 4.78, p = 0.00011); RSC vs. OPA (t (27) = -3.85, p = 0.002). OPA differed in its retinal motion response from PPA (t (27) = 3.93, p = 0.0016) and RSC (t (27) = 2.92, p = 0.021). Only RSC and OPA responded to pursuit, with the former differing significantly from PPA and RSC (t (32) = -3.56, p = 0.0035). All p-values are Bonferroni corrected for 3 comparisons.

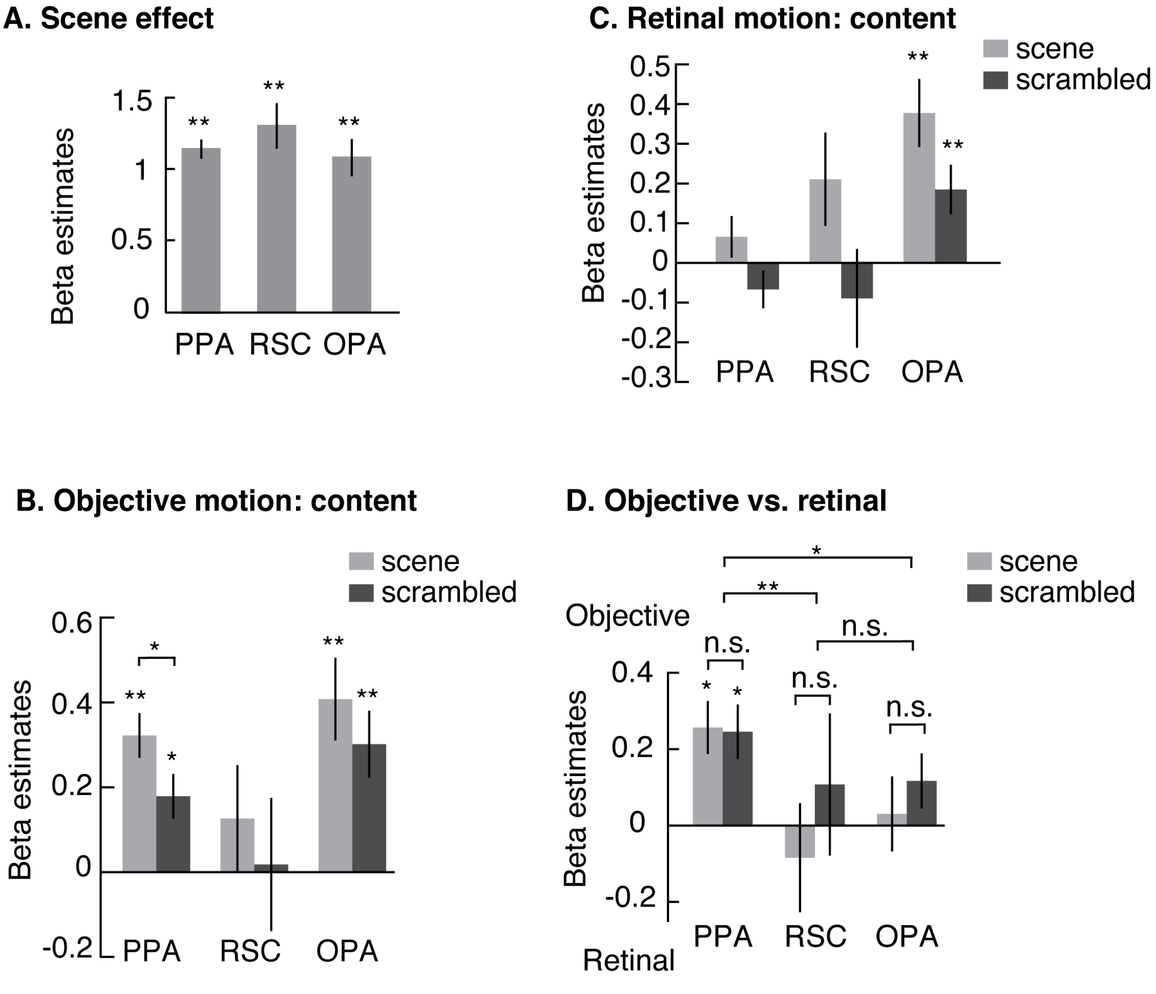
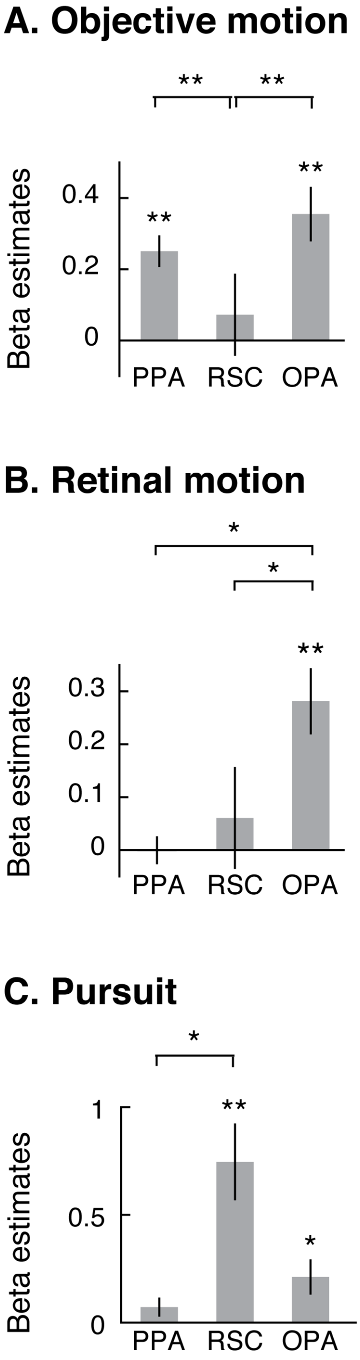


Figure 4. Effect of scene content on ROI responses. (A) Main effect of content, scene versus scrambled across all conditions (B) Responses to objective motion for scene and scrambled content. The figure shows responses to objective motion, separately for scene and scrambled backgrounds. (C) Responses to retinal motion for scene and scrambled content. The figure shows responses to retinal motion, separately for scene and scrambled backgrounds. (D) Objective versus retinal motion preferences across ROIs. Objective versus retinal motion responses, shown for scenes and scrambled backgrounds separately. Brackets across ROIs refer to tests calculated for the average response across scene and scramble * : $p < 0.05$, ** : $p < 0.001$, Bonferroni corrected. Plots show mean ± standard error of mean (SEM).

Figure 4C shows that PPA had a higher preference to objective versus retinal motion than both, OPA (t (27) = 2.964, p = 0.019, Bonferroni corrected) and RSC (t (32) = 5.012, p = $0.57 * 10^{-4}$, Bonferroni corrected).


A. Objective motion

B. Retinal motion

C. Pursuit

Figure 5. Responses of PPA, RSC, and OPA to objective motion, retinal motion and pursuit. (A) Main effect of objective motion. (B) Main effect of retinal motion. Note that this is the interaction of objective motion and pursuit. (C) Main effect of pursuit. Note that data of each ROI was normalized. ** : $p < 0.001$, * : $p < 0.05$, Bonferroni corrected. Plots show mean ± standard error of mean (SEM).

In summary, objective motion responses were unique to PPA and OPA, and retinal motion responses were unique to OPA. PPA was the only ROI that showed significant objective motion preference over retinal motion. RSC stood out with strong pursuit responses, and PPA was the only region lacking them (Figure 5C).

With regards to content-motion interactions, PPA was the only region with content-motion interaction, with higher activation to motion during scenes than scramble. This interaction was not present in the comparison between objective motion with retinal motion. Importantly, the objective motion responses of PPA and OPA cannot be accounted for only by moving scene-content, as they were highly significant also for moving Fourier scramble.

Adaptation Index

In a final analysis, we provide a comparison point to a related study. Golomb and colleagues used a similar paradigm yet with crucial differences. That study used saccades and still images rather than pursuit and continuous motion, respectively, and spatial updating (i.e. saccades) occurred in the absence of a stimulus (Golomb et al., 2011). All their results were quantified statistically in the form of adaptation indices, so we applied the same formula to our data to allow for a comparison. We did so for scene conditions only as they did not have scramble conditions, and we used the (+/+/sce) as most responsive condition as baseline. Using range-normalized data, we calculated the adaptation index for each ROI, each subject and each condition with scene background as follows:

Adaptation index= ((+/+) - condition) / ((+/+) + condition)

Figure 6 shows AIs for all ROIs and conditions. For PPA, the (+/-) condition (t (33) = 3.97, p = 0.003) and (-/-) condition (t (33) = 4.83, p = 0.0003) showed a significant adaptation whereas no adaptation was found for (-/+) condition (t (33) = 0.24, p = 0.8). All p-values are Bonferroni-Holm corrected for 9 tests.

Additionally, we also calculated the significance of adaptation indices for RSC and OPA. For RSC, there was a significant adaptation for (-/-) condition (t (31) = 3.91, p = 0.003), but there was no adaptation for (-/+) (t (31) = 2.56, p = 0.08) or (+/-) conditions (t (31) = -0.87, p = 0.39). OPA showed similar adaptation responses to RSC; there was a significant adaptation for (-/-) (t (21) = 3.08, p = 0.034) condition but no adaptation was found for (-/+) (t (21) = -1.52, p = 0.14) or (+/-) (t (21) = 0.26, p = 0.8) conditions.

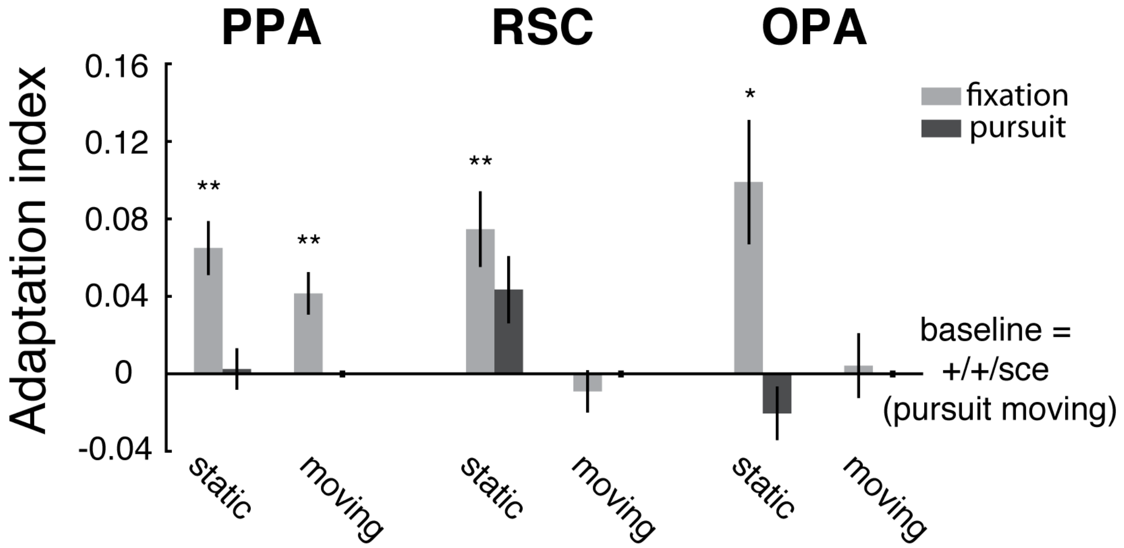


Figure 6. Adaptation indices for PPA, OPA and RSC. The baseline was calculated using (+/+) condition during scenes. Adaptation indices are only calculated during scenes. * : $p < 0.05$; ** : $p < 0.005$ (Bonferroni-Holm corrected). Plots show mean ± standard error of mean (SEM).

Whole brain Analyses

Beyond the ROI analyses, we performed additional random-effects analyses across all voxels of the brain in order to test whether other regions outside the selected ROIs responded to objective versus retinal motion contrast. Group level whole-brain analysis revealed that voxels overlapping with PPA showed activation at uncorrected levels ($p < 0.05$ uncorrected), as well as voxels near posterior occipital sulcus (POS), coinciding with motion sensitive region V3A (figure 7). Previously, V3A was shown to be sensitive to objective versus retinal motion contrast when random dots were used as stimuli (Fischer et al., 2012a). The activation overlapping with PPA has peak T-statistic values 3.26 and 1.90, respectively in right (30, -44, -6) and left PPA (-26, -48, -8).

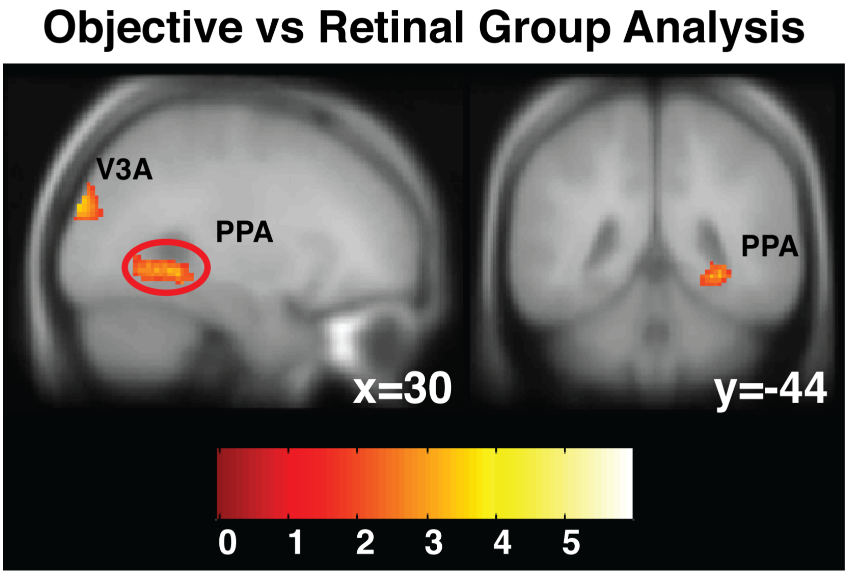


Figure 7. Objective versus retinal motion group data across all subjects. Voxels overlapping with PPA were activated at uncorrected thresholds, shown here inside the red circle in the sagittal slice in the left hemisphere and in the coronal slice in the right hemisphere. The dorsal activity near posterior occipital sulcus (POS) coincides with V3A (Fischer et al., 2012a). Peak t-statistic values 3.26 in right and 1.90 in left PPA (not shown here). For illustration only, the map is shown at $p < 0.05$ uncorrected.

Behavioral Data

Participants performed a character back-matching task intended to maintain vigilance and balance attention across conditions during the main experiment. The mean correct response rate was near ceiling with 0.85 ± 0.07 (mean ± S.D.). The mean response time was 0.57 ± 0.14 s (mean ± S.D.). We analyzed response times using a 3-way repeated measures ANOVA with the factors objective motion, pursuit and scene. There were only significant effects of pursuit (F (1,16) = 5.06, p = 0.039) but no other main effects (objective motion: F (1,16) = 0.005, p = 0.95; scene: F (1,16) = 0.39, p = 0.54), nor interactions (objective and pursuit (i.e., retinal motion): F (1,16) = 0.93, p = 0.35; objective and scene: F (1,16) = 0.00029, p = 0.99; pursuit and scene (1,16) = 1.58, p = 0.23; objective, pursuit and scene: F (1,16) = 0.98, p = 0.34). There was also no difference between objective and retinal motion (t (1,16) = -.81, p = 0.43).

Eye tracking Data

We analyzed eye-position error using the same 3-way ANOVA as used for the behavioral data. Like in behavior, the only significant effect was the main effect of pursuit (F(1,67) = 682.38, $p = 7.55 * 10^{-37}$), with no other main effects (objective motion: F(1,67) = 0.055, p = 0.82; scene: F(1,67) = 0.00014, p = 0.99), or interactions (retinal motion: F(1,67) = 0.28, p = 0.60; objective motion and scene: F(1,67) = 0.14, p = 0.71; pursuit and scene: F(1,67) = 2.80, p = 0.10; objective motion, pursuit and scene: F (1,67) = 0.16, p = 0.69). There was no objective versus retinal difference (t (67) = 0.13, p = 0.90).

3.5. Discussion

The aim of this study was to determine to which extent scene responsive regions PPA (Aguirre et al., 1998; Epstein & Kanwisher, 1998), RSC (Maguire, 2001) and OPA (Dilks et al., 2013; Grill-Spector, 2003; Hasson et al., 2003; Nakamura et al., 2000) encode visual motion cues in eye-centered (i.e. retinal) or world-centered (i.e. objective) reference frames. We used a previously established paradigm that combined pursuit eye-movements with on-screen motion (Fischer et al., 2012a) to determine retinal- and objective motion responses in independently defined scene-selective ROIs. To dissociate scene-dependent effects from low-level motion responses, we repeated all conditions using natural scenes and their Fourier scrambled versions, allowing examining motion-content interactions.

We found that each of the scene responsive regions had unique and dissociable response profiles. In particular, PPA was unique in its sensitivity to objective motion while not responding at all to retinal motion or to pursuit eye movements. While PPA's response to objective motion was enhanced by scene-content, its invariance to retinal motion and its response to objective motion were both preserved also for scrambled content. In contrast, OPA was the only scene region responsive to retinal motion (along with objective motion), and RSC responded strongly to pursuit but not to either motion type. Apart from providing a functional segregation of scene regions in terms of motion reference frames, the present results show that PPA is able to represent a stable environment during self-induced retinal motion.

Controversy ventral stream reference frame

The visual system uses different spatial representations in order to keep a stable vision of the world and these representations are reported in various brain areas. For instance, retinotopic representations are reported in early visual areas as well as higher level visual regions (Gardner, Merriam, Movshon, & Heeger, 2008; Golomb & Kanwisher, 2012) whereas world-centered spatial representations are thought to be encoded in dorsal regions, parietal regions and high level motion areas (Crespi et al., 2011; d'Avossa et al., 2007; Fischer et al., 2012a, 2012b; Galletti, Battaglini, & Fattori, 1993; Ilg et al., 2004). The spatial representation in ventral regions is rather under debate; both retinotopic (Amano, Wandell, & Dumoulin, 2009; Larsson & Heeger, 2006) and spatiotopic (McKyton & Zohary, 2007) maps are reported in LOC (lateral occipital cortex) and previous research indicate both eye centered (Golomb et al., 2011; Ward et al., 2010) and world centered (MacEvoy & Epstein, 2007) spatial representation in scene selective parahippocampal cortex.

Choice of reference frame in scene regions

Previous studies demonstrated viewpoint dependency and specificity in scene representation of PPA in context of static scene views (Epstein et al., 2003; Epstein, Parker, & Feiler, 2007; MacEvoy & Epstein, 2007; Park & Chun, 2009; Park, Chun, & Johnson, 2010). The objective motion contrast used here is – conceptually – comparable to the kind of viewpoint changes of static scene views used in previous studies, since both involved a gradual, even if step-wise, lateral translation of the scene. It is worth noting that none of the previous studies mentioned involved motion or pursuit eye movements, since they were mostly carried out during eye fixation. The problem lies within the never changing position of eyes. When the eyes are fixated, the motion of the scenes in the background results in a combined objective and retinal motion. Therefore, it is not possible to differentiate whether the mechanism behind the mentioned PPA responses, which was described as viewpoint dependency or specificity, are retinal or world-centered viewpoint dependency. In their 2011 study, Golomb and her colleagues tried to address this problem and they found evidence for both reference frames, but concluded that the results in PPA are mainly driven by retinotopic responses (Golomb et al., 2011). Using an adaptation paradigm, they reported that PPA was adapting to successive views of the same scene when the participants did saccades across the scenes, which resulted in matching retinal input (Golomb et al., 2011). We calculated adaptation indices similar to theirs to directly compare our results. Since our experimental design did not include a novel scene condition (and it is not possible to have novel scenes condition with pursuit), we used +/+ scenes condition as baseline. This was feasible because this condition covered both types of motion (smooth pursuit eye movements and objective motion) and elicited the highest beta responses in PPA.

Similar to the results of Golomb et al (Golomb et al., 2011), our results also revealed that PPA showed significant adaptation when eyes moved over static scenes but not when eyes were fixated and scene was moved. Moreover, our results reporting PPA responses during fixation is also in agreement with their results in PPA during fixation. However, the main inconsistency between our and their results arises from

PPA adaptation responses during eye movements. In our experiment, pursuit during scene motion ((+/+) condition during scenes) resulted in highest beta responses and there was no adaptation whereas they reported significant adaptation during saccades on scrolling scenes. This difference between our results and theirs might be rising from the use of saccades in their study versus the use of smooth pursuit eye movements in our experiment. Although smooth pursuit and saccades are traditionally thought to have partially overlapping neural networks, it is known that these networks are not identical. For instance, smooth pursuit eye movements and saccades are processed in different subregions of frontal eye fields (Rosano et al., 2002). Moreover, pursuit related neurons are found in regions such as V5/MT, MST and VIP (Ilg, 2008; Thier & Ilg, 2005). V3A is also thought to have a role in encoding pursuit (Fischer et al., 2012a), while it does not have any role in encoding saccades. Another difference between saccades and pursuit eye movements is that saccadic suppression generates stable visual input by actively suppressing the vision, particularly via suppression of magnocellular pathway (Burr, Morrone, & Ross, 1994; Thiele, Henning, Kubischik, & Hoffmann, 2002). Thus, it is possible that saccadic eye movements and smooth pursuit eye movements engage different brain regions. Another main difference between the experimental design of present study and the study of Golomb et al (Golomb et al., 2011) is that in their experiment, the saccades were executed during blank screen, preventing remapping of the actual scene, whereas in our experiment smooth pursuit eye movements were carried out on scenes, allowing for direct continuous remapping. Lastly, it is also possible that using saccades and pursuit eye movements during the stimuli might be resulting in perceptually different inputs. Further studies are needed in order to directly compare the effect of eye movement types during dynamic scene perception.

In OPA and RSC, our results indicate adaptation only during (-/-) condition, and no adaptation was found for the other conditions. Golomb et al (Golomb et al., 2011) reported that they found only retinotopic adaptation in OPA and RSC, noting that the responses they observed in OPA and RSC were much noisier than the responses observed in PPA. It is interesting to note that in the present study, similar to their results, RSC did not adapt to overlapping scenes when eyes fixated whereas this condition was previously reported to result in significant adaptation in RSC (Park & Chun, 2009). As they discussed, this difference could be driven by different tasks used in our experiment and abovementioned previous studies(Golomb et al., 2011). OPA lacked adaptation during any type of motion. This fits well with previous view about OPA being a lower-tier scene responsive region.

Similarity to V3A

The objective motion preference of PPA was so pronounced, also with regard to the complete absence of retinal motion, that it bears similarity to that of dorsal motion areas V3A and V6 that have a similar response profile (Fischer et al., 2012a).

Content – motion interactions

Our finding that PPA responses showed an interaction between objective motion and scene content, in that it showed higher responses to motion during scenes than during scrambled images, extend our previous findings, which were carried out by moving scenes and scrambled images during fixation (Korkmaz Hacialihafiz & Bartels, 2015). One could argue that the objective motion response in PPA could be explained by the fact that objective motion conditions revealed more of the scene image than static conditions. This is however an unlikely explanation. First, the preference for objective motion was still significant in PPA in context of scrambled images lacking any scene content. Second, the *difference* contrast of (objective – retinal) was equally large for scene and for scramble, implying that content played no role whatsoever here. Third, the displacement of the background images was a mere 1.98 visual degrees in either direction, with a field of view of 19 degrees, hence only revealing marginally new scene content, in the periphery where it is hard to recognize. It is thus highly unlikely that motion responses in PPA can be explained by increased scene content only.

Responses in OPA

OPA was the only scene responsive region that showed significant retinal motion responses, and could be differentiated by this response profile from PPA and RSC, which lacked retinal motion responses entirely. OPA responded to both objective and retinal motion equally. These findings, and the fact that OPA lacks adaptation during both retinal and objective motion, are compatible with the previously proposed theory that OPA constitutes a comparably early level in the hierarchy amongst scene responsive regions due to its physical location and close proximity to dorsal regions (Dilks, Julian, Kubilius, Spelke, & Kanwisher, 2011; Dilks et al., 2013; MacEvoy & Epstein, 2007). This view was further supported by a number of studies reporting low-level features of OPA.

For instance, OPA has been shown to have lower visual field bias (Silson, Chan, Reynolds, Kravitz, & Baker, 2015) and its receptive field sizes were smaller compared to other scene responsive ROIs (MacEvoy & Epstein, 2007; Silson et al., 2015).

No pursuit in PPA

Another notable result is that PPA did not respond to pursuit at all, whereas OPA and RSC did. This shows an extreme degree of invariance to motion in the peripheral field of view invariantly induced by pursuit, and potentially powerful mechanisms cancelling eye-movement related jitter everywhere in its visual representations. Overall, this is compatible with the view of PPA representing scenes (and any visual signal) in a manner robustly independent of precise fixation points. This, together with PPA's unique representation of motion only in world- but not eye-centered coordinates would clearly facilitate perception of a continuous world from discontinuous views. Stable scene representations are also supported by the ability to extrapolate views through boundary extension, a property PPA shares with RSC (Park, Intraub, Yi, Widders, & Chun, 2007).

Responses in RSC

The lack of motion responses in RSC is comparable to previous studies reporting that RSC has viewpoint invariant properties (Park & Chun, 2009; Park et al., 2010) and that RSC shows little change in activity when the stimuli presented in different locations (other than central fixation point) (MacEvoy & Epstein, 2007; Ward et al., 2010). However, the pursuit response in RSC, especially when considered together with the lack of motion responses, is rather interesting. Previously, correlations between the activity of frontal eye fields (FEF) and Brodmann areas that constitute retrosplenial cortex were found during resting state (Hutchison et al., 2012). In primates, retrosplenial cortex is found to have projections to peripheral vision regions in MT and MST (Palmer & Rosa, 2006). Moreover, a recent fMRI study in humans showed that in comparison to other scene regions PPA and OPA, RSC showed more peripheral bias (Baldassano, Fei-Fei, & Beck, 2016) and RSC is located near the peripheral visual regions in V1 and V2 (Nasr et al., 2011). During pursuit condition, peripheral visual changes are not controlled in the present study. So, pursuit responses in RSC could have been driven by peripheral bias in RSC responses during scene representation. More studies are needed in order to understand the mechanisms behind the pursuit related responses in RSC.

3.6. Conclusion

In conclusion, our results reveal a novel dissociation between scene responsive regions PPA, OPA and RSC in their responses to real-world motion. In particular, we found that PPA could be differentiated from OPA and RSC in objective motion preference to retinal motion of PPA and no preference in OPA and RSC. OPA, given its location and probably mid-level position in scene processing hierarchy, was responsive to both objective and retinal motion. RSC, which is thought to be involved in higher-level functions such as spatial navigation and which was shown to have viewpoint independent responses, lacked motion responses but was activated by smooth pursuit eye movements. Altogether, these findings shed light into our understanding of how PPA, OPA and RSC responses are adjusted by real world motion. Present findings suggest a role for PPA in providing input for stable visual perception, by separating real-world motion from self-induced retinal motion, similar to the selectivity previously shown in V3A and partially in V6 (Fischer et al., 2012a).

Acknowledgements

This work was funded by the Centre for Integrative Neuroscience Tübingen through the German Excellence Initiative (EXC307) and by the Max Planck Society, Germany.

Conflict of Interest: The authors declare no competing financial interests.

3.7. References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An Area within Human Ventral Cortex Sensitive to "Building" Stimuli: Evidence and Implications. *Neuron, 21*(2), 373-383. doi: http://dx.doi.org/10.1016/S0896-6273(00)80546-2

Amano, K., Wandell, B. A., & Dumoulin, S. O. (2009). Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *Journal of neurophysiology, 102*(5), 2704-2718.

Baldassano, C., Fei-Fei, L., & Beck, D. M. (2016). Pinpointing the peripheral bias in neural scene-processing networks during natural viewing. *Journal of Vision, 16*(2), 9-9. doi: 10.1167/16.2.9

Bartels, A., Zeki, S., & Logothetis, N. K. (2008). Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. *Cerebral cortex, 18*, 705-717. doi: http://dx.doi.org/10.1093/cercor/bhm107

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using the MarsBar toolbox for SPM 99. *Neuroimage, 16*(2), S497.

Burr, D. C., Morrone, M. C., & Ross, J. (1994). Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature, 371*(6497), 511-513.

Crespi, S., Biagi, L., d'Avossa, G., Burr, D. C., Tosetti, M., & Morrone, M. C. (2011). Spatiotopic Coding of BOLD Signal in Human Visual Cortex Depends on Spatial Attention. *PloS one, 6*(7), e21661. doi: 10.1371/journal.pone.0021661

d'Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nature neuroscience, 10*(2), 249-255. doi: 10.1038/nn1824

Desjardins, A. E., Kiehl, K. A., & Liddle, P. F. (2001). Removal of confounding effects of global signal in functional MRI analyses. *Neuroimage, 13*(4), 751-758. doi: http://dx.doi.org/10.1006/nimg.2000.0719

Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 31*, 11305-11312. doi: http://dx.doi.org/10.1523/JNEUROSCI.1935-11.2011

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 33*, 1331-1336a. doi: http://dx.doi.org/10.1523/JNEUROSCI.4081-12.2013

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron, 37*(5), 865-876. doi: http://dx.doi.org/10.1016/s0896-6273(03)00117-x

Epstein, R., Higgins, J. S., & Thompson-Schill, S. L. (2005). Learning Places from Views: Variation in Scene Processing as a Function of Experience and Navigational Ability. *Journal of Cognitive Neuroscience, 17*(1), 73-83. doi: http://dx.doi.org/10.1162/0898929052879987

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598-601. doi: http://dx.doi.org/10.1038/33402

Epstein, R., Parker, W., & Feiler, A. (2007). Where Am I Now? Distinct Roles for Parahippocampal and Retrosplenial Cortices in Place Recognition. *The Journal of*

Neuroscience, 27(23), 6141-6149. doi: http://dx.doi.org/10.1523/jneurosci.0799-07.2007

Fischer, E., Bulthoff, H. H., Logothetis, N. K., & Bartels, A. (2012a). Human areas V3A and V6 compensate for self-induced planar visual motion. *Neuron, 73*(6), 1228-1240. doi: 10.1016/j.neuron.2012.01.022

Fischer, E., Bulthoff, H. H., Logothetis, N. K., & Bartels, A. (2012b). Visual motion responses in the posterior cingulate sulcus: a comparison to V5/MT and MST. *Cerebral cortex (New York, N.Y. : 1991), 22*, 865-876. doi: 10.1093/cercor/bhr154

Fox, C. J., Iaria, G., & Barton, J. J. (2009). Defining the face processing network: optimization of the functional localizer in fMRI. *Hum Brain Mapp, 30*(5), 1637-1651. doi: http://dx.doi.org/10.1002/hbm.20630

Galletti, C., Battaglini, P. P., & Fattori, P. (1990). 'Real-motion' cells in area V3A of macaque visual cortex. *Exp Brain Res, 82*(1), 67-76.

Galletti, C., Battaglini, P. P., & Fattori, P. (1993). Parietal neurons encoding spatial locations in craniotopic coordinates. *Exp Brain Res, 96*(2), 221-229.

Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *The Journal of Neuroscience, 28*(15), 3988-3999.

Golomb, J. D., Albrecht, A. R., Park, S., & Chun, M. M. (2011). Eye Movements Help Link Different Views in Scene-Selective Cortex. *Cerebral Cortex, 21*(9), 2094-2102. doi: http://dx.doi.org/10.1093/cercor/bhq292

Golomb, J. D., & Kanwisher, N. (2012). Higher Level Visual Cortex Represents Retinotopic, Not Spatiotopic, Object Location. *Cerebral Cortex, 22*(12), 2794-2810. doi: 10.1093/cercor/bhr357

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology, 13*(2), 159-166. doi: http://dx.doi.org/10.1016/s0959-4388(03)00040-0

Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-Scale Mirror-Symmetry Organization of Human Occipito-Temporal Object Areas. *Neuron, 37*(6), 1027-1041. doi: http://dx.doi.org/10.1016/S0896-6273(03)00144-2

Hutchison, R. M., Gallivan, J. P., Culham, J. C., Gati, J. S., Menon, R. S., & Everling, S. (2012). Functional connectivity of the frontal eye fields in humans and macaque monkeys investigated with resting-state fMRI. *Journal of neurophysiology, 107*(9), 2463-2474.

Ilg, U. J. (2008). The role of areas MT and MST in coding of visual motion underlying the execution of smooth pursuit. *Vision research, 48*(20), 2062-2069. doi: 10.1016/j.visres.2008.04.015

Ilg, U. J., Schumann, S., & Thier, P. (2004). Posterior parietal cortex neurons encode target motion in world-centered coordinates. *Neuron, 43*(1), 145-151. doi: 10.1016/j.neuron.2004.06.006

Korkmaz Hacialihafiz, D., & Bartels, A. (2015). Motion responses in scene-selective regions. *NeuroImage, 118*, 438-444. doi: http://dx.doi.org/10.1016/j.neuroimage.2015.06.031

Larsson, J., & Heeger, D. J. (2006). Two retinotopic visual areas in human lateral occipital cortex. *J Neurosci, 26*(51), 13128-13142. doi: 26/51/13128 [pii]

10.1523/JNEUROSCI.1657-06.2006

MacEvoy, S. P., & Epstein, R. A. (2007). Position Selectivity in Scene- and Object-Responsive Occipitotemporal Regions. *Journal of Neurophysiology, 98*(4), 2089-2098. doi: 10.1152/jn.00438.2007

Maguire, E. A. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavian Journal of Psychology, 42*(3), 225-238. doi: http://dx.doi.org/10.1111/1467-9450.00233

McKyton, A., & Zohary, E. (2007). Beyond retinotopic mapping: the spatial representation of objects in the human lateral occipital complex. *Cerebral Cortex, 17*(5), 1164-1172.

Murray, S. O., & Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral occipital complex. *Nat Neurosci, 7*(1), 70-74. doi: http://dx.doi.org/10.1038/nn1161

Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., . . . Zilles, K. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing. *Brain, 123*(9), 1903-1912. doi: http://dx.doi.org/10.1093/brain/123.9.1903

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. H. (2011). Scene-selective cortical regions in human and nonhuman primates. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 31*, 13771-13785. doi: 10.1523/JNEUROSCI.2792-11.2011

Palmer, S. M., & Rosa, M. G. P. (2006). A distinct anatomical network of cortical areas for analysis of motion in far peripheral vision. *European Journal of Neuroscience, 24*(8), 2389-2405.

Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage, 47*(4), 1747-1756. doi: http://dx.doi.org/10.1016/j.neuroimage.2009.04.058

Park, S., Chun, M. M., & Johnson, M. K. (2010). Refreshing and integrating visual scenes in scene-selective cortex. *Journal of cognitive neuroscience, 22*, 2813-2822. doi: http://dx.doi.org/10.1162/jocn.2009.21406

Park, S., Intraub, H., Yi, D.-J., Widders, D., & Chun, M. M. (2007). Beyond the edges of a view: boundary extension in human scene-selective visual cortex. *Neuron, 54*, 335-342. doi: 10.1016/j.neuron.2007.04.006

Rosano, C., Krisky, C. M., Welling, J. S., Eddy, W. F., Luna, B., Thulborn, K. R., & Sweeney, J. A. (2002). Pursuit and Saccadic Eye Movement Subregions in Human Frontal Eye Field: A High-resolution fMRI Investigation. *Cerebral Cortex, 12*(2), 107-115. doi: 10.1093/cercor/12.2.107

Sato, N., & Nakamura, K. (2003). Visual response properties of neurons in the parahippocampal cortex of monkeys. *J Neurophysiol, 90*(2), 876-886. doi: http://dx.doi.org/10.1152/jn.01089.2002

Silson, E. H., Chan, A. W.-Y., Reynolds, R. C., Kravitz, Dwight J., & Baker, C. I. (2015). A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *The Journal of Neuroscience, 35*(34), 11921-11935.

Sulpizio, V., Committeri, G., & Galati, G. (2014). Distributed cognitive maps reflecting real distances between places and views in the human brain. *Frontiers in Human Neuroscience, 8*, 716. doi: 10.3389/fnhum.2014.00716

Sulpizio, Valentina, Committeri, Giorgia, Lambrey, Simon, Berthoz, Alain, & Galati, Gaspare. (2013). Selective role of lingual/parahippocampal gyrus and retrosplenial complex in spatial memory across viewpoint changes relative to the environmental reference frame. *Behavioural Brain Research, 242*(0), 62-75. doi: http://dx.doi.org/10.1016/j.bbr.2012.12.031

Thiele, A., Henning, P., Kubischik, M., & Hoffmann, K. P. (2002). Neural mechanisms of saccadic suppression. *Science, 295*(5564), 2460-2462.

Thier, P., & Ilg, U. J. (2005). The neural basis of smooth-pursuit eye movements. *Curr Opin Neurobiol, 15*(6), 645-652.

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

Ward, E. J., MacEvoy, S., & Epstein, R. (2010). Eye-centered encoding of visual space in scene-selective regions. *Journal of Vision, 10*(14). doi: http://dx.doi.org/10.1167/10.14.6

Zhang, T., Heuer, H. W., & Britten, K. H. (2004). Parietal area VIP neuronal responses to heading stimuli are encoded in head-centered coordinates. *Neuron, 42*(6), 993-1001.

4. Motion regions are modulated by scene content

Authors: Didem Korkmaz Hacialihafiz and Andreas Bartels*

Vision and Cognition Lab, Centre of Integrative Neuroscience, University of Tübingen, Germany

*Corresponding author:

Andreas Bartels

Vision and Cognition Lab

Centre for Integrative Neuroscience,

University of Tübingen,

Otfried-Müller-Str. 25

72076 Tübingen

Germany

Email: andreas.bartels@tuebingen.mpg.de

Phone: +49 7071 2989168

4.1. Abstract

Creating a stable perception of the world during pursuit eye movements is one of the everyday roles of visual system. Some motion regions have been shown to differentiate between motion in the external world from that generated by eye movements. However, in most circumstances, perceptual stability is consistently related to content: the surrounding scene is typically stable. However, no prior study has examined to which extent motion responsive regions are modulated by scene content, and whether there is an interaction between content and motion response. In the present study we used a factorial design that has previously been shown to reveal regional involvement in integrating efference copies of eye-movements with retinal motion to mediate perceptual stability and encode real-world motion. We then added scene content as a third factor, which allowed us to examine to which extent real-motion, retinal motion, and static responses were modulated by meaningful scenes versus their Fourier scrambled counterpart. We found that motion responses in human motion responsive regions V3A, V6, V5+/MT+ and cingulate sulcus visual area (CSv) were all modulated by scene content. Depending on the region, these motion-content interactions differentially depended on whether motion was self-induced or not. V3A was the only motion responsive region that also showed responses to still scenes. Our results suggest that contrary to the two-pathway hypothesis, scene responses are not isolated to ventral regions, but also can be found in dorsal areas.

4.2. Introduction

The visual system encounters different types of motion in dynamic scenes every day and processes visual scenes. Prior studies on motion processing are mostly based on abstract stimuli, like gratings or random dot displays (Born & Bradley, 2005; Boussaoud, Ungerleider, & Desimone, 1990; Erickson & Thier, 1991; Galletti & Fattori, 2003; Goossens, Dukelow, Menon, Vilis, & van den Berg, 2006; Gu, DeAngelis, & Angelaki, 2007; Huk, Dougherty, & Heeger, 2002; Maciokas & Britten, 2010; Smith, Wall, Williams, & Singh, 2006). However, most of these previous studies on motion processing have not used natural scenes, except for one that differentiated self-motion and object motion during movie viewing (Bartels, Zeki, & Logothetis, 2008). Hence, little is known how scene-content influences motion processing and whether these motion regions' responses are modulated by natural scene content.

Compared to V5+/MT+, which is a well-studied, low-level region in the motion processing hierarchy (Dubner & Zeki, 1971; Zeki et al., 1991), higher-level motion responsive regions such as V6, V3A, or CSv are involved in processing of more complex motion, for instance self-induced visual motion (Fischer, Bulthoff, Logothetis, & Bartels, 2011), integration of self motion cues with vestibular signals (Chowdhury, Takahashi, DeAngelis, & Angelaki, 2009; Gu et al., 2007), or full-field flow compatible with ego-motion (Arnoldussen, Goossens, & van den Berg, 2011; Goossens et al., 2006).

Previous studies showed content related responses in motion processing regions. For instance, human V5/MT has object responses and shows an interaction between object content and motion (Kourtzi, Bulthoff, Erb, & Grodd, 2002; Kourtzi & Kanwisher, 2000). V5/MT, as well as another motion region V3A, were shown to have object selective, size dependent and viewpoint dependent responses (Konen & Kastner, 2008). Further, V3A is shape sensitive (Denys et al., 2004; Grill Spector, Kushnir, Edelman, Itzchak, & Malach, 1998) and involved in form processing (Schira, Fahle, Donner, Kraft, & Brandt, 2004). Importantly, V3A was shown to have a role in scene segmentation (Scholte, Jolij, Fahrenfort, & Lamme, 2008). Additionally, another higher-level motion area V6, that is neighbouring V3A, analyses form and movement in visual field (Galletti et al., 2001). Despite the number of studies pointing out responses related to shape, form or object processing, the effect of scene content on motion regions is not truly known yet.

In this fMRI study, we were interested in whether motion responsive regions modulated by scene content and if they are, how their motion responses depend on scene content. To investigate these questions, we designed stimuli according to a previously established 2 x 2 factorial design with the factors objective motion (on/off) and pursuit (on/off) and this design led us to distinguish objective 'real' motion from retinal motion during smooth pursuit eye movements (Fischer, Bulthoff, Logothetis, & Bartels, 2012). In addition, we also added another factor for scene content (gray scale landscape and cityscape scenes or Fourier scrambled versions of these scenes). To balance attention across all conditions, participants performed a central character-matching task at all times. We performed GLM whole-brain analyses as well as region of interest (ROI) analyses. Motion responsive regions were identified using a dedicated localizer scan.

We found that all motion responsive regions tested in this study, namely V5+/MT+, V3A, V6 and CSv, indeed showed scene responses in context of motion. More interestingly, only V3A was responsive to still scenes. The responses in V5+/MT+ were modulated by scene content during both objective and retinal motion, whereas in V3A and V6, the responses were only modulated by scene content during retinal motion. In CSv, although significant responses were present for scenes, we did not observe any interaction between motion and scene content. These results show the importance of naturalistic stimuli use in understanding the visual system and its adaptation to everyday natural scenes.

4.3. Materials and Methods

Participants

17 healthy participants with normal or corrected-to normal vision (9 female, 1 left-handed, between the age of 20 and 36) gave written informed consent before participating in this study. The study was approved by ethics committee of the University Hospital of Tübingen. All participants were given instructions about the experiment and the task before going into the scanner.

Experimental Setup

This study consisted of one main experiment, one functional localizer for identifying motion regions V5+/MT+, V3A, V6 and CSv and one structural scan.

The gamma corrected visual stimuli was back-projected onto a screen via a projector outside the scanner room. The visual field of the screen was 19 x 15 visual degrees.

The main experiment was programmed using Psychtoolbox-3 (Brainard 1997, Kleiner, Brainard et al. 2007) whereas the functional localizer experiment that is used for localizing motion regions was prepared using Cogent Graphics v.1.29 developed by John Romaya at the Wellcome Department of Imaging Neuroscience (http://www.vislab.ucl.ac.uk/cogent.php). All stimuli was then presented using a windows PC and MATLAB 7.10.0 (The Mathworks, Natick, MA, 2010) (MATLAB, 2010).

Main Experiment

The main experiment consisted of eight conditions forming a 2 x 2 x 2 factorial design with the factors objective motion (on/off), pursuit (on/off) and scene (on/off), resulting in the eight conditions (Figure 1). The first two factors were described in a previous study (Fischer et al., 2012).

We picked 32 images of outdoor scenes (both landscape and cityscape) and converted them to grayscale, identical to the images used in one of our previous studies (Korkmaz Hacialihafiz & Bartels, 2015). These grayscale images and their phase-scrambled versions composed the stimuli. In order to balance horizontal inequalities in the images, half of the images were left right flipped duplicates of the other half of the images. All images were adjusted so that they had equal contrast and luminance

(luminance: 144 cd/m2, contrast: 32.4 cd/m2 root-mean-square (RMS) contrast, resulting in an average Michelson contrast of 0.9004 ± 0.0925). We used images that were large enough to give a feeling of moving across the screen.

In order to construct phase-scrambled versions of the images, Fourier transformation was applied and images were reconstructed with random phases. This resulted in preservation of low-level features of the image such as luminance, contrast and spatial frequencies while removing scene content.

The stimuli were presented in a block design. Each run consisted of 33 blocks per run. Each block lasted 12 seconds. The eight conditions were pseudorandomized and back matched so that each condition was preceded by all the other conditions equally. The sequence that allowed this back matching was then divided into two runs. We did this twice in order to obtain 4 runs in total. Each participant took part in 4 runs in total. This way, each condition preceded by each condition in equal frequency across two runs. Moreover, one additional block was added to the beginning of each run in order to initially counterbalance the first block. The images were randomly chosen for each block and only one image was used for an entire block. Stimuli followed a sine trajectory, extending across 4 cycles per block, in order to have a smooth horizontal motion. The velocity varied between 0 and 3.08 deg/s, yielding a mean velocity of 2.53 deg/s and the motion extended up to 1.98 visual degrees in each direction. The starting direction of motion was pseudorandomized and counterbalanced across runs. Each run started with 6.9 seconds of gray screen with fixation and ended with 10 seconds of gray screen with fixation (luminance of gray screen(s): 144 cd/m2). Each run lasted a total of 412.9 seconds. During the experiment, there was a gray fixation disk (width: 0.74 deg, luminance: 282 cd/m2) present at all times on the center of the screen, with the fixation task described below.

Fixation Task

There was a 1-back character- matching task at all times in both main experiment and localizer experiment to provide fixation and balanced attention of participants. On the fixation disk, a randomly chosen alphabetical character (a-z) was presented for 1 second each with 83 ms blank intervals in between. Every 3 to 8 presentations, a repetition of the presented character occurred, where participants were required to press a button once they see the repetition. The timings of button presses were recorded and used in GLM analyses as a regressor of no interest.

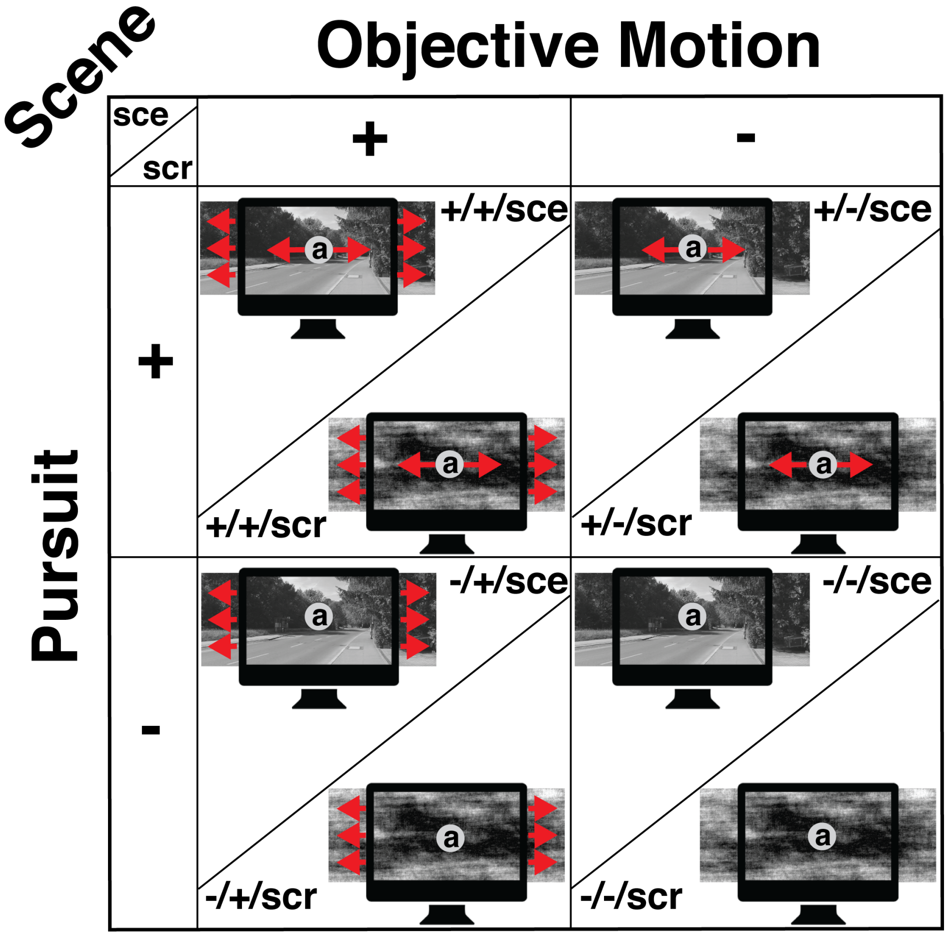


Figure 1: Stimuli. Eight conditions were obtained by a 2 x 2 x 2 factorial design with the following factors: objective motion (on/off), pursuit (on/ off) and scene (scene/ scrambled). . In the “±/±” notation, the first position refers to pursuit, the second to objective motion. “+” refers to presence and “-“ to absence. Objective motion was horizontal motion of the background image (scenes or scrambled images) and pursuit was horizontal motion of the fixation disk. There was a one-back character-matching task inside the fixation disk (shown larger for illustration).

Functional Localizer

Visual stimuli for the functional motion localizer consisted of random dot patterns. 7 conditions were present in this localizer and each condition was presented 6 times during the session in a pseudorandom history-matched manner. The conditions were as following: 3D fullfield motion (coherent motion), random motion, right and left hemifield 3D fullfield motion (left or right 1/3rd of the screen), 2D lateral motion with synched pursuit (coherent motion), smooth pursuit with static background and static dots with static fixation task. Each block lasted 12 seconds. Participants performed a 1-back character-matching task, identical to the task in main experiment. Motion areas V3A, V6 and CSv were localized as described previously (Fischer et al., 2011, 2012) and V5+/MT+ was localized using random motion.

Due to technical problems, we could not use right and left hemifield 3D fullfield motion conditions in order to define MT and MST separately, as established previously. Instead, V5+/MT+ was localized using responses to random dot-motion versus static dots. CSv was localized as described previously, using responses to fullfield coherent motion with coherently moving fixation dot versus random dot-motion with still fixation(Fischer et al., 2011). V6 was also localized using the same contrast. V3A was localized using responses to coherent 2D motion versus moving fixation dot on static background consisting of dots. All regions are defined using an individual p-value.

Data Acquisition

T2* weighted functional images were acquired using a 64-channel phased-array head coil in a Siemens Magnetom PRISMA 3T scanner (Siemens, Erlangen, Germany). The voxel size was 3 x 3 x 3 mm^3 and TR was 2.3 seconds, while TE was 35 ms and flip angle was 79°. The images included 32 slices, in an ascending order. In order to allow T1 equilibration, the first 3 volumes of data (the first 4 volumes for motion localizer) were discarded. We also collected anatomical images for each participant using T1-weighted images (1 x 1 x 1 mm^3 resolution).

FMRI Data Preprocessing and Statistical Analysis

SPM5 toolbox in MATLAB 7.10.0 was used in order to preprocess functional images with the following steps: reslicing and realignment, followed by coregistration of the structural image to the mean functional image, normalization of the data to the Montreal neurological institute (MNI) space and finally spatial smoothing with 6 mm full-width at half maximum Gaussian kernel for single participants and 12 mm for group level analyses.

Data of each participant were analysed separately using the GLM (general linear model) in SPM5. We modelled each condition and button presses, as well as regressors of no interest, which were six motion realignment parameter series and one additional regressor for global signal variance (Desjardins, Kiehl, & Liddle, 2001; Van Dijk et al., 2010). The global signal variance regressor was orthogonalized to the conditions of interest. The data were high pass filtered using a cut-off value of 128s. In addition, the beta images from the first level GLMs of each participant were used for group level analyses.

The ROI analyses were done by defining ROIs using independent localizer for each participant separately and then extraction of mean beta values for each ROI and for each participant. We used MarsBaR toolbox in order to define ROIs (Brett, Anton, Valabregue, & Poline, 2002). Beta values were range normalized between 0 and 1 for each ROI and participant separately. For 4 runs and 8 conditions, the minimum of all these 32 beta values were subtracted from all 32 beta values and then all of them were divided by maximum of these 32 beta values, for each participant and each ROI separately. Repeated measures ANOVAs, as well as paired t-tests were conducted in order to analyze the effects of conditions using statistical analysis software IBM SPSS Statistics version 22.0. Mauchly's sphericity test results were considered for the definition of violation of sphericity and Greenhouse-Geisser correction was used in case of violation of sphericity.

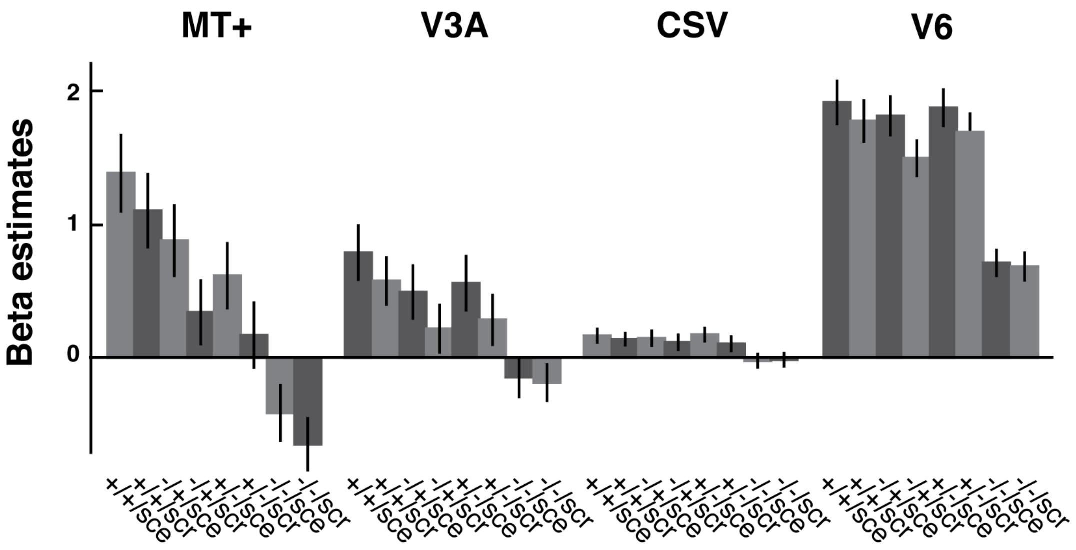
Eye Tracking

Eye tracking of participants during the main experiment was done using an infrared camera based eye tracker system (Eye-Trac 6; Applied Science Laboratories). The steps of preprocessing included blink removal, smoothing of x and y positions using a running average window of 200 milliseconds. We calculated the fixation accuracy by the root mean square error of actual eye position relative to the fixation disk for each condition across participants and runs. Repeated measures ANOVAs were facilitated in order to analyse eye-tracking data

4.4. Results

After independently localizing motion responsive regions V5+/MT+, V3A, V6 and CSv, we analyzed their responses to scene content. We localized V5+/MT+ in 30 hemispheres, V3A in 29 hemispheres, V6 in 28 hemispheres and CSv in 25 hemispheres. For all ROIs, raw and normalized mean beta responses are shown separately in figure 2.

A. Raw data



B. Normalized data

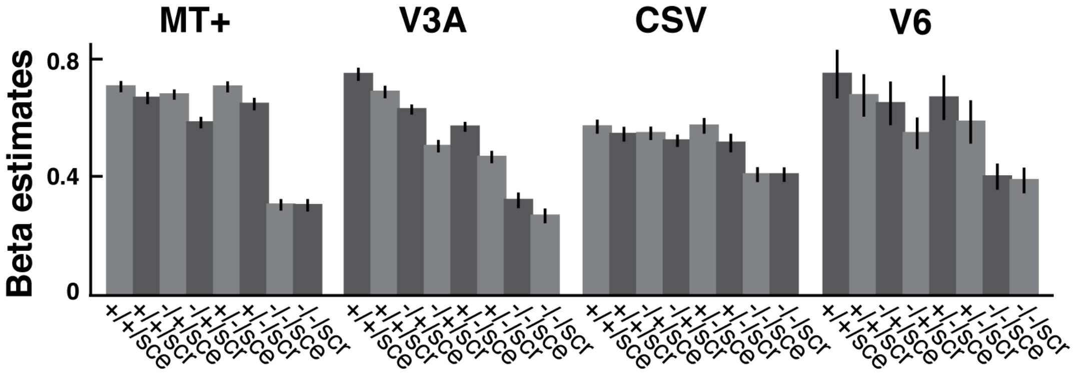


Figure 2: Responses to all conditions across ROIs. (A) Raw beta estimates in motion ROIS V5+/MT+, V3A, CSv and V6. (B) Normalized beta estimates (see methods). In the "±/±" notation, the first position refers to pursuit, the second to objective motion. "+" refers to presence and "-" to absence. scr: scramble, sce: scene. Plots show mean ± standard error of mean (SEM).

A 4 x 2 x 8 repeated-measures ANOVA with the factors ROI, hemisphere, planar motion, pursuit and scene was conducted in order to test for hemisphere effect. For each ROI we pooled data from both hemispheres since there was no effect of hemisphere (F (1,5) = 0.209, p = 0.667) or any interactions including hemisphere as a factor (hemisphere and ROI: F (6,30) = 0.721, p = 0.636, hemisphere and planar motion: F (1,5) = 0.280, p = 0.620, hemisphere and pursuit: F (1,5) = 0.212, p = 0.664, ROI, hemisphere and planar motion: F (6,30) = 0.643, p = 0.695, ROI, hemisphere, planar motion and pursuit: F (6,30) = 0.980, p = 0.456, hemisphere and scene: F (1,5) = 0.014, p = 0.911, ROI, planar motion, hemisphere and scene: F (6,30) = 0.288, p = 0.938, hemisphere, pursuit and scene: F (1,5) = 1.526, p = 0.272 and hemisphere, ROI, planar motion, pursuit and scene: F (6,30) = 1.172, p = 0.347) .

Next, we analyzed scene responses in these ROIs and then we analyzed content related motion responses, meaning the interactions between scene and different motion types in all ROIs separately.

Scene responses

We tested scene responses of all ROIs using paired t-tests, using the contrast for all conditions with scenes compared to all conditions with scramble images. All ROIs had a significant scene response (V5+/MT+: t (29) = 8.44, p = 0.11 * 10^{-7}, V3A: t (28) = 7.99, p = 0.32 * 10^{-7}, CSv: t (24) = 2.2, p = 0.038, V6: t (27) = 5.99, p = 0.4 * 10^{-5}, all Bonferroni-Holm corrected for 4 comparisons). Either scenes themselves or motion responses could drive the scene responses in these regions. In order to investigate this, we compared scene versus scramble during still ((-/-/+) versus (-/-/-)) responses using paired t-tests on each motion responsive ROI separately. Only V3A showed a significant difference between responses to scenes and responses to scrambled images during still (t (28) = 4.41, p = 0.0006, Bonferroni-Holm corrected for four comparisons). This difference was not present in other motion responsive regions we investigated (V6 (t (27) = 0.87, p = 0.39), V5+/MT+ (t (29) = 0.69, p = 0.49) and CSv (t (24) = 0.36, p = 0.72). Figure 3A shows scene responses of all ROIs. Next, we tested whether V3A could be differentiated from other regions in its responses to still scenes. Indeed, V3A can be differentiated from all regions by its scene responses (V3A vs. V5+/MT+: t (27) = 2.47, p = 0.02, V3A vs. V6: t (25) = 2.61, p = 0.03, V3A vs. CSv: t (22) = 3.36, p = 0.009, all corrected using Bonferroni- Holm correction for 3 comparisons). Figure 3B shows responses to still scenes in all ROIs.

We also tested scene versus scramble responses during background motion with eye fixation ((-/+/sce) vs. (-/+/scr)). As seen in figure 3C, V5+/ MT+ (t (29) = 8.11, p = 0.25 * 10^{-7}), V3A (t (28) = 6.72, p = 0.81 * 10^{-6}), and V6 (t (27) = 4.27, p = 0.0004) significantly responded to scenes with background motion during fixation, whereas CSv did not (t (24) = 1.08, p = 0.29) (Corrected using Bonferroni-Holm correction for four comparisons).

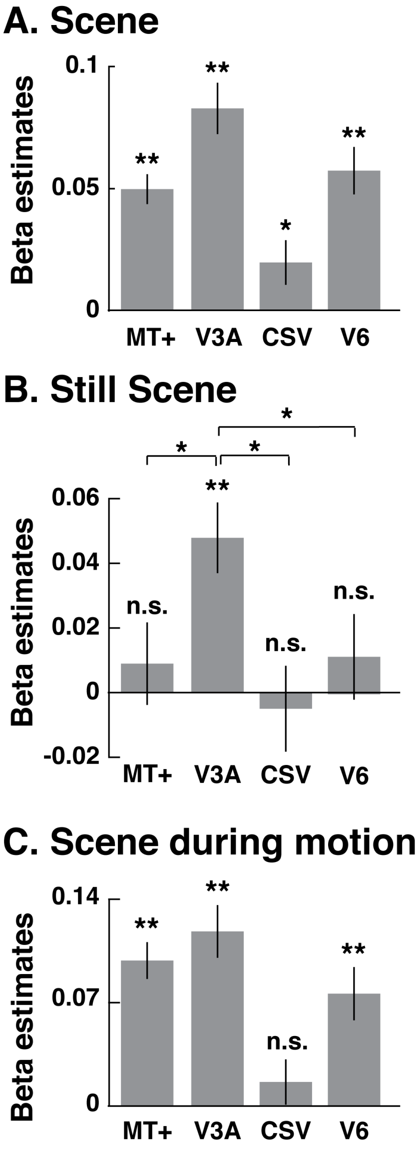


Figure 3: Scene responses in V5+/MT+, V3A, V6 and CSv. (A) Main effect of scenes. (B) Responses to still scenes vs. still scramble images. (C) Responses to moving scenes vs moving scramble images during fixation (-/+/sce vs. -/+/scr). **: p < 0.001, *: p < 0.05. Bonferroni-Holm corrected. Plots show mean ± standard error of mean (SEM).

Content effect on motion responses

Next, we tested the interactions between motion and scene content in the regions of interests.

First, we tested objective motion responses during scenes, during scrambled images and their interaction. As expected, in all regions there were significant objective motion responses during scenes (V5+/MT+: t (29) = 15.51, p = $0.17 * 10^{-13}$, V3A: t (28) = 11.95, p = $0.18 * 10^{-10}$, V6: t (27) = 7.59, p = $0.26 * 10^{-6}$, CSv: t (24) = 5.2, p = $0.13 * 10^{-3}$) and scrambled images (V5+/MT+: t (29) = 10.33, p = $0.29 * 10^{-9}$, V3A: t (28) = 11.78, p = $0.23 * 10^{-10}$, V6: t (27) = 9.7, p = $0.22 * 10^{-8}$, CSv: t (24) = 5.74, p = $0.42 * 10^{-4}$). However, there was a significant interaction between objective motion and scene content only in V5+/MT+ (t (29) = 2.94, p = 0.024), although in V3A and V6, there was a trend for higher objective motion responses during scenes compared to during scramble (All corrected for 12 comparisons using Bonferroni-Holm correction).

Next, we tested retinal motion responses during scenes, scrambles and their interaction. All regions showed significant retinal motion responses during scenes (V5+/MT+: t (29) = 15.28, p = $0.25 * 10^{-13}$, V3A: t (28) = 5.24, p = $0.14 * 10^{-3}$, V6: t (27) = 4.58, p = $0.65 * 10^{-3}$, CSv: t (24) = 4.94, p = $0.43 * 10^{-3}$). Interestingly, only V5+/MT+ (t (29) = 9.99, p = $0.75 * 10^{-9}$) and CSv (t (24) = 3.63, p = 0.005) had significant retinal motion responses during scramble images while V3A (t (28) = 0.49, p = 0.627) and V6 (t (27) = 1.28, p = 0.211) did not. All regions except CSv showed a significant interaction between retinal motion and scene content (V5+/MT+: t (29) = 4.74, p = $0.41 * 10^{-3}$, V3A: t (28) = 4.03, p = 0.0024, V6: t (27) = 3.06, p = 0.02, CSv: t (24) = 1.57, p = 0.129) (All corrected for 12 comparisons using Bonferroni-Holm correction).

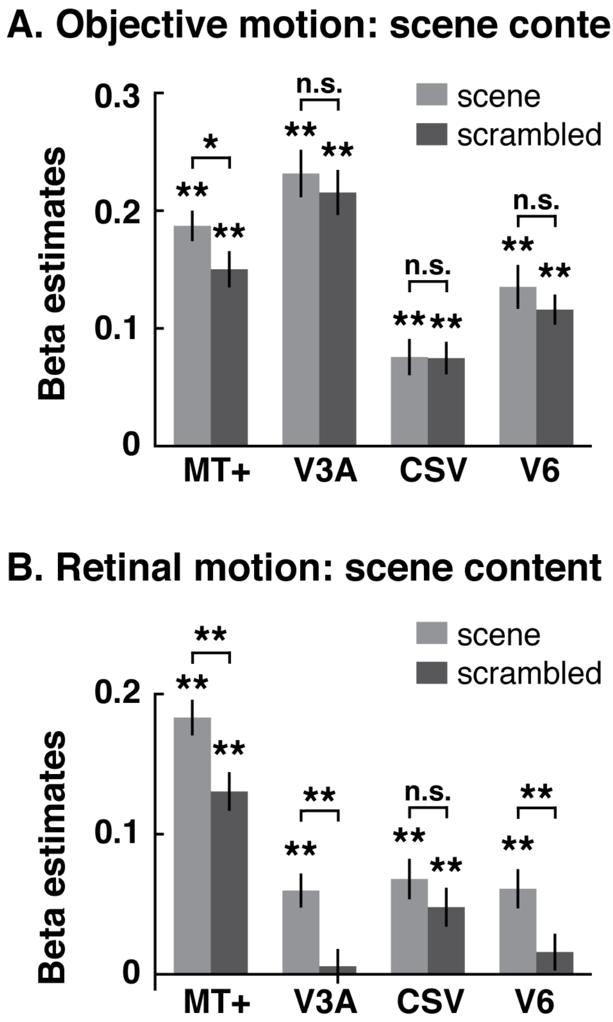


Figure 4. Objective and retinal motion preferences across ROIs. (A) Objective motion responses, shown for scenes and scrambled backgrounds separately and their interactions. (B) Retinal motion responses, shown for scenes and scrambled backgrounds separately and their interactions. **: $p < 0.001$, *: $p < 0.05$, Bonferroni-Holm corrected. Plots show mean ± standard error of mean (SEM).

Whole brain Analyses

Since V3A is responsive to still scenes, we wanted to check if it overlaps with scene responsive areas in the group level analysis. We first calculated the contrast still scenes vs. still scramble images in group level, with $p < 0.05$ uncorrected. Next, we defined V3A using the previously established contrast (Fischer et al., 2012); objective vs. retinal motion during both scenes and scramble images in the group level using $p < 0.001$ uncorrected. Figure 5 shows the overlap between these two contrasts.

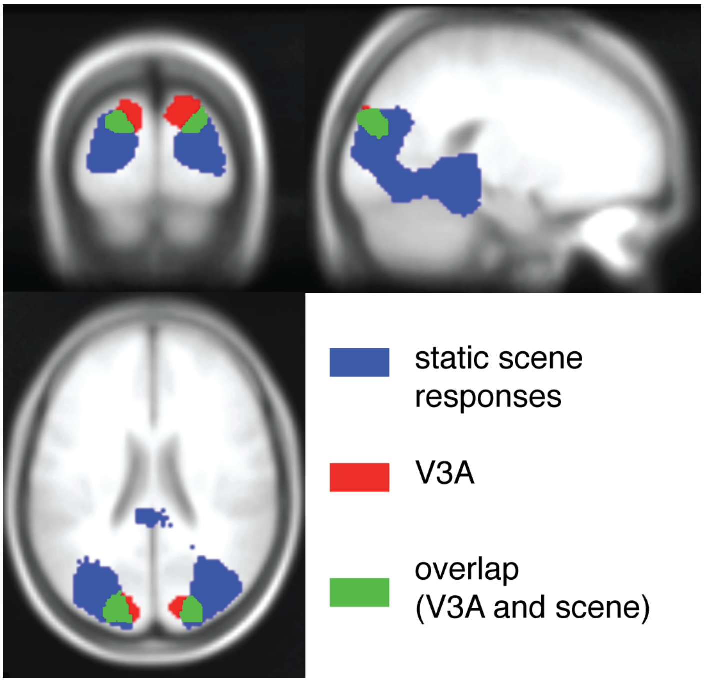


Figure 5: Whole brain results showing overlap between V3A and scene responsive areas. The contrasts were created using (a) still scene vs still scrambled images, shown by blue and (b) objective versus retinal motion to define V3A (during both scenes and scramble images), shown by red and their interaction is shown by green.

Behavior Data

Participants performed a character back-matching task during the main experiment. The mean correct response rate was 0.85 ± 0.07 (mean ± std), whereas mean response time was 0.57 ± 0.14 s (mean ± std). We analyzed response times using a 3-way repeated measures ANOVA with the factors objective motion, pursuit and scene. There were only significant effects of pursuit (F (1,16) = 5.06, p = 0.039) but no other main effects (objective motion: F (1,16) = 0.005, p = 0.95; scene: F (1,16) = 0.39, p = 0.54), nor interactions (objective and pursuit (i.e., retinal motion): F (1,16) = 0.93, p = 0.35; objective and scene: F (1,16) = 0.00029, p = 0.99; pursuit and scene F (1,16) = 1.58, p = 0.23; objective, pursuit and scene: F (1,16) = 0.98, p = 0.34).

Eye tracking Data

Eye tracking data were collected for each participant during scanning and were preprocessed as described in methods. Following preprocessing, RMSE of eye position relative to the fixation disk was calculated and these RMSE were used for calculation and comparison of fixation accuracy. The average RMSE across participants and runs and conditions is 1.54 ± 0.73 deg (mean ± std). We used 2 x 2 x 2 repeated measures ANOVA with factors objective motion, pursuit and scene using the RMSE data mentioned above. There was an effect of pursuit (F (1, 67) = 682.38, p = 0.75 * 10^{-36}), but there were no effects of objective motion (F (1, 67) = 0.06, p = 0.82) or scene (F (1, 67) = 0.0001, p = 0.99). Moreover, there were no interaction between objective motion and pursuit (which gives us retinal motion) (F (1, 67) = 0.28, p = 0.56), between objective motion and scene (F (1, 67) = 0.14, p = 0.71), between pursuit and scene (F (1, 67) = 2.8, p = 0.099) and between objective motion, pursuit and scene (F (1, 67) = 0.16, p = 0.69).

4.5. Discussion

Here, we sought answers to the following questions: are motion responsive regions modulated by scene content and if they are, how do their responses to objective and retinal motion change with scene content modulation? We used a 2 x 2 x 2 factorial design with the factors being real world scenes versus their scrambled versions, horizontal panning motion and smooth pursuit eye movements. We chose horizontal panning motion and pursuit eye movements, as they are frequently found and are natural in daily life. We examined well-known motion responsive regions V5+/MT+, V3A, V6, and CSv.

We found that the motion responses of all motion regions showed a scene preference, whereas only V3A also responded significantly to still scene images compared to still scrambles. Moreover, V5+/MT+, V3A and V6 showed motion and scene content interaction whereas this was absent in the responses of CSv.

Scene responses in V3A

V3A was the only region with significant scene responses even in the absence of any type of motion. V3A is an intermediate-tier, retinotopic region with relatively large receptive field size and representing both upper and lower visual fields (Tootell et al., 1997). So, it is unlikely that responses to static scenes could be driven by possible local visual field irregularities between scenes and scrambled images. Human V3A is neighbouring a scene responsive region, occipital place area (OPA), which is also known as transverse occipital sulcus (TOS) (Nasr et al., 2011). Interestingly, some recent studies have shown that OPA partially overlaps with V3A (Bettencourt & Xu, 2013; Silson, Groen, Kravitz, & Baker, 2016). Additionally, a recent study found that V3A is connected to ventral regions (hV4/ VO-1) via a major white matter pathway in human brain, namely vertical occipital fasciculus (VOF) (Takemura et al., 2015). More research is needed to understand the exact role of V3A in scene processing.

Although V3A is known as a motion-processing region, its exact role is still under investigation. Previous studies have shown that V3A processes three-dimensional structure and depth cues and as well as shapes defined by colour or motion (Georgieva, Peeters, Kolster, Todd, & Orban, 2009; Paradis et al., 2000; Self & Zeki, 2005). Other findings include that V3A has a role in shape integration, contributes to contour integration, is modulated by context (Aspell, Wattam-Bell, Atkinson, & Braddick, 2010; Schira et al., 2004), has shape sensitive responses and is involved in object processing (Denys et al., 2004; Grill Spector et al., 1998) and has a role in not only in boundary detection, but also in scene segmentation and texture segregation (together with other early level visual areas) (Kastner, De Weerd, & Ungerleider, 2000; Scholte et al., 2008). The responses to still scenes in V3A shown here could be driven by continuous contours, which exist in scene images but absent scrambled images, or by higher-level image features such as shapes or objects within the scenes.

Our results also showed an interaction between scene and retinal motion in V3A, meaning that V3A's scene responses were higher during motion. Interestingly, while V3A showed significant responses to retinal motion during scenes, it did not show any significant retinal motion responses during scrambled images. We believe this could be driven by viewpoint changes across scenes, since in a previous study, V3A, as well as MT, was shown to have viewpoint specific object selective responses (Konen & Kastner, 2008).

V3A has many connections with parietal regions and is thought to provide information about object location and motion to these regions. This information is probably used by parietal regions during self-object interaction such as reaching and grasping. While encoding object location and motion, V3A is involved in depth perception and having a representation of visual scenes could be useful with this role of V3A.

Content-dependent responses in motion processing regions

V5+/MT+ showed significant interaction between scene and objective motion and between scenes and retinal motion, meaning that, it responded more to motion (both objective and retinal motion) when there were scenes in the background, compared to when there were scrambled images in the background. It is unlikely that these responses are due to differences of spatial frequency in lower and upper parts of the visual scene images since human V5/MT and MST show no upper-lower visual field bias (Kolster, Peeters, & Orban, 2010). In addition to its well-known motion related responses such as responses to optic flow, speed and direction selective responses, V5/MT also shows object and shape sensitivity as well as viewpoint specific object responses (Denys et al., 2004; Kolster et al., 2010; Konen & Kastner, 2008; Kourtzi et al., 2002; Kourtzi & Kanwisher, 2000). Our results are parallel to the aforementioned findings regarding content effect in V5+/MT+.

Interestingly, V6 showed scene related responses during retinal-motion. Previous studies on V6 showed its involvement in optic flow responses and processing of self-motion related cues (Cardin & Smith, 2010; Fischer et al., 2012; Pitzalis et al., 2010). Additionally, V6 contains representations of both upper and lower visual field (Pitzalis et al., 2006). V6 is highly connected to parietal regions and its visuomotor neighbor V6A. V6A and parietal regions are particularly interested in reaching and grasping. In relation to this, a study showed that V6 is more responsive to near visual field compared to far visual field, and suggested that these responses are related to object locations for reaching (Quinlan & Culham, 2007). It is possible that scene content responses are also related to distance encoding in V6 in order to provide input to parietal regions for reaching and grasping. Thus, motion during scene generates more response compared to motion during scrambled images.

CSv only showed scene responses but no interaction or no scene response during motion or no motion. It is possible that scene responses in CSv are modulated by eye movements, but since pursuit condition is not well controlled, we did not investigate pursuit related responses in this study. CSv is located in posterior cingulate cortex and has been shown to contain information about heading direction during self-motion (Furlan, Wann, & Smith, 2014). CSv also showed vestibular responses (Smith, Wall, & Thilo, 2012). Furthermore, CSv has been shown to integrate eye movements with retinal motion (Fischer et al., 2011). Cingulate sulcus has been previously shown to have place category responses (Epstein & Higgins, 2007). Natural scene images, as we used in this experiment, resembles everyday experience for the visual cortex in a way that it provides realistic input (compared to scrambled images) and perceptually, provides more important cues regarding heading and self-motion. So, in this context, CSv responses could be explained by more engagement of CSv during eye movements in natural scenes since this resembles perceptually more natural input regarding self-motion extraction or heading direction.

Our result showing scene related higher activation in every motion responsive region is rather interesting. Scenes related responses can be used during depth perception via textures or perspective whereas this is absent in scrambled images.

Previous studies showed that natural stimuli are preferred by the visual cortex (Kayser, Kording, & Konig, 2004). However, more studies on contextual effects of natural scene stimuli on visual processing regions would provide a better insight about the effects seen here.

Low-level versus high-level interpretations

One can think that the results shown here are due to confound regarding low level differences across scenes and their scrambled versions. Phase scrambling was used in order to conserve low-level image features such as spatial frequency while eliminating contextual effect, thus making the high level aspect of the image (such as scenes as in here, but also used for objects and faces) unrecognizable. Traditionally, applying phase scrambling while keeping luminance and contrast equal across images and their scrambled versions is thought to result in no response from early visual cortex but engage higher level regions' (or ventral regions') responsiveness. However, a number of studies raised concerns about using phase-scrambled images in this way. For instance, a comparison of the contrasts of natural scenes and their scrambled counterparts resulted in higher number of peaks in the histogram of natural images (Dumoulin, Dakin, & Hess, 2008). Also, scrambling scatters local constructions within the image all over (Kay, Winawer, Rokem, Mezer, & Wandell, 2013). Related to these concerns, modifications to phase scrambling and even different scrambling methods have been proposed (Ales, Farzin, Rossion, & Norcia, 2012; Stojanoski & Cusack, 2014). Hence, it cannot be ruled out that our findings regarding higher scene responses in motion responsive regions might be related to higher-order or localized differences in low-level image features such as lines or contours. Numerous studies have investigated how different level features are processed in human brain. More related to our findings, extra-striate visual cortex has been shown to have contour-based responses to scene images (Dumoulin et al., 2008). Another study showed that a comparison of lines and edges to phase scrambled images created higher responses for lines and edges in most of the visual areas, even early visual cortex (Perna, Tosetti, Montanaro, & Morrone, 2008). For ventral scene-responsive regions, a series of recent studies provided evidence that low-level features may account for their previously reported preference to certain high-level categories (Nasr, Echavarria, & Tootell, 2014; Nasr & Tootell, 2012; Rajimehr, Devaney, Bilenko, Young, & Tootell, 2011). These studies argued that low-level features typically associated to scenes, such as cardinal orientations, rectilinearity, and high spatial frequencies alone selectively activate PPA. However, more recent evidence showed that even when all low-level features are controlled for, PPA still prefers high-level interpretations of features perceived as spatial arrangements (Bryan, Julian, & Epstein, 2016; Schindler & Bartels, 2016). Clearly, further more detailed studies are needed in order to clarify the underlying mechanism of scene related responses in motion responsive regions.

4.6. Conclusion

In conclusion, V5+/MT+, V3A, V6 and CSv had content effect due to scene content during motion. V3A also had scene responses during still scenes. These results contribute to our understanding of how V5+/MT+, V3A, V6 and CSv responses are modulated by scene content.

These results support the view that unlike the traditional theories about completely segregated dorsal ‘what’ and ventral ‘where’ streams, these two pathways functionally interact. Consistent with this view, our results could be interpreted as V3A taking part in analysing the 3D overlay of the visual scenes, which can be useful when calculating the motion of objects in depth. Further studies are needed to investigate the role of motion regions, especially V3A, in detail during scene viewing.

Acknowledgements

This work was funded by the Centre for Integrative Neuroscience Tübingen through the German Excellence Initiative (EXC307) and by the Max Planck Society, Germany.

Conflict of Interest: The authors declare no competing financial interests

4.7. References

Ales, J. M., Farzin, F., Rossion, B., & Norcia, A. M. (2012). An objective method for measuring face detection thresholds using the sweep steady-state visual evoked response. *Journal of Vision, 12*(10), 18-18. doi: 10.1167/12.10.18

Arnoldussen, D. M., Goossens, J., & van den Berg, A. V. (2011). Adjacent visual representations of self-motion in different reference frames. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 11668-11673. doi: 10.1073/pnas.1102984108

Aspell, J. E., Wattam-Bell, J., Atkinson, J., & Braddick, O. J. (2010). Differential human brain activation by vertical and horizontal global visual textures. *Experimental brain research, 202*(3), 669-679.

Bartels, A., Zeki, S., & Logothetis, N. K. (2008). Natural Vision Reveals Regional Specialization to Local Motion and to Contrast-Invariant, Global Flow in the Human Brain. *Cerebral Cortex, 18*(3), 705-717. doi: 10.1093/cercor/bhm107

Bettencourt, K. C., & Xu, Y. (2013). The Role of Transverse Occipital Sulcus in Scene Perception and Its Relationship to Object Individuation in Inferior Intraparietal Sulcus. *Journal of Cognitive Neuroscience, 25*(10), 1711-1722. doi: http://dx.doi.org/10.1162/jocn_a_00422

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT *Annual Review of Neuroscience* (Vol. 28, pp. 157-189). Palo Alto: Annual Reviews.

Boussaoud, D., Ungerleider, L. G., & Desimone, R. (1990). Pathways for Motion Analysis - Cortical Connections of the Medial Superior Temporal and Fundus of the Superior Temporal Visual Areas in the Macaque. *Journal of Comparative Neurology, 296*(3), 462-495.

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using the MarsBar toolbox for SPM 99. *Neuroimage, 16*(2), S497.

Bryan, P. B., Julian, J. B., & Epstein, R. (2016). Rectilinear edge selectivity is insufficient to explain the category selectivity of the parahippocampal place area. *Frontiers in human neuroscience, 10*.

Cardin, V., & Smith, A. T. (2010). Sensitivity of human visual and vestibular cortical regions to egomotion-compatible visual stimulation. *Cerebral cortex, 20*(8), 1964-1973. doi: 10.1093/cercor/bhp268

Chowdhury, S. A., Takahashi, K., DeAngelis, G. C., & Angelaki, D. E. (2009). Does the middle temporal area carry vestibular signals related to self-motion? *The Journal of neuroscience : the official journal of the Society for Neuroscience, 29*(38), 12020-12030. doi: 10.1523/JNEUROSCI.0004-09.2009

Denys, K., Vanduffel, W., Fize, D., Nelissen, K., Peuskens, H., Van Essen, D., & Orban, G. A. (2004). The processing of visual shape in the cerebral cortex of human and nonhuman primates: a functional magnetic resonance imaging study. *J Neurosci, 24*(10), 2551-2565.

Desjardins, A. E., Kiehl, K. A., & Liddle, P. F. (2001). Removal of confounding effects of global signal in functional MRI analyses. *Neuroimage, 13*(4), 751-758. doi: http://dx.doi.org/10.1006/nimg.2000.0719

Dubner, R., & Zeki, S. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research, 35*, 528-532.

Dumoulin, S. O., Dakin, S. C., & Hess, R. F. (2008). Sparsely distributed contours dominate extra-striate responses to complex scenes. *Neuroimage, 42*(2), 890-901. doi: 10.1016/j.neuroimage.2008.04.266

Epstein, R., & Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb Cortex, 17*(7), 1680-1693. doi: http://dx.doi.org/10.1093/cercor/bhl079

Erickson, R. G., & Thier, P. (1991). A Neuronal Correlate of Spatial Stability during Periods of Self-Induced Visual-Motion. *Experimental Brain Research, 86*(3), 608-616.

Fischer, E., Bulthoff, H. H., Logothetis, N. K., & Bartels, A. (2011). Visual Motion Responses in the Posterior Cingulate Sulcus: A Comparison to V5/MT and MST. *Cerebral cortex*. doi: 10.1093/cercor/bhr154

Fischer, E., Bulthoff, H. H., Logothetis, N. K., & Bartels, A. (2012). Human areas V3A and V6 compensate for self-induced planar visual motion. *Neuron, 73*(6), 1228-1240. doi: 10.1016/j.neuron.2012.01.022

Furlan, M., Wann, J. P., & Smith, A. T. (2014). A Representation of Changing Heading Direction in Human Cortical Areas pVIP and CSv. *Cerebral Cortex, 24*(11), 2848-2858. doi: 10.1093/cercor/bht132

Galletti, C., & Fattori, P. (2003). Neuronal mechanisms for detection of motion in the field of view. *Neuropsychologia, 41*(13), 1717-1727.

Galletti, C., Gamberini, M., Kutz, D. F., Fattori, P., Luppino, G., & Matelli, M. (2001). The cortical connections of area V6: an occipito-parietal network processing visual information. *Eur J Neurosci, 13*(8), 1572-1588.

Georgieva, S., Peeters, R., Kolster, H., Todd, J. T., & Orban, G. A. (2009). The processing of three-dimensional shape from disparity in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 29*, 727-742. doi: 10.1523/JNEUROSCI.4753-08.2009

Goossens, J., Dukelow, S. P., Menon, R. S., Vilis, T., & van den Berg, A. V. (2006). Representation of Head-Centric Flow in the Human Motion Complex. *The Journal of Neuroscience, 26*(21), 5616-5627. doi: 10.1523/jneurosci.0730-06.2006

Grill Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *NEURON. Neuron., 21*(1), 191-202.

Gu, Y., DeAngelis, G. C., & Angelaki, D. E. (2007). A functional link between area MSTd and heading perception based on vestibular signals. *Nature Neuroscience, 10*(8), 1038-1047.

Huk, A. C., Dougherty, R. F., & Heeger, D. J. (2002). Retinotopy and Functional Subdivision of Human Areas MT and MST. *The Journal of Neuroscience, 22*(16), 7195-7205.

Kastner, S., De Weerd, P., & Ungerleider, L. G. (2000). Texture Segregation in the Human Visual Cortex: A Functional MRI Study. *Journal of Neurophysiology, 83*(4), 2453-2457.

Kay, K. N., Winawer, J., Rokem, A., Mezer, A., & Wandell, B. A. (2013). A Two-Stage Cascade Model of BOLD Responses in Human Visual Cortex. *PLoS Comput Biol, 9*(5), e1003079. doi: 10.1371/journal.pcbi.1003079

Kayser, C., Kording, K. P., & Konig, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Curr Opin Neurobiol, 14*(4), 468-473.

Kolster, H., Peeters, R., & Orban, G. A. (2010). The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. *The Journal of* *neuroscience : the official journal of the Society for Neuroscience, 30*, 9801-9820. doi: 10.1523/JNEUROSCI.2069-10.2010

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature neuroscience, 11*, 224-231. doi: 10.1038/nn2036

Korkmaz Hacialihafiz, D., & Bartels, A. (2015). Motion responses in scene-selective regions. *NeuroImage, 118*, 438-444. doi: http://dx.doi.org/10.1016/j.neuroimage.2015.06.031

Kourtzi, Z., Bulthoff, H. H., Erb, M., & Grodd, W. (2002). Object-selective responses in the human motion area MT/MST. *Nat Neurosci, 5*(1), 17-18. doi: http://www.nature.com/neuro/journal/v5/n1/suppinfo/nn780_S1.html

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience, 12*(1), 48-55.

Maciokas, J. B., & Britten, K. H. (2010). Extrastriate Area MST and Parietal Area VIP Similarly Represent Forward Headings. *Journal of Neurophysiology, 104*(1), 239-247. doi: 10.1152/jn.01083.2009

MATLAB. (2010). *version 7.10.0* Natick, Massachusetts: The MathWorks Inc.

Nasr, S., Echavarria, C. E., & Tootell, R. B. H. (2014). Thinking Outside the Box: Rectilinear Shapes Selectively Activate Scene-Selective Cortex. *The Journal of Neuroscience, 34*(20), 6721-6735. doi: 10.1523/JNEUROSCI.4802-13.2014

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. H. (2011). Scene-selective cortical regions in human and nonhuman primates. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 31*, 13771-13785. doi: 10.1523/JNEUROSCI.2792-11.2011

Nasr, S., & Tootell, R. B. H. (2012). A cardinal orientation bias in scene-selective visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 32*, 14921-14926. doi: http://dx.doi.org/10.1523/JNEUROSCI.2036-12.2012

Paradis, A. L., Cornilleau-Pérès, V., Droulez, J., Van De Moortele, P. F., Lobel, E., Berthoz, A., . . . Poline, J. B. (2000). Visual perception of motion and 3-D structure from motion: an fMRI study. *Cerebral cortex (New York, N.Y. : 1991), 10*, 772-783.

Perna, A.., Tosetti, M., Montanaro, D., & Morrone, M. C. (2008). BOLD response to spatial phase congruency in human brain. *Journal of Vision, 8*(10), 15-15. doi: 10.1167/8.10.15

Pitzalis, S., Galletti, C., Huang, R. S., Patria, F., Committeri, G., Galati, G., . . . Sereno, M. I. (2006). Wide-field retinotopy defines human cortical visual area v6. *J Neurosci, 26*(30), 7962-7973.

Pitzalis, S., Sereno, M. I., Committeri, G., Fattori, P., Galati, G., Patria, F., & Galletti, C. (2010). Human v6: the medial motion area. *Cerebral cortex, 20*(2), 411-424. doi: 10.1093/cercor/bhp112

Quinlan, D. J., & Culham, J. C. (2007). fMRI reveals a preference for near viewing in the human parieto-occipital cortex. *NeuroImage, 36*(1), 167-187. doi: 10.1016/j.neuroimage.2007.02.029

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. H. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS biology, 9*, e1000608. doi: 10.1371/journal.pbio.1000608

Schindler, A., & Bartels, A. (2016). Visual high-level regions respond to high-level stimulus content in the absence of low-level confounds. *NeuroImage, 132*, 520-525.

Schira, M. M., Fahle, M., Donner, T. H., Kraft, A., & Brandt, S. A. (2004). Differential Contribution of Early Visual Areas to the Perceptual Process of Contour Processing. *Journal of Neurophysiology, 91*(4), 1716-1721.

Scholte, H. S., Jolij, J., Fahrenfort, J. J., & Lamme, V. A. F. (2008). Feedforward and recurrent processing in scene segmentation: electroencephalography and functional magnetic resonance imaging. *Journal of cognitive neuroscience, 20*, 2097-2109. doi: 10.1162/jocn.2008.20142

Self, M. W., & Zeki, S. (2005). The integration of colour and motion by the human visual brain. *Cerebral cortex, 15*, 1270-1279. doi: http://dx.doi.org/10.1093/cercor/bhi010

Silson, E. H., Groen, I. A., Kravitz, D. J., & Baker, C. I. (2016). Evaluating the correspondence between face-, scene-, and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. *Journal of vision, 16*(6), 14-14.

Smith, A. T., Wall, M. B., & Thilo, K. V. (2012). Vestibular inputs to human motion-sensitive visual cortex. *Cerebral cortex (New York, N.Y. : 1991), 22*, 1068-1077. doi: 10.1093/cercor/bhr179

Smith, A. T., Wall, M. B., Williams, A. L., & Singh, K. D. (2006). Sensitivity to optic flow in human cortical areas MT and MST. *European Journal of Neuroscience, 23*(2), 561-569.

Stojanoski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision, 14*(12), 6-6. doi: 10.1167/14.12.6

Takemura, H., Rokem, A., Winawer, J., Yeatman, J. D., Wandell, B. A., & Pestilli, F. (2015). A Major Human White Matter Pathway Between Dorsal and Ventral Visual Cortex. *Cerebral Cortex*. doi: 10.1093/cercor/bhv064

Tootell, R. B. H., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., . . . Dale, A. M. (1997). Functional analysis of V3A and related areas in human visual cortex. *J Neurosci, 17*(18), 7060-7078.

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

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *The Journal of Neuroscience, 11*(3), 641-649.

5. Speed tuning to real-world- and retinal motion in cortical motion regions

Authors: Didem Korkmaz Hacialihafiz and Andreas Bartels*

Vision and Cognition Lab, Centre of Integrative Neuroscience, University of Tübingen, Germany

*Corresponding author:

Andreas Bartels
Vision and Cognition Lab
Centre for Integrative Neuroscience,
University of Tübingen,
Otfried-Müller-Str. 25
72076 Tübingen
Germany

Email: andreas.bartels@tuebingen.mpg.de
Phone: +49 7071 2989168

5.1. Abstract

Motion signals can arise for two reasons in the retina: due to self-motion or due to real motion in the environment. Prior studies on speed tuning always measured joint responses to real and retinal motion, and for some of the more recently identified human motion processing regions, speed tuning has not been examined in at all. We localized motion regions V3A, V6, V5/MT, MST and cingulate sulcus visual area (CSv) in 20 human participants, and then measured their responses to motion velocities from 1-24 degrees per second. Importantly, we used a pursuit paradigm that allowed us to quantify responses to objective and retinal motion separately. In order to provide optimal stimulation, we used stimuli with natural image statistics derived from Fourier scrambles of natural images. The results show that all regions increased responses with higher speeds for both, retinal and objective motion. V3A stood out in that it was the only region whose slope of the speed-response function for objective motion was higher than that for retinal motion. V6, V5/MT, MST and CSv did not differ in objective and retinal speed slopes, even though V5/MT and MST tended to respond more to objective motion at all speeds. These results reveal highly similar speed tuning functions for early and high-level motion regions, and support the view that human V3A encodes primarily objective rather than retinal motion signals.

5.2. Introduction

Motion perception is a crucial function of the visual cortex. Perceiving the speed of motion is important for many everyday tasks of the visual system, such as pursuing a target or interacting with a moving object, or even for compensating self-induced retinal motion with non-retinal signals. Although visual motion perception and processing have been widely studied, our understanding of speed processing is still limited. Most prior studies on speed processing focused on V5/MT both in humans and primates. Majority of V5/MT neurons were shown to be speed selective ((Maunsell & Van Essen, 1983b; Perrone & Thiele, 2001) but also see (Priebe, Cassanello, & Lisberger, 2003)) and V5/MT plays an important role in speed perception and discrimination, which was shown by lesion studies (Dursteler & Wurtz, 1988; Newsome, Wurtz, Dürsteler, & Mikami, 1985; Orban, Saunders, & Vandenbussche, 1995; Pasternak & Merigan, 1994; Yamasaki & Wurtz, 1991). Neurons are clustered according to preferred speed with no columnar organisation for speed in V5/MT (Liu & Newsome, 2003). The range of preferred speeds of neurons were reported differently in different studies, such as between 2-256 deg/s (Maunsell & Van Essen, 1983b) or between 5-150 deg/s (Rodman & Albright, 1987). The peak of preferred speeds across neurons was reported as 32 deg/s (Maunsell & Van Essen, 1983b) and speed representation in V5/MT is thought to be in logarithmic scale (Nover, Anderson, & DeAngelis, 2005). Speed perception is shown to be consistent with average responses from V5/MT neurons of macaques (M. M. Churchland & Lisberger, 2001; Priebe & Lisberger, 2004), although it has also been shown in monkeys that overall sensitivity of V5/MT neurons to speed discrimination is less than the behavioural responses (Liu & Newsome, 2005). Despite the vast number of studies to characterize speed responses of V5/MT neurons in nonhuman primates, our knowledge about the speed responses in human V5/MT is limited.

Previous studies in human psychophysics about speed discrimination show that humans can discriminate even small differences in speed when the speed is between 4-32 deg/s (Beauchamp, Cox, & DeYoe, 1997; Orban, de Wolf, & Maes, 1984). Speed coding in human V5/MT using an adaptation paradigm was shown by fMRI findings (Lingnau, Wall, & Smith, 2009) and application of TMS (transcranial magnetic stimulation) to human V5/MT induced deficits in speed perception, revealing an important role of V5/MT in speed perception in humans (McKeefry, Burton, Vakrou, Barrett, & Morland, 2008). Using fMRI, Chawla and colleagues found that optimal speed responses in human V5/MT were between 7 deg/s and 30 deg/s in one study (Chawla, Phillips, Buechel, Edwards, & Friston, 1998) and between 4 deg/s and 8 deg/s in another study (Chawla et al., 1999). Additionally, some researchers investigated how other image features effect the speed perception. For instance, in both humans and monkeys, it has been shown that changes in contrast result in both altered perception of speed and altered speed tuning, more specifically low-contrast causes lower speed perception and shifts in neural responses to lower speed tuning and firing rate (Krekelberg, Boynton, & van Wezel, 2006).

Some previous studies examined speed-dependent responses also in another motion responsive region, V3A (Arnoldussen, Goossens, & van den Berg, 2011; Chawla et al., 1999; Chawla et al., 1998; McKeefry et al., 2008). In macaques, almost all V3A neurons are reported to have speed sensitivity, which is present for a wide range of speeds, even faster than 50 deg/s (Galletti, Battaglini, & Fattori, 1990). In human, fMRI results showed that V3A has speed responses similar to MT, although in V3A, optimal speed range is reported to be between 4-16 deg/s (Chawla et al., 1999; Chawla et al., 1998) and application of TMS to V3A has been shown to impair speed perception (McKeefry et al., 2008). Particularly, V3A is responsive to headcentric speed of motion (Arnoldussen et al., 2011). Apart from its speed tuning, human V3A is just as sensitive to motion as V5+/MT+ (Bartels, Zeki, & Logothetis, 2008; Tootell et al., 1997). Importantly though, it differs substantially from the human V5/MT+ complex in that it is primarily driven by real-world motion, with nearly no response to retinal motion per se (Fischer, Bulthoff, Logothetis, & Bartels, 2012a). Consistent with this, in monkey, it contains a large fraction of real-motion neurons, similar to MST (Galletti & Battaglini, 1989; Galletti et al., 1990; Ilg, Schumann, & Thier, 2004).

All prior studies examining speed responses measured joint responses to objective and retinal motion, since participants fixated while speed of the (objective) background motion varied. This induced motion on the retina that was the result of physical motion on the screen. However, during eye- and head- movements, the visual system continuously compares retinal motion with efference copies of the eye, in order to estimate world-centered motion. This allows it to create a stable perception of the world despite retinal motion, or to infer object-motion also when eye-movements reduce retinal motion to zero. Previously, human visual areas V3A and V6 were shown to compensate for self-induced retinal motion and encode almost exclusively objective motion during smooth pursuit eye movements (Fischer et al., 2012a). To a weaker extent, also CSv was shown to discount self-induced retinal motion (Fischer, Bulthoff, Logothetis, & Bartels, 2012b). CSv responds to egomotion related visual signals (Cardin & Smith, 2010; Fischer et al., 2012b; Wall & Smith, 2008), heading direction (Furlan, Wann, & Smith, 2014), and vestibular signals (Smith, Wall, & Thilo, 2012).

V6 is located in parieto-occipital sulcus (POS) (Pitzalis et al., 2006), shows retinotopic organisation both in humans and macaque (Galletti, Fattori, Gamberini, & Kutz, 1999; Pitzalis et al., 2006; Pitzalis et al., 2010) with a stronger representation of periphery, although fovea is also represented (Pitzalis et al., 2006) and responds preferentially to wide field stimulation, and responds to ego-motion compatible coherent motion (Cardin & Smith, 2010; Pitzalis et al., 2006; Pitzalis et al., 2013; Pitzalis et al., 2010). V6, together with V3A, distinguishes real motion from self-induced retinal motion (Arnoldussen et al., 2011; Fischer et al., 2012a). Moreover, V6 is involved in processing of head-centric translation and rotation speed (Arnoldussen et al., 2011; Arnoldussen, Goossens, & van den Berg, 2015). For none of these regions speed responses have been measured separately for retinal and objective motion.

Here we used a pursuit paradigm that allowed us to do just that (Fischer et al., 2012a): participants fixated a disc that was either stationary or moved along a circular trajectory around the screen. At the same time, either a stationary background was shown or one that moved on the same circular trajectory. This 2 x 2 factorial design led to four conditions that allowed separating responses to retinal or objective motion in the absence of pursuit-related confounds. We applied this paradigm at six different levels of speed (1, 2, 4, 8, 16, and 24 degrees per second) in order to obtain speed-tuning profiles for retinal and objective motion for separately localized visual motion regions V5/MT, MST, V3A, V6 and CSv. The background stimulus was derived from Fourier scrambles of natural images in order to match natural image statistics. This provides a closer match to the spatio-temporal exposure in real-life and has been shown to provide more optimal stimuli for the visual system (Kayser, Kording, & Konig, 2004; Vinje & Gallant, 2000). Further, speed responses in MT were enhanced when the stimuli had multiple spatial frequencies and this was thought to be a result of natural scenes containing multiple spatial frequencies (Priebe et al., 2003; Priebe, Lisberger, & Movshon, 2006). We found that only in V3A, the slope of objective motion speed was siginificantly higher than the slope of retinal motion. Additionally, all regions were modulated by the speed of both types of motion.

5.3. Materials and Methods

Participants

20 healthy participants with normal or corrected vision (14 female, 7 male, 4 left handed, age between 18 and 37 (average: 26), 1 author) took part in this study. All participants gave written informed consent before the experiment and were compensated for their participation. The study was approved by the local ethics committee of the University Hospital of Tübingen.

Main Experiment

Stimuli and Paradigm

We applied a 2 x 2 factorial pursuit paradigm that allowed us to measure responses to objective and retinal motion separately (Fischer et al., 2012a). This paradigm was applied using 6 different levels of speed (1, 2, 4, 8, 16, and 24 degrees per second). The factors pursuit (on/off) and objective planar motion (on/off) resulted in the following 4 conditions: fixation on a static background, fixation on a moving background, pursuit on a static background, and pursuit on a moving background. In the last condition, the pursuit trajectory was locked to that of the background, thus nulling retinal motion.

The motion trajectory of the background and pursuit followed a circular path with a radius of 4 degrees (1/4th of the screen height, with screen dimensions of 22 x 16 degrees). The motion radius was chosen such that the area of controlled visual stimulation was maximal: the nearest border to the screen edge was at all times further away than 4 visual degrees, leading to controlled visual stimulation within at least 8 x 8 visual degrees (Fischer et al., 2012a). The rotation direction and starting point of the fixation cross was randomized and counter balanced within each participant. Note that the upper speed limit of 24 deg/s was set following piloting by the authors showing that this speed could still be reliably pursued on its circular trajectory, while considerably higher speeds could not.

For the background, 100 pink noise images were created from 100 natural scene images using Fourier phase scrambling. The images were converted to grayscale and contrast and luminance were matched across the images before phase scrambling. Stimuli were chosen randomly for each trial and back-projected with 1024 x 768 pixels resolution and 120 Hz refresh rate. The size of the images was 2048 x 1536 pixels, with 737.5 cd/m2 luminance and 147.5 root-mean-square (RMS) contrast.

Procedure

The experiment was presented in a block design manner; each condition lasted 12 s, and was preceded by 1 s fixation and followed by 2 s fixation, leading to an inter-trial interval of 15 s. The fixations at the beginning and end of each trials were presented in order to minimize confounds related to eye movements at the beginning of each trial since the location of starting point in each trial was randomized. Each condition was shown 4 times per run, yielding a total number of 17 blocks including one initial additional block for counterbalancing. We used back-matched pseudorandom sequences (Brooks, 2012) such that each condition was preceded equally often by all conditions. The initial block inserted for counterbalancing was discarded from further analyses. Each run started with 1.74 s grey screen with fixation and ended with 10 s of grey screen with fixation. There was a grey fixation disk (width: 0.74 degrees, luminance: 1153.7 cd/m2) with a fixation task (see below) at all times. The order of runs was also pseudorandomized for half of the participants and the flipped version of this run sequence was used for the other half.

In total six runs were acquired from each participant, each run containing one speed level. The sequence of speed levels was random and counterbalanced across participants.

The main experiment was presented using Psychtoolbox 3.0 (Brainard, 1997; Kleiner et al., 2007) and Matlab 7.10.0 (MATLAB, 2010).

Motion Localizer

In order to localize V5/MT, MST, CSv, V3A and V6, we used an independent motion localizer that was previously described (Fischer et al., 2012a, 2012b). It consisted of 7 conditions presented 12 s each in 7 counterbalanced repetitions: 3D full field motion (coherent expanding/contracting motion), random motion (with trajectories matched to 3D motion), right and left hemifield 3D full field motion (left or right 2/5th of the screen), 2D planar motion with synched pursuit, static background with pursuit, and static baseline. In each condition except baseline, random patterns of black and white dots on a grey background were used as stimuli. The baseline consisted of blank grey screen. As in the main experiment, there was a gray fixation disk (width: 0.74 degrees) with the fixation task (see below). The stimuli were presented using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN and Cogent Graphics developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience (http://www.vislab.ucl.ac.uk/ cogent_graphics.php) and Matlab 7.10.0 (MATLAB, 2010).

Region of interest (ROI) definitions

ROIs were defined using the MarsBar toolbox as follows: MST was defined as ipsilateral response within the V5+/MT+ complex using hemifield motion versus baseline contrast, V5/MT was defined as contralateral response during the same contrast excluding MST voxels. V6 and CSv were localized using 3D coherent motion versus random motion. V3A was localized using 2D lateral motion with pursuit versus smooth pursuit with static background. For each participant, each region was localized using an individual p-value, and when ROIs could not reliably be detected it was not defined at all (Fox, Iaria, & Barton, 2009; Murray & Wojciulik, 2004).

Fixation Task

In order to balance attention across conditions, participants were required to do a 1-back character-matching task, which was a randomly presented sequence of alphabetical characters displayed one at a time on the fixation disk. There was a repeating character between every 3 to 8 character presentations, which were reported by participants via button press. The timing of these button presses was included in the GLM analysis as a regressor.

Image Acquisition

Data were acquired with a Siemens Magnetom PRISMA 3 Tesla scanner using a 64-channel phased-array head coil (Siemens, Erlangen, Germany). A gradient echo sequence consisting of T2* weighted images with the following parameters was used for functional scans: TR = 0.87 s, TE = 30 ms, flip angle = 57°, Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) g-factor = 2, multi-band factor = 4, voxel size = 2 x 2 x 2 mm^3. 56 slices were acquired in an interleaved order. T1 weighted anatomical images were acquired with a resolution of 1 x 1 x 1 mm^3. The first 2 functional volumes of each run were discarded for T1 equilibration. Field map images were also acquired in order to correct for B_0 field distortions.

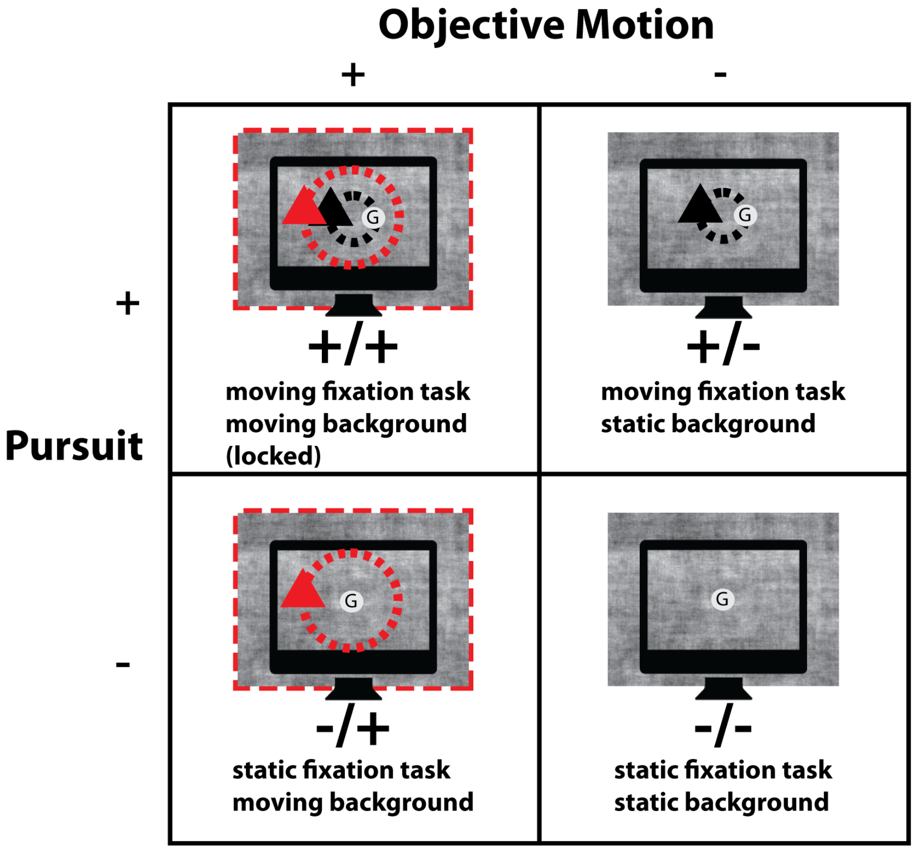


Figure 1: Four stimulus conditions presented in the main experiment, which were produced from a 2 x 2 factorial design with factors “pursuit” (on/off) and “objective motion” (on/off). The gray fixation disk with a one-back character-matching task was present at all times. Motion was planar on a circular trajectory. Movement of the fixation task resulted in pursuit whereas movement of the background on the same trajectory produced objective motion. The four conditions were as follows: +/+: moving fixation task with moving background (both trajectories were locked), -/+: static fixation task with moving background, +/-: moving fixation task with static background and -/-: both fixation task and background are static.

FMRI Data Preprocessing

Data were analysed using SPM8 (www.fil.ion.ucl.ac.uk/spm/) and MATLAB 8.4 (2014b). Preprocessing steps were as follows: the functional images were realigned and the anatomical image was coregistered to mean functional image. The anatomical image was also normalized to MNI space. Functional images were smoothed with a 4 mm full-width at half maximum Gaussian kernel.

Statistical Analysis

After preprocessing, each run of each participant (corresponding to one speed level) was analysed separately using the general linear model (GLM) approach. Regressors included one regressor for each of the four conditions, and one regressor for button presses. In addition, the following regressors of no interest were included: six motion realignment regressors, and an additional regressor of global mean signal, which was orthogonalized with respect to the conditions of interest in the design matrix (Desjardins, Kiehl, & Liddle, 2001; Van Dijk et al., 2010). High pass filtering with 128 s cut-off value was applied.

For each participant and each ROI, beta values were extracted for each condition and each speed level. For each participant, the resulting 24 (4 x 6) beta-values per ROI were z-normalized prior to contrast calculations and random effects ANOVA calculations: the mean was subtracted, followed by division by the standard deviation. This yielded 24 z-values for each participant and each ROI.

For each of the repeated measures ANOVAs conducted, Mauchly's sphericity test was applied and in case of violation of sphericity, Greenhouse-Geisser correction was used.

Statistical Contrasts

'Objective motion' was defined as the contrast of both conditions containing background motion versus both that did not, i.e. ((-/+) plus (+/+)) versus ((+/-) plus (-/-)). Note that pursuit was matched (equally present on either side).

'Retinal motion' was defined as both conditions containing retinal motion versus both that did not ((-/+) plus (+/-)) versus ((+/+) plus (-/-)). Again pursuit was matched.

'Motion diff': the *difference* between objective and retinal motion boils down to ((+/+) versus (+/-)) that again has matched pursuit conditions. This difference reliably and near-exclusively activates V3A in every participant (Fischer et al., 2012a).

'Motion sum' is the sum of objective and retinal motion ((-/+) versus (-/-)). This equals moving versus static background during fixation, and corresponds to the contrast used by previous motion studies, including those on speed processing. The responses to this contrast were calculated only for visualization and as a reference.

‘Pursuit’ was defined as ((+/+) plus (+/-)) versus ((-/+) plus (-/-)). This contrast contains several poorly controlled contributing factors, i.e. in addition to pursuit-controlling neural mechanisms also peripheral retinal motion beyond the projection screen that is not controlled, and a higher fixational error compared to non-pursuit (Fischer et al., 2012a). This contrast was therefore not analysed further, but included in illustrations as a reference.

The above contrasts were calculated for each speed separately, using the z-normalized beta-values.

Eye Tracking

An infrared camera based eye tracker (Eye-Trac 6; Applied Science Laboratories) was used in order to record eye position during the experiment together with Viewpoint Eyetracker software (Arrington Research, Scottsdale, USA) with a sampling rate of 60 Hz. Eye tracking data was analysed by the following steps: preprocessing was done by removing blinks and smoothing with a 200 ms running average window. The data was grouped according to the conditions. Root mean square error (RMSE) of eye position relative to the fixation disk was used to calculate fixation accuracy. This calculation was done separately for each condition, participant and speed. The same tests as conducted for the ROI data were applied to eye-tracking data.

5.4. Results

We explored the responses to 6 different levels of speed, separately for objective motion and retinal motion, in independently localized motion responsive ROIs V3A, V6, V5/MT, MST and CSv.

For reference, Figure 2 plots values for the sum of objective and retinal motion. The sum (i.e. moving versus static background during fixation) corresponds to the contrast used in previous studies on motion and speed processing (e.g. (Chawla et al., 1999; Chawla et al., 1998).

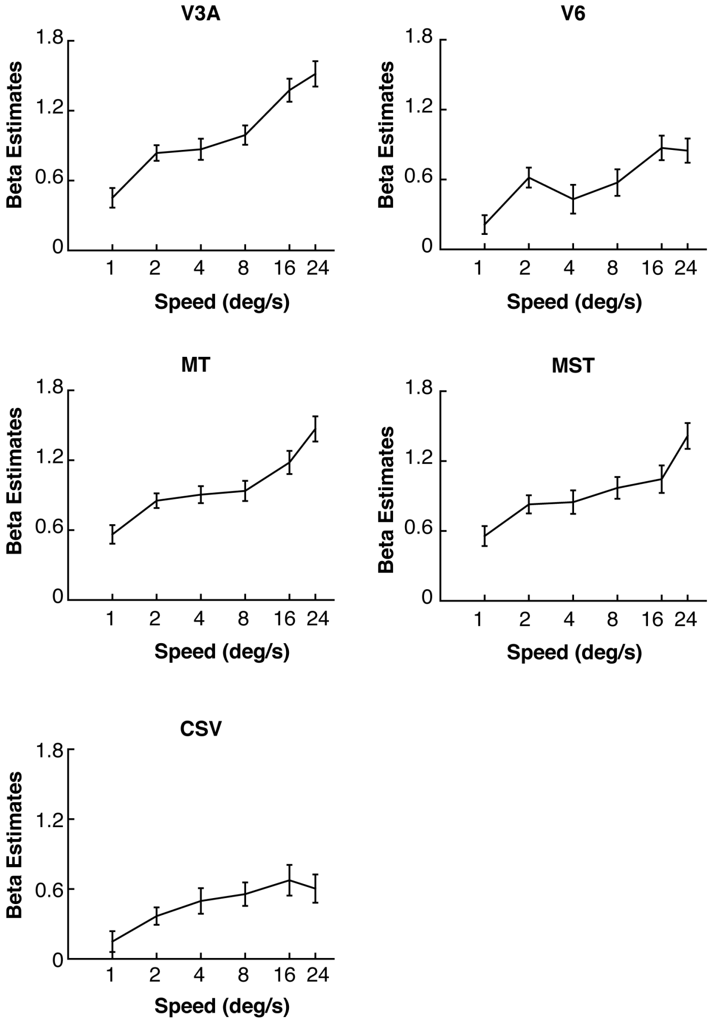


Figure 2. Contrast for 'motion sum'. (-/+) vs. (-/-), as a function of speed. Here, values for moving background vs. static during fixation as a function of speed are shown for each ROI separately. This contrast would result in the sum of objective and retinal motion and was used in previous studies. Note that all data are z-normalized. Error bars show standard error of mean (SEM).

Figure 3 plots contrast values for objective and retinal motion (see methods for contrast definitions) as a function of speed. Contrast values were calculated for each speed separately using z-normalized beta-values (see methods). The difference between objective and retinal motion, i.e. preference for objective over retinal motion, has previously been shown to strongly activate V3A (Fischer et al., 2012a).

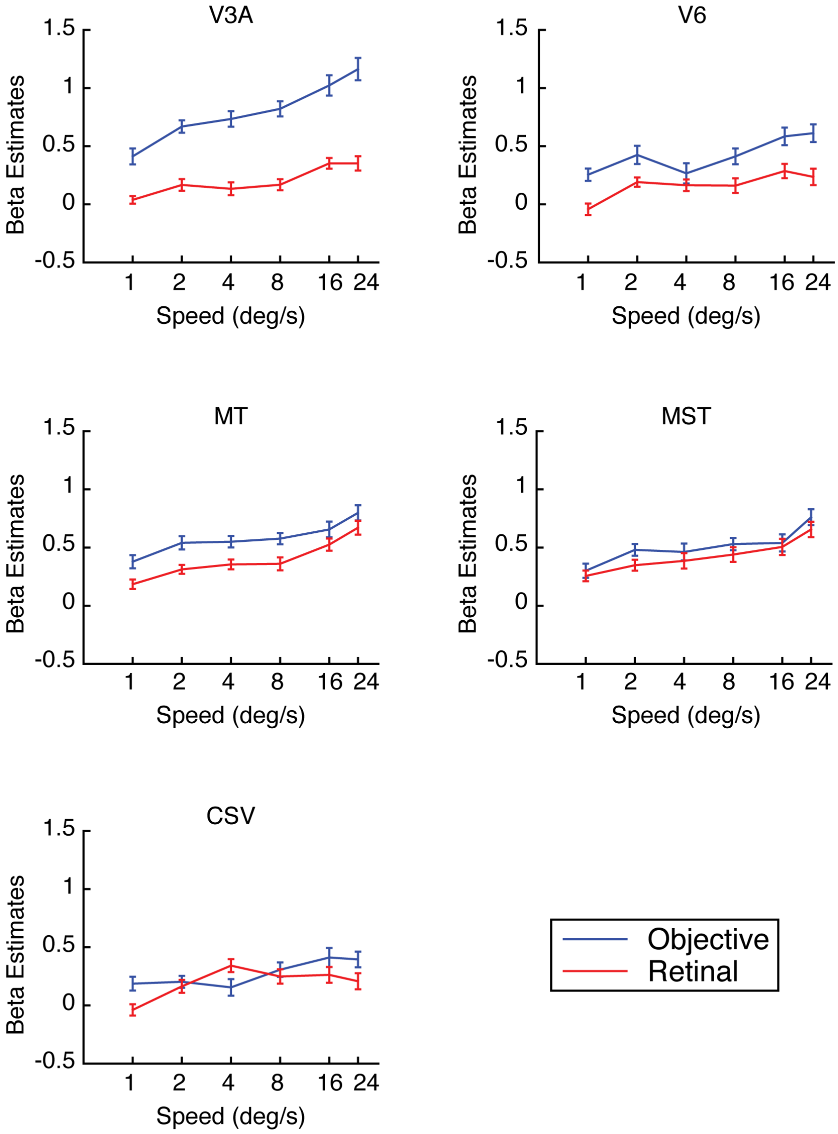


Figure 3. Contrast values for objective and retinal motion as a function of speed. Note that all data are z-normalized. Error bars show standard error of mean (SEM).

The results plotted in Figure 3 suggest three main findings. First, all regions tended to increase their responses with increasing speeds for both, objective and retinal motion. Second, V3A, V6 and to a lesser extent also V5/MT and MST had a higher overall response (i.e. offset) of objective relative to retinal motion. Third, the *slope* of the objective and retinal speed respons differed in V3A, with a steeper slope for objective motion.

In order to quantify these raw observations, we modelled each response using a simple linear fit to obtain a measure for the slope (b) and one for the offset (a), i.e. y = b*x +a, with y being the normalized contrast values, and x being the speed and it was scaled between -0.5 and +0.5. Previous single-unit recording studies for V5/MT reported a logarithmic speed tuning in motion responsive regions (Nover et al., 2005; Priebe et al., 2003), and, compatible with those, prior fMRI speed tuning studies also found log-like tuning in the speed range tested here (Chawla et al., 1999).

We hence directly tested two regression models: one modelled linear speed levels (x-axis: 1, 2, 4, 8, 16, 24 deg/s) and the other logarithmic scaling of speed (x-axis: 0, 1, 2, 3, 4, 4.585). We then tested which one provided a better fit by calculating the correlation coefficient for each, and comparing them in paired t-tests (after Fischer-Z transforming r-values to z-values) for each ROI across participants (Table 1).

Table 1: Results of t tests comparing the fitness of regression models between linear scale and logarithmic scale of speed. For objective motion, retinal motion and pursuit, separately, we fitted regression models using either linear scale (1, 2, 4, 8, 16, 24 deg/s) or logarithmic scale (0, 1, 2, 3, 4, 4.585 deg/s) for speed. Next, for each ROI and motion type, we compared the correlation coefficients using t tests. P values that are significant after Bonferroni correction are shown in bold.

p values	**CSV**	**V3A**	**V5/MT**	**MST**	**V6**
objective	0.77 (t(37) = -0.3)	0.328 (t(37) = -1)	0.157 (t(39) = -1.4)	0.155 (t(37) = -1.5)	0.189 (t(37) = -1.3)
retinal	0.322 (t(37) =1.)	0.823 (t(37) = -0.23)	0.964 (t(39) = -0.05)	0.984 (t(37) = -0.02)	**0.00056** (t(37) = 3.8)
pursuit	0.77 (t(37) = 0.3)	0.027 (t(37) = 2.3)	0.326 (t(39) = 1)	0.525 (t(37) = 0.6)	0.954 (t(37) = 0.06)

After applying Bonferroni correction for 15 comparisons, the only significant result was for V6, showing that retinal motion was fitted better by logarithmic scaling.

Judging from the low t-values for all other tests, there was no clear preference for either type of scaling, even though objective motion values tended towards linear scaling for all ROIs. We used logarithmic scaling of speed for the rest of the analysis.

Next, we tested regression slopes for each motion type and every ROI against zero, and compared regression slopes between objective and retinal motion for each ROI using paired t-tests. Independently of this, we also tested the mean signal (offset) in the same way. Figure 4A shows the regression slopes for objective and retinal motion for all ROIs, figure 4B mean signals. All t-test results were corrected for 15 comparisons using Bonferroni-Holm correction. As shown in table 2, the slope for objective motion was significantly different than zero for all, whereas for retinal motion, it was only significant for V5/MT and but not for or V6. The comparison of objective and retinal motion regression slopes was only significant for V3A. As shown in table 3, mean responses to objective and retinal motion were significant for all ROIs and V3A, V5/MT and V6 had significantly higher mean responses to objective motion than retinal motion.

Table 2. Paired t-test results for slopes

p values	**CSV**	**V3A**	**V5/MT**	**MST**	**V6**
Objective slope	**0.0001** (t(37)=4.33)	**$0.3 * 10^{-9}$** (t(37)=8.38)	**$0.2 * 10^{-6}$** (t(39)=6.26)	**$0.2 * 10^{-5}$** (t(37)=5.59)	**$0.2 * 10^{-4}$** (t(37)=4.87)
Retinal slope	0.005 (t(37)=2.95)	**$0.2 * 10^{-4}$** (t(37)=4.87)	**$0.7 * 10^{-7}$** (t(39)=6.61)	**$0.2 * 10^{-4}$** (t(37)=4.93)	0.004 (t(37)=3.03)
Objective slope vs. retinal slope	0.59 (t(37)=0.54)	**0.0009** (t(37)=3.58)	0.24 (t(39)=-1.19)	0.99 (t(37)=-0.001)	0.4 (t(37)=0.85)

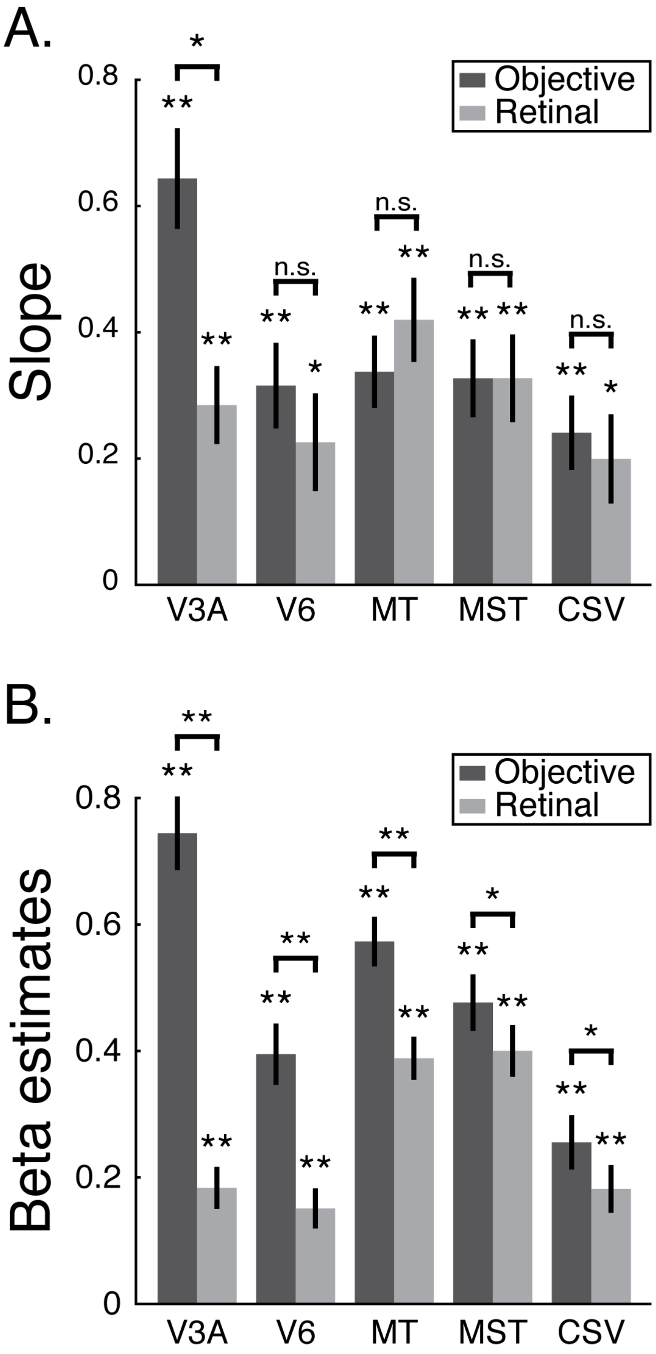


Figure 4: Regression slopes and mean signal for objective and retinal motion. (A) Regression slopes calculated for objective and retinal motion, for each ROI separately. Every ROI was modulated by both objective motion and retinal motion speed. Only in V3A, objective motion speed slope was significantly higher than the slope of retinal motion speed. (B) Mean signal (offset) for objective and retinal motion. For every ROI, the mean objective motion and mean retinal motion was significantly greater than zero and mean objective motion was significantly higher than mean retinal motion. Note that all data are z-normalized. ** $p<0.001$, * $p<0.05$, Bonferroni-Holm corrected. Error bars show standard error of mean (SEM).

Table 3. Paired t-test results for mean responses

p values	CSV	V3A	V5/MT	MST	V6
Mean objective	**$0.1 * 10^{-6}$** (t(37)=6.47)	**$0.3 * 10^{-15}$** (t(37)=13.51)	**$0.9 * 10^{-18}$** (t(39)=16.05)	**$0.4 * 10^{-13}$** (t(37)=11.53)	**$0.9 * 10^{-10}$** (t(37)=8.76)
Mean retinal	**$0.5 * 10^{-5}$** (t(37)=5.32)	**$0.3 * 10^{-6}$** (t(37)=6.16)	**$0.2 * 10^{-14}$** (t(39)=12.69)	**$0.4 * 10^{-12}$** t(37)=10.67)	**$0.4 * 10^{-5}$** (t(37)=5.36)
Mean objective vs mean retinal	0.01 (t(37)=2.71)	**$0.2 * 10^{-11}$** (t(37)=10.01)	**$0.1 * 10^{-4}$** (t(39)=5.09)	p=0.04 (t(37)=2.13)	**$0.7 * 10^{-7}$** (t(37)=6.64)

Behavioral Data

During the experiment, participants were required to perform a character back-matching task at all times. Average rate of correct responses was 0.83 ± 0.04 (mean ± std) and mean response time was 0.52 ± 0.12 s (mean ± std Similar to fMRI data analysis, for each participant, we first calculated correct response rate during objective motion and retinal motion separately. Next, for each participant, we fit two separate GLMs to the correct response data with 2 regressors each; one for the speed and the other being all ones. The speed regressor was first calculated in logarithmic scale (0, 1, 2, 3, 4, 4.585 deg/s) and then scaled between -0.5 and +0.5. This way the first regressor modeled the slope while the second one was modeling for the mean correct response. For each participant, we calculated slope of correct responses during objective and retinal motion speeds separately.

Using paired t-tests we tested the correct response rate slopes during objective and retinal motion. We also did additional t-tests to compare objective and retinal motion slopes. None of the test results reached significance (objective motion: t (19) = 1.41, p = 0.176; retinal: t (19) = 0.08, p = 0.936; objective vs. retinal: t (19) = 0.54, p = 0.594). The slopes for task responses during objective and retinal motion speed are shown in figure 5A.

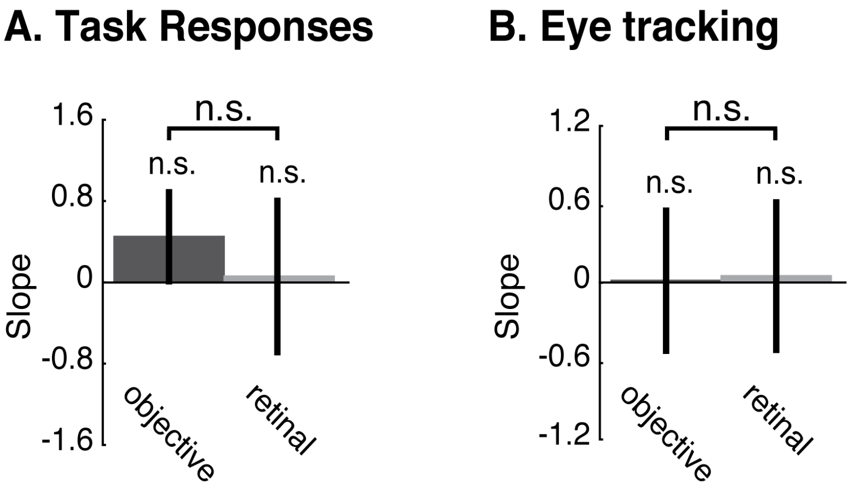


Figure 5. Regression slopes for behavioural and eye tracking data. (A) Regression slopes calculated for correct task responses during objective motion speed contrast and retinal motion speed contrast separately. There is no significant effect of speed during retinal and objective motion and there is no significant difference between task performance during objective and retinal motion speed. (B) Regression slopes calculated for RMSE of eye position relative to the fixation disk during objective motion speed and retinal motion contrast separately. Here, there is no significant effect of objective motion speed or retinal motion speed on RMSE of eye position and there is no significant difference between objective and retinal motion speed on RMSE of eye position. Error bars show standard error of mean (SEM).

Eye Tracking Data

We collected eye-tracking data for 13 of the 20 participants. After preprocessing of the data, we calculated RMSE of eye position relative to the fixation disk and used this for comparing fixation accuracy across conditions and different levels of speed. The mean RMSE ± SEM across participants for each speed and condition is shown in table 4.

Table 4: RMSE of eye position for each condition and speed.. Values illustrate mean ± standard error of mean (SEM)

Conditions	**1 deg/s**	**2 deg/s**	**4 deg/s**	**8 deg/s**	**16 deg/s**	**24 deg/s**
(+/+)	2.24 ± 0.30	2.02±0.21	2.13±0.21	2.44±0.29	2.83±0.19	3.65±0.15
(-/+)	1.35 ± 0.21	1.57±0.18	1.57±0.24	1.28±0.22	1.43±0.25	1.49±0.19
(+/-)	2.01 ± 0.22	2.26±0.25	2.28±0.27	2.25±0.21	2.86±0.20	3.56±0.10
(-/-)	1.75 ± 0.31	1.31±0.18	1.66±0.25	1.23±0.18	1.57±0.29	1.44±0.22

For each participant, using RMSE of eye position, we calculated regression slopes for objective and retinal motion speed separately. We calculated RMSE of eye position during objective and retinal motion for each participant separately and fit GLMs to calculate slopes, similar to fMRI and behavioral data analyses.

Next, we did paired t tests to test the significance of objective and retinal motion slopes as well as to compare objective and retinal motion slopes. None of t tests yielded a significant result (objective motion: $t(12) = 1.13$, $p = 0.28$; retinal motion: $t(12) = 0.18$, $p = 0.859$; objective vs. retinal motion: $t(12) = 0.56$, $p = 0.589$). The slopes for eye position during objective and retinal motion are shown in figure 5B.

5.5. Discussion

Many motion processing regions contain neurons that differentiate between world-centered (objective) and eye-centered (retinal) motion (Dicke, Chakraborty, & Thier, 2008; Erickson & Thier, 1991; Galletti et al., 1990; Ilg et al., 2004; Zhang, Heuer, & Britten, 2004). Correspondingly, recent fMRI studies showed that net BOLD responses in human motion regions can vary strongly by their preferences to both types of motion (Arnoldussen et al., 2011; Fischer et al., 2012a). Speed dependency of either type of motion has never been examined before. Prior speed studies (that measured that sum of both motion types) are over 15 years old, used only 3 participants, and no distraction task, and may contain attention-speed interactions, and potentially low-level motion-streak effects due to older projection technology (Chawla et al., 1999; Chawla et al., 1998). Finally, speed tuning is entirely unknown for several regions that only recently came into the spotlight of human fMRI research, such as CSv and V6.

Hence, in this study, we re-addressed the question of speed dependent responses across independently localized visual motion areas V3A, V6, V5/MT, MST and CSv, in 20 participants. We used stimuli with natural image statistics in order to stimulate the motion system with stimulus properties it evolved for. In addition, we used a pursuit paradigm to address this question that enabled us to separate objective and retinal motion while matching eye movements across both types of motion (Fischer et al., 2012a). Finally, we used a projector with of 120 Hz refresh rate that had been specifically selected for low motion-streak effects.

We found that all regions increased responses monotonically as a function of speeds for both, retinal and objective motion, for speeds between 1-24 deg/s. In contrast to prior fMRI studies, we did not observe an inversed 'u' shape response for either type of motion. This was most likely due to our upper limit of 24 deg/s, whereas electrophysiology studies show that in many motion regions optimal speed tuning is between 32 deg/s or far beyond (see below). We were constrained here to 24 deg/s, as beyond this reliable pursuit performance would have been difficult to achieve.

The responses in all regions were fit well using either a linear or logarithmic speed-tuning function. The slopes of the speed-response functions did not differ between objective and retinal motion in all ROIs, except for V3A. V3A stood out in that its objective motion slope was significantly higher than that for retinal motion. In V3A, also the mean response to objective motion was by far higher than that to retinal motion, with a significant difference also being present in V6, V5/MT, and MST and CSv, in descending order.

One could argue that the speed responses here could have resulted from eye movement related biases between conditions. It is hence important to note that for both, objective and retinal motion, pursuit-versus-non-pursuit conditions were entirely matched and cancelled out. Moreover, our eye tracking and behavioral data results show no confound; there were no effects of speed during objective or retinal motion.

To our knowledge this is the first systematic study of speed tuning conducted using modern imaging standards, eye-tracking, a demanding attention task, and including human motion regions V6 and CSv. In addition, it is the first to segregate objective and retinal motion speed tuning.

Speed Encoding in V5/MT and MST

In this study, V5/MT and MST showed significant speed tuning to both objective and retinal motion. Moreover, neither V5/MT nor MST seemed to differentiate between speed of objective motion and speed of retinal motion in terms of response slopes, but both regions had marginally higher overall responses to objective motion.

As shown by single neuron recordings and lesion studies, motion responsive regions V5/MT has a role in speed perception and processing of speed related motion signals as most MT neurons are speed responsive (Dursteler & Wurtz, 1988; Krekelberg et al., 2006; Liu & Newsome, 2005; Maunsell & Van Essen, 1983b; Newsome et al., 1985; Orban, Saunders, et al., 1995; Pasternak & Merigan, 1994; Perrone & Thiele, 2001; Yamasaki & Wurtz, 1991). Moreover, different studies reported different preferred range of speed in monkey V5/MT neurons, for instance between 2 and 256 deg/s (Maunsell & Van Essen, 1983b) or 5-250 deg/s (Rodman & Albright, 1987) while the optimal speed is shown to be around 32 deg/s in many studies (Cheng, Hasegawa, Saleem, & Tanaka, 1994; Mikami, Newsome, & Wurtz, 1986; Newsome, Mikami, & Wurtz, 1986). Our results therefore fit to the previous findings since we do not see any optimal range of speed and only observe monotonic increase in speed responses when tested between 1 to 24 deg/s. Testing higher speed ranges might provide the optimal speed range in human.

Similarly, single neuron studies in monkey MST also showed speed tuning in these neurons (Duffy & Wurtz, 1997). These authors tested speed tuning of MST neurons for radial and circular motion types and found that in about 2/3rd of the tested neurons, responses were modulated by speed when tested for speed range between 10-80 deg/s (Duffy & Wurtz, 1997). Previous studies reported that MST neurons preferred high stimulus speeds both during laminar motion and optic flow (A. K. Churchland & Lisberger, 2005; Duffy & Wurtz, 1997; Kawano, Shidara, Watanabe, & Yamane, 1994; Orban, Lagae, Raiguel, Xiao, & Maes, 1995; Tanaka & Saito, 1989). While Orban and colleagues reported the optimal speed range of MST neurons between 15-20 deg/s (Orban, Lagae, et al., 1995), Duffy and colleagues found that approximately 40% of the neurons tested showed increasing response profile with increasing speed (Duffy & Wurtz, 1997). Additionally, Kawano and colleagues reported the optimal speed for MST single neuron responses as 160 deg/s (Kawano et al., 1994). These results reveal that while there seems to be no consensus on optimal speed range of MST neurons, it is possible that similar to the case in V5/MT, we also did not reach optimal speed range in MST.

Single neuron studies showed that MT processes retinal signals whereas MST responses take both retinal and extraretinal signals into account (Erickson & Thier, 1991; Galletti & Fattori, 2003; Ilg & Thier, 2003). In monkey, a large fraction of neurons in MSTd have been shown to compensate for speed of pursuit during eye movements whereas neurons in V5/MT are generally more responsive to retinal motion (Chukoskie & Movshon, 2009; Inaba, Miura, & Kawano, 2011; Lee, Pesaran, & Andersen, 2007; Shenoy, Crowell, & Andersen, 2002). For instance, MSTd neurons were shown to compensate for pursuit speed when tested for pursuit the speed of 2.58, 5.05 and 9.22 deg/s (Shenoy et al., 2002). Moreover, the same study showed that MSTd uses both and extraretinal input to compensate for pursuit, since the preferred pursuit and retinal motion directions are opposite of each other (Shenoy et al., 2002). Similarly, Lee and colleagues found that MSTd neurons also compensate for changes in translation speed (Lee et al., 2007).

Perceptual effects

The question thus arises why here V5/MT showed higher mean objective motion compared to retinal motion. One possible account may have to do with perceptual effects, which, in contrast to spiking activity, are often dominantly reflected in fMRI signal (Bartels, Logothetis, & Moutoussis, 2008). In humans, V5/MT+, as well as V3A, was previously shown to have a role in speed perception (McKeefry et al., 2008) and V5/MT+ was shown to have headcentric speed tuning during optic flow stimuli (Arnoldussen et al., 2011). Also previous electrophysiology studies showed that MT responses correlate with speed perception (Krekelberg & van Wezel, 2013; Liu & Newsome, 2005). Additionally, the speed of background motion during eye fixation has been shown to be perceived higher than the speed of retinal motion of smooth pursuit eye movements on a static background, even though the retinal speed induced in both conditions are exactly the same (Aubert-Fleischl effect) (Aubert, 1886; Dichgans, Körner, & Voigt, 1969). Regarding this, it is possible that in the present study, the speed during objective and retinal motion conditions were perceived differently, meaning the speed during (-/+) condition was perceived higher than the speed during (+/-) condition. Additionally, the visual system is thought to be more tolerable to retinal motion in the opposite direction of the pursuit, since this type of motion is natural during pursuit and could be arising from the eye movements itself and less tolerable to retinal motion in the same direction as pursuit, as this could indicate 'real' object motion (Lindner, Schwarz, & Ilg, 2001). This was shown by previous studies; when brief injections of background motion was introduced during smooth pursuit eye movements, it only caused an increase in eye velocity when it is in the same direction as pursuit, but did not cause any increase when it is in opposite direction (Lindner et al., 2001). Additionally, V5/MT and MST are densely connected to each other and MST is higher in the visual processing hierarchy than V5/MT (Boussaoud, Ungerleider, & Desimone, 1990; Maunsell & Van Essen, 1983a; Ungerleider & Desimone, 1986). Another possibility is that feedback from MST to V5/MT could be resulting in higher objective motion responses compared to retinal motion responses.

Relation to prior fMRI studies

Two fMRI studies showed that when combined objective/retinal motion speed was used with the range of speed between 3.7 and 61.6 degrees and between 1 and 32 degrees, V5/MT (and V3A) showed nonlinear (inverted U shape) speed responses and their optimal speed range was between 7 deg/s and 30 deg/s and between 4 deg/s and 8 deg/s in two separate studies (Chawla et al., 1999; Chawla et al., 1998).

However, our results disagree with the optimal speed ranges reported by Chawla and colleagues (Chawla et al., 1999; Chawla et al., 1998). According to their reports, speed range used in the present study, which is between 1-24 deg/s, should be more or less the optimal range to result in inverted U-shape speed responses, similar to their findings. Although some previous single neuron studies report a similar inverted U-shaped speed response, the range of speeds reported in those studies are much wider and the for the majority of cells, optimal speed is considerably higher than those reported by Chawla and colleagues (Chawla et al., 1999; Chawla et al., 1998; Cheng et al., 1994; Mikami et al., 1986; Rodman & Albright, 1987). One can think the stimuli used as one main difference. In both mentioned studies, Chawla and colleagues used dots moving radially from centre towards the edges of the screen (optic expansion flow) (Chawla et al., 1999; Chawla et al., 1998). However, they do not report how many dots were implemented or whether new dots appeared at the centre of the screen. Inherent to any expansion flow display is that dot-appearances, typically foremost near the centre of the screen, scale with speed. In this case, the inverted-U shape could simply be an artefact of temporal frequency tuning. It is well known that V5/MT is responsive to flickering static stimuli (Malonek, Tootell, & Grinvald, 1994; Tootell et al., 1995). Further, in a previous fMRI study, Singh and colleagues (Singh, Smith, & Greenlee, 2000) investigated spatial and temporal frequency tuning of visual areas in human brain and found a very similar inverse-U shape response tuning for temporal frequency. Additionally, Chawla et al. studies did not have any attention task (apart from fixation) and this could lead to potential attentional load differences across speed (Chawla et al., 1999; Chawla et al., 1998). Previously, attention to speed was shown to strengthen the responses in MT (Beauchamp et al., 1997). One should also keep in mind that mere data quality, field-strength, and standards for number of participants (3 participants vs. 20) have changed dramatically (Chawla et al., 1999; Chawla et al., 1998).

Speed Encoding in V3A

It is well known that V3A is a part of motion processing network (Galletti et al., 1990; Tootell et al., 1997). In humans, V3A has the second highest motion responses after MT and motion responses in human V3A are more similar to that of monkey V3, not V3A (Tootell et al., 1997). However, speed tuning of V3A is not extensively studied for different types of motion. In humans, attention to speed of motion activated V3A

(Sunaert, Van Hecke, Marchal, & Orban, 2000). In the present study, V3A responses show speed tuning for both objective and retinal motion types. These results are in accord with previous studies that showed that V3A have a crucial role for perceiving stimulus speed (McKeefry et al., 2008; Pitzalis, Strappini, De Gasperis, Bultrini, & Di Russo, 2012). It is possible that we explored only the lower end of the speed range, since V3A neurons in macaque monkey are shown to be sensitive to a wide range of speeds and they are even activated at speeds higher than 50 deg/s (Galletti et al., 1990).

Previous studies on single neuron responses showed that V3A neurons showed real-motion preference (Galletti et al., 1990) and approximately half of the V3A neurons in monkey were found to be gaze dependent (Galletti & Battaglini, 1989). In this study, V3A was the only region to significantly differentiate between the speed of objective and retinal motion. Moreover, V3A prefers objective motion to retinal motion at all speeds. This is compatible with previous studies about real motion responses in V3A (Fischer et al., 2012a; Galletti & Fattori, 2003). Even during 3D motion, V3A has been shown to have strong self- motion responses and encodes headcentric speed of rotation (Arnoldussen et al., 2011, 2015). We conclude that during 2D motion, V3A is speed tuned to both objective and retinal motion, and it can differentiate between the speed of objective motion and retinal motion.

Speed Responses in V6

V6 is a motion responsive region with large-field responses, and in particular responds to egomotion compatible motion such as 3D flow (Cardin & Smith, 2011; Fattori, Pitzalis, & Galletti, 2009; Pitzalis et al., 2006; Pitzalis et al., 2013; Pitzalis et al., 2010). V6 neurons in macaque respond to a wide variety of speeds, reported between 0 deg/s to 900 deg/s (Galletti, Fattori, Battaglini, Shipp, & Zeki, 1996) and a previous study reported that it showed higher responses for the fast speed when tested for 3 deg/s and 25 deg/s (Pitzalis et al., 2012). Further, Arnoldussen et al. (Arnoldussen et al., 2011) showed that V6, together with V3A and MT+, is encoding headcentric speed of motion. We found that V6 showed speed-tuned responses to objective motion, and to a lesser extent to retinal motion and that it did not show differential responses between objective and retinal motion speed tuning.

Similar to V3A, V6 is also responsive to real motion (Fischer et al., 2012a; Galletti & Fattori, 2003). V6 is thought to be encoding extrapersonal space since majority of neurons in macaque V6 are reported to be eye position sensitive and some neurons are reported to be ‘real-position’ cells (Galletti, Battaglini, & Fattori, 1995). In the present study, we used 2D planar motion, whereas most of the previous studies mentioned here used 3D coherent motion. When 2D planar motion was used, V6 responses were supressed during retinal motion (Fischer et al., 2012a). Objective motion preference found in V6 in this study is in accord with egocentric / headcentric motion preference of V6.

In humans, V6 is shown to prefer near field stimuli (Quinlan & Culham, 2007). Together with this and its connections to V6A and other parietal regions that have a role in grasping and to visual cortex (Galletti et al., 2001), V6 is thought to have a role in encoding the movement of graspable objects (Galletti & Fattori, 2003; Galletti et al., 2001). Being able to differentiate objective and retinal motion, as well as preference for high speeds, can both be explained by the preference for near visual field and encoding of graspable object motion. Objects that are closer to us seem to move faster than objects that are further away. Thus, the higher speed preference in V6 is in line with its role in encoding for graspable objects that are in peripersonal space.

Speed Responses in CSv

CSv is a recently defined visual region in dorsal posterior cingulate sulcus (dPCC) that is responsive to complex motion and it is specialized in processing self motion related signals and parsing optic flow (Antal, Baudewig, Paulus, & Dechent, 2008; Cardin & Smith, 2010; Fischer et al., 2012b; Wall & Smith, 2008). Although there are studies in macaque posterior cingulate sulcus showing visual responses (Dean, Crowley, & Platt, 2004), it is not clear whether there is a homologue of human CSv in primates. Thus, there are not so many electrophysiology studies that are directly comparable.

Previously CSv has been shown to monitor eye position during saccades and smooth pursuit eye movements (Olson, Musil, & Goldberg, 1996). CSv was also shown to compensate for self-induced retinal eye movements (Fischer et al., 2012b). Our results regarding higher mean objective motion responses compared to mean retinal motion responses in CSv are consistent with these studies. A recent fMRI study, which was conducted using 3D optic flow motion, reported no significant speed tuning for retinal or headcentric motion in CSv, whereas there was a significant speed tuning response for pursuit (Arnoldussen et al., 2011). While our results seem to disagree with those findings, it is possible that eye position related signals used for compensation of eye movements could be involved differently in calculations regarding planar motion and more complex motion such as heading related optic flow signals (for a detailed discussion, please see (Fischer et al., 2012a)).

To our knowledge, this is the first study to systematically investigate the speed tuning in CSv. We found that CSv showed significant speed tuning for both objective motion and to a lesser extent for retinal motion, but its speed tuning did not differentiate between these two types of motion. Although the mean objective motion response in CSv was significantly higher than the mean retinal motion response, this was not consistent cross all speeds (Figure 3).

5.6. Conclusion

In conclusion, our results provide the speed tuning of motion responsive regions during objective and retinal motion. Furthermore, V3A is the only motion responsive region that shows different speed tuning to objective and retinal motion, even though all motion responsive regions we investigated show speed tuning to objective motion. These results support the view that human V3A encodes primarily objective rather than retinal motion signals even during a range of motion speeds.

Acknowledgements

This work was funded by the Centre for Integrative Neuroscience Tübingen through the German Excellence Initiative (EXC307) and by the Max Planck Society, Germany.

Conflict of Interest: The authors declare no competing financial interests.

5.7. References

Antal, A., Baudewig, J., Paulus, W., & Dechent, P. (2008). The posterior cingulate cortex and planum temporale/parietal operculum are activated by coherent visual motion. *Visual neuroscience, 25*, 17-26. doi: 10.1017/S0952523808080024

Arnoldussen, D. M., Goossens, J., & van den Berg, A. V. (2011). Adjacent visual representations of self-motion in different reference frames. *Proceedings of the National Academy of Sciences of the United States of America, 108*(28), 11668-11673. doi: 10.1073/pnas.1102984108

Arnoldussen, D. M., Goossens, J., & van den Berg, A. V. (2015). Dissociation of retinal and headcentric disparity signals in dorsal human cortex. *Frontiers in Systems Neuroscience, 9*, 16. doi: 10.3389/fnsys.2015.00016

Aubert, H. (1886). Die bewegungsempfindung. *Pflügers Archiv European Journal of Physiology, 39*(1), 347-370.

Bartels, A., Logothetis, N. K., & Moutoussis, K. (2008). fMRI and its interpretations: an illustration on directional selectivity in area V5/MT. *Trends in neurosciences, 31*, 444-453. doi: 10.1016/j.tins.2008.06.004

Bartels, A., Zeki, S., & Logothetis, N. K. (2008). Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. *Cerebral cortex, 18*, 705-717. doi: http://dx.doi.org/10.1093/cercor/bhm107

Beauchamp, M. S., Cox, R. W., & DeYoe, E. A. (1997). Graded effects of spatial and featural attention on human area MT and associated motion processing areas. *Journal of Neurophysiology, 78*(1), 516-520.

Boussaoud, D., Ungerleider, L. G., & Desimone, R. (1990). Pathways for Motion Analysis - Cortical Connections of the Medial Superior Temporal and Fundus of the Superior Temporal Visual Areas in the Macaque. *Journal of Comparative Neurology, 296*(3), 462-495.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision, 10*(4), 433-436.

Brooks, J. L. (2012). Counterbalancing for serial order carryover effects in experimental condition orders. *Psychological Methods, 17*(4), 600-614. doi: 10.1037/a0029310

Cardin, V., & Smith, A. T. (2010). Sensitivity of human visual and vestibular cortical regions to egomotion-compatible visual stimulation. *Cerebral cortex, 20*(8), 1964-1973. doi: 10.1093/cercor/bhp268

Cardin, V., & Smith, A. T. (2011). Sensitivity of human visual cortical area V6 to stereoscopic depth gradients associated with self-motion. *Journal of neurophysiology, 106*(3), 1240-1249. doi: 10.1152/jn.01120.2010

Chawla, D., Buechel, C., Edwards, R., Howseman, A., Josephs, O., Ashburner, J., & Friston, K. J. (1999). Speed-dependent responses in V5: A replication study. *Neuroimage, 9*(5), 508-515.

Chawla, D., Phillips, J., Buechel, C., Edwards, R., & Friston, K. J. (1998). Speed-dependent motion-sensitive responses in V5: an fMRI study. *Neuroimage, 7*(2), 86-96.

Cheng, K., Hasegawa, T., Saleem, K. S., & Tanaka, K. (1994). Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *Journal of Neurophysiology, 71*(6), 2269-2280.

Chukoskie, L., & Movshon, J. A. (2009). Modulation of visual signals in macaque MT and MST neurons during pursuit eye movement. *Journal of neurophysiology, 102*(6), 3225-3233. doi: 10.1152/jn.90692.2008

Churchland, A. K., & Lisberger, S. G. (2005). Discharge properties of MST neurons that project to the frontal pursuit area in macaque monkeys. *J Neurophysiol, 94*(2), 1084-1090.

Churchland, M. M., & Lisberger, S. G. (2001). Shifts in the Population Response in the Middle Temporal Visual Area Parallel Perceptual and Motor Illusions Produced by Apparent Motion. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 21*(23), 9387-9402.

Dean, H. L., Crowley, J. C., & Platt, M. L. (2004). Visual and saccade-related activity in macaque posterior cingulate cortex. *J Neurophysiol, 92*(5), 3056-3068. doi: 10.1152/jn.00691.2003

00691.2003 [pii]

Desjardins, A. E., Kiehl, K. A., & Liddle, P. F. (2001). Removal of confounding effects of global signal in functional MRI analyses. *Neuroimage, 13*(4), 751-758. doi: http://dx.doi.org/10.1006/nimg.2000.0719

Dichgans, J., Körner, F., & Voigt, K. (1969). Vergleichende Skalierung des afferenten und efferenten Bewegungssehens beim Menschen: Lineare Funktionen mit verschiedener Anstiegssteilheit. *Psychologische Forschung, 32*(4), 277-295.

Dicke, P. W., Chakraborty, S., & Thier, P. (2008). Neuronal correlates of perceptual stability during eye movements. *Eur J Neurosci, 27*(4), 991-1002. doi: EJN6054 [pii]

10.1111/j.1460-9568.2008.06054.x

Duffy, C. J., & Wurtz, R. H. (1997). Medial superior temporal area neurons respond to speed patterns in optic flow. *The Journal of neuroscience, 17*(8), 2839-2851.

Dursteler, M. R., & Wurtz, R. H. (1988). Pursuit and optokinetic deficits following chemical lesions of cortical areas MT and MST. *Journal of Neurophysiology, 60*(3), 940-965.

Erickson, R. G., & Thier, P. (1991). A Neuronal Correlate of Spatial Stability during Periods of Self-Induced Visual-Motion. *Experimental Brain Research, 86*(3), 608-616.

Fattori, P., Pitzalis, S., & Galletti, C. (2009). The cortical visual area V6 in macaque and human brains. *Journal of Physiology-Paris, 103*(1–2), 88-97. doi: http://dx.doi.org/10.1016/j.jphysparis.2009.05.012

Fischer, E., Bulthoff, H. H., Logothetis, N. K., & Bartels, A. (2012a). Human areas V3A and V6 compensate for self-induced planar visual motion. *Neuron, 73*(6), 1228-1240. doi: 10.1016/j.neuron.2012.01.022

Fischer, E., Bulthoff, H. H., Logothetis, N. K., & Bartels, A. (2012b). Visual motion responses in the posterior cingulate sulcus: a comparison to V5/MT and MST. *Cerebral cortex (New York, N.Y. : 1991), 22*, 865-876. doi: 10.1093/cercor/bhr154

Fox, C. J., Iaria, G., & Barton, J. J. (2009). Defining the face processing network: optimization of the functional localizer in fMRI. *Hum Brain Mapp, 30*(5), 1637-1651. doi: http://dx.doi.org/10.1002/hbm.20630

Furlan, M., Wann, J. P., & Smith, A. T. (2014). A Representation of Changing Heading Direction in Human Cortical Areas pVIP and CSv. *Cerebral Cortex, 24*(11), 2848-2858. doi: 10.1093/cercor/bht132

Galletti, C., & Battaglini, P. P. (1989). Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *J Neurosci, 9*(4), 1112-1125.

Galletti, C., Battaglini, P. P., & Fattori, P. (1990). 'Real-motion' cells in area V3A of macaque visual cortex. *Exp Brain Res, 82*(1), 67-76.

Galletti, C., Battaglini, P. P., & Fattori, P. (1995). Eye position influence on the parieto-occipital area PO (V6) of the macaque monkey. *Eur J Neurosci, 7*(12), 2486-2501.

Galletti, C., & Fattori, P. (2003). Neuronal mechanisms for detection of motion in the field of view. *Neuropsychologia, 41*(13), 1717-1727.

Galletti, C., Fattori, P., Battaglini, P. P., Shipp, S., & Zeki, S. (1996). Functional demarcation of a border between areas V6 and V6A in the superior parietal gyrus of the macaque monkey. *Eur J Neurosci, 8*(1), 30-52.

Galletti, C., Fattori, P., Gamberini, M., & Kutz, D. F. (1999). The cortical visual area V6: brain location and visual topography. *Eur J Neurosci, 11*(11), 3922-3936.

Galletti, C., Gamberini, M., Kutz, D. F., Fattori, P., Luppino, G., & Matelli, M. (2001). The cortical connections of area V6: an occipito-parietal network processing visual information. *Eur J Neurosci, 13*(8), 1572-1588.

Ilg, U. J., Schumann, S., & Thier, P. (2004). Posterior parietal cortex neurons encode target motion in world-centered coordinates. *Neuron, 43*(1), 145-151. doi: 10.1016/j.neuron.2004.06.006

Ilg, U. J., & Thier, P. (2003). Visual tracking neurons in primate area MST are activated by smooth-pursuit eye movements of an "imaginary" target. *J Neurophysiol, 90*(3), 1489-1502.

Inaba, N., Miura, K., & Kawano, K. (2011). Direction and speed tuning to visual motion in cortical areas MT and MSTd during smooth pursuit eye movements. *Journal of neurophysiology, 105*(4), 1531-1545. doi: 10.1152/jn.00511.2010

Kawano, K., Shidara, M., Watanabe, Y., & Yamane, S. (1994). Neural activity in cortical area MST of alert monkey during ocular following responses. *Journal of Neurophysiology, 71*(6), 2305-2324.

Kayser, C., Kording, K. P., & Konig, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Curr Opin Neurobiol, 14*(4), 468-473.

Kleiner, M., Brainard, D. H., Pelli, D., Ingling, A., Murray, R. M., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*(14), 1.1-16.

Krekelberg, B., Boynton, G. M., & van Wezel, R. J. (2006). Adaptation: from single cells to BOLD signals. *Trends Neurosci, 29*(5), 250-256.

Krekelberg, B., & van Wezel, R. J. (2013). Neural mechanisms of speed perception: transparent motion. *Journal of neurophysiology, 110*, 2007-2018. doi: 10.1152/jn.00333.2013

Lee, B., Pesaran, B., & Andersen, R. A. (2007). Translation speed compensation in the dorsal aspect of the medial superior temporal area. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 27*(10), 2582-2591. doi: 10.1523/jneurosci.3416-06.2007

Lindner, A., Schwarz, U., & Ilg, U. J. (2001). Cancellation of self-induced retinal image motion during smooth pursuit eye movements. *Vision research, 41*(13), 1685-1694.

Lingnau, A., Wall, M. B., & Smith, A. T. (2009). Speed encoding in human visual cortex revealed by fMRI adaptation Hiroshi Ashida. *9*, 1-14. doi: 10.1167/9.13.3.Introduction

Liu, J., & Newsome, W. T. (2003). Functional organization of speed tuned neurons in visual area MT. *Journal of neurophysiology, 89*, 246-256. doi: 10.1152/jn.00097.2002

Liu, J., & Newsome, W. T. (2005). Correlation between speed perception and neural activity in the middle temporal visual area. *The Journal of neuroscience : the* *official journal of the Society for Neuroscience, 25*, 711-722. doi: 10.1523/JNEUROSCI.4034-04.2005

Malonek, D., Tootell, R. B. H., & Grinvald, A. (1994). Optical Imaging Reveals the Functional Architecture Of Neurons Processing Shape and Motion In Owl Monkey Area MT. *Proceedings of the Royal Society (London) B, 258*(1352), 109-119.

MATLAB. (2010). *version 7.10.0* Natick, Massachusetts: The MathWorks Inc.

Maunsell, J. H., & Van Essen, D. C. (1983a). The connections of the middle temporal area and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience, 3*(12), 2563-2586.

Maunsell, J. H., & Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J Neurophysiol, 49*(5), 1148-1167.

McKeefry, D. J., Burton, M. P., Vakrou, C., Barrett, B. T., & Morland, A. B. (2008). Induced deficits in speed perception by transcranial magnetic stimulation of human cortical areas V5/MT+ and V3A. *J Neurosci, 28*(27), 6848-6857. doi: 28/27/6848 [pii]

10.1523/JNEUROSCI.1287-08.2008

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex I. Mechanisms of speed and directional selectivity in extrastriate area MT. *Journal of Neurophysiology, 55*, 1308-1327.

Murray, S. O., & Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral occipital complex. *Nat Neurosci, 7*(1), 70-74. doi: http://dx.doi.org/10.1038/nn1161

Newsome, W. T., Mikami, A., & Wurtz, R. H. (1986). Motion Selectivity In Macaque Visual-Cortex .3. Psychophysics and Physiology Of Apparent Motion. *Journal of Neurophysiology, 55*(6), 1340-1351.

Newsome, W. T., Wurtz, R. H., Dürsteler, M. R. , & Mikami, A. (1985). Deficits in visual motion processing following ibotenic acid lesions of the middle temporal visual area of the macaque monkey. *Journal of Neuroscience, 5*(3), 825-840.

Nover, H., Anderson, C.H., & DeAngelis, G. C. (2005). A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. (1529-2401 (Electronic)).

Olson, C. R., Musil, S. Y., & Goldberg, M. E. (1996). Single neurons in posterior cingulate cortex of behaving macaque: eye movement signals. *J Neurophysiol, 76*(5), 3285-3300.

Orban, G. A., de Wolf, J., & Maes, H. (1984). Factors influencing velocity coding in the human visual system. *Vision research, 24*(1), 33-39.

Orban, G. A., Lagae, L., Raiguel, S., Xiao, D., & Maes, H. (1995). The speed tuning of medial superior temporal (MST) cell responses to optic-flow components. *Perception, 24*(3), 269-285.

Orban, G. A., Saunders, R. C., & Vandenbussche, E. (1995). Lesions Of the Superior Temporal Cortical Motion Areas Impair Speed Discrimination In the Macaque Monkey. *European Journal Of Neuroscience, 7*(11), 2261-2276.

Pasternak, T., & Merigan, W. H. (1994). Motion perception following lesions of the superior tempora lsulcus in the monkey. *Cerebral Cortex, 4*, 247-259.

Perrone, J. A., & Thiele, A. . (2001). Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nat Neurosci, 4*(5), 526-532.

Pitzalis, S., Galletti, C., Huang, R. S., Patria, F., Committeri, G., Galati, G., . . . Sereno, M. I. (2006). Wide-field retinotopy defines human cortical visual area v6. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 26*, 7962-7973. doi: 10.1523/JNEUROSCI.0178-06.2006

Pitzalis, S., Sdoia, S., Bultrini, A., Committeri, G., Di Russo, F., Fattori, P., . . . Galati, G. (2013). Selectivity to translational egomotion in human brain motion areas. *PloS one, 8*, e60241. doi: 10.1371/journal.pone.0060241

Pitzalis, S., Sereno, M. I., Committeri, G., Fattori, P., Galati, G., Patria, F., & Galletti, C. (2010). Human v6: the medial motion area. *Cerebral cortex (New York, N.Y. : 1991), 20*, 411-424. doi: 10.1093/cercor/bhp112

Pitzalis, S., Strappini, F., De Gasperis, M., Bultrini, A., & Di Russo, F. (2012). Spatio-temporal brain mapping of motion-onset VEPs combined with fMRI and retinotopic maps. *PloS one, 7*, e35771. doi: 10.1371/journal.pone.0035771

Priebe, N. J., Cassanello, C. R., & Lisberger, S. G. (2003). The Neural Representation of Speed in Macaque Area MT/V5. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 23*(13), 5650-5661.

Priebe, N. J., & Lisberger, S. G. (2004). Estimating Target Speed from the Population Response in Visual Area MT. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 24*(8), 1907-1916. doi: 10.1523/JNEUROSCI.4233-03.2004

Priebe, N. J., Lisberger, S. G., & Movshon, J. A. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *The Journal of Neuroscience, 26*(11), 2941-2950.

Quinlan, D. J., & Culham, J. C. (2007). fMRI reveals a preference for near viewing in the human parieto-occipital cortex. *NeuroImage, 36*(1), 167-187. doi: 10.1016/j.neuroimage.2007.02.029

Rodman, H. R., & Albright, T. D. (1987). Coding of visual stimulus velocity in area MT of the macaque. *Vision Research, 27*(12), 2035-2048.

Shenoy, K. V., Crowell, J. A., & Andersen, R. A. (2002). Pursuit speed compensation in cortical area MSTd. *J Neurophysiol, 88*(5), 2630-2647.

Singh, K. D., Smith, A. T., & Greenlee, M. W. (2000). Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage, 12*(5), 550-564.

Smith, A. T., Wall, M. B., & Thilo, K. V. (2012). Vestibular inputs to human motion-sensitive visual cortex. *Cerebral cortex (New York, N.Y. : 1991), 22*, 1068-1077. doi: 10.1093/cercor/bhr179

Sunaert, S., Van Hecke, P., Marchal, G., & Orban, G. A. (2000). Attention to speed of motion, speed discrimination, and task difficulty: an fMRI study. *Neuroimage, 11*(6 Pt 1), 612-623.

Tanaka, K., & Saito, H. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology, 62*(3), 626-641.

Tootell, R. B. H., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., . . . Dale, A. M. (1997). Functional analysis of V3A and related areas in human visual cortex. *J Neurosci, 17*(18), 7060-7078.

Tootell, R. B. H., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., . . . Belliveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal Of Neuroscience, 15*(4), 3215-3230.

Ungerleider, L. G., & Desimone, R. (1986). Cortical connections of visual area MT in the macaque. *Journal of Comparative Neurology, 248*, 190-222.

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

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science, 287*(5456), 1273-1276.

Wall, M. B., & Smith, A.T. (2008). The representation of egomotion in the human brain. *Current biology : CB, 18*, 191-194. doi: 10.1016/j.cub.2007.12.053

Yamasaki, D. S ., & Wurtz, R. W. (1991). Recovery of function after lesions in the superior temporal sulcus in the monkey. *Journal of Neurophysiology, 66*(3), 651-673.

Zhang, T., Heuer, H. W., & Britten, K. H. (2004). Parietal area VIP neuronal responses to heading stimuli are encoded in head-centered coordinates. *Neuron, 42*(6), 993-1001.