Vision Research 226 (2025) 108523

Contents lists available at ScienceDirect

Vision Research

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

# Tracking cortical entrainment to stages of optic-flow processing

Cai Wingfield [a,b,c,1], Andrew Soltan [d,e,f], Ian Nimmo-Smith [a], William D. Marslen-Wilson [a,g], Andrew Thwaites [a,g,h,*,1]

[a] MRC Cognition and Brain Sciences Unit, Cambridge, UK
[b] Department of Psychology, Lancaster University, Lancaster, UK
[c] Institute for Data and AI, University of Birmingham, Birmingham, UK
[d] Oxford Cancer & Haematology Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[e] Department of Oncology, University of Oxford, Oxford, UK
[f] Institute of Biomedical Engineering, University of Oxford, Oxford, UK
[g] Department of Psychology, University of Cambridge, Cambridge, UK
[h] Department of Speech, Hearing & Phonetic Sciences, University College London, London, UK

ARTICLE INFO

*Keywords:*
Electroencephalography
Magnetoencephalography
Motion perception
Optic flow
Model expression
Entrainment

ABSTRACT

In human visual processing, information from the visual field passes through numerous transformations before perceptual attributes such as motion are derived. Determining the sequence of transforms involved in the perception of visual motion has been an active field since the 1940s. One plausible family of models are the spatiotemporal energy models, based on computations of motion energy computed from the spatiotemporal features the visual field. One of the most venerated is that of Heeger (1988), which hypotheses that motion is estimated by matching the predicted spatiotemporal energy in frequency space. In this study, we investigate the plausibility of Heeger's model by testing for evidence of cortical entrainment to its components. Entrainment of cortical activity to these components was estimated using measurements of electro- and magnetoencephalographic (EMEG) activity, recorded while healthy subjects watched videos of dots moving left and right across their visual field. We find entrainment to several components of Heeger's model bilaterally in occipital lobe regions, including representations of motion energy at a latency of 80 ms, overall velocity at 95 ms, and acceleration at 130 ms. We find little evidence of entrainment to displacement. We contrast Heeger's biologically inspired model with alternative baseline models, finding that Heeger's model provides a closer fit to the observed data. These results help shed light on the processes through which perception of motion arises in the visual processing stream.

## 1. Introduction

As information travels through the human visual system, it is subjected to a variety of transformations. The cornea and the lens alter the spectral content of incoming light (Bone & Sparrock, 1971), and the excitation patterns from photoreceptor cones are integrated by neuronal cells in the inner retinal layer to form different chromatic channels (Lee & Silveira, 2016). After passing through the lateral geniculate nucleus (LGN), visual information reaches the primary visual cortex, from which point a wide array of transforms relating to color, depth perception, shape, and object recognition are initiated (Chalupa & Werner, 2004; Kaas & Collins, 2003).

A set of transformations of particular interest are those that relate to the motion of objects in the visual field. A foundational framework for studying motion perception is the Lu–Sperling three-systems theory (Lu & Sperling, 1995, 1996), in which first-order motion (characterized by spatiotemporal changes in luminance), second-order motion (changes in contrast, depth, and other perceptual image features), and third-order motion (changes in figure salience maps and most binocular effects) are addressed by a hierarchy of neural-processing systems (Lu & Sperling, 2001; see also Vaina et al., 1999). In non-human primates, regions including the medial temporal (MT) and medial superior temporal

*Abbreviations:* BOLD, Blood-Oxygen Level Dependent; EMEG, electro- and magneto-encephalographic; KID, Kymata identifier; LGN, lateral geniculate nucleus; MT, medial temporal area; MST, medial superior temporal area; MRI, magnetic resonance imaging; V1, primary visual cortex.

* Corresponding author at: Department of Speech, Hearing & Phonetic Sciences, University College London, London, UK.
*E-mail address:* acgt2@cam.ac.uk (A. Thwaites).

[1] These authors contributed equally.

https://doi.org/10.1016/j.visres.2024.108523
Received 1 September 2023; Received in revised form 2 August 2024; Accepted 13 November 2024
Available online 26 November 2024

(MST) areas exhibit sensitivity to different forms of coherent motion stimuli (Mikami et al., 1986; Duffy & Wurtz, 1991; McKeefrey et al., 1997) more so than to contrast (Newsome & Paré, 1986) or size (Tanaka & Saito, 1989). The posterior medial temporal visual area is widely studied as a region key to the processing of visual motion (Nishimoto & Gallant, 2011; Britten et al., 1993), while analogous regions and pathways have been identified in humans. The MT/MST complex has been functionally mapped using non-invasive neuroimaging techniques (Bradley, 2001), with further specialized sub-regions localized (e.g., Smith et al., 2006, Ahlfors et al., 1999; Morrone et al., 2000; Greenlee, 2000).

Often, patterns of coherent visual motion such as translation, rotation, expansion, and contraction correspond to the movement of an object, or the direction of motion of the observer (Gibson, 1950; Koenderink & van Doorn, 1977). For example, movement in the direction of facing corresponds to an expansion of patterns in the visual field, as objects which are far away and visible near the center of the field get nearer and then move past the observer, visible in the periphery. Similarly, an observer rotating about a vertical axis corresponds to patterns to translating horizontally across the visual field. Together, these kinds of motion patterns are known as "optic flow" fields (Koenderink, 1986). They comprise components of a framework of biological visual motion processing related to relative motion of observer and scene (Orban et al., 1995; Nakayama & Loomis; 1974, Tanaka & Saito, 1989; Duffy & Wurtz, 1991), and tracking moving objects (Masson et al., 2009; Miles & Sheliga, 2009).

As well as orienting the observer within the environment, optic-flow processing is crucial for discerning targets and obstacles, requirements for both biological organisms and machine vision (Lee, 1980; Duchon et al., 1998; Diekmann et al., 2009; Warren & Rushton, 2009; Janai et al., 2020). Many sighted organisms, including primates, have visual systems which are tuned to the perception of optic flow (Srinivasan & Zhang, 2004; Newsome & Paré, 1986; Maunsell & van Essen, 1983; Braddick et al., 2016). Thus, an understanding of the neurobiological underpinnings of visual motion perception often begins with an investigation into neural responses to optic flow fields (Boussaoud et al., 1990). In primates including humans, the MT and MST regions exhibit speed- and direction-selective sensitivity to different components of optic flow (McKeefrey et al., 1997; Maunsell & van Essen; 1983; Heeger et al., 1999), receiving inputs from simpler upstream motion detectors. A parallel stream of research has investigated models of optic-flow processing for computer vision as an engineering problem, without explicit regard for biological plausibility (e.g., Horn & Schunck, 1981; Fennema & Thompson, 1978; Nagel, 1987; Farnebäch, 2000; Fleet & Weiss, 2006; Fortun et al., 2015).

Entrainment is the phenomenon wherein neuronal or cortical activity is observed to "track" a certain attribute of a stimulus, either directly, or after some transformative processing (Ding et al., 2014; Ding & Simon, 2014). To investigate entrainment, one needs a characterization of the hypothesized transformation to which neural responses become entrained. There are many accounts of the sequence of transformations underlying the processing of optic-flow stimuli in the human visual system (e.g., Adelson & Bergen, 1985; Heeger 1988; Perrone & Stone, 1994; Bowns, 2011; 2018; Nishimoto & Gallant, 2011; Barron, Fleet & Beauchemin, 1994; Orchard & Etienne-Cummings, 2014, for overview), and there has been a lot of experimental work characterizing the emergence of human optic-flow processing within the three-systems framework (see Lu & Sperling, 2001, for a review). Nevertheless, thanks to recent advances in millisecond-resolution measurement and analysis has allowed for the mapping, the identification of entrainment is an effective way to probe the sequence of transformations that visual information undergoes as it is processed in the human brain. In primates, evidence of entrainment to elements of optic-flow stimuli has been found in neural activity in macaque MT, measured using intracellular recordings (Lisberger & Movshon, 1999; Nishimoto & Gallant, 2011). In humans, evidence of entrainment to translation, rotation and expansion optic flow components has also been reported in blood-oxygen level dependent (BOLD) signals, both through block-designed contrast studies (e.g. Morrone et al, 2001; Smith et al, 2006; Braddick et al, 2001; Rees et al., 2000), and correlation of periodic stimuli (Engel, et al. 1994; Engel et al, 1997; Smith et al., 1998). As with non-human primates, the cortical locations of human motion entrainment are predominantly V5/MT/MST, the primary visual cortex, or nearby regions.

In the natural environment, neural computations relating to visual-motion processing are applied to rapidly changing visual input which changes at the sub-second timescale. Thus, neuroimaging studies which measure electromagnetic activity, such as electroencephalography (EEG) and magnetoencephalography (MEG), or their combination (EMEG), are especially well suited to investigating these processes. For example, Ahlfors et al. (1999) found that an abrupt change in the direction of a coherent motion stimulus evoked responses 130–170 ms and 260 ms later in MT-adjacent regions, and in frontal regions at about 180 ms, while Fesi et al. (2014) found high EEG response at frequency components which matched that of periodically modulated optic flow stimuli. However, to our knowledge, none have yet fully investigated entrainment to functions of continuously varying coherent-motion optic-flow stimuli in human visual cortex.

In the current study, we look for evidence of cortical entrainment to two stages of the Heeger (1988) model of motion processing, a popular and influential example of motion-energy modelling. During the first stage of the model, the moment-to-moment oriented motion energy is derived for spatiotemporal filters of varying widths, in the second, these energies are integrated to estimate the moment-to-moment velocity for each region of the visual field. In addition to these two components, we consider two further transformations: the moment-to-moment *difference* or change in velocity for each region of the visual field (an analogue to acceleration), and the cumulative sum of the velocity (an analogue to displacement) over time. In this study the stimuli moved only left and right horizontally, so we extract only the horizontal component.

We proceed using the approach of Thwaites et al., (2015; 2016; 2017; 2018), a procedure which performs a systematic search through time-varying neuroimaging data for the signature of a particular hypothesized computation (see Section 2.4). In this context, a *computation* is a transformation of an input signal (e.g. a stimulus) into an output signal (i.e., the result when applied to the input). The predicted output signal is determined by a functional specification for that computation, and as such amounts to a specific, testable hypothesis regarding the nature of neural processes. The signatures of these computations come in the form of evidence for cortical activity which is entrained to the predicted output of the computation. Thwaites' method naturally affords the selection of a model of best fit from among a set of competitor models; we therefore compare components of Heeger's (1988) biologically inspired model to the computer-vision (not biologically inspired) model of Horn & Schunck (1981), which we use as a baseline model for optic-flow processing.

We use this procedure to search for evidence of entrainment to components of Heeger's motion-energy model in the neural activity of regions of the cortex (striate, extra striate, and regions beyond), as estimated though EMEG, and the baseline competitor model of Horn & Schunck. Specifically, we aim to determine (1) whether entrainment to

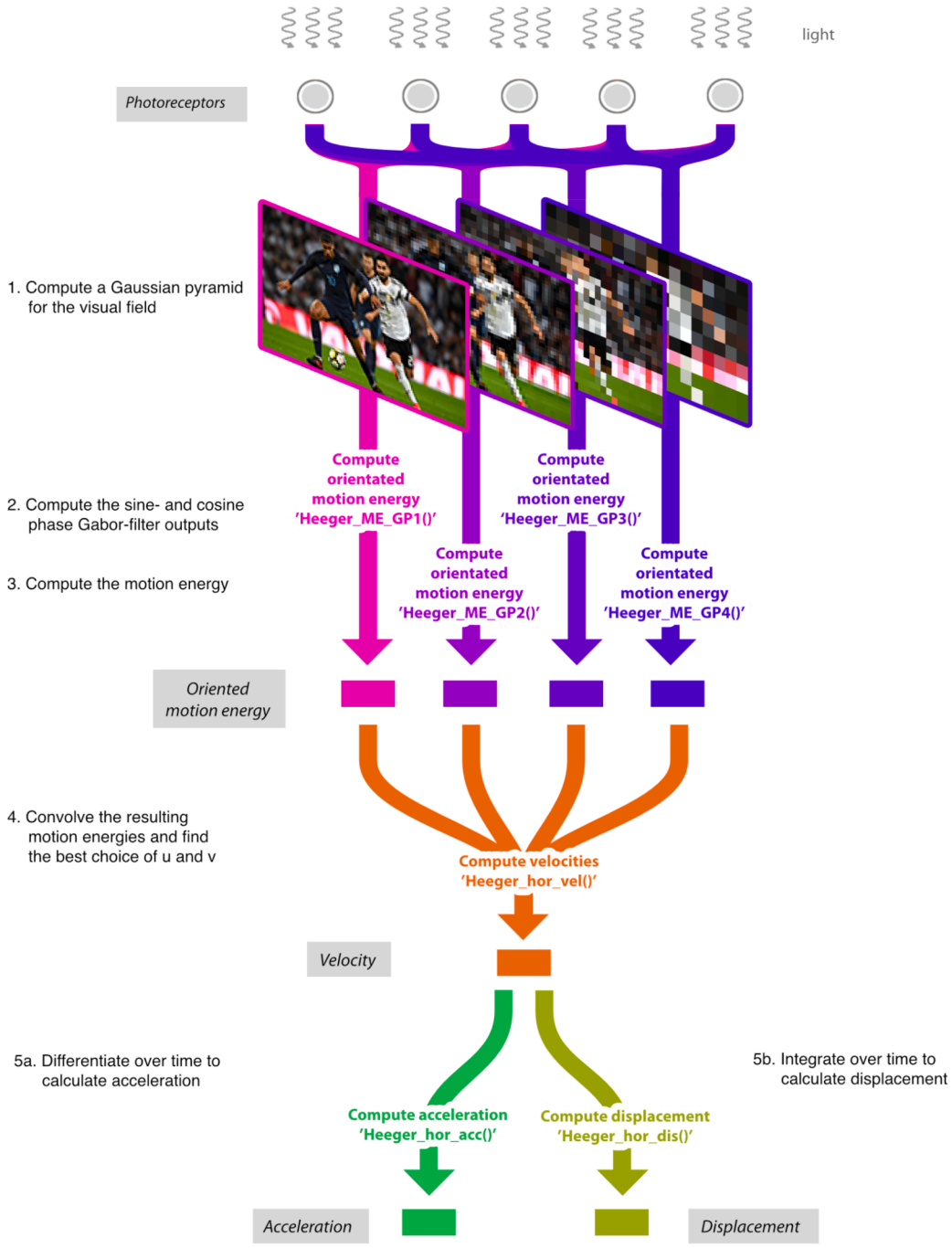


**Fig. 1.** Schematic of the Heeger 1988 model of motion perception. The Heeger model of motion perception 1) computes a Gaussian pyramid for the image, and, for each layer 2) computes the sine- and cosine-phase Gabor-filter outputs using separable convolutions, 3) computes the motion energy as the squared sum of the sine- and cosine-phase Gabor-filter outputs. 4) Convolves the resulting motion energies with a Gaussian to approximate the integral in Parseval's Theorem and find the best choice of u and v using a minimization technique. 5) Computes the acceleration and displacement by differentiating and integrating velocity respectively. Pictured with colored blocks are the components of the Heeger model tested in this study: motion energy for four layers of the Gaussian pyramid (yellow-to-orange), estimated velocity (blue), acceleration (green) and displacement (purple).

these components occurs, and if so: (2) which models provide the best prediction of entrainment, and (3) the latencies and (3) location of the entrainment.

In addition to the static graphic presentation of results in this paper, a dynamic, interactive representation of this study's results can be viewed on the online Kymata Atlas (https://kymata.org). For ease of reference, each transformation in this paper is assigned a Kymata ID (KID), with which it can be found in the Kymata Atlas web interface.

## 2. Methods

### 2.1. Defining candidate models

The method of Thwaites et al. (2016) follows a framework in which information processing is modelled as a function taking time-varying signals as input (in this case the visual field) and produces a time-varying signal as output (the predication of neural activity). The model must be characterized by a function *f*, taking the form:

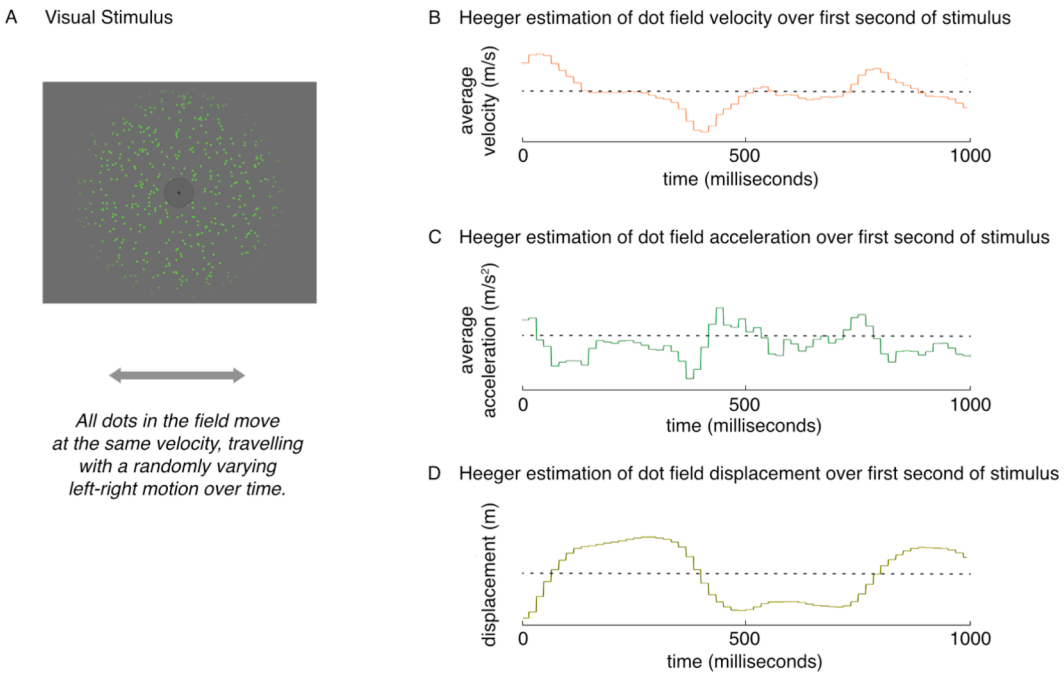


**Fig. 2.** The stimulus and three of the Heeger (1988) model predictions. A. The stimulus consisted of moving colored dots on a grey background, with a fixation cross in the center. B–D. Estimations of the dot field's average velocity, acceleration and displacement (relative to the fixation cross).

$$f(x_1, x_2, x_3, \ldots, x_t) = (y_1, y_2, y_3, \ldots, y_t) \quad (1)$$

where input ($x_1$,…,$x_t$) and output ($y_1$,…,$y_t$) are both time-courses of length *t*, where *f* is bound by a set of formal requirements ensuring it is well-defined (Davis et al., 1994), and the additional requirement that $y_i$ is independent of all $x_k$ where $k > i$ (i.e., excluding non-causal or future-sensitive functions).

### 2.2. The Heeger model of motion perception

In this study we discuss component computations from the Heeger (1988) model of motion perception, one of a broad family of *spatiotemporal energy* models. Spatiotemporal energy models hypothesize that the representation of motion is derived from Fourier components of the visual field, that is, derived from the distribution of energy over many spatiotemporal frequency filters. As a model of the brain's processing of visual motion, this account is supported by contemporary studies, including the presence of cells in macaque MT sensitive to specific spatiotemporal frequencies (Perrone & Thiele, 2001; Simoncelli & Heeger, 2001), and experimental evidence that suggests similar specific sensitivities in humans (e.g. Adelson & Movshon, 1982). The Adelson and Bergen (1985) model first calculates *oriented energy* from a set of spatiotemporal filters, which gives the magnitude of the energy at a set of spatiotemporal frequencies, and then constructs *opponent energy* as the sum of two oriented energy units tuned to the same speed, but facing opposite directions. In addition, such models can be confounded by variations in the spatial- or temporal-frequency content of the stimulus or variations in its contrast (Adelson & Bergen, 1985; Heeger, 1988). Heeger's (1988) model aimed to overcome such issues by retaining the intuition behind spatiotemporal filters, but using the filters to estimate the slope of the plane in the frequency space corresponding to the velocity of movement in the visual field. Heeger achieves this using a least-squares minimization of the error between observed motion energy and that predicted by target velocity vectors in the visual field (described in more detail below).

We select Heeger's model in this study as a simple and widely used, biologically inspired model of optic-flow processing, representative of the motion-energy modelling approach, which is also implemented in readily available peer-reviewed software (Raudies, 2013).

Fig. 1 gives an overview of the transforms involved in the Heeger (1988) model. The complete pipeline of transforms for the Heeger model, and their derivations, can be found in Heeger's (1988) schema. The implementation we use here is based on Raudies' (2013) implementation, and the model's free parameters are set to those values suggested in Heeger's original schema.[2] In the current study, we test seven components of the model (marked in Fig. 1 in color), outlined in the following sections.

#### 2.2.1. Motion energies from Gaussian pyramid layers 1–4 (Heeger_ME_GP1–4)

Lower-frequency spatial filters have their greatest outputs for patterns moving at greater velocities and vice versa. So, for a spatiotemporal energy model to capture both fast and slow velocities, filters with different spatial frequency tunings are required. Heeger suggests implementing this using a Gaussian pyramid approach (Adelson et al., 1984), which achieves the same end: reducing the resolution of the visual field by half at each level of the pyramid is equivalent to lowering the spatial frequency of the filter applied to visual field at full resolution. We follow this suggestion: our implementation therefore begins by computing a Gaussian pyramid over the image; in our implementation we use four layers.

Separable sine- and cosine-phased Gabor filters are defined at different spatial orientations (Heeger, 1987). Spatiotemporal Gabor filters are the combination of sinusoidal gratings with Gaussian kernels:

[2] In the original formulation, Heeger describes a potential issue where a very bright, low-contrast image can cause filter outputs to be dominated by the mean value (which should be irrelevant to perceived motion) rather than contrast frequencies. Heeger notes this problem could be addressed by subtracting the mean luminance. We exclude this mean-luminance correction step since the average luminance of our moving-dot stimulus is low in comparison to its contrast signal.

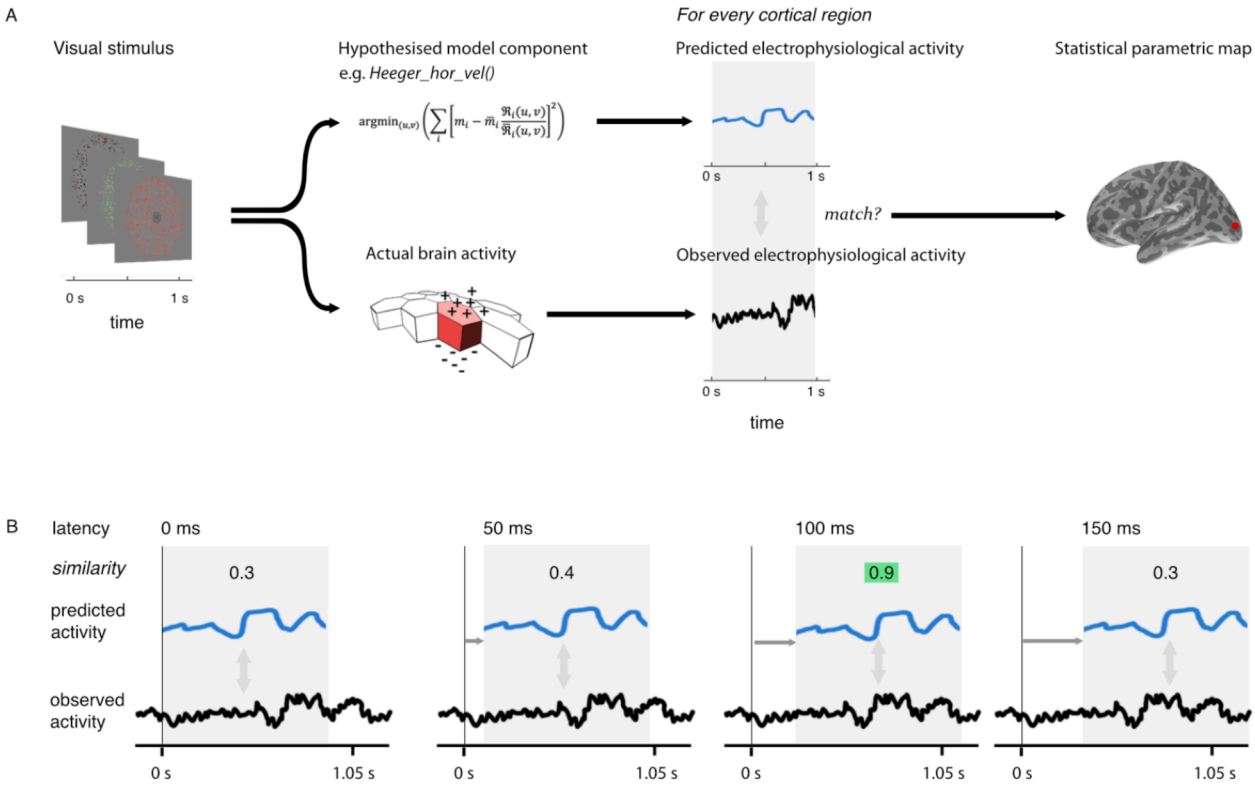


**Fig. 3. Technique overview.** First (A), the electrophysiological activity of the brain in response to a given stimulus (measured using EMEG) is matched to the pattern of neural activity predicted by the model being evaluated. Predicted and observed activity are tested for similarity and the resulting statistical parametric map displays the regions where the match is statistically significant. Second (B), this procedure is repeated at different lags (illustrated here from 0–150 ms) between the onset of the observed neural activity and the onset of the predicted output. We record the lag at which the similarity is greatest (highlighted). This produces a statistical parametric map that changes over time.

$$\text{Gabor}(x, y, t) = \frac{1}{\sqrt{2\pi^3}\sigma_x\sigma_y\sigma_t} \bullet e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}+\frac{t^2}{\sigma_t^2}\right)} \bullet \sin 2\pi(\omega_x+\omega_y+\omega_t) \tag{2}$$

with the $\sigma_i$ and $\omega_i$ giving the spatiotemporal spread and frequency of the filter.[3]

These Gabor filters are then convolved with the spatiotemporal stimulus and the motion energies for this layer of the pyramid are computed as the squared sum of the Gabor filter outputs.[4]

### 2.2.2. Horizontal velocity (Heeger_Hor_Vel)

The expected motion energy $\Re$ for a given velocity $(u, v)$ is

$$\Re(u, v) = e^{\left(\frac{-4e^2\sigma_x^2\sigma_y^2\sigma_z^2(u\omega_x+v\omega_y+\omega_t)}{(u\sigma_x\sigma_t)^2+(v\sigma_y\sigma_t)^2+(\sigma_x\sigma_y)^2}\right)} \tag{3}$$

(see Heeger, 1988; Baron et al., 1994; Raudies, 2013 for derivation).

The difference between expected and observed motion energies for a given velocity versus the output of the filters is minimized.

$$(\text{Heeger\_Hor\_Vel}, \text{Heeger Vert\_Vel}) = \text{argmin}_{(u,v)}\left(\sum_i\left[m_i - \overline{m}_i\frac{\Re_i(u,v)}{\overline{\Re}_i(u,v)}\right]^2\right) \tag{4}$$

where $m_i$ are the motion energies for different filters, $\overline{m}_i$ are the summed filter outputs for the same spatial orientations, $\Re_i$ is the predicted motion energy for $(u, v)$, and $\overline{\Re}_i$ is the prediction of $\overline{m}_i$ (Raudies, 2013). Heeger (1988) used a Gauss–Newton gradient descent method. In this study, the stimuli move only left and right horizontally, so we extract only the horizontal component of these velocities, i.e. Heeger_Hor_Vel only.[5] The velocity of the stimuli dots is constant over the entire field of vision, so we take the average value over the field. This average value is shown in Fig. 2B over a 1-second period in response to an example stimulus (Fig. 2A).

### 2.2.3. Horizontal acceleration (Heeger_Hor_Acc)

In the current study, we extend the Heeger model with an acceleration component, following Lisberger & Movshon (1999) and Price et al. (2005). We define Heeger acceleration[6] as the derivative of the Heeger velocity:

$$\text{Heeger\_Hor\_Acc}(t) = \frac{d}{dt}\text{Heeger\_Hor\_Vel(t)} \tag{5}$$

[3] As Heeger notes (1988; p. 283), centring the temporal component of the Gabor filter to the moment of interest results in the filter accessing information that has not yet been observed, thereby breaking our definition of a causally consistent model set out in section 2.1. One solution is to replace the Gabor filter with a more complex filter that only accesses historical information (e.g., Adelson & Bergen, 1985). In the current study we shift the zero-point in Eq. (2) to be the point at which 98% of the information is historic (i.e., two standard deviations of the filter width, equivalent to 8.3 ms).

[4] Entries for all tested models can be found in the Kymata Atlas. For ease of reference, we provide the Kymata IDs (KIDs) of each of the models tested here. Heeger_ME_GP1, KID: KWMVH; Heeger_ME_GP2, KID: L5DN4; Heeger_ME_GP3, KID: YJTMV; Heeger_ME_GP4, KID: E8Q98.

[5] Heeger_Hor_Vel, KID: FAF47.

[6] *Heeger_Hor_Acc*, KID: 5T9QY.

The value of this variable over the first 1 s of the stimulus is shown in Fig. 2C. We can compute this value using the discrete derivative because our stimulus is discrete. Again, the acceleration of the stimuli dots is constant over the field of vision, so we take the average value over the field.

#### 2.2.4. Horizontal displacement (Heeger_Hor_Dis)

Displacement can be estimated by integrating over the cumulative values of the Heeger velocity:

$$\text{Heeger\_Hor\_Dis}(t) = \int_0^t \text{Heeger\_Hor\_Vel}(T)\, dT \tag{6}$$

We compute this integral by cumulative summation[7] of the discrete values of Heeger_Hor_Vel. The value of this variable over the first 1 s of the stimulus is shown in Fig. 2D. Like Heeger_Hor_Acc above, this component is an extension to Heeger's original model.

### 2.3. The Horn–Schunck (1981) method for optic flow estimation

As a baseline alternative model for optic flow estimation, we use the (non-biologically inspired) model of Horn & Schunck (1981). The Horn–Schunck model is foundational within the field of machine vision (over 17,000 citations on Google Scholar at time of writing), and while it has been superseded by more modern models which overcome some of its inherent limitations, it is suitably contemporaneous to Heeger's (1988) model. While not required for the method of Thwaites et al. (2015), the inclusion of a "competitor" computational model provides a useful point of comparison: as well as showing when and where a given model predicts cortical entrainment, it shows when and where that model out-performs other possible alternatives. We provide a specification for the model in Appendix A, and adapt the software implementation of Raudies (2013; see also Bruhn et al., 2006).

### 2.4. The analysis procedure

The analysis procedure and methods follow Thwaites et al. (2018), and as such Sections 2.4 and 2.5 are reproduced from this work, with minor amendments to reflect the difference in models tested. A more detailed explanation of the statistical properties of the method can be found in Thwaites et al. (2018) and Thwaites et al. (2015).

The reconstructed distributed source current of the cortex is specified as the current of 10,242 cortical regions (sources), spaced uniformly over the cortex of each hemisphere. The testing procedure involves examining each of these sources, looking for evidence that the current predicted by a model is similar to the current observed (Fig. 3A). This procedure is repeated at 5 ms intervals (Fig. 3B) across a range of time-lags (–200 to 800 ms), covering the range of plausible latencies (0 to 800 ms) and a short, pre-stimulation range (–200 to 0 ms) during which we expect to see no entrainment. This produces a statistical parametric map that changes over time as the lag is varied, revealing the changes in similarity of a given model's predicted output with observed activity over the cortical surface. Evidence of this similarity is expressed as a *p*-value, which is generated through the match–mismatch technique described in Thwaites et al. (2015), where evidence for similarity is described as significant if the *p*-value falls below a pre-defined threshold, α*. We refer to the observation of significant matches at a specific lag and location as *model expression*.

Setting α* so that it accurately reflects what we know about the data being tested can be difficult. In the current study, some of the measurements used in the tests would be dependent on other measurements (because of spatial and temporal similarities between neighboring sources and lags). However, it is very difficult, if not impossible, to get accurate estimations of, for instance, the spatial dependencies between sources. In the present study, rather than accept assumptions about the dependencies that are hard to justify, we use Šidák correction (Šidák, 1967), consistent with a conservative assumption that each source and lag are independent. As a result, the type II error rate is likely to be high, making the reported results very conservative. We used the formula for the familywise false alarm rate from Thwaites et al. (2015) to set α* at approximately $3 \times 10^{-13}$; *p*-values greater than this are deemed to be not significant.

The results are presented as *expression plots* (see section 3), which show the latency at which each of the 10,242 sources per hemisphere best matched the output of the tested model (marked as a stem). On the y-axis is the evidence of the match at this latency: if any of the sources have evidence, at their best latency, indicated by a *p*-value lower than α*, their stems are colored to match the model being tested.

The expression plots also allow us to chart which models are most likely to be entrained at a particular source through a model selection procedure, using p-values as a proxy for model likelihood. By default, the expression plot displays only the best model (i.e. the one with the lowest *p*-value) for a source. In this way, the displacement, velocity, acceleration and motion-energy functions defined by the Heeger model compete not only with each other, but also with their counterparts in the Horn–Schunck model, and there is no a priori reason why one model should be preferred over another. It is important to note that this model selection procedure does not indicate that any one model is *significantly* better than another for some source. It indicates only that one model fits the data to a higher degree than another, even if this evidence may not differ strongly between models. We take this approach as we are only interested in the trend of which models explain the activity best in each source, and to disambiguate between models which may be correlated over time.

### 2.5. MEG and EEG methods and materials

#### 2.5.1. Experiment design

*Participants:* 15 right-handed participants (7 men, mean age = 24 years, range = 18–30) were recruited. All gave informed consent and were paid for their participation. The study was approved by the Peterborough and Fenland Ethical Committee (UK) and was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

*Stimuli*: Each participant was shown an identical stimulus, a pattern of randomly placed colored dots on a grey-masked annular field with a black fixation cross in the center to control eye movements (Fig. 1A). Each frame of the stimulus had a resolution of 1280 × 1024 pixels, with each dot being 10 × 10 pixels, the central fixation mask being 16 × 16 pixels and the circular edge of the annular field extending to upper and lower edges of the frame, with 100 pixels of linear falloff of the colored dots at the outer edge of the annulus to minimize edge artefacts. Within the annular mask, the horizontal displacement and color of these dots fluctuated pseudo-randomly. The stimulus lasted 7 min, allowing it to be split later in the analysis procedure into 400 segments of length 1000 ms each (trials). The first and last 10 s of stimulus were intentionally excluded from the analysis to avoid the sudden appearance and disappearance of the stimulus during the first and last trial. The screen size was 0.5 × 0.4 m. It was placed 1.3 m from the participant, resulting in a viewing angle of 22°. The screen's white-light luminance was 162 cd/m$^2$.

The moment-to-moment position of the dots was controlled by adjusting their horizontal displacement relative to their original starting position. The frequency of these fluctuations ranged between 4 and 40 Hz. The resulting horizontal acceleration, velocity and displacement (relative the fixation cross) for the first 5 s of the stimulus, as predicted by the Heeger model, is shown in Fig. 2B–D. The movement over time of the dots were constructed to be pseudo-periodic to avoid the predictions of the derivatives/integrals being correlated with one another (see

[7] *Heeger_Hor_Dis*, KID: AUY3G.

discussion).

All stimulus frames are included in the supplementary materials,[8] along with colorimeter readings taken from the projector screen.

#### 2.5.2. Procedure

Each participant received one practice stimulus lasting 20 s. The continuous 6 min and 40 s stimulus was played 4 times to the participant, with instructions to fixate on the cross in the middle of the screen. After each presentation, the participant could rest, playing the next presentation when ready, using a button box. Presentation of stimuli was controlled with Matlab, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). The stimuli were presented on a Panasonic PT-D7700 DLP projector, with a refresh rate of 60 Hz.

To keep them alert during the experiment while not introducing visual task-dependent effects, the participants listened passively to an audio podcast (a BBC radio talk-show) while watching the stimuli. Note that while participants were instructed to centrally fixate throughout the experimental duration, and with breaks interspersed every few minutes, eye movements were not further controlled during the experiment. In future studies, it could be effective to include a secondary active task based on rapid serial visual presentation at the point of fixation, to avoid eye-wandering effects.[9] However, eye movements were measured using EOG, with these measurements included in the data-cleaning procedures (see below).

#### 2.5.3. Modelling the predicted Heeger entrainment responses

The outputs of each of the seven different Heeger models (Fig. 1B–D) were obtained by feeding the same visual stimulus shown to the participants into the equations given in section 2.2.

#### 2.5.4. MEG recording

Continuous MEG data were recorded using a 306 channel Triux system (MEGIN, Espoo, Finland) containing 102 identical sensor triplets (two orthogonal planar gradiometers and one magnetometer) in a hemispherical array situated in a magnetically shielded room. The position of the head relative to the sensor array was monitored every 200 ms by four head-position indicator coils attached to the scalp. Simultaneous EEG was recorded from 70 Ag-AgCl electrodes placed in an elastic cap (EASYCAP GmbH, Herrsching-Breitbrunn, Germany) according to the 10/20 system, using a nose electrode as reference. Vertical and horizontal electrooculograms (EOG) were also recorded. All data were sampled at 1 kHz with a band-pass filter from 0.03–330 Hz. A 3D digitizer (Fastrak Polhemus Inc., Colchester, VA) recorded the locations of the EEG electrodes, the head-position indicator coils and approximately 50–100 “headpoints” along the scalp, relative to three anatomical fiducials (the nasion and left and right pre-auricular points).

#### 2.5.5. Data pre-processing

Static MEG bad channels were detected and excluded from subsequent analyses (MaxFilter version 2, MEGIN, Stockholm, Sweden). Compensation for head movements and a temporal extension of the signal–space separation technique (Taulu et al., 2005) were applied to the MEG data. Static EEG bad channels were visually detected and removed from the analysis (MNE version 2.7, Martinos Center for Biomedical Imaging, Boston, Massachusetts). The EEG data were re-referenced to the average over all channels. The continuous data were low-pass filtered to 100 Hz (zero-phase shift, overlap-add, FIR filtering). The recording was split into 400 epochs of 1000 ms duration. Each epoch included the 200 ms from before the epoch onset and 800 ms after the epoch finished (taken from the epoch previous and subsequent) to allow for the testing of different latencies. Epochs in which the EEG or EOG exceeded 200 μV, or in which the value on any gradiometer channel exceeded 2000 fT/m, were rejected from both EEG and MEG datasets (between 5 % and 15 %, depending on the participant, mean: 8 %). Epochs that were not rejected were averaged over all four stimulus repetitions.

#### 2.5.6. Source reconstruction

The location of the cortical current sources was estimated using minimum-norm estimation (MNE; Hämäläinen and Ilmoniemi, 1994), neuro-anatomically constrained by MRI images obtained using a GRAPPA 3D MPRAGE sequence (TR = 2250 ms; TE = 2.99 ms; flip-angle = 9 degrees; acceleration factor = 2) on a 3 T Tim Trio (Siemens, Erlangen, Germany) with 1 mm isotropic voxels. For each participant a representation of their cerebral cortex was constructed using FreeSurfer 5.3 (Martinos Center for Biomedical Imaging, Boston, Massachusetts) from their individual MR images. The forward model was calculated with a three-layer boundary element model using the outer surface of the scalp and the outer and inner surfaces of the skull identified in the structural MRI. Anatomically constrained source activation reconstructions at the cortical surface were created by combining MRI, MEG, and EEG data. The MNE representations were downsampled to 10,242 vertices per hemisphere, roughly 3 mm apart, to improve computational efficiency. Representations of individual subjects were aligned using a spherical morphing technique (Fischl et al., 1999). Source activations for each trial were averaged over participants. We employed a loose-orientation constraint (0.2) to improve the spatial accuracy of localization. Sensitivity to neural sources was improved by calculating a noise covariance matrix based on a 1 s pre-stimulus period. Reflecting the reduced sensitivity of MEG sensors for deeper cortical activity (Hauk et al., 2011), sources located on the cortical medial wall and in subcortical regions were not included in the analyses reported here.

The entrainment testing procedure (section 2.4) was performed using the open-source Kymata-Core software package (Thwaites et al., in prep.) on these participant-average source reconstructions.

#### 2.5.7. Visualization

The cortical slices in Fig. 4. use the visualization software MRIcron (Georgia State Center for Advanced Brain Imaging, Atlanta, Georgia) with results mapped to the high-resolution *colin27* brain (Schmahmann et al., 2000).

## 3. Results

An interactive representation of all models tested in this paper can be viewed using https://kymata.org/atlas.

### 3.1. Heeger_ME_GP1–4

The regions where expression for the Heeger_ME_GP1, Heeger_ME_GP2, Heeger_ME_GP3 and Heeger_ME_GP4 were the most significant of the models tested — and below the α* threshold — were located bilaterally at 80 ms (Fig. 4A), centered in regions in the lateral occipital sulcus, occipital lobe, cuneus cortex and lingual gyrus.

The locations (namely lateral occipital sulcus, cuneus cortex and occipital lobe regions) are approximate in all cases, as a consequence of the error introduced by the point-spread function inherent in EMEG source localization (see discussion). This spatial ambiguity means it is often not possible to narrow these locations to specific visual areas, beyond noting that the entrainment is centered around the occipital pole (primary visual cortex V1) and middle temporal (MT/V5) regions.

### 3.2. Heeger_Hor_Vel

The regions where expression for the Heeger_Hor_Vel model was the most significant of the models tested — and crossing the α* threshold —

[8] Available at https://osf.io/gxfqb/.

[9] We thank an anonymous reviewer for this suggestion.

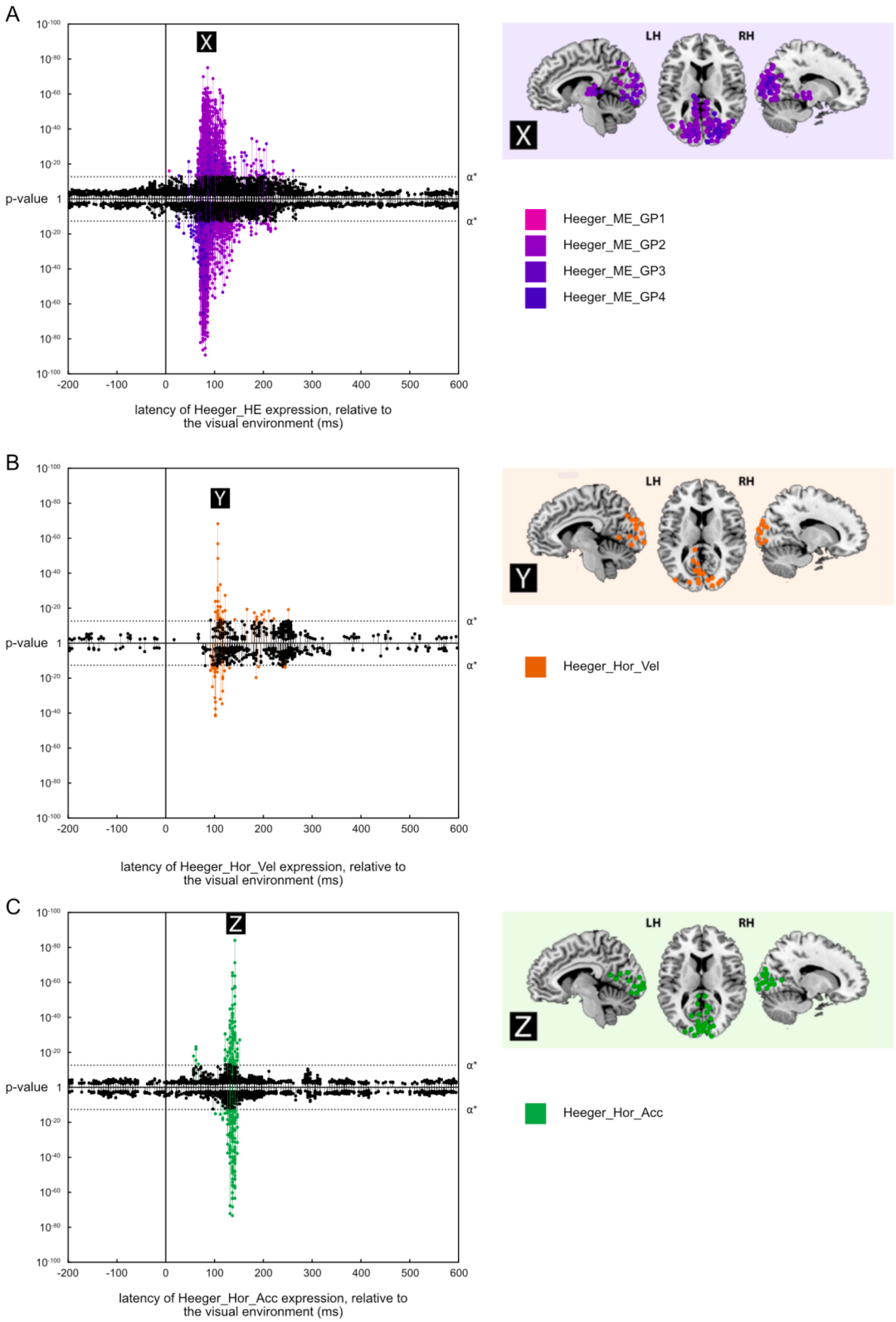


**Fig. 4.** Expression plots for the seven components of the Heeger model of visual motion perception. A) Plot for the expression for the Heeger_ME_GP1, Heeger_ME_GP2, Heeger_ME_GP3 and Heeger_ME_GP4 components of the Heeger 1988 model of motion perception, across processing lags from −200 to + 600 ms, relative to the visual environment. Results for the left and right hemispheres are plotted separately, mirror-wise across the mid-line. The minimum p-values at a given source, over all latencies, are marked as stems. Stems at or crossing the p-threshold value ($p \simeq 3 \times 10^{-13}$) indicate significant expression of the one of these components (orange) at that .

Source. B) Plot for the expression for the Heeger_hor_vel component (blue). C) Plot for the expression for the Heeger_Hor_acc component (green). The peaks for all seven component's significant expression are marked X (at 80 ms, The Heeger_ME_GP components), Y (at 95 ms, Heeger_Hor_Vel) and Z (130 ms, Heeger_Hor_Acc). The cortical locations of significant sources are indicated on the coronal and sagittal slices to right of the plots. These plots use the model-selection procedure (see section 2.4) so that each source appears at most once in Figures 4A, 4B, 4C, 5 and 6

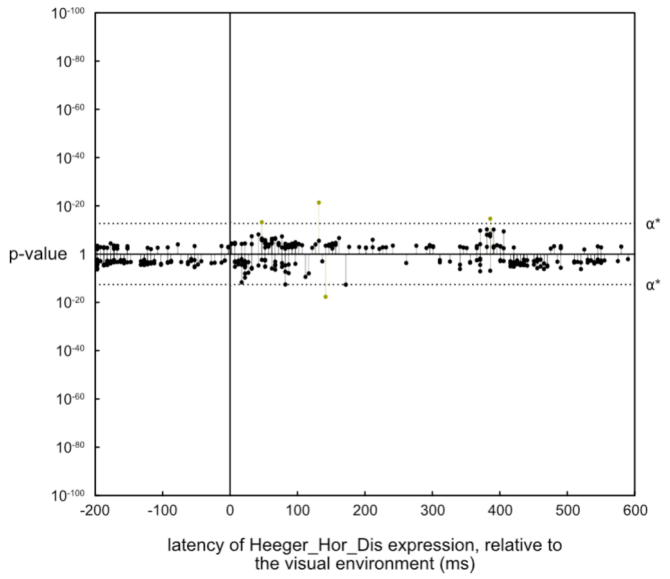


**Fig. 5. Expression plots for Heeger_Hor_Dis.** Plot for the expression for Heeger_Hor_Dis across processing lags from −200 to + 600 ms, relative to the visual environment. Little significant entrainment is observed.

were located bilaterally at 95 ms (Fig. 4B), centered in regions surrounding the occipital lobe.

### 3.3. *Heeger_Hor_Acc*

The regions where expression output of the Heeger_Hor_Acc model was the most significant of the models tested, and crossed the α* threshold, were located bilaterally at 130 ms (Fig. 4C). Like Heeger_Hor_Vel, these regions were centered in lingual gyrus, cuneus cortex and occipital lobe regions.

### 3.4. *Heeger_Hor_Dis*

Only four timepoints indicated a significant preference for Heeger_Hor_Dis, and even then only relatively weakly (Fig. 5).

### 3.5. *HornSchunck_Hor_Dis, HornSchunck_Hor_Vel and HornSchunck_Hor_Acc*

There was some evidence of entrainment to the HornSchunck_Hor_Vel model, though it was overall far less pronounced than entrainment to Heeger_Hor_Vel, crossing the threshold of significance in only 11 vertices across the whole brain. The peak latency for entrainment was also in the region of 100 ms, with the most significant evidence at 105 ms at a vertex located in the lateral occipital sulcus. A later single significant vertex in the left precuneus exhibited entrainment at a latency of 235 ms. There was no evidence of any significant entrainment to HornSchunck_Hor_Dis or HornSchunck_Hor_Acc. Expression plots are shown in Fig. 6(A).

## 4. Discussion

This study was concerned with whether entrainment to components of a motion-energy-based model for velocity, acceleration and displacement could be located, and whether the latency of the entrainment for each of these components could be determined. To this end, we made specific predictions regarding optic-flow-component entrainment using Heeger's (1988) model. Our results indicate that bilateral regions in the polar occipital cortex, where the afferent pathways project into the cortex, do exhibit entrainment to the first three of these: motion energies, velocity and acceleration.

The inherent insolvability of the inverse problem during EMEG source reconstruction (Grave de Peralta-Menendez et al., 1996; Grave de Peralta-Menendez and Gonzalez-Andino, 1998) means that substantial "point spread" of localization data may be present. Therefore, while entrainment to oriented motion energy is found in the MST/MT/V5 regions, including the inferior parietal lobule and the lateral occipital sulcus, there is significant entrainment of these components for lingual gyrus and cuneus cortex for the motion energies and acceleration. The reader should be cautious about narrowing down the location of the entrainment to these components more narrowly; improvements in source reconstruction (through the gathering of more data or improved inverse techniques) may reduce this error in the future.

The entrainment to different motion-energy components observed is consistent with numerous studies reporting the existence of motion specific neurons in macaques MT tuned to specific velocities (Maunsell & van Essen, 1983; Rodman & Albright, 1987; Perrone & Thiele, 2001). Indeed, entrainment studies on these MT neurons indicate a firing rate latency of around 90 ms (Lisberger & Movshon, 1999), broadly consistent with our latency of 80 ms, with the latency mismatch likely stemming from our use of 10 ms time steps coupled with noise induced by the low signal-to-noise-ratio. While entrainment in the current study was found in an analogous area (human MT/V5), entrainment was also observed strongly in lingual gyrus and cuneus cortex (see Fig. 4A), although, as noted above, is not clear whether this is a true result or an EMEG source reconstruction error.

The entrainment of cortical regions to stimulus velocity, while plausible, is more difficult to link with previous work. Although numerous studies have shown a correlation between neural activity and velocity (e.g. intracellular recordings: Lisberger & Movshon,1999; and functional MRI: Engel et al.,1997; Smith et al, 1998), most use sinusoidal changes in motion as the basis of their experimental paradigm, making it difficult to establish whether the neural activity is tracking motion energy, velocity or acceleration. (By using pseudo-periodic changes in motion, this study is able to ascribe this activity to velocity alone.).

The current study considered an expanded Heeger (1988) model in which acceleration and displacement is derived from velocity. Our findings suggest that acceleration of motion is indeed calculated subsequent to velocity, and the location of its entrainment to neural activity is located in the lingual gyrus, cuneus cortex and occipital lobe regions at 130 ms. The encoding of acceleration in the human cortex remains something of an unknown quantity. Lisberger and Movshon (1999) demonstrated that individual MT neurons in the macaque encode stimulus velocity but not acceleration in their firing rates, and they and others interpret this to mean that the acceleration of visual motion can be encoded in the population response of these neurons (Lisberger & Movshon, 1999; Price et al. 2005). While our results cannot distinguish between population responses and individual neurons, if acceleration is indeed derived from velocity as these authors suggest, then acceleration must be encoded at a later latency to velocity, as is observed in this study.

The current study also hypothesized that displacement in the visual field was calculated from velocity through its integration. As noted in Fig. 5, almost no significant entrainment was found to the output of this transform. Why might this be? An object's moment-by-moment displacement is certainly a plausible calculation for a mammalian brain to compute – potentially as a way to keep track of objects relative to the body – so why do we find little evidence for this process in the current study? There are a number of possibilities. Our analysis is only testing for a very specific relationship between output of the transform with cortical activity of a region. If the output of this transform is encoded in a more complicated fashion, then our analysis procedure would not find it. Additionally, velocity and acceleration may be automatic, low-level qualities, while tracking displacement requires conscious effort. In this case, seven minutes of stimulus time may mean that participants stop consciously tracking the dot field's displacement.

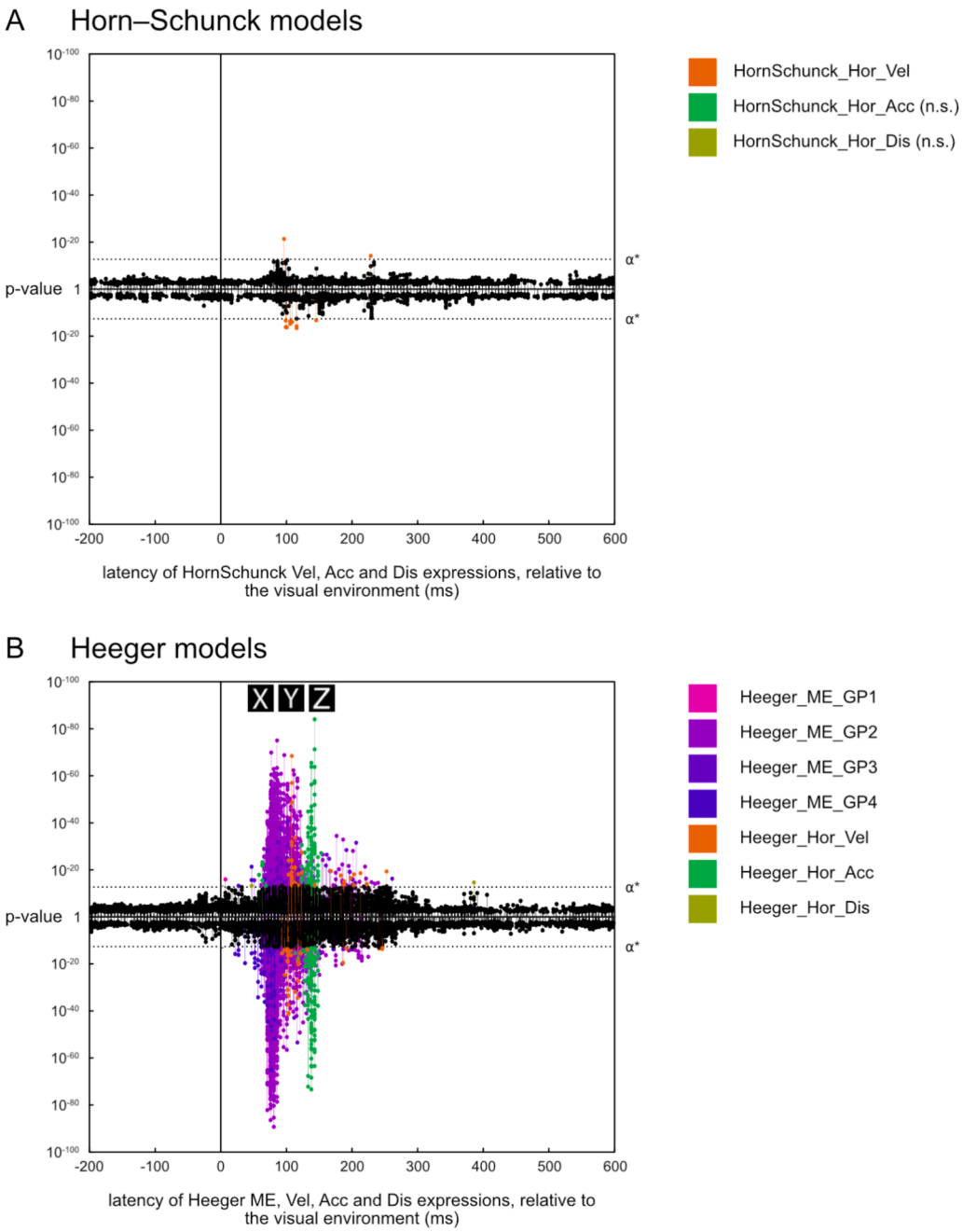


**Fig. 6. Plots for displacement, velocity and acceleration in the Horn–Schunck model, and the Heeger models also tested. A)** Plot for the expression for HornSchunck_Hor_Dis, HornSchunck_Hor_Vel and HornSchunck_Hor_Acc across processing lags from −200 to + 600 ms, relative to the visual environment. Only HornSchunck_Hor_Vel had some significant vertices. **B)** Expression distributions for Heeger_ME_GP1–4, Heeger_hor_vel, and Heeger_hor_acc, superimposed on a single plot. Fig. 4A–C are superimposed to show Heeger_hor_vel outputs lagging behind Heeger_ME_GP1–4 components, and Heeger_hor_acc lagging both. As in Fig. 4, the peaks for the components' significant expression are marked X, Y and Z.

Perhaps the most interesting result is the latency differences shown by the seven components from sequential stages of the motion-energy model. The findings in this study indicate a specific sequence of transforms: motion energy from the Gabor filters is calculated first, the output of which is used as an input to the calculation of velocity, and then (through Lisberger and Movshon) passed to a transform characterizing acceleration (Fig. 6B). Yet the latencies hold other points of interest, notably the fact that output of all four motion energy pyramid layers are expressed at the same lag. This seems plausible, indeed perhaps even necessary: the subsequent velocity estimate is derived from the output of each layer, and if each layer had a different latency, then the computation of velocity would have to be delayed until the last had been computed, thereby delaying the computation of velocity itself. Having all energies computed at the same time cuts this delay down—which is likely useful from an evolutionary standpoint.

The cross-correlation method of Thwaites et al. (2016) does not require a baseline model to test for significant evidence of entrainment, and the evidence of entrainment to components of a motion-energy model of optic-flow processing can be understood alone. Nevertheless, by way of comparison we also searched for entrainment to baseline estimates of displacement, velocity and acceleration derived from the model of Horn & Schunck (1981), and used the method of Thwaites et al. (2016) to select the better model at each time and location throughout the brain. Overall, Heeger's motion-energy model showed better overall

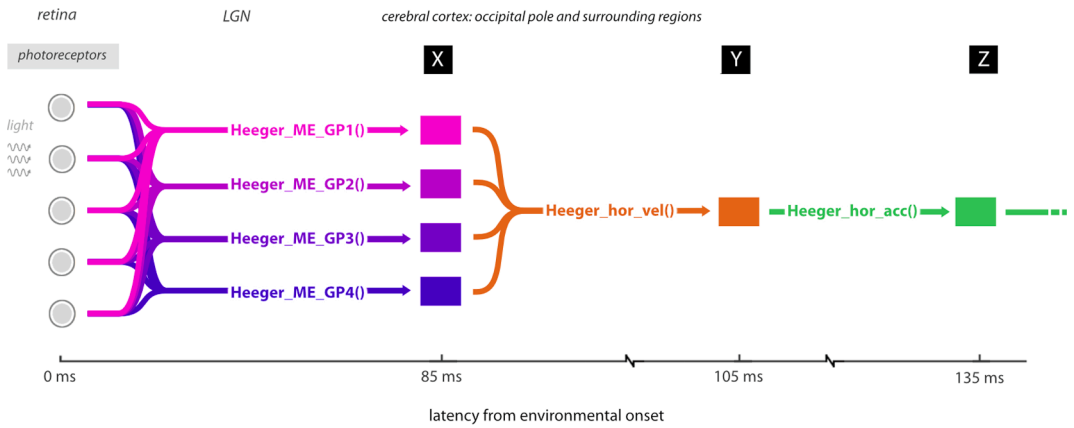


**Fig. 7.** Implied pathway of motion perception information, as modelled by Heeger 1988. The interpretation of the pathways suggested by the findings of this study. Information from the visual field undergoes a number of sequential transforms in order to calculate velocity and acceleration. First, motion energy is computed for several different sets of spatiotemporal filters, in parallel, with the output of each of these sets of filters expressed at 80 ms latency in the occipital pole of the cortex. From these motion energy estimates, the velocity is calculated, with the output expressed at 95 ms, followed by a transformation of acceleration is then computed, with the output of this transform at 130 ms, again in the occipital pole. As with previous figures, the peak latencies of each component are marked X, Y and Z.

entrainment than those of Horn & Schunck; neither Horn–Schunck displacement nor acceleration showed significant evidence of entrainment. While Horn–Schunck velocity did show some minor points of significance (Fig. 6A), the Heeger model showed much greater levels of significance, and was overall the far better model (Fig. 6B). The timing and locations of both are similar, suggesting that both are picking up the same effect.

Overall, the results of this study support the view that the oriented motion energy of the visual field is computed at different spatial frequencies, each of which is computed in parallel (with the outputs of these computations exhibiting entrainment at 80 ms delay in the MT/V5 and occipital lobe regions; Fig. 7). These motion energies are then convolved, with the velocity found through a process of minimization (the result of which is expressed at 95 ms latency) after which this information is differentiated to calculate the acceleration (which exhibits entrainment at 130 ms latency). Like motion energy, both are expressed in the regions surrounding the occipital poles.

It should be emphasized that this characterization—based on the plausibilities of the models tested in this study—is likely a simplification of the full picture of visual-motion processing. Our source-localization technique is unable to estimate neural activity in the retina, leaving its role in these computations unspecified, even though this role appears to be both significant (e.g. Sabbah et al., 2017; Ölveczky et al., 2003; Manookin et al, 2018) and complex (Gollisch & Meister, 2010; Masland & Martin, 2007). The simplicity of our current stimuli has also excluded other aspects of motion processing that might be characterized. For instance, the strict left–right motion of the dot field means we are unable to characterize how the orientation of the filters interacts with motion in more than one dimension, while the uniformity of the motion over the full visual field means we are unable to investigate how the computation of motion in the periphery of the visual field differs and interacts with that from the center.

The findings reported in this study open a number of avenues for further work. Perhaps the most important is additional components of the Heeger model; including computations prior to motion energy, and additional orientations beyond the horizontal. Our prior work on entrainment to different color-appearance models (Thwaites et al., 2018) made some progress in this direction, but there is more research needed to evidence a fully computational account of the pipeline of stimulus transformations from retinal activity to motion perception. A major limitation of this study is the simplicity and homogeneity of the stimulus. In order to make the effect size as large as possible for this initial study, we constrained the entire field of vision to carry the same motion pattern and constrained that motion to horizontal motion only. However, this makes it impossible to distinguish entrainment to general translational motion, and entrainment in a simpler left–right motion in particular. To address this in future work, we would include coherent two-dimensional translational motion, other two-dimensional motion such as expansion patterns, and examine motion which varied in direction across the visual field, allowing us to map the retinotopic aspect of motion representation in the cortical current. This will also allow us to test a wider range of models of optic-flow processing.

### 4.1. Summary

The results from this study suggest that motion energy, computed from Gabor filters sized at different spatial frequencies, takes place in the cortex, with entrainment to the output of this calculation occurring at 80 ms latency bilaterally in occipital lobe regions. Subsequent to this computation, these components are combined to form a representation of velocity, which is entrained at 95 ms, followed by a calculation of the acceleration is also computed, observed at 130 ms, both also in the occipital lobe regions. Entrainment to the perception of motion displacement was marginal at best. The locations of the significant entrainment are only approximate due to the inherent error in source estimation of EMEG data and more work is needed in improving the accuracy of these reconstructions in order to improve confidence in these locations.

## CRediT authorship contribution statement

**Cai Wingfield:** Formal analysis, Investigation, Software, Visualization, Writing – original draft, Writing – review & editing. **Andrew Soltan:** Writing – original draft. **Ian Nimmo-Smith:** Conceptualization, Methodology, Supervision, Writing – original draft. **William D. Marslen-Wilson:** Conceptualization, Funding acquisition, Supervision, Writing – original draft. **Andrew Thwaites:** Writing – original draft.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

AT designed the study and wrote the analysis code. CW wrote the model code. CW and AT ran the analysis and contributed to the drafting of the manuscript equally, with support from AS and INS. This work was

supported by an ERC Advanced Grant (Neurolex) to WMW, by MRC Cognition and Brain Sciences Unit (CBU) funding to WMW (U.1055.04.002.00001.01), and by a grant from Neuroverse FZCO to AT. Computing resources were provided by the MRC-CBU, and office space for AT by the Centre for Speech, Language and the Brain. We would also like to thank Anirudh Lakra for helpful comments on a draft version of this manuscript, and Russell Thompson, Anastasia Klimovich-Smith, Alexandra Woolgar, Gary Chandler and Clare Cook for invaluable support and suggestions.

*Declaration*

Cai Wingfield and Andrew Thwaites report article publishing charges were provided by Neuroverse FZCO. Cai Wingfield reports a relationship with Neuroverse FZCO that includes consulting or advisory fees and travel reimbursement. Andrew Thwaites reports a relationship with Neuroverse FZCO that includes employment, consulting or advisory fees, equity and stock ownership, funding or grants, and travel reimbursement.

## Appendix A:. The Horn–Schunck (1981) model of optic flow determination

Not all methods for determining optic flow are biologically inspired, or indeed biologically plausible. While the importance and relevance of optic-flow determination was already understood (e.g., Nakayama & Loomis, 1974), the earliest computational accounts for how it could be achieved came from research into computer graphics, computer vision, and computational signal processing (e.g., Netravali & Robbins, 1979; Nagel, 1978).

The Horn–Schunck (1981) model operates under the constraint that the movement of the image patterns varies smoothly (i.e., that the optic flow field is differentiable), a constraint which is satisfied by our uniformly translational stimulus.

The Horn–Schunck (1981) model provides an iterative solution to estimating motion vector ($u$, $v$). This solution is a sequence of vectors ($u_n$, $v_n$) defined as

$$u_{n+1} = \bar{u}_n - E_x \frac{\left(E_x \bar{u}_n + E_y \bar{v}_n + E_t\right)}{a^2 + E_x^2 + E_y^2}$$

$$v_{n+1} = \bar{v}_n - E_y \frac{\left(E_x \bar{u}_n + E_y \bar{v}_n + E_t\right)}{a^2 + E_x^2 + E_y^2}$$

where $\bar{u}$ and $\bar{v}$ are local motion averages at position ($i$,$j$) and time $k$

$$\bar{u}_{i,j,k} = \frac{1}{6}\left(u_{i\pm1,j,k} + u_{i,j\pm1,k}\right) + \frac{1}{12}\left(u_{i\pm1,j\pm1,k}\right)$$

$$\bar{v}_{i,j,k} = \frac{1}{6}\left(v_{i\pm1,j,k} + v_{i,j\pm1,k}\right) + \frac{1}{12}\left(v_{i\pm1,j\pm1,k}\right)$$

(using the shorthand $\pm 1$ in an index to indicate the sum of values using combinations of + 1 and –1 in that index; see Horn & Schunck, 1981, p. 6); where $E_x$, $E_y$ and $E_t$ are partial derivatives with respect to $x$, $y$ and $t$ of the brightness $E$ of the image plane. Horn & Schunck (1981; p. 6 for derivation) approximate these partial derivatives using

$$E_x \approx \frac{1}{4}\left(E_{i,j+1,k} - E_{i,j,k} + E_{i+1,j+1,k} - E_{i+1,j,k} + E_{i,j+1,k+1} - E_{i,j,k+1} + E_{i+1,j+1,k+1} - E_{i+1,j,k+1}\right)$$

$$E_y \approx \frac{1}{4}\left(E_{i+1,j,k} - E_{i,j,k} + E_{i+1,j+1,k} - E_{i,j+1,k} + E_{i+1,j,k+1} - E_{i,j,k+1} + E_{i+1,j+1,k+1} - E_{i,j+1,k+1}\right)$$

$$E_t \approx \frac{1}{4}\left(E_{i,j,k+1} - E_{i,j,k} + E_{i+1,j,k+1} - E_{i+1,j,k} + E_{i,j+1,k+1} - E_{i,j+1,k} + E_{i+1,j+1,k+1} - E_{i+1,j+1,k}\right)$$

and $a^2$ is a weighting factor. The sequence of $u_n$ and $v_n$ yields approximations of $u$ and $v$ which minimize the error

$$\xi^2 = a^2(\bar{u} - u)^2 + a^2(\bar{v} - v)^2 + \left(E_x u + E_y v + E_t\right)^2$$

We follow Raudies (2013), who adapted Bruhn et al.'s (2006) discretized implementation of the Horn–Schunk model. We follow Raudies' (2013) choices for free parameters, and selected 15 iterations as a balance between optimally conservative recommendations and available computational resources.

Given that our stimulus moved only in the horizontal axis and all dots had the same velocity, we defined Horn_Schunck_hor_vel($t$) as the average $u$ over the entire image frame.[10] Following this, and similarly to the Heeger model, we defined a displacement function

$$\text{Horn\_Schunck\_hor\_dis}(t) = \int_0^t \text{Horn\_Schunck\_hor\_vel}(T)\ \ dT$$

approximated using a cumulative sum,[11] and an acceleration function

[10] *Horn_Schunck_Hor_Vel*, KID: FPWMD.
[11] *Horn_Schunck_Hor_Dis*, KID: QJUNT.

$$\text{Horn\_Schunck\_hor\_acc}(t) = \frac{d}{dt}\text{Horn\_Schunck\_hor\_vel}(t) \quad dT$$

approximated using a discrete derivative.[12] As before, in both cases we took the average vector over the image frame.

## Data availability

The data used in this study (including stimuli, EMEG recordings and pre-processed data) can be found at https://kymata.org/datasets, available under a CC-BY license.

## References


Adelson, E. H., Anderson, C. H., Bergen, J. R., Burt, P. J., & Ogden, J. M. (1984). Pyramid methods in image processing. *RCA engineer, 29*(6), 33–41.

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, 2*, 284–299.

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature, 300*, 523–525.

Ahlfors, S. P., Simpson, G. V., Dale, A. M., Belliveau, J. W., Liu, A. K., Korvenoja, A., Virtanen, J., Huotilainen, M., Tootel, R., Aronen, H., & Ilmoniemi, R. J. (1999). Spatiotemporal activity of a cortical network for processing visual motion revealed by MEG and fMRI. *Journal of Neurophysiology, 82*(5), 2545–2555.

Barron, J. L., Fleet, D. J., & Beauchemin, S. S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision, 12*(1), 43–77.

Bone, R. A., & Sparrock, J. M. B. (1971). Comparison of macular pigment densities in the human eye. *Vision Research, 11*, 1057–1064. https://doi.org/10.1016/0042- 6989(71)90112-X

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. https://doi.org/10.1002/cne.902960311

Bowns, L. (2011). Taking the energy out of spatio-temporal energy models of human motion processing: The Component Level Feature Model. *Vision Research, 51*(23–24), 2425–2430. https://doi.org/10.1016/j.visres.2011.09.014

Bowns, L. (2018). Motion estimation: A biologically inspired model. *Vision Research, 150*, 44–53. https://doi.org/10.1016/j.visres.2018.07.003

Braddick, O. J., O'Brien, J. M., Wattam-Bell, J., Atkinson, J., Hartley, T., & Turner, R. (2001). Brain areas sensitive to coherent visual motion. *Perception, 30*(1), 61–72.

Bradley, D. C. (2001). Motion perception: Psychological and neural aspects. In Smelser, N. J., & Baltes, P. B. (Eds.). (2001). International Encyclopedia of the social & behavioral sciences (Vol. 11). 10099–10105.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436. https://doi.org/10.1163/156856897X00357

Britten, K., Shadlen, M., Newsome, W., & Movshon, J. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience, 10*(6), 1157–1169. https://doi.org/10.1017/S0952523800010269

Bruhn, A., Weickert, J., Kohlberger, T., & Schnörr, C. (2006). A Multigrid Platform for Real-Time Motion Computation with Discontinuity-Preserving Variational Methods. *International Journal of Computer Vision, 70*, 257–277. https://doi.org/10.1007/s11263-006-6616-7

Chalupa, L. M., & Werner, J. S. (2004). *The Visual Neurosciences*. Cambridge, MA, US: MIT Press.

Davis, M., Sigal, R., & Weyuker, E. J. (1994). *Computability, complexity, and languages: Fundamentals of theoretical computer science*. Burlington, MA: Morgan Kaufmann Pub.

Diekmann, V., Jürgens, R., & Becker, W. (2009). Deriving angular displacement from optic flow: A fMRI study. *Experimental brain research, 195*(1), 101.

Ding, N., Chatterjee, M., & Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage, 88*, 41–46. https://doi.org/10.1016/j.neuroimage.2013.10.054

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2014.00311

Duchon, A. P., Kaelbling, L. P., & Warren, W. H. (1998). Ecological robotics. *Adaptive Behavior, 6*(3–4), 473–507. https://doi.org/10.1177/105971239800600306

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. https://doi.org/10.1152/jn.1991.65.6.1329

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E., & Shadlen, M. N. (1994). fMRI of human visual cortex. *Nature, 369*, 525. https://doi.org/10.1038/369525a0

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex, 7*, 181–192.

Farnebäck, G. (2000). Fast and accurate motion estimation using orientation tensors and parametric motion models. In Pattern Recognition, 2000. Proceedings. 15th International Conference on (Vol. 1, pp. 135–139). IEEE.

Fennema, C. L., & Thompson, W. B. (1979). Velocity determination in scenes containing several moving objects. *Computer graphics and image processing, 9*(4), 301–315.

Fesi, J. D., Thomas, A. L., & Gilmore, R. O. (2014). Cortical responses to optic flow and motion contrast across patterns and speeds. *Vis. Res., 100*, 56–71.

Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999). High-resolution inter-subject averaging and a coordinate system for the cortical surface. *Human Brain Mapping, 8*, 272–284.

Fleet, D. J., & Weiss, Y. (2006). Optical flow estimation. *Handbook of mathematical models in computer vision*, 237–257.

Fortun, D., Bouthemy, P., & Kervrann, C. (2015). Optical flow modeling and computation: A survey. *Computer vis and image understanding., 134*, 1–21. https://doi.org/10.1016/j.cviu.2015.02.008

Gibson, J. (1950). *The Perception of the Visual World*. Cambridge: Riverside Press.

Gollisch, T., & Meister, M. (2010). Eye smarter than scientists believed: Neural computations in circuits of the retina. *Neuron, 65*, 150–164. https://doi.org/10.1016/j.neuron.2009.12.009

Grave de Peralta-Menendez, R., Gonzalez-Andino, S. L., & Lutkenhoner, B. (1996). Figures of merit to compare linear distributed inverse solutions. *Brain Topography, 9*, 117–124. https://doi.org/10.1007/BF01200711

Grave de Peralta-Menendez, R., & Gonzalez-Andino, S. L. (1998). A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem. *IEEE Transactions on Biomedical Engineering, 45*, 440–448. https://doi.org/10.1109/10.664200

Greenlee, M. W. (2000). Human cortical areas underlying the perception of optic flow: Brain imaging studies. *International Review of Neurobiology, 44*, 269–292. https://doi.org/10.1016/S0074-7742(08)60746-1

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing, 32*, 35–42. https://doi.org/10.1007/BF02512476

Hauk, O., Wakeman, D. G., & Henson, R. (2011). Comparison of noise-normalized minimum norm estimates for MEG analysis using multiple resolution metrics. *NeuroImage, 54*, 1966–1974. https://doi.org/10.1016/j.neuroimage.2010.09.053

Heeger, D. J. (1987). *Models for Motion Perception*. University of Pennsylvania. PhD thesis, CIS Department.

Heeger, D. J. (1988). Optical flow using spatiotemporal filters. *International journal of computer vision, 1*(4), 279–302.

Heeger, D. J., Boynton, G. M., Demb, J. B., Seidemann, E., & Newsome, W. T. (1999). Motion opponency in visual cortex. *Journal of Neuroscience, 19*(16), 7162–7174.

Horn, B. K., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence, 17*(1–3), 185–203. https://doi.org/10.1016/0004-3702(81)90024-2

Janai, J., Güney, F., Behl, A., & Geiger, A. (2020). Computer vision for autonomous vehicles: Problems, datasets and state of the art. *Foundations and Trends in Computer Graphics and Vision, 12*(1–3), 1–308.

Kaas, J. H. and Collins C. E. (2003) The Primate Visual System CRC Press.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36*, 14. https://doi.org/10.1177/03010066070360S101

Koenderink, J. J. (1986). Optic flow. *Vision research, 26*(1), 161–179. https://doi.org/10.1016/0042-6989(86)90078-7

Koenderink, J. J., & van Doorn, A. J. (1977). How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow. *Kybernetik, 77*, 224.

Lee, D. N. (1980). The optic flow field: The foundation of vision. *Philosophical Transactions of the Royal Society B, 290*(1038), 169–179.

Lee, B. B., & Silveira, L. C. L. (2016). In *Cone opponency: An efficient way of transmitting chromatic information*. Human Color Vision, Cham: Springer. https://doi.org/10.1007/978-3-319-44978-4_4.

Lisberger, S. G., & Movshon, J. A. (1999). Visual motion analysis for pursuit eye movements in area MT of macaque monkeys. *The Journal of Neuroscience, 19*, 2224–2246.

Lu, Z. L., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research, 35*(19), 2697–2722. https://doi.org/10.1016/0042-6989(95)00025-U

Lu, Z. L., & Sperling, G. (1996). Three systems for visual motion perception. *Current Directions in Psychological Science, 5*(2), 44–53. https://doi.org/10.1111/1467-8721.ep10772720

Lu, Z. L., & Sperling, G. (2001). Three-systems theory of human visual motion perception: Review and update. *JOSAA, 18*(9), 2331–2370. https://doi.org/10.1364/JOSAA.18.002331

Manookin, P. (2018). and Linehan. *Neural Mechanisms Mediating Motion Sensitivity in Parasol Ganglion Cells of the Primate Retina Neuron, 97*, 1327–1340. https://doi.org/10.1016/j.neuron.2018.02.006

Masland, R. H., & Martin, P. R. (2007). The unsolved mystery of vision. *Current Biology, 17*, R577–R582. https://doi.org/10.1016/j.cub.2007.05.040


[12] *Horn_Schunck_Hor_Acc,* KID: 572F7.

Masson, G. S., Montagnini, A., & Ilg, U. J. (2009). When the Brain Meets the Eye: Tracking Object Motion. In U. Ilg, & G. Masson (Eds.), *Dynamics of Visual Motion Processing*. Boston, MA: Springer. https://doi.org/10.1007/978-1-4419-0781-3_8.

Maunsell, J. H., & van Essen, D. C. J. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Neurophysiol., 49*, 1127–1147. https://doi.org/10.1152/jn.1983.49.5.1127

McKeefry, D. J., Watson, J. D. G., Frackowiak, R. S. J., Fong, K., & Zeki, S. (1997). The activity in human areas V1/V2, V3, and V5 during the perception of coherent and incoherent motion. *NeuroImage, 5*(1), 1–12.

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT. *Journal of Neurophysiology, 55*(6), 1308–1327.

Miles, F. A., & Sheliga, B. M. (2009). Motion Detection for Reflexive Tracking. In U. Ilg, & G. Masson (Eds.), *Dynamics of Visual Motion Processing*. Boston, MA: Springer. https://doi.org/10.1007/978-1-4419-0781-3_7.

Morrone, M. C., Tosetti, M., Montanaro, D., Fiorentini, A., Cioni, G., & Burr, D. C. (2000). A cortical area that responds specifically to optic flow, revealed by fMRI. *Nature Neuroscience, 3*(12), 1322–1328.

Nagel, H. H. (1978). Analysis Techniques for image sequence. In Proc. 4th IJCPR (pp. 186–211).

Nagel, H. H. (1987). On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence, 33*(3), 299–324.

Nakayama, K., & Loomis, J. M. (1974). Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis. *Perception, 3*(1), 63–80.

Netravali, A. N., & Robbins, J. D. (1979). Motion-compensated television coding: Part I. *Bell System Technical Journal, 58*(3), 631–670.

Newsome, W. T., & Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience, 8*(6), 2201–2211. https://doi.org/10.1523/JNEUROSCI.08-06-02201.1988

Nishimoto, S., & Gallant, J. L. (2011). A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *Journal of Neuroscience, 31*(41), 14551–14564. https://doi.org/10.1523/JNEUROSCI.6801-10.2011

Ölveczky, B. P., Baccus, S. A., & Meister, M. (2003). Segregation of object and background motion in the retina. *Nature, 423*, 401–408. https://doi.org/10.1038/nature01652

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. https://doi.org/10.1068/p240269

Orchard, G., & Etienne-Cummings, R. (2014). Bioinspired visual motion estimation. *Proceedings of the IEEE, 102*(10), 1520–1536.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442. https://doi.org/10.1163/156856897X00366

Perrone, J. A., & Stone, L. S. (1994). A model of self-motion estimation within primate extrastriate visual cortex. *Vision Research, 34*, 2917–2938. https://doi.org/10.1016/0042-6989(94)90060-4

Perrone, J. A., & Thiele, A. (2001). Speed skills: Measuring the visual speed analyzing properties of primate MT neurons. *Nature Neuroscience, 5*, 526–532. https://doi.org/10.1038/87480

Price, N. S. C., Ono, S., Mustari, M. J., & Ibbotson, M. R. (2005). Comparing Acceleration and Speed Tuning Macaque MT: Physiology and Modeling. *Journal of Neurophysiology, 94*, 3451–3464. https://doi.org/10.1152/jn.00564.2005

Raudies, F. (2013). *Optic flow. Scholarpedia, 8*(7), 30724. https://doi.org/10.4249/scholarpedia.30724

Rees, G., Friston, K., & Koch, C. (2000). A direct quantitative relationship between the functional properties of human and macaque V5. *Nature Neuroscience, 3*, 716–723.

Rodman, H. R., & Albright, T. D. (1987). Coding of visual stimulus velocity in area MT of the macaque. *Vision Research, 27*, 2035–2048. https://doi.org/10.1016/0042-6989(87)90118-0

Sabbah, S., Gemmer, J. A., Bhatia-Lin, A., Manoff, G., Castro, G., Siegel, J. K., Jeffery, N., & Berson, D. M. (2017). A retinal code for motion along the gravitational and body axes. *Nature, 546*, 492–497. https://doi.org/10.1038/nature22818

Schmahmann, J. D., Doyon, J., Toga, A. W., Petrides, M., & Evans, A. C. (2000). *MRI Atlas of the Human Cerebellum*. San Diego, CA: Academic Press.

Simoncelli, E. P., & Heeger, D. J. (2001). Representing retinal image speed in visual cortex. *Nature Neurosci., 4*, 461–462. https://doi.org/10.1038/87408

Smith, A. T., Greenlee, M. W., Singh, K. D., Kraemer, F. M., & Hennig, J. (1998). The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging (fMRI). *The Journal of Neuroscience, 18*, 3816–3830.

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.

Srinivasan, M. V., & Zhang, S. (2004). Visual motor computations in insects. *Annual Review of Neuroscience, 27*, 679–696.

Šidák, Z. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association, 62*(318), 626–633. https://doi.org/10.1080/01621459.1967.10482935

Tanaka, K., & Saito, H. A. (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. https://doi.org/10.1152/jn.1989.62.3.626

Taulu, S., Simola, J., & Kajola, M. (2005). Applications of the signal space separation method. *IEEE T Signal Proces., 53*, 3359–3372. https://doi.org/10.1109/TSP.2005.853302

Thwaites, A., Glasberg, B. R., Nimmo-Smith, I., Marslen-Wilson, W. D., & Moore, B. C. J. (2016). Representation of Instantaneous and Short-Term Loudness in the Human Cortex. *Frontiers in Neuroscience, 10*, 183. https://doi.org/10.3389/fnins.2016.00183

Thwaites, A., Nimmo-Smith, I., Fonteneau, E., Patterson, R. D., Buttery, P., & Marslen-Wilson, W. D. (2015). Tracking cortical entrainment in neural activity: Auditory processes in human temporal cortex. *Front Comp Neurosci., 1–14*. https://doi.org/10.3389/fncom.2015.00005

Thwaites, A., Schlittenlacher, J., Nimmo-Smith, I., Marslen-Wilson, W. D., & Moore, B. C. J. (2017). Tonotopic representation of loudness in the human cortex. *Hearing Research, 344*, 244–254. https://doi.org/10.1016/j.heares.2016.11.015

Thwaites, A., Wingfield, C., Wieser, E., Soltan, A., Marslen-Wilson, W. D., & Nimmo-Smith, I. (2018). Tracking cortical entrainment to the CIECAM02 and CIELAB color appearance models in the human cortex. *Vision Research., 145*, 1–10. https://doi.org/10.1016/j.visres.2018.01.011

Thwaites, A., Wieser, E., Soltan, A., Nimmo-Smith, I., Wingfield, C., Lakra, A., Woolgar, A., Parish, O., Yang, C., Zhang, C. (in prep.) The Kymata Atlas, a graph of information processing in the human cortex.

Vaina, L. M., Cowey, A., & Kennedy, D. (1999). Perception of first- and second-order motion: Separable neurological mechanisms? *Human Brain Mapping, 7*(1), 67–77. https://doi.org/10.1002/(SICI)1097-0193(1999)7:1<67::AID-HBM6>3.0.CO;2-K

Warren, P. A., & Rushton, S. K. (2009). Optic flow processing for the assessment of object movement during ego movement. *Current Biology, 19*(18), 1555–1560. https://doi.org/10.1016/j.cub.2009.07.057