Decoding neural responses to motion-in-depth using EEG

Marc M. Himmelberg[1,3], Federico G. Segala[1], Ryan T. Maloney[1], & Alex R. Wade[1,2]

[1] Department of Psychology, University of York, Heslington, York, YO10 5DD

[2] York Biomedical Research Institute, University of York, Heslington, York, YO10 5DD

[3] Department of Psychology, New York University, New York, NY, 10003

Corresponding Author: Marc M. Himmelberg, Department of Psychology, 6 Washington Place, New York University, NY, 10003 (marc.himmelberg@nyu.edu)

Abstract

Two potential stereoscopic cues that underlie the perception of motion-in-depth (MID) are changes in retinal disparity over time (CD) and interocular velocity differences (IOVD). These cues have independent spatiotemporal sensitivity profiles, suggesting that they are processed by distinct neural mechanisms. Here, we test whether we can accurately decode the patterns of electroencephalogram (EEG) signals that are generated in response to CD- and IOVD-isolating stimuli. Our results show that the decoder could distinguish between signals generated from CD and IOVD cues moving towards or away in depth. Critically, decoding performance could not be accounted for by differences in signal due to behavioural keyboard responses or by static disparity cues. Results from a cross-trained decoder indicated the presence of a relatively late neural mechanism that codes stereomotion irrespective of the low-level cues. These findings support the hypothesis that while CD and IOVD stimuli must drive very different initial neuronal mechanisms, they converge on a more central, amodal 'motion-in-depth' computation.

Keywords: Motion-in-depth, EEG, decoding, CD, IOVD

Introduction

To navigate the visual environment successfully, the primate visual system must interpret signals that indicate movement through three-dimensional space, or motion-in-depth (MID). Stereoscopic vision offers two potential cues to MID (Cumming & Parker, 1994; Harris, Nefs, & Grafton, 2008; Rashbass & Westheimer, 1961; Regan, 1993). First, the change in disparity over time (CD) can be computed from the temporal derivative of interocular differences (disparities) in retinal position. In this case, position is computed first, and temporal differences are computed second: objects moving towards or away from the observer have opposite-signed CD signals. Alternatively, the interocular velocity difference (IOVD) can be derived by reversing the order of the computation above: retinal velocities are computed first and the interocular difference between these velocities also provides a signed indication of 3D object velocity.

Electrophysiological studies in macaque have shown that CD and IOVD stimuli both drive responses in macaque MT, while functional magnetic resonance neuroimaging (fMRI) studies have similarly identified that such cues reliably activate human homologue MT+ and nearby regions (Czuba, Huk, Cormack, & Kohn, 2014; Likova & Tyler, 2007; Rokers, Cormack, & Huk, 2009; Sanada & DeAngelis, 2014). However, psychophysical measurements of CD and IOVD suggest that these mechanisms have different speed sensitivities, adapt independently, and engage systems with different spatial resolutions (Giesel, Wade, Bloj, & Harris, 2018; Joo, Czuba, Cormack, & Huk, 2016). Although these two mechanisms predominately activate the same neural substrates, they may be processed via two computationally distinct pathways.

Here, we ask whether there are differences in the overall pattern of neural signals generated in response to CD and IOVD cues. We used electroencephalography (EEG) to measure neural signals from 64 channels across the scalp, generated in response to well-isolated CD and IOVD stimulus cues that approached and receded in depth. We then tested whether a multivariate pattern classifier could use the overall pattern of EEG signals to accurately decode cue type (CD or IOVD) and depth direction (towards or away). The ability to decode these cues

would suggest that the evoked EEG responses are relatively independent of each other, a finding that is consistent with the hypothesis that CD and IOVD processes are carried by two distinct mechanisms.

Our CD stimuli carried a confound: while they had the same time-averaged depth, stimuli moving away from the observer necessarily have a closer starting point and more distant end point. Depth was therefore confounded with motion direction and our CD motion decoding could, in theory, be based on the time-averaged disparity around the stimulus onset. To examine this, we ran a control experiment using dynamically-updating CD stimuli that were located either near or far in depth, with no stereomotion information. This allowed us to examine the separate contributions of stereomotion and static CD depth information to decoding accuracy.

Although we equated low-level stimulus features as far as possible, CD and IOVD stimuli necessarily have fundamentally different low-level features. For example, individual CD elements are static and have shorter lifetimes than the moving IOVD elements. As expected, the decoder could accurately distinguish between EEG responses to CD and IOVD cues very early in the response timecourse. However, we also found accurate *within-cue* discrimination of 3D-motion direction: patterns of neural response were significantly different for the towards- and away- conditions for both IOVD and CD stimuli, suggesting that these could drive subtly different cortical 3D-direction selective cell populations - or drive similar populations with differences in synchrony or coherence. We validated these findings in three additional analyses that examined: the contribution of static disparity information to CD decoding performance, whether decoding could be influenced by motor responses generated from behavioural keyboard press, and whether a cross-trained decoder could accurately decode CD cues after being trained using IOVD signals, and vice versa.

Methods

Participants

In the first experiment, twelve healthy participants (mean age 24 years, eight males) were recruited from the University of York. All participants completed a pre-

screening to ensure they had functional stereoscopic vision. Two participants did not pass the pre-screening and did not participate in the experiment. Thus, ten participants (eight males) completed a one-hour EEG session. In a second 'static disparity' control experiment, nine participants (mean age 27 years, four males, including three participants from the previous experiment) completed a one-hour EEG session. All participants had normal or corrected-to-normal vision and provided written consent before participating in the study. Experiments were conducted in accordance with the Declaration of Helsinki and the experimental protocol was approved by the ethics committee at the University of York Department of Psychology.

Stimuli

CD and IOVD Stimuli

The stimuli were very similar to those used in previous papers (Kaestner et al., 2019; Maloney et al., 2018). Briefly, MID stimuli consisted of random dot patterns that were designed to experimentally dissociate CD- or IOVD-based mechanisms. The low-level properties of the dots were matched between cue types as far as possible. In all conditions, the dots were anti-aliased and had a Gaussian profile ($\sigma = 0.05°$). The black and white dots were positioned at a density of 1 dot per 1° of visual angle^2 on a mean luminance grey background, with a Michelson contrast of 100% (50:50 black:white). Stimuli were presented using MATLAB 2014a (The Mathworks Inc., Massachusetts, USA) and the Psychophysics Toolbox Version 3 (Brainard, 1997).

For CD stimuli, MID information is carried by changes in the disparities of pairs of dots presented to the left and right eye. To remove any velocity cues, dots can be replaced at new random positions on each video refresh. Thus, CD stimuli consisted of temporally uncorrelated dynamic random dot patterns, where each frame consisted of a set of binocularly correlated, randomly positioned dot patterns, only differing in a lateral shift in retinal disparity between the left and right eye (see Figure 1A). CD stimuli changed unidirectionally in disparity (i.e. the stimulus moved either towards or away in depth). This shift in disparity followed a single linear ramp that began at the far point and finished at the near point for towards motion, and the opposite points for away motion. The mean stimulus-averaged depth relative to fixation for both stimuli was

therefore zero. The near and far points were identical for each with a peak binocular disparity of ±32 arcmin which is below the fusion limit for typical human observers (Howard & Rogers, 2002; Norcia & Tyler, 1984).

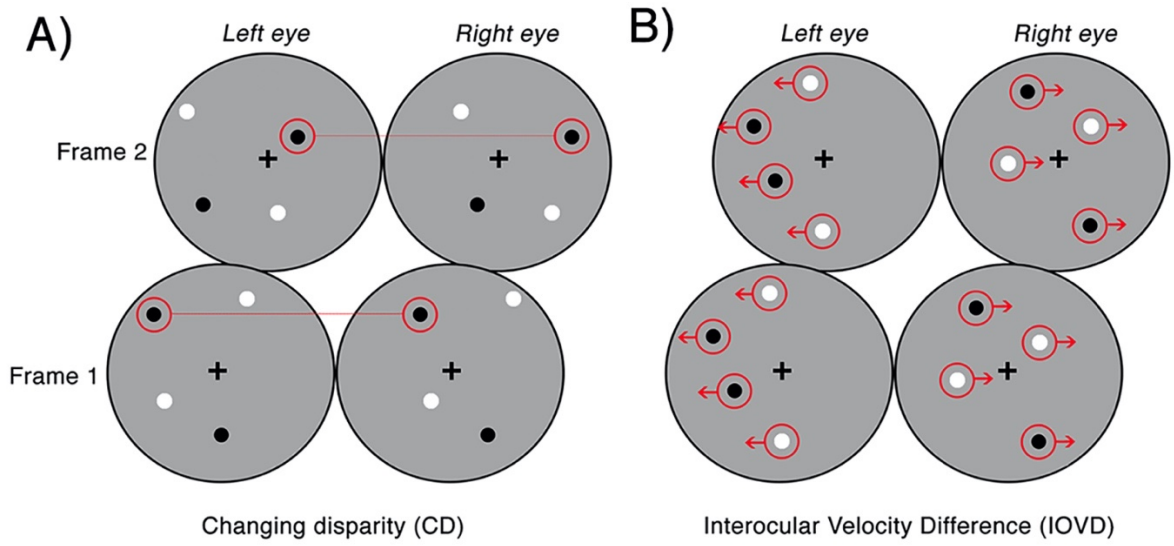

Figure 1 Examples of the CD and IOVD stimuli presented in two consecutive frames. Panel A) illustrates CD, where the dots are positioned randomly and paired between the left and right eyes, and only differ in their lateral shift in disparity across each video frame. Panel B) illustrates IOVD, where the dots laterally shift through time and there is no pairing of dots between the left and right eyes.

For IOVD stimuli, MID information is carried by the relative interocular velocities of retinal features across time. No fine-grained interocular matches are required (or indeed possible). Our IOVD stimuli consisted of fields of moving dots with the left and right eyes seeing uncorrelated dot patterns. The patterns moved coherently across time and the velocity of this motion was equal and opposite in the two eyes (see Figure 1B). The lateral shift of the dots that defined the unidirectional motion-in-depth was, again, in the form of a linear ramp. To maintain uniform dot density and equal visual transients across the stimulus, equal numbers of dots reached the end of their lifetime (50ms) and were 'reborn' in a new position on the stimulus for each frame of the stimulus. For IOVD stimuli, the maximum monocular lateral displacement was ±128 arcmin.

To ensure that no disparity information leaked into IOVD stimuli, the spatial positioning of the IOVD dot patterns was organised so that they fell into two horizontal strips that were two dot widths wide (~0.5°). The strips alternated across the eyes to ensure that they did not coincide on the two retinae (Shioiri, Nakajima, Kakehi, & Yaguchi, 2008; Shioiri, Saisho, & Yaguchi, 2000). On the rare occasion when a dot fell near the border of a horizontal strip in the left eye that was close to the dot near the border of the horizontal strip in the right eye, these dots were assigned opposite contrast polarity to disrupt any potential disparity signals (Maloney et al., 2018).

All stimuli were presented within an annulus that had an inner radius of 1° and an outer radius of 6° (see Figure 2). The contrast of the dots appearing at the edge of the annulus was smoothed with a cosine ramp that was 0.5° wide. To provide relative depth information, a circular central fixation lock (0.4° radius) surrounded the fixation cross, while a circular outer fixation lock (11.75° radius) surrounded the edge of the display. These locks were split into black and white quadrants that served as a set of nonius lines to assist in gaze stabilisation and the fusion of the images presented to the two retinae. An example of a single frame from the stimulus is presented in Figure 2. Animated examples of CD and IOVD stimuli are available in supplementary materials of previous work (Kaestner et al., 2019).

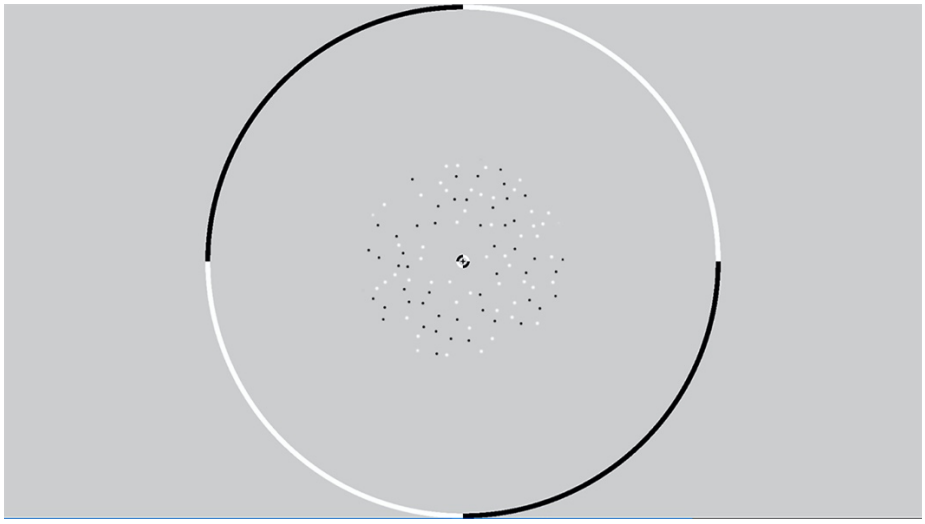

Figure 2 An example from a single frame of the random dot stereogram pattern that is presented within a circular fixation lock. The dots are presented in an annulus with an inner radius of 1° and an outer radius of 6°.

Static disparity stimuli

Using fixed far and near end-points for CD stimuli means that decoding between CD stimuli moving towards and away in depth could potentially rely on static disparity information from the beginning of the stimulus, rather than true MID information. To test this, we ran a control experiment that examined whether the decoder could distinguish between dynamically-updating but depth-fixed CD stimuli located at either the near or far positions in depth. These disparity stimuli had identical durations (250 ms), and dot update rates to the MID CD stimuli, however stereomotion velocity was set to zero.

Stimulus display

Stimuli were presented on a VIEWpixx/3D (VPixx Technologies, Saint-Bruno, Quebec, Canada) display (1920 x 1200 pixels,120 Hz refresh rate). The display had a mean luminance of 100 candela/m^2 and was gamma corrected using a photospectrometer (Jaz; Ocean Optics, Largo, FL, USA). Subjects viewed the display from 57 cm and were required to wear wireless liquid crystal display (LCD) shutter googles that were controlled by an infrared emitter (NVIDIA GeForce 3D; NVIDIA, Santa Clara, CA, USA). Here, binocular separation of the stimuli, with minimal crosstalk, was achieved by synchronising the VIEWpixx refresh rate (120 Hz, thus 60

Hz per eye) with the toggling of the LCD shutter goggles (Baker, Kaestner, & Gouws, 2016).

Experimental design

Participants' 3D vision was tested using the TNO Stereo Test, 19th edition (Lameris Ootech, NL) to ensure that they had functional stereo-acuity (with thresholds at or below 120 arcsec). During the EEG cap set-up, participants were presented with a demonstration of each of the four stimulus conditions on the VIEWpixx display, whereby the MID oscillated towards and away continuously according to the linear ramp. This allowed them to practise identifying CD and IOVD cues and direction. These demonstrations were presented until the participant was confident in identifying all four stimulus conditions. Following this, participants completed an additional two practice runs of the experiment itself. During testing, participants were allowed to return to practice runs if necessary.

Participants completed six blocks of 220 stimulus trials. Within each block the four stimulus conditions (CD towards, CD away, IOVD towards, and IOVD away) were randomised and occurred with equal frequency. For each participant, we then retained 210 of these trials to account for dropped triggers during testing. In the static disparity experiment, each block contained two stimulus conditions, CD near or CD far, that were also randomised and occurred with equal frequency. Participants were given a one-minute break between each block. For each trial, the stimulus probe had a duration of 250 ms, which corresponded to one full linear ramp (i.e. the stimulus followed a unidirectional trajectory over the 250 ms duration) for each probe event. The stimulus probe was followed by an inter-trial interval. The length of this interval was selected randomly from a uniform distribution (lower bound = 1 s, upper bound = 3 s). Participants were instructed to maintain their focus on the central fixation cross. For each trial, participants were instructed to respond using their right hand, via keyboard press, whether the stimulus was moving towards (2 on the number pad) or away (8 on the number pad) from them, and the reaction time of this keyboard press was recorded. In first three of the six blocks of the static disparity stimuli experiment participants did not give a response. However, in the second three blocks they were instructed to

respond, via the keyboard, whether they perceived the stimulus as being near or far in depth. Participants received audio feedback in the form of a tone to indicate incorrect responses only. Participants were not shown their overall response accuracy.

EEG recording and collection

EEG data were collected at 1 KHz from 64 electrodes that were distributed across the scalp according to the 10/20 system in an ANT WaveGuard EEG cap and digitized using the ASA Software (ANT Neuro, Hengelo, NL). The ground electrode was placed posterior to electrode *FPz* and the signal from each channel was referenced to a whole-head average. Eye-blinks were recorded using two vertical electrooculogram electrodes. Artefacts were removed using custom software that identified and removed bins with high slew rates or high overall noise levels in the eye channels. These signals were amplified, and the onset of each stimulus trial was recorded in the EEG trace using a low-latency digital trigger. Data were exported to MATLAB 2018a for offline analysis using customised script.

EEG pre-processing and bootstrapping

For each participant, EEG data were epoched into 1000 ms epochs at -200 ms to 800 ms, relative to stimulus onset at 0 ms (see Figure 3). A low-pass filter of 30Hz was applied to remove high frequency noise, including line noise. Each epoch was then down-sampled from 1000 Hz to 125 via linear interpolation. Thus, we retained a sampling pool of 1260 EEG epochs (210 trials x 6 blocks) that included each of the four stimulus conditions. From this sampling pool, we bootstrapped the data of 10 epochs from each condition which were averaged together to create a mean bootstrapped epoch for each condition. For each participant, this process was repeated 21 times so that there were 21 mean bootstrapped EEG epochs, for each of the 64 electrodes and for each condition. Bootstrap resampling and averaging were done in parallel across all EEG channels. Data were then z-scored to standardize across electrodes. This process was repeated for 1000 iterations of a bootstrapped support vector machine (SVM) classifier, with each iteration generating a unique set of 21 x 21 mean bootstrapped EEG epochs by averaging together new sets of sample epochs from the sampling pool to generate input data for each iteration of the SVM.

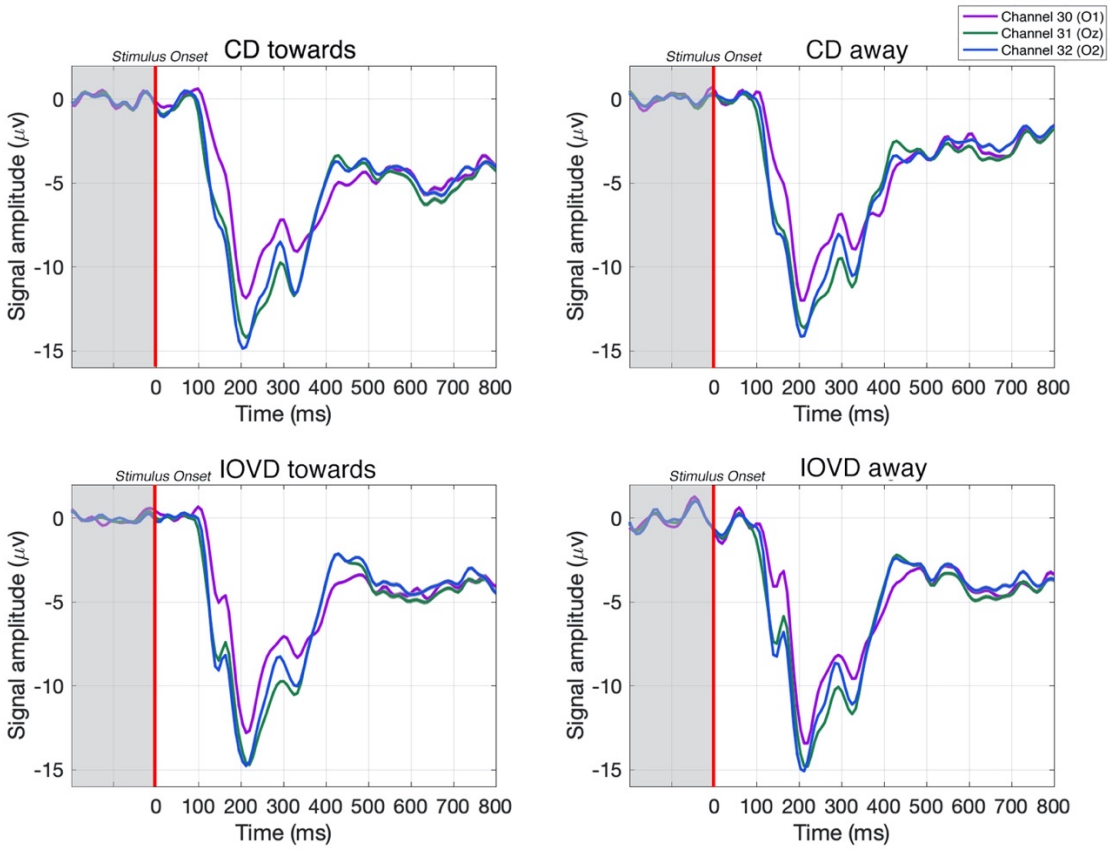

Figure 3 Examples of EEG waveforms from a single participant, taken from 3 occipital electrodes in response to each stimulus type over 210 trials. Each epoch is -200 to 800 ms and the stimulus is presented between 0 – 250 ms.

Bootstrapping keyboard response data

While EEG can measure responses to bottom-up static- and dynamic-depth stimuli in visual and parietal cortex that begin shortly after stimulus onset, the scalp distribution at the time of keyboard response is typically located over motor areas (Cottereau, Ales, & Norcia, 2014; Cottereau, McKee, Ales, & Norcia, 2011; Cottereau, McKee, & Norcia, 2013). As we decoded using all the electrodes across the scalp, we tested whether these motor responses (due to using the middle or index finger on the same hand to indicate if the stimulus was moving towards or away) had any influence on decoding performance. If our SVM was less accurate at decoding responses after they were pooled into keyboard-response conditions (with balanced stimulus types), rather than stimulus conditions, it would indicate that the EEG signals associated with the motor response have little influence on decoding accuracy and that our overall decoding results are driven predominately by sensory inputs.

We pooled our data into epochs labelled with either a 'towards' or an 'away' keyboard press response. This resulted in eight pools of data, for each participant: four in which the keyboard direction response matched the stimulus, four in which it was erroneous. We then performed a decoding analysis on keyboard button press using datasets containing equal numbers of 'towards' and 'away' stimuli. Nine participants were included in this analysis, as one participant had only a single epoch in one of these pools (as they were near perfect in identifying stimulus direction) and was not included. For each of the eight data pools, we bootstrapped 210 epochs by taking four random epochs from each data pool and averaging them together. Thus, each of the eight data pools was now equal in size (210 simulated epochs), with four pools containing data made from 'towards keyboard press' epochs and the other four pools containing data made from 'away keyboard press' epochs. The subsequent analysis was identical to our main decoding analysis, repeatedly bootstrapping 21 'towards keypress' epochs and 21 'away keypress' epochs for each of 1000 iterations of the classifier.

Decoding EEG signals using a support vector machine (SVM)

To decode MID information from our EEG data, we bootstrapped a pairwise SVM classifier with a linear kernel in MATLAB 2018a using the LIBSVM toolbox, ver. 3.23 (Chang & Lin, 2011). In the case of our EEG data, decoding could only occur at above-chance levels if different stimuli produce patterns of electrical activity at the scalp that differ in some consistent manner.

Decoding was performed at each of the 125 time points (from -200 ms to 800 ms), using data from all 64 electrodes. EEG data were bootstrapped repeatedly (as described above) and run through 1000 iterations of the SVM to derive a mean decoding accuracy at each time point, for each subject. Decoding accuracy was estimated using a leave-one-out (LOO) procedure, where the SVM is trained on all but one of the epochs and the left out epoch serves as the test data - the SVM is trained using 40 epochs while 1 epoch is reserved to be used as the test data (Boser, Guyon, & Vapnik, 1992; Cortes & Vapnik, 1995; Wang, 2018). The trial opposite to the omitted trial was withheld from the analysis. Thus, the SVM is trained to decode stimulus cue

by discriminating between neural responses arising from a pair of stimulus conditions. The trained decoder then predicts the class of the omitted test data (Grootswagers, Wardle, & Carlson, 2017). Chance baselines were verified by shuffling class labels and confirmed that SVM performance was not driven by artefacts in the signal processing pipeline (see Supplementary Figure 1 and Supplementary Figure 2 for decoding accuracies after shuffling labels). In an additional analysis we 'cross-trained' the SVM classifier. Here, the SVM is trained to decode a pair of stimulus conditions (i.e. CD towards vs CD away) and outputs a classification model. The decoder can then use this model to decode new, unseen data. In the additional analysis, we train the SVM to decode CD direction and use the output model to decode IOVD direction, and vice versa.

In principle, time points at which the decoder performs at above-chance (as assessed from the distribution of bootstrapped SVM classification accuracies) must have an overall pattern of neural signals that differs significantly between stimulus conditions. However, the large number of time points (125) and electrodes (64) tested can inflate the false discovery rate. To reduce the number of multiple comparisons, we used a non-parametric cluster correction to determine significant clusters across time and space (Maris & Oostenveld, 2007). Here, one sample t-tests were used to compare decoding accuracy to baseline accuracy (50%) at each of the 125 time points, using a Bonferroni-correction with a cluster significance threshold of $p < .05$.

Decoding was performed using six binary comparisons. First, we tested if we could decode CD and IOVD cue responses when pooled across direction. Next, we tested if could decode towards and away directions when pooled across MID cue type. Following this, we asked if we could decode stimulus direction *within* MID cue type, and between MID cue type *within* stimulus direction; therefore, comparing between CD towards against CD away cues, IOVD towards against IOVD away cues, CD towards against IOVD towards cues, and CD away against IOVD away cues.

Following this, we ran three support analyses. First, to examine the reliance of the decoder on static disparity information, we ran an analysis that tested whether we

could decode static disparity stimuli at either near or far positions in depth. Second, we used a simulated dataset in which EEG responses were pooled into keyboard-response conditions, rather than stimulus conditions, to test if we could decode keyboard response (i.e. either towards or away on the keyboard) irrespective of stimulus cue type or direction. Third, we cross-trained the decoder and tested if it could distinguish between CD direction cues using a decoding model that was trained using IOVD direction data, and vice versa.

Results

Behavioural responses

The overall percentage of correct responses and the mean reaction time to the CD and IOVD stimulus conditions (pooled across direction) and to the four separate stimulus conditions are presented in Figure 4.

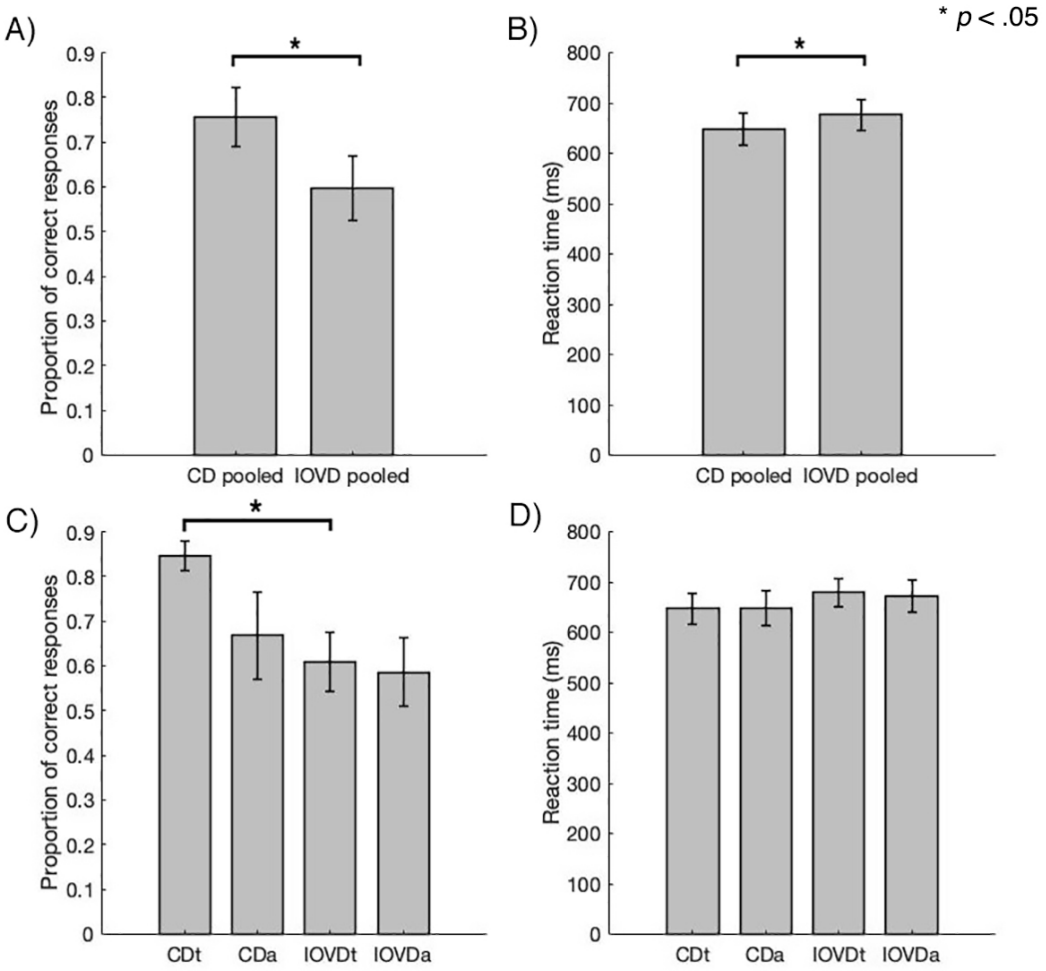

Figure 4 Bar plots of behavioural responses. In panel A) we present the mean proportion of correct responses after data are pooled into CD and IOVD conditions and in B) we present the mean reaction time after data are pooled into CD and IOVD conditions. In panel C) we present the mean proportion of correct responses comparing across all four stimulus conditions, and in panel C) we present the reaction time comparing across all four stimulus conditions.

First, a paired samples *t*-test was used to examine differences in the proportion of correct responses when comparing between CD and IOVD stimuli (pooled across direction). Participants were significantly better at discriminating the direction of CD stimuli when compared to IOVD stimuli, $t(19) = 2.269$, $p < .05$, with participants 15.9% more accurate in identifying the direction of CD stimuli (see Figure 4A). This analysis was repeated for reaction time, which revealed a significant difference in reaction time when comparing between CD and IOVD stimuli, $t(19) = -3.141$, $p < .05$, with

participants being on average 28 ms faster in identifying CD cues when compared to IOVD cues (see Figure 3B).

Next, we split our data into the four stimulus conditions and a four-way repeated measures ANOVA was performed to assess whether there were differences in the proportion of correct responses between the four conditions. Mauchly's test of sphericity was met ($\chi2(5) = 9.107$, $p = .107$). The ANOVA found that there was a main effect of stimulus condition on the proportion of correct responses $F(3,27) = 3.341$, $p < .05$, $\eta_p^2 = .271$, power = .693. Pairwise comparisons revealed that participants were more accurate at identifying CD towards responses when compared to IOVD towards responses ($p < .05$). No other comparisons were significant.

Similarly, a four-way repeated measures ANOVA was performed to assess whether there were differences in the reaction time between the four stimulus conditions. Mauchly's test of sphericity was met ($\chi2(5) = 6.580$, $p = .257$), however the ANOVA found that there was no significant difference in reaction time when comparing between the four stimulus conditions, $F(3,27) = 2.743$, p = .063, $\eta_p^2 = .234$, power = .598.

EEG Results

Decoding between cue type and direction: CD and IOVD stimuli give rise to different neural response patterns

Although we took care to match the low-level cue properties of the CD and IOVD stimuli as far as possible, they still differed in some respects. CD stimuli (deliberately) contained no coherent monocular motion energy while the dots in the IOVD stimuli travelled short distances on the retina before being refreshed. These differences are intrinsic to the stimulus cues themselves – a MID system that isolates CD will necessarily rely on different low-level aspects of the stimulus than an IOVD pathway, and these differences in stimulus properties cannot be avoided.

As presented in Figure 5, EEG responses to the different types of MID cue (irrespective of motion direction) could be decoded relatively early. Decoding occurred

above chance at 120 ms after stimulus onset and continued at above chance levels across the remaining time course (Bonferroni corrected, $p < .05$). Decoding accuracy peaked 224 ms after stimulus onset at 84%. The decoding performance for IOVD vs CD at 120 ms is the earliest difference that we see in this study. This result confirms the observation that systematic, low-level aspects of the MID stimuli drive different neural populations, or alternatively, drive similar neural populations that generate unique responses.

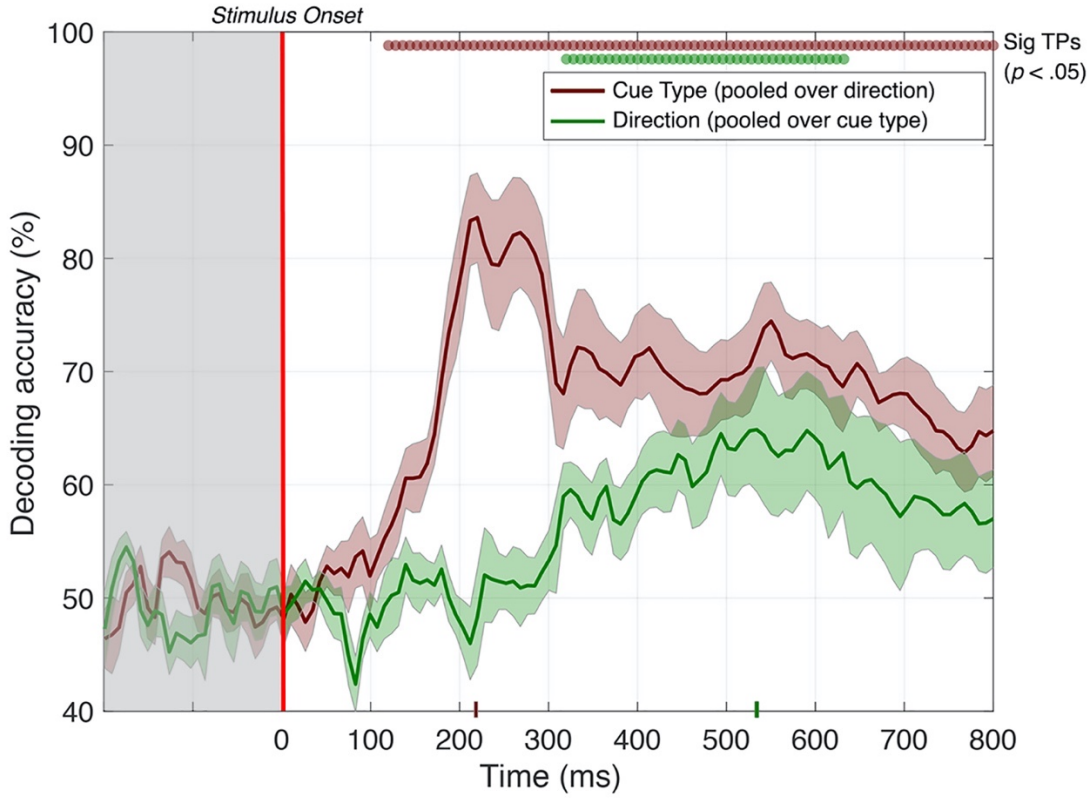

Figure 5 Decoding accuracy across 125 time points for cue type (pooled over direction) in red and direction (pooled over cue type) in green when all 64 electrodes are included as features. Cue type could be decoded from 120 ms - 800 ms after stimulus onset, while direction could be decoded from 320 – 632 ms after stimulus onset. The coloured ticks on the x-axis represent the time point of peak accuracy for each condition. Red and green circles indicate the points of time when the Bonferroni-corrected t-tests were significant ($p < .05$). Shaded error bars represent ±1SE of the bootstrapped mean (1000 iterations).

The ability to decode between CD and IOVD may be based on very low-level differences that are present at the earliest stage of cortical processing of MID. In Figure 6 we present a scalp montage of the support vector (SV) weights at the time of peak decoding accuracy. The weights indicate the mean contribution of each electrode to decoding performance across 1000 iterations of the bootstrapped decoder. The distribution of signal contributions appears to be complex and is not limited to occipital regions, however some interesting features are visible. In the CD vs IOVD (pooled across direction) distribution (Figure 6A) there is a significant input from both visual cortex and a band of activity in more superior regions that we hypothesise may be driven by the frontal eye fields, perhaps due to vergence in the IOVD condition (Giesel et al., 2017; Heinen, Rowland, Lee, & Wade, 2006).

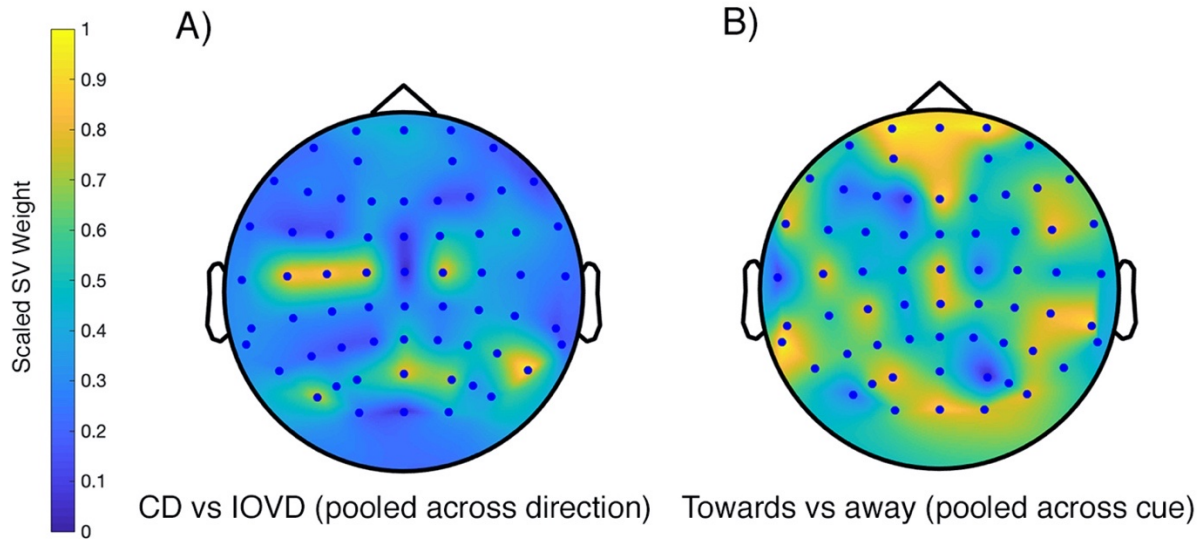

Figure 6 Distribution of mean standardised SV weights for CD vs IOVD (pooled across direction) and for towards vs away (pooled across MID cue) across 1000 decoder iterations, at their respective peak time point. In panel A) we present the mean SV weights at 224 ms for CD vs IOVD (pooled across direction) decoding, and in B) we present the mean SV weights at 536 ms for towards vs away (pooled across MID cue) decoding.

Due to the fundamental differences in CD and IOVD stimulus cues themselves, it is perhaps unsurprising that we are able to identify overall differences in the EEG response that they generate. However, our results show that we can also decode the *direction* of motion through depth (i.e. either moving towards or away), independent of

cue type. As illustrated in Figure 5 (green line), after collapsing over CD and IOVD cue types, we were able to decode towards vs away motion direction at above-chance levels from 320 ms until 632 ms after stimulus onset, peaking at 65% at 536 ms ($p <$ .05). The mean towards vs away (pooled MID cue) SV scalp distribution (Figure 6B) is more complex. This may be due to the peak occurring at a later time point when compared to the CD vs IOVD distribution. The complexity of this distribution could reflect contributions from multiple visual areas and feedback signals, although much of the contribution does still appear to arise in early visual cortex.

Decoding within cue-type and direction

Decoding performance pooled across stimulus cue type, and direction, can be interpreted as CD and IOVD cue properties, and different motion directions, recruiting different neural populations. Alternatively, these cues could activate similar neural populations but generate unique responses that differ in their coherence or synchrony. However, because of the way that we pooled the data, it is possible that decoding performance is driven by only one of the two pooled elements at any particular time point.

To address this, we asked whether we could decode cue type *within* individual motion directions and whether we could decode motion direction *within* individual cue types. Our data (Figure 7) show that we could do so at above-chance levels. Decoding between CD towards and IOVD towards stimuli (Figure 7, blue line) was significant from 128 ms and peaked at 82% at 264 ms. Decoding between CD away and IOVD away stimuli (Figure 7, orange line) was significant from 136 ms and peaked at 78% at 224 ms. Performance in decoding between cue-type within motion direction starts at a very similar time point, peaks early and stays high, regardless of direction.

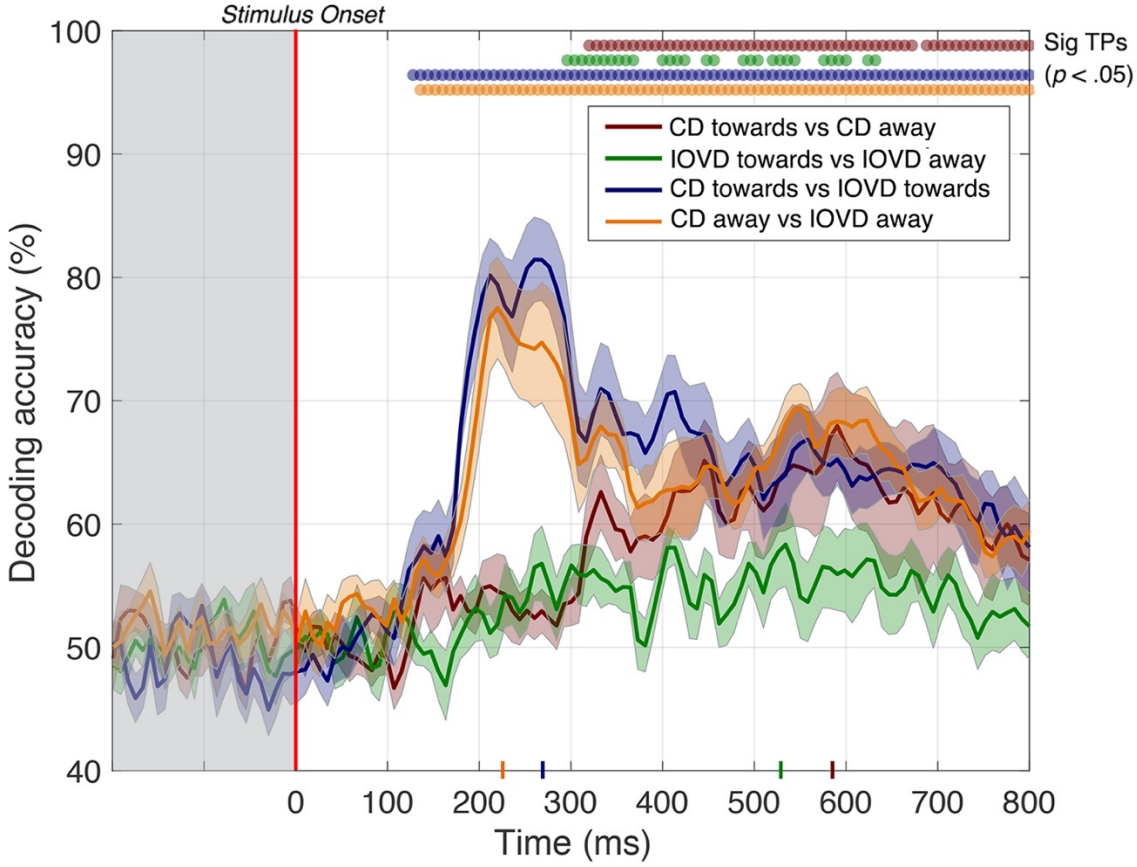

Figure 7 Pairwise decoding accuracy across 125 time points for four stimulus conditions. CD vs IOVD cue type (within motion direction) can first be decoded from 128 ms (towards) and 136 ms (away) after stimulus onset, while stimulus direction (within cue type) can first be decoded at 320 ms (CD) and 296 ms (IOVD) after stimulus onset. Red, green, blue, and orange dots indicate time points when the Bonferroni-corrected t-tests were significant for each condition ($p < .05$) and the coloured ticks on the x-axis represent the time point of peak decoding performance. Shaded error bars represent ±1SE of the bootstrapped mean (1000 iterations).

Most surprisingly though, we are able to obtain above-chance performance in decoding the motion direction of both CD and IOVD stimuli ($p < .05$) (Figure 7). CD towards vs CD away stimuli could be decoded from 320 ms - 672 ms, then again from 688 ms – 800 ms, with a peak decoding accuracy of 68% at 592 ms. IOVD towards against IOVD away stimuli could be intermittently decoded from 296 ms post-stimulus onset, with a decoding peak of 58% at 536 ms (Figure 7, green line). The decoder

could first distinguish between IOVD direction 24 ms before CD direction and IOVD direction decoding accuracy (536 ms) peaked 56 ms before CD direction decoding accuracy (592 ms).

Scalp montages illustrating mean electrode contributions for each of the four conditions across 1000 bootstrapped runs of the decoder, at their respective peak time of decoding, are presented in Figure 8. Again, the distribution of signal contributions is complex and is not confined to the visual cortex. As expected, the distribution for CD vs IOVD comparisons for towards (Figure 8C) and away (Figure 8D) are similar, and like the pooled distributions in Figure 6A, significant contributions appear from early visual cortex and there is a band of activity in more frontal regions that may be driven by the frontal eye fields (Heinen et al., 2006). Signal contributions for the within direction comparison (Figure 8A and B) are complex (again, perhaps due to the later time point of this montage), however there is clearly a large influence of occipital regions. Notably, for the CD towards vs CD away scalp distribution (Figure 8A), there is strong response over the occipital cortex, with a second peak over superior regions of the scalp. This distribution is not dissimilar to those from a previous experiment that found occipital and superiorly located responses at a similar time point for CD motion stimuli (Cottereau et al., 2014).

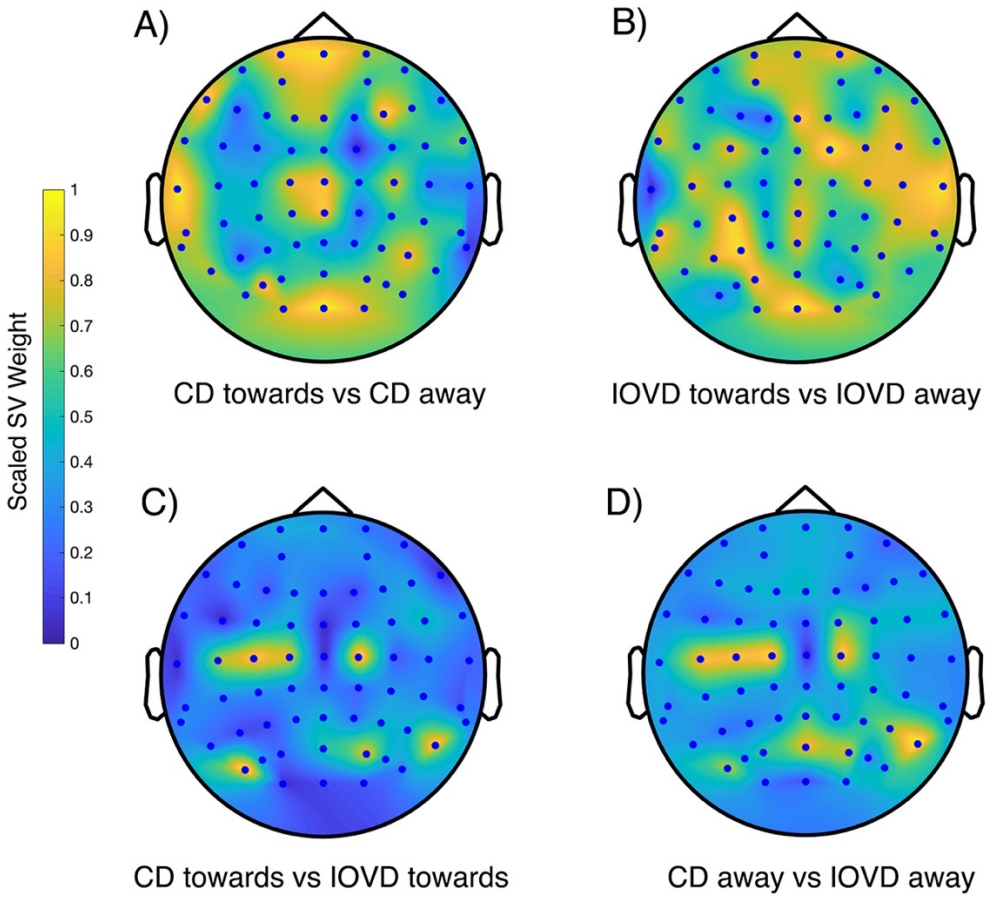

Figure 8 Distribution of mean standardised SV weights at their peak time point of decoding across 1000 iterations of the decoder, plotted onto a scalp montage for the four comparisons. In panel A) we present the mean SV weights for decoding CD towards vs CD away responses (592 ms), in panel B) for IOVD towards vs IOVD away responses (536 ms), in panel C) for CD towards against IOVD towards (264 ms), and in panel D) for CD away against IOVD away (224 ms).

The ability of EEG to resolve these responses to motion direction is intriguing. The ability to decode depth direction from EEG data suggests that these responses could potentially come from separate networks of depth-direction selective neurons.

Decoding static disparity stimuli: Static depth information contributes to early stages of MID decoding, but not late stages

A potential confound to our experiment is that the CD stimulus begins at one extreme in depth (i.e. either near or far) before traversing towards or away from the observer. It is therefore possible that decoding performance for the CD towards vs CD away condition relies on static disparity information from a single time point, rather than MID cue information. To address this, we ran a second experiment to test how a decoder would perform on CD stimuli at fixed depths. These stimuli were dynamic random dot stereograms presented for 250 ms at one of two fixed depth locations – near or far. If the decoder *cannot* differentiate between these two conditions, it might suggest that decoding performance relies on MID information to decode CD stimuli, rather than differences in static disparity information at each time point. Conversely, if there are time points where near and far static disparity stimuli can be decoded, this static information must have some contribution to MID CD decoding performance.

Our data (Figure 9) show that the timecourse of decoding accuracy was different compared to that found for CD MID stimuli in Figure 7. We observed a short set of time points where the decoder could distinguish between near and far depths. Irrespective of keyboard responses, decoding was possible between 280 – 366 ms (Figure 9, with keyboard response: orange line), and 288 – 312 ms (Figure 9, without keyboard response: purple line). It was also possible (briefly) much later at 632 - 648 ms but only when participants were not required to give a keyboard response (Figure 9, purple line). Accuracy was, overall, far lower (~58%) when compared to the original comparison using MID stimuli (see Figure 7, red line).

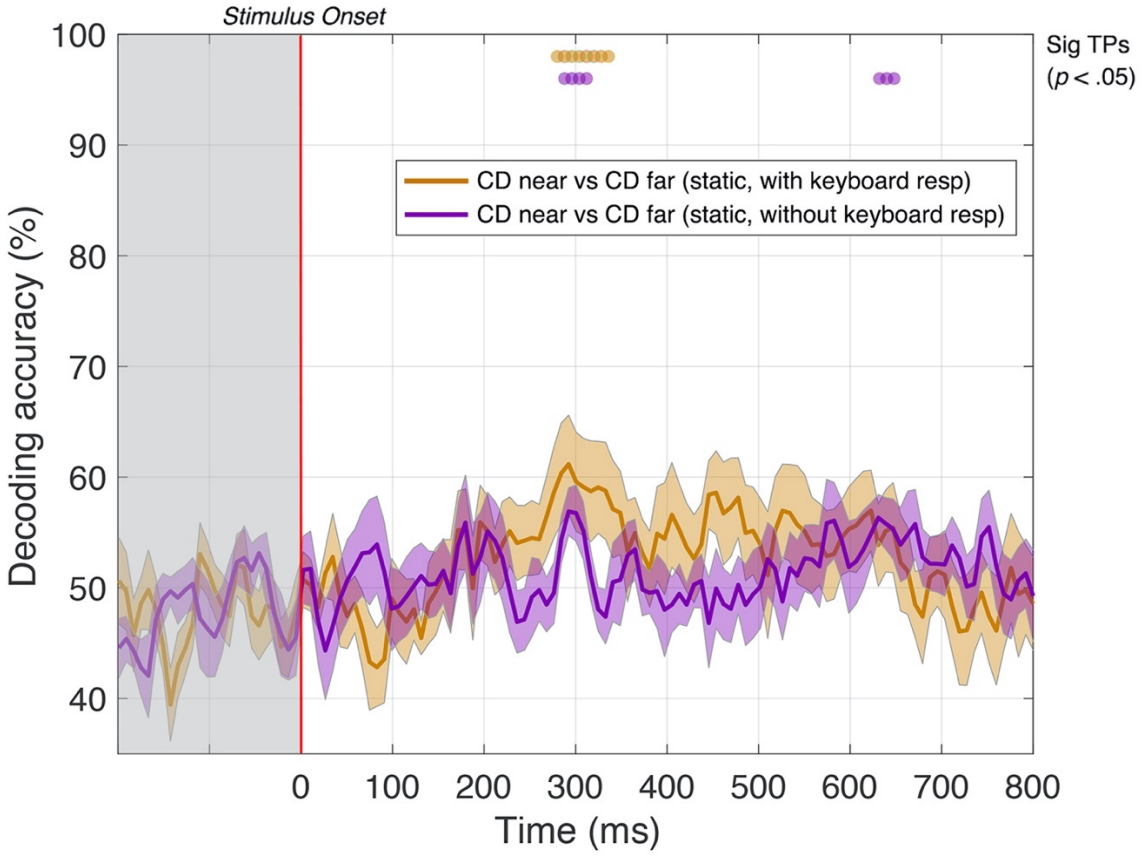

Figure 9 Pairwise decoding accuracy across 125 time points for static disparity stimuli that are presented either near or far in depth. The orange line represents trials when participants were required to respond, via keyboard, whether they perceived the stimulus as either near or far in depth. The purple line represents trials when no keyboard response was required. Orange and purple dots indicate time points when the Bonferroni-corrected t-tests were significant for the 2 conditions ($p < .05$). Shaded error bars represent ±1SE of the bootstrapped mean (1000 iterations).

Generally, it appears that the decoder can distinguish between static disparity stimuli located near or far in depth for a short 'early' period around 300 ms. This differs from MID CD stimuli, where decoding was also possible from around 300 ms but *continued* to be significant across the rest of the time course. It may be that transient static depth information contributes to early stages of MID depth decoding, while the later stages of decoding may be attributed to the smooth motion information that was not present in these stimuli.

Decoding keyboard responses: motor responses do not contribute to decoding performance

On each trial, participants were required to respond, via the keypad, whether they perceived the stimulus as moving towards or away in depth. The signals generated from pressing one of the two keys could, potentially, influence decoding accuracy. To address this, we ran the decoder on simulated data that differed only according to how participants responded on the keyboard during the trial. The results (Figure 10) show that it is not possible to decode keyboard responses independently of the stimulus, with accuracy hovering around the 50% chance baseline across the majority of the 125 time points. A similar analysis, where we use keyboard response data to decode conscious perception of stimulus direction, is presented in Supplementary Information Figure S3. These results suggest that stimulus decoding performance is likely to be driven by MID cue properties, rather than differences in signal from the motor cortex in response to keyboard button press.

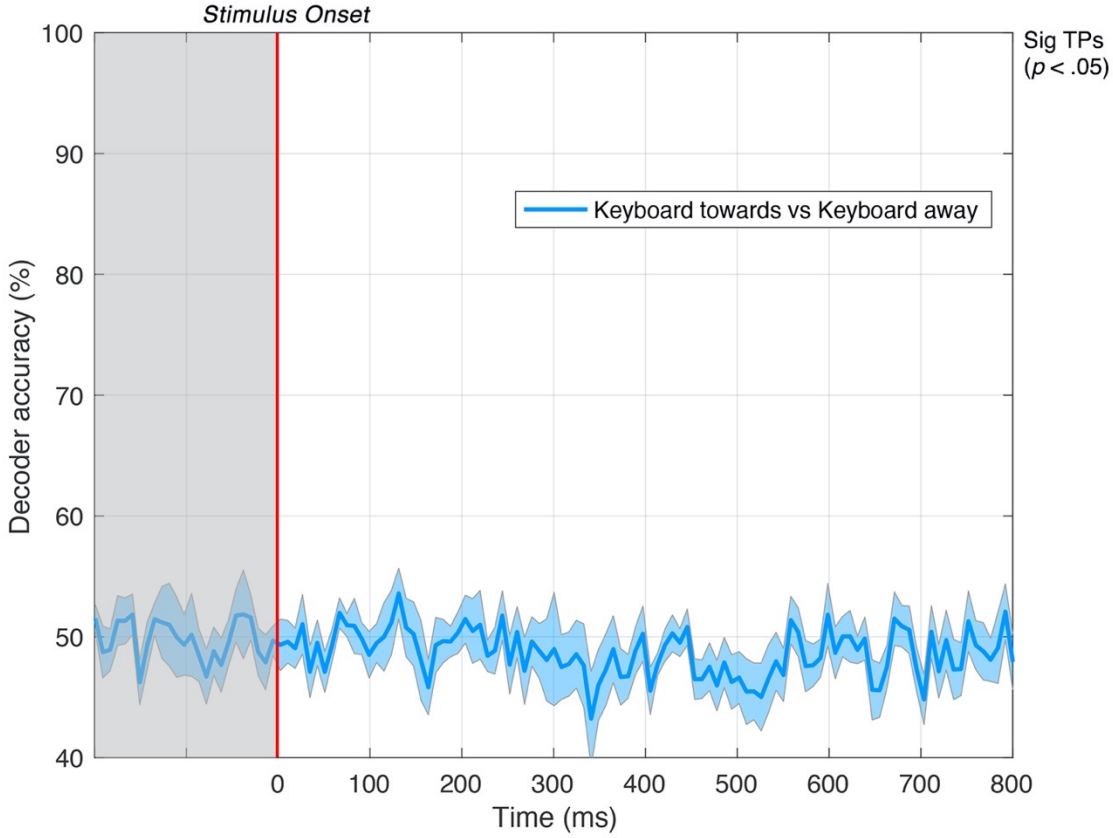

Figure 10 Pairwise decoding accuracy across 125 time points for keyboard response. Keyboard response (either pressing 2 on the keypad to indicate towards or 8 on the

keypad to indicate away) cannot be accurately decoded: accuracy is at chance across all time points. Shaded error bars represent ±1SE of the bootstrapped mean (1000 iterations).

Decoding from a cross-trained decoder: early performance is associated with cue differences while later performance could extend from a shared MID mechanism

Next, we examined whether decoder performance relied on unique CD and IOVD signals. Here, we trained the decoder using CD responses, and then cross-validated against IOVD responses and vice versa. If these signals are truly distinct and independent, one might expect accuracy to fall around 50% chance, as a decoder trained on data from one cue (i.e. CD) should not be helpful in decoding data from the other cue (i.e. IOVD). However, if cross-trained decoding is possible, it suggests some commonality in the pattern of responses generated from CD and IOVD. We note that because IOVD stimuli have no unique depth, this cross-validation information must be independent of confounds from pure disparity cues.

We could decode motion direction from the cross-trained classifier for significant periods after ~500 ms. Specifically, IOVD direction could first be decoded from a CD-trained classifier at 512 ms after stimulus onset (56% accuracy) with intermittent significance through to 800 ms (see Figure 11, purple line, $p < .05$). A similar result was found for decoding CD direction from an IOVD-trained classifier: significant decoding first occurred at 488 ms (again, 56% accuracy) through to 552 ms, and again was intermittently significant through to 648 ms (see Figure 11, green line, $p < .05$). Importantly, decoding was not possible during the stimulus presentation period (0 – 250 ms) or soon thereafter. Here, the ability of the cross-trained decoder to perform *within* cue direction decoding from a classifier trained on another cue type suggests that there must be *some* shared MID computation at later stages of processing that is agnostic to low-level cue properties.

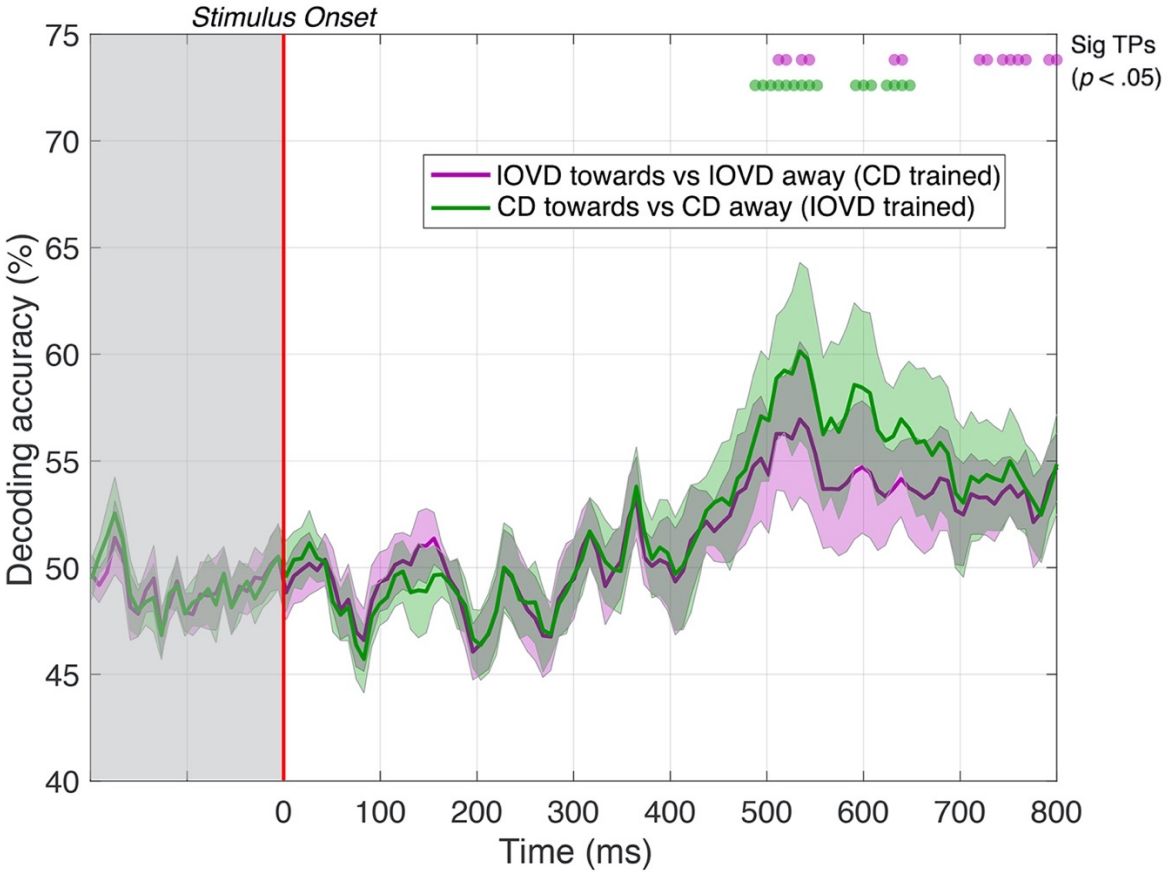

Figure 11 Pairwise decoding accuracy of a cross-trained decoder across 125 time points, comparing between CD and IOVD directions. The purple line represents decoding accuracy for IOVD towards vs IOVD away after decoder is trained using CD direction data. The green line represents decoding accuracy for CD towards vs CD away after the decoder is trained using IOVD direction data. Both CD and IOVD direction can be decoded intermittently from around 500 ms. Purple and green dots indicate time points when the Bonferroni-corrected *t*-tests were significant for the 2 conditions ($p < .05$). Shaded error bars represent ±1SE of the bootstrapped mean (1000 iterations).

Discussion

We have examined whether, and when, we can decode EEG signals generated in response to CD- and IOVD-isolating stimuli, moving either towards or away in depth. We find that both cue *type* (CD and IOVD) and motion *direction* (towards and away) can be decoded based on the distinct pattern of neural responses measured at the scalp. We also find that overall performance in decoding CD direction is not primarily

due to the presence of static disparity information, although this information might contribute to early parts of the EEG time course. Further, we could not decode data that was partitioned into keyboard response conditions, suggesting that decoding performance is likely to be driven by MID stimulus cues rather than by motor responses. Finally, a cross-trained decoder could not decode EEG signals at early stages of the time course but could at later stages.

Decoding MID cues

CD and IOVD mechanisms are thought to drive independent computations, and engage systems with different spatial and temporal resolutions (Czuba, Rokers, Huk, & Cormack, 2010; Shioiri et al., 2008; Wardle & Alais, 2013). Accurate decoding between CD and IOVD cues suggests that the neural signals that are processed downstream to these initial computations are also relatively independent from each other. Similarly, we could decode between CD and IOVD cues that were moving in the same direction in depth, indicating that the high performance of the decoder was indeed driven by differences in the properties of CD and IOVD cues, rather than motion direction. We interpret these results as evidence that CD and IOVD cues are processed through two distinct computational mechanisms and they may be further processed by different pathways in the visual cortex (Harris et al., 2008; Rokers et al., 2009). Decoding accuracy was high, sitting around 84% at its peak. Although we made every attempt to match the low-level features of our CD and IOVD stimuli they were, necessarily, different. For example, CD monocular dot fields contain no monocular coherent motion energy and have a short lifetime, refreshing on every frame (i.e. every 16.6 ms across the two eyes), whereas IOVD dots travel short distances and have a relatively longer lifetime of 50 ms. Thus, it is likely that intrinsic differences in CD and IOVD cue properties contributed to the inter-cue decoding accuracy that we observed.

Decoding MID direction

Importantly, the decoder could distinguish between signals generated from stimuli moving either towards or away in depth within CD or IOVD cue types. It is well-known that neurons in primary visual cortex and MT+ are selective for 2D motion-direction (Born & Bradley, 2005; Hubel & Wiesel, 1959, 1968). Recent studies have

also found evidence for neurons that are selective for 3D-motion direction, with both fMRI and psychophysical evidence for 3D direction-selective adaptation (Joo et al., 2016; Rokers et al., 2009). At the retinal level, the stimuli for the two motion directions (i.e. towards or away for a single cue type) are similar. Decoding motion direction in the case of the CD stimulus must depend on a mechanism that can compute a temporal derivative from disparity selective neurons in primary visual cortex. Decoding in the case of the IOVD stimulus must be driven, at a minimum, by differential responses from monocular motion-selective neurons or by neurons that receive input from those populations. Most models of MID processing would posit that these direction selective cell populations are intermingled within individual visual areas and so the ability to resolve these responses using EEG is intriguing. A common suggestion within the field of EEG multivariate decoding is that neurons with different stimulus sensitivities are not co-mingled at random, but rather, are arranged in macroscale structures (such as columns) whose electrical signals might be differentiated – perhaps because of additional selectivity imposed by the curvature of the cortical surface. An alternative model might suggest that towards and away motion directions drive similar cell populations, however they might do so with differences in coherence or synchrony, and this is what decoding performance is based on. Overall, these data provide some support for the existence of clusters of 3D-motion direction selective neurons, although we cannot identify the spatial scale of this clustering. It may be that our results are driven by direction selective neurons organised into columnar-scale structures within individual visual areas (much as 2D direction selective neurons are), or that common populations of neurons process motion direction information but give rise to unique signals that potentially differ in their timing, synchrony, or coherence.

Decoding static disparity

To examine the contribution of the time-averaged disparity information to MID CD decoding performance, we ran a second experiment that used depth-fixed random dot stereogram stimuli with an identical distribution of starting positions, durations, and dot update rates to our original experiment. Although it was possible to decode static disparity information from the EEG signal at a few time points around 300 ms, decoding performance differed notably from MID CD performance, which also occurred at 300

ms but then remained above chance for the remainder of the EEG time course. Although static depth information may drive an early phase of MID CD decoding (i.e. the time point where decoding was possible for both experiments, around 300 ms) later stages of MID decoding (beyond 300 ms) appear to be attributable to motion through depth rather than depth *per se*.

Cross trained decoder

Intriguingly, we found that a cross-trained decoder was able to discriminate between CD and IOVD direction at later stages of the EEG signal, beyond ~500 ms. The cross-trained decoder's performance at later stages of the EEG time course could depend on signals originating from a late-stage mechanism involved in MID information processing that is agnostic to differences in CD and IOVD cue properties. One candidate cortical location for this general stereomotion processing computation is the region around the human MT+ complex which includes neurons known to be sensitive to both lateral motion defined from a variety of cues, as well as 3D motion (Czuba et al., 2014; Huk, 2012; Kaestner et al., 2019; Likova & Tyler, 2007; Rokers et al., 2009; Sanada & DeAngelis, 2014).

Post-stimulus response and keyboard analysis

We found that cue type and direction could be accurately decoded well beyond the stimulus-offset at 250 ms. Even brief stimuli can produce a complex temporally sustained pattern of cortical activity (presumably including feed-forward and feedback signals, and phase-resetting of endogenous rhythms) that persists for hundreds of milliseconds after stimulus offset (Bertamini, Rampone, Oulton, Tatlidil, & Makin, 2019; Mullinger, Cherukara, Buxton, Francis, & Mayhew, 2017; Mullinger, Mayhew, Bagshaw, Bowtell, & Francis, 2013; Stevenson, Brookes, & Morris, 2011). Importantly, the post-stimulus decoding seen in this study cannot be accounted for by different keyboard motor responses as the decoder was unable to differentiate between towards and away keyboard responses. This has further implications. By decoding keyboard responses, we are effectively decoding the *percept* of 3D motion direction, that is, observers perceiving the stimulus either moving towards or away in depth (as well as occasional 'button errors'). Thus, these results suggest that it is not possible to decode

signals relating to the conscious percept of 3D motion direction for MID cues – although some brain regions (including motor cortex) do presumably carry such signals.

Conclusions

MID cues (CD and IOVD) and 3D-motion direction (towards and away) can be accurately decoded from the distinct pattern of EEG signals that they generate across the scalp. We have interpreted our findings with the hypothesis that CD and IOVD are processed through two computationally distinct mechanisms that begin in the retina and may be further processed via independent large-scale networks in the visual cortex. Further, these data are consistent with reports suggesting the existence of 3D-motion direction selective neurons that may exist at a finer scale and are intermingled across the visual cortex. We found that static disparity information may play a small, early role in CD direction decoding, however this information cannot account entirely for the decoding performance we see in MID CD decoding. Finally, we found that early stages of the EEG signal may rely upon MID cue properties, while later stages of the signal could potentially be driven by a more central, amodal motion-in-depth mechanism.

Acknowledgements: M.M.H was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 641805. A.R.W was supported by UK Biotechnology and Biological Science Research Council (BBSRC) grant number BB/M002543/1. The authors are grateful to Miaomiao Yu for assistance in setting up the EEG system.

Author contributions: A.R.W and R.T.M conceptualized the experiments, M.M.H and F.G.S performed the experiments, M.M.H and A.R.W analysed the data, M.M.H wrote the paper, and M.M.H, R.T.M and A.R.W revised the paper. M.M.H and F.G.S contributed equally to this paper.

Declaration of interest: None.

Data and code availability: All data and code analysed in the current study may be requested from the authors and will be made available on OSF at the time of publication.

References

Baker, D. H., Kaestner, M., & Gouws, A. D. (2016). Measurement of crosstalk in stereoscopic display systems used for vision research. *J Vis*, *16*(15), 14. https://doi.org/10.1167/16.15.14

Bertamini, M., Rampone, G., Oulton, J., Tatlidil, S., & Makin, A. D. J. (2019). Sustained response to symmetry in extrastriate areas after stimulus offset: An EEG study. *Scientific Reports*, *9*(1), 4401. https://doi.org/10.1038/s41598-019-40580-z

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*, *28*, 157–189. https://doi.org/10.1146/annurev.neuro.26.041002.131052

Boser, B., Guyon, I., & Vapnik, V. (1992). A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, 144–152. https://doi.org/10.1145/130385.130401

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Chang, C., & Lin, C. (2011). LIBSVM -- A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1–27. https://doi.org/10.1145/1961189.1961199

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*, 273–297. https://doi.org/10.1023/A:1022627411411

Cottereau, B. R., Ales, J. M., & Norcia, A. M. (2014). The evolution of a disparity decision in human visual cortex. *NeuroImage*, *92*, 193–206. https://doi.org/10.1016/j.neuroimage.2014.01.055

Cottereau, B. R., McKee, S. P., Ales, J. M., & Norcia, A. M. (2011). Disparity-tuned population responses from human visual cortex. *Journal of Neuroscience*, *31*(3), 954–965.

Cottereau, B. R., McKee, S. P., & Norcia, A. M. (2013). Dynamics and cortical distribution of neural responses to 2D and 3D motion in human. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00549.2013

Cumming, B. G., & Parker, A. J. (1994). Binocular mechanisms for detecting motion-in-depth. *Vision Research*, *34*(4), 483–495.

Czuba, T. B., Huk, A. C., Cormack, L. K., & Kohn, A. (2014). Area MT encodes three-dimensional motion. *J Neurosci*, *34*(47), 15522–15533. https://doi.org/10.1523/JNEUROSCI.1081-14.2014

Czuba, T. B., Rokers, B., Huk, A. C., & Cormack, L. K. (2010). Speed and eccentricity tuning reveal a central role for the velocity-based cue to 3D visual motion. *J Neurophysiol*, *104*(5), 2886–2899. https://doi.org/10.1152/jn.00585.2009

Giesel, M., Harris, J., Yakovleva, A., Wade, A., Bloj, M., & Norcia, A. (2017). Comparison of horizontal vergence responses to changing disparity and inter-ocular velocity differences. *Journal of Vision*, *17*(10), 412.

Giesel, M., Wade, A. R., Bloj, M., & Harris, J. M. (2018). Investigating Human Visual Sensitivity to Binocular Motion-in-Depth for Anti- and De-Correlated Random-Dot Stimuli. *Vision*, *2*(4), 41. https://doi.org/10.3390/vision2040041

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Harris, J. M., Nefs, H. T., & Grafton, C. E. (2008). Binocular vision and motion-in-depth. *Spatial Vision*, *21*(6), 531–547.

Heinen, S. J., Rowland, J., Lee, B.-T., & Wade, A. R. (2006). An Oculomotor Decision Process Revealed by Functional Magnetic Resonance Imaging. *Journal of Neuroscience*, *26*(52), 13515–13522. https://doi.org/10.1523/JNEUROSCI.4243-06.2006

Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth* (Vol. 2). Toronto: I. Porteous.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, *148*(3), 574–591. https://doi.org/10.1113/jphysiol.1959.sp006308

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and the functional architecture of monkey striate cortex. *Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Huk, A. C. (2012). Multiplexing in the primate motion pathway. *Vision Research*, *62*, 173–180.

Joo, S. J., Czuba, T. B., Cormack, L. K., & Huk, A. C. (2016). Separate perceptual and neural processing of velocity- and disparity-based 3D motion signals. *J Neurosci*, *36*(42), 10791–10802. https://doi.org/10.1523/JNEUROSCI.1298-16.2016

Kaestner, M., Maloney, R. T., Wailes-Newson, K. H., Bloj, M., Harris, J. M., Morland, A. B., & Wade, A. R. (2019). Asymmetries between achromatic and chromatic extraction of 3D motion signals. *Proceedings of the National Academy of Sciences*, *116*(27), 13631–13640. https://doi.org/10.1073/pnas.1817202116

Likova, L. T., & Tyler, C. W. (2007). Stereomotion processing in the human occipital cortex. *NeuroImage*, *38*(2), 293–305. https://doi.org/10.1016/j.neuroimage.2007.06.039

Maloney, R. T., Kaestner, M., Bruce, A., Bloj, M., Harris, J. M., & Wade, A. R. (2018). Sensitivity to velocity- and disparity-based cues to motion-In-depth with and without spared stereopsis in binocular visual impairment. *Investigative Ophthalmology & Visual Science*, *59*(11), 4375–4383. https://doi.org/10.1167/iovs.17-23692

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mullinger, K. J., Cherukara, M. T., Buxton, R. B., Francis, S. T., & Mayhew, S. D. (2017). Post-stimulus fMRI and EEG responses: Evidence for a neuronal origin hypothesised to be inhibitory. *NeuroImage*, *157*, 388–399. https://doi.org/10.1016/j.neuroimage.2017.06.020

Mullinger, K. J., Mayhew, S. D., Bagshaw, A. P., Bowtell, R., & Francis, S. T. (2013). Poststimulus undershoots in cerebral blood flow and BOLD fMRI responses are modulated by poststimulus neuronal activity. *Proceedings of the National*

Academy of Sciences, *110*(33), 13636–13641. https://doi.org/10.1073/pnas.1221287110

Norcia, A. M., & Tyler, C. W. (1984). Temporal frequency limits for stereoscopic apparent motion processes. *Vision Research*, *24*(5), 395–401.

Rashbass, C., & Westheimer, G. (1961). Independence of conjugate and disjunctive eye movements. *J Physiol*, *159*, 361–364.

Regan, D. (1993). Binocular correlates of the direction of motion in depth. *Vision Research*, *33*(16), 2359–2360. https://doi.org/10.1016/0042-6989(93)90114-C

Rokers, B., Cormack, L. K., & Huk, A. C. (2009). Disparity-and velocity-based signals for three-dimensional motion perception in human MT+. *Nature Neuroscience*, *12*(8), 1050–1055.

Sanada, T. M., & DeAngelis, G. C. (2014). Neural representation of motion-in-depth in area MT. *J Neurosci*, *34*(47), 15508–15521. https://doi.org/10.1523/JNEUROSCI.1072-14.2014

Shioiri, S., Nakajima, T., Kakehi, D., & Yaguchi, H. (2008). Differences in temporal frequency tuning between the two binocular mechanisms for seeing motion in depth. *J Opt Soc Am A Opt Image Sci Vis*, *25*(7), 1574–1585. (18594613).

Shioiri, S., Saisho, H., & Yaguchi, H. (2000). Motion in depth based on inter-ocular velocity differences. *Vision Res*, *40*(19), 2565–2572. (10958908).

Stevenson, C. M., Brookes, M. J., & Morris, P. G. (2011). β-Band correlates of the fMRI BOLD response. *Human Brain Mapping*, *32*(2), 182–197. https://doi.org/10.1002/hbm.21016

Wang, L. (2018). *Support Vector Machines: Theory and Applications (Studies in Fuzziness and Soft Computing)*. Berlin: Springer-Verlag.

Wardle, S. G., & Alais, D. (2013). Evidence for speed sensitivity to motion in depth from binocular cues. *Journal of Vision*, *13*(1). https://doi.org/10.1167/13.1.17