Decoding the neuronal correlates of motion-in-depth from EEG

Marc M. Himmelberg[1], 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

Corresponding Author: Marc M. Himmelberg, Department of Psychology, the University of York, Heslington, YO10 5DD (marc.himmelberg@york.ac.uk)

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 neuronal mechanisms. Here, we test whether, and when, we can accurately decode the patterns of electroencephalogram (EEG) signals that are generated in response to CD- and IOVD-isolating stimuli. We found that the decoder could accurately distinguish between signals generated from CD and IOVD cues that moved 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. A cross-trained decoder showed that early decoding performance might depend upon fundamentally different low-level CD and IOVD stimulus features, while later performance might depend upon a mechanism that is agnostic to these low-level features. Overall, these data are consistent with the hypothesis that CD and IOVD are processed initially through distinct neuronal mechanisms but may drive a more central, amodal 'motion-in-depth' computation.

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[1–4]. 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[5–8]. However, psychophysical measurements of CD and IOVD suggest that these mechanisms have different speed sensitivities, adapt independently, and engage systems with different spatial resolutions[9,10]. Although these two mechanisms predominately activate the same neural substrates, it appears that they may be processed via two computationally distinct pathways.

Here, we ask whether there are differences in the overall pattern of neuronal signals generated in response to CD and IOVD cues. To test this, we used

electroencephalography (EEG) to measure neuronal signals from 64 channels across the scalp, generated in response to well-isolated CD and IOVD 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 stimulus 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, providing support for the hypothesis that CD and IOVD processes are carried by two distinct mechanisms. In a second experiment, we tested whether the decoder could accurately decode dynamically-updating CD stimuli that were located either near or far in depth, with no stereomotion information. This allowed us to validate the contribution of static CD depth information to decoding accuracy.

Overall, the decoder could accurately distinguish between EEG responses to CD and IOVD cues. As these two stimulus types differ at a fundamental level (the CD cue having shorter dot lifetime than the IOVD cue), this was, perhaps, unsurprising. However, we also found accurate *within-cue* discrimination of 3D-motion direction: patterns of neuronal response were significantly different for the towards- and away-conditions, suggesting that these drive subtly different cortical neuronal populations - or drive similar populations with differences in synchrony or coherence. Although participants were faster and more accurate in identifying the direction of CD cues, stimulus direction decoding accuracies for the two cue types reached significance at approximately the same time. We validate these findings in three additional analyses that examined: the contribution of static CD 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, 8 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 (8 males) completed a 1-hour EEG session. In a second 'static CD' control experiment, nine participants (mean age 27 years, 4 males, including 3 participants from the previous experiment) completed a 1-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[11,12]. 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[13].

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. For the towards motion, the ramp in disparity was defined by a positive slope in the right eye and a negative slope in the left eye, and the opposite was true for away motion. The near and far points were identical for each trial. The peak binocular disparity was ±32 arcmin: this is the upper limit for typical human observers: disparities greater than this do not permit haplopia to be achieved, being outside the traditionally-defined extent of Panum's fusional area[14].

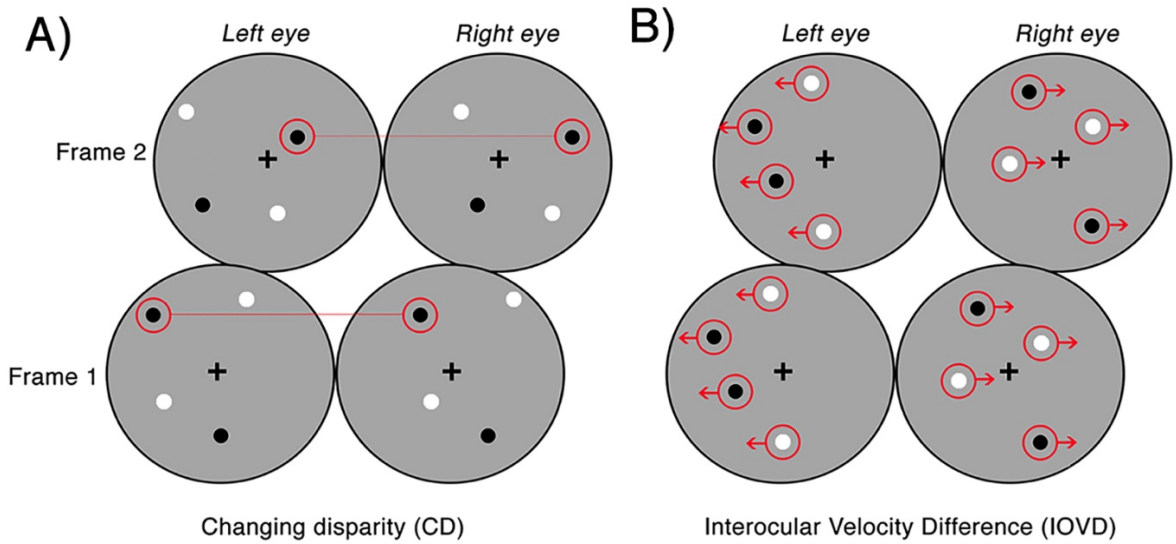

Figure 1 Examples of the CD and IOVD stimuli presented in two consecutive frames. Panel A) illustrates CD, where the dots are positioned randomly and paired between the left and right eyes, and only differ in their lateral shift in disparity across each video frame. Panel B) illustrates IOVD, where the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

Conversely, for the IOVD stimuli, the 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 essentially 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[15,16]. 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[12].

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 the supplementary materials of previous work[11].

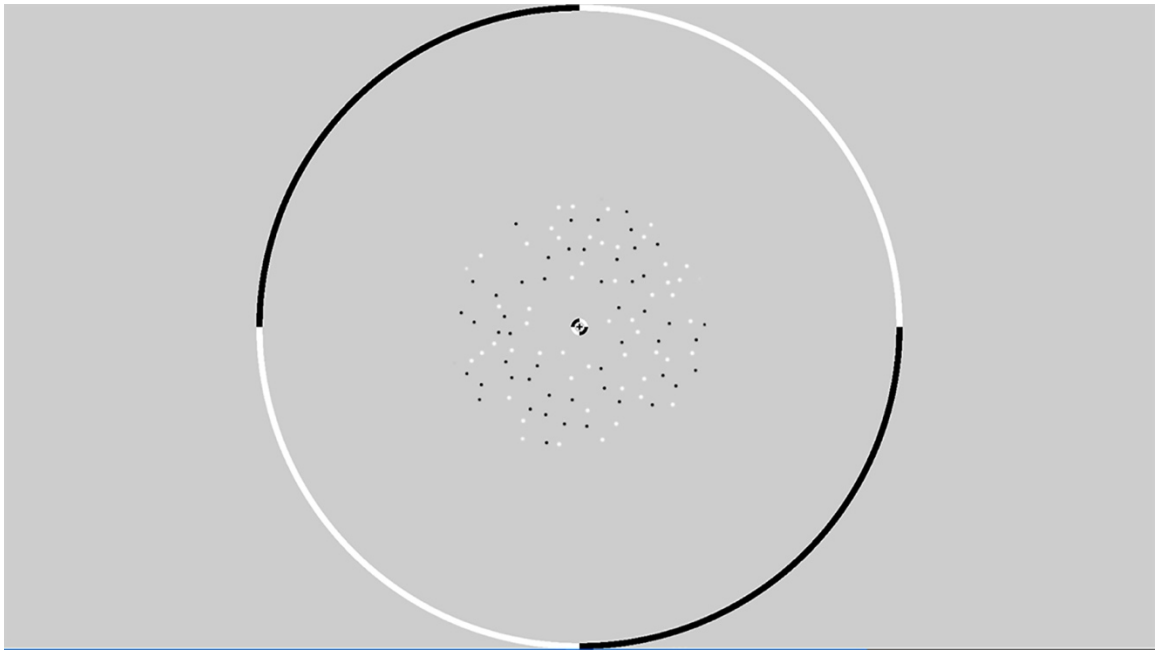

Figure 2 An example from a single frame of the random dot stereogram pattern that is presented within a circular fixation lock. The dots are presented in an annulus with an inner radius of 1° and an outer radius of 6°.

Static CD stimuli

Decoding between CD stimuli moving towards and away in depth might rely on static disparity information at each time point, rather than MID CD cues. To test this, we ran a second control experiment to examine whether the decoder could distinguish between static disparity cues located at either near or far positions in depth (i.e. the first frame of the CD towards cue or the first frame of the CD away cue). Static CD stimuli had an identical distribution of starting positions (i.e. either near or far), durations (250 ms), and dot update rates to the MID CD stimuli, however stereomotion velocity was set to zero. To maintain continuity and foster comparisons with the rest of this report, we use the term ‘static CD’ to describe the stimulus in this experiment, whole acknowledging that technically ‘static CD’ is a misnomer.

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[17].

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 stimuli oscillated towards and away continuously according to the linear ramp. This allowed them to practise identifying CD and IOVD cues and the motion directions. 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 CD 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 CD 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.

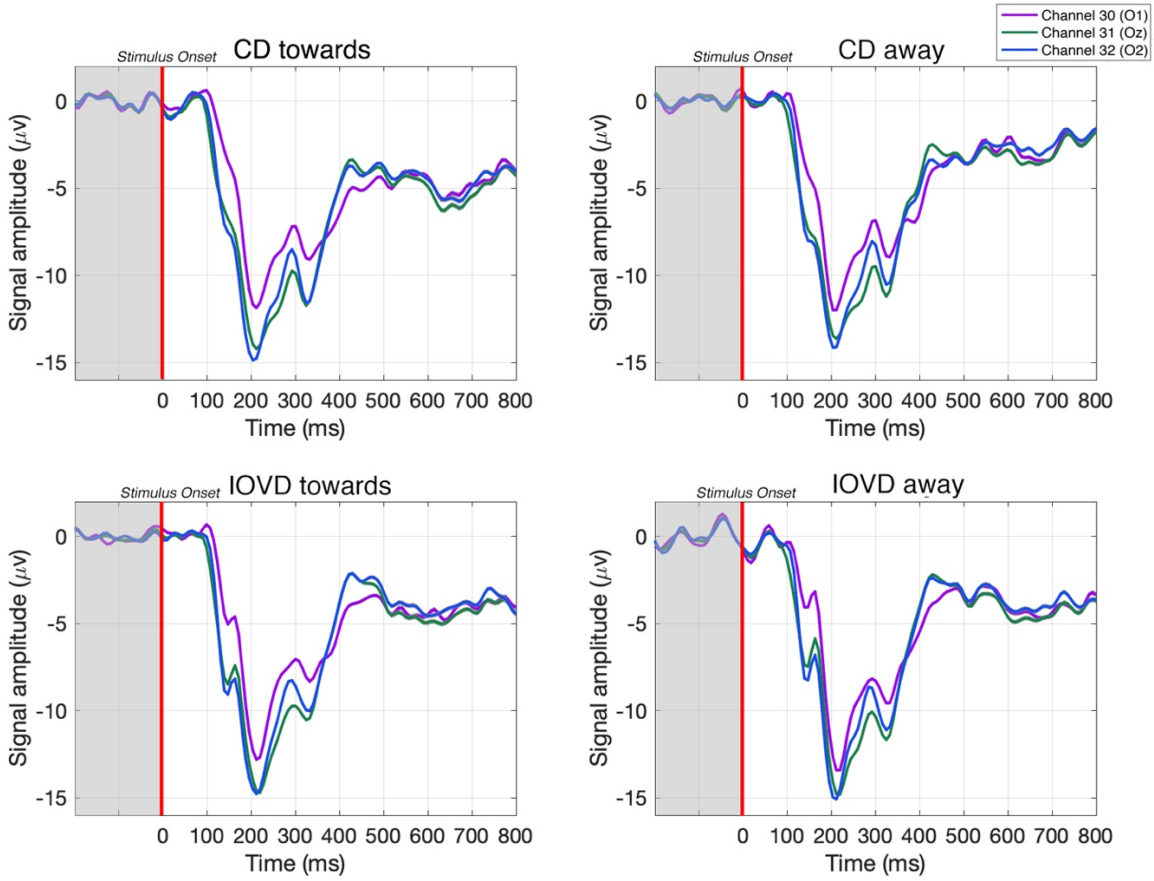

Figure 3 Examples of EEG waveforms from a single participant, taken from 3 occipital electrodes in response to each stimulus type over 210 trials. Each epoch is -200 to 800 ms and the stimulus is presented between 0 – 250 ms.

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[18–20]. As we decoded using all the electrodes across the scalp, we tested whether these motor responses (due to using one of two fingers to indicate if the stimulus was moving towards or away) had any influence on decoding performance. If our SVM is less accurate at decoding responses after they are pooled into keyboard-response conditions (with balanced stimulus types), rather than stimulus conditions, it suggests that the EEG signal that is elicit from making a keyboard response does not influence decoding accuracy, and we can conclude that any differences in the pattern of EEG signals are driven predominately by sensory MID signals.

We pooled our data into epochs that were associated with either a 'towards' or an 'away' keyboard press response. This resulted in eight unequal pools of data for each participant: four pools in which the keyboard direction response matched the stimulus, four pools in which it was erroneous. We then performed a decoding analysis on keyboard button press using bootstrapped datasets that contained 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 (i.e. we generated 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[21]. In the case of our EEG data, decoding could only occur at above-chance levels if different stimuli elicit 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[22–24]. The trial opposite to the omitted trial was withheld from the analysis. Thus, the SVM is trained to decode stimulus cue by discriminating between neuronal responses arising from a pair of stimulus conditions. The trained decoder then predicts the class of the omitted test data[25]. 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 cross-trained 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 statistically from the distribution of bootstrapped SVM classification accuracies) must have an overall pattern of neuronal 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[26]. 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 CD information, we ran an analysis that tested whether we could decode static CD 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.

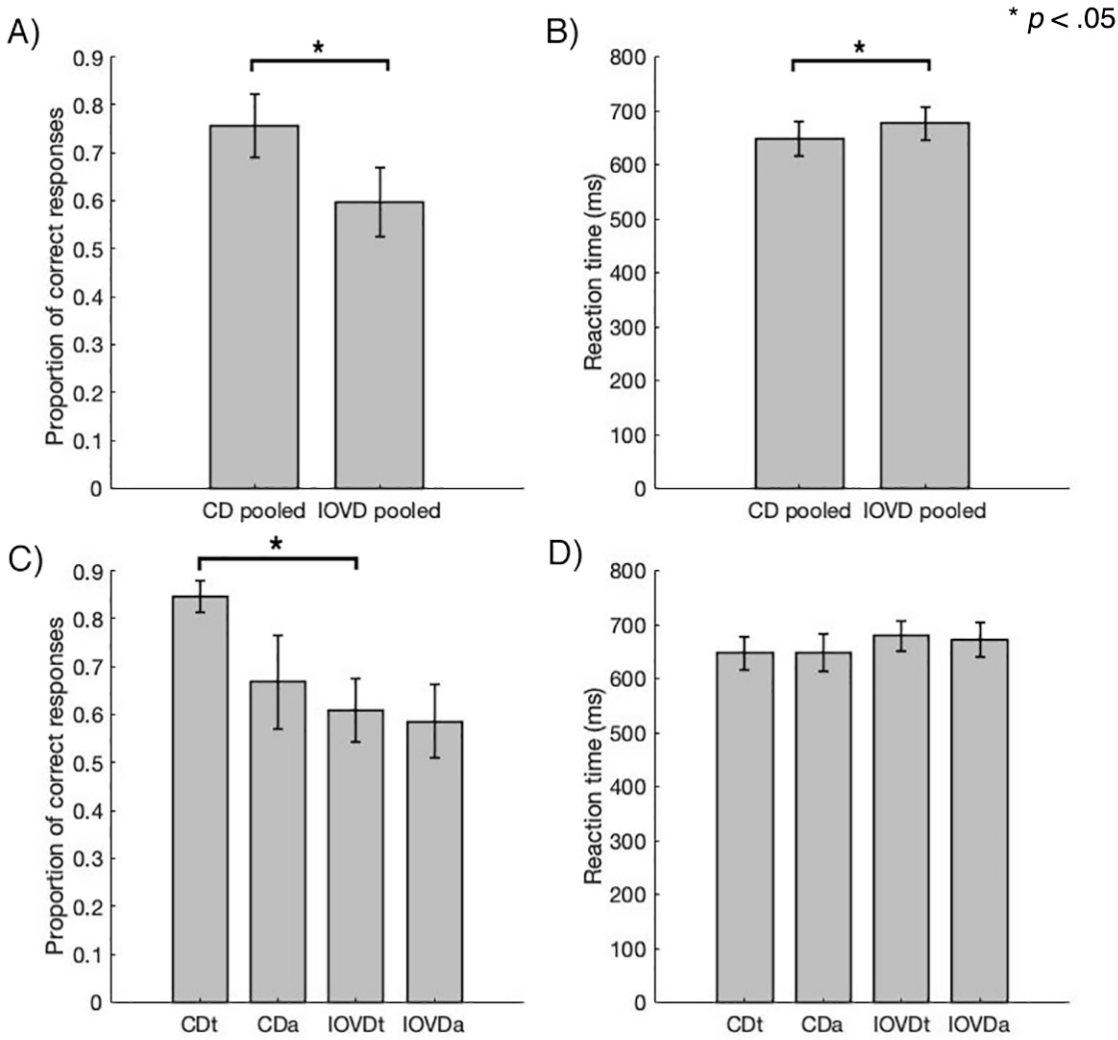

Figure 4 Bar plots of behavioural responses. In panel A) we present the mean proportion of correct responses after data are pooled into CD and IOVD conditions and in B) we present the mean reaction time after data are pooled into CD and IOVD conditions. In panel C) we present the mean proportion of correct 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 that a significant difference in reaction time when comparing between CD and IOVD stimuli, $t(19) = -3.141$, $p < .05$, with participants 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 neuronal response patterns

Although we took care to match the low-level properties of the CD and IOVD stimuli as far as possible, they still differed in some important 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 aspects of the low-level stimulus than an IOVD pathway, and these differences in low-level stimulus properties cannot be avoided.

As expected, EEG responses to the different types of MID cue (irrespective of motion direction) could be decoded at well above chance levels, suggesting that CD and IOVD cues may drive different neuronal populations, or alternatively, drive similar populations but elicit different neuronal responses (Figure 5, red line). Decoding occurred at chance until 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.

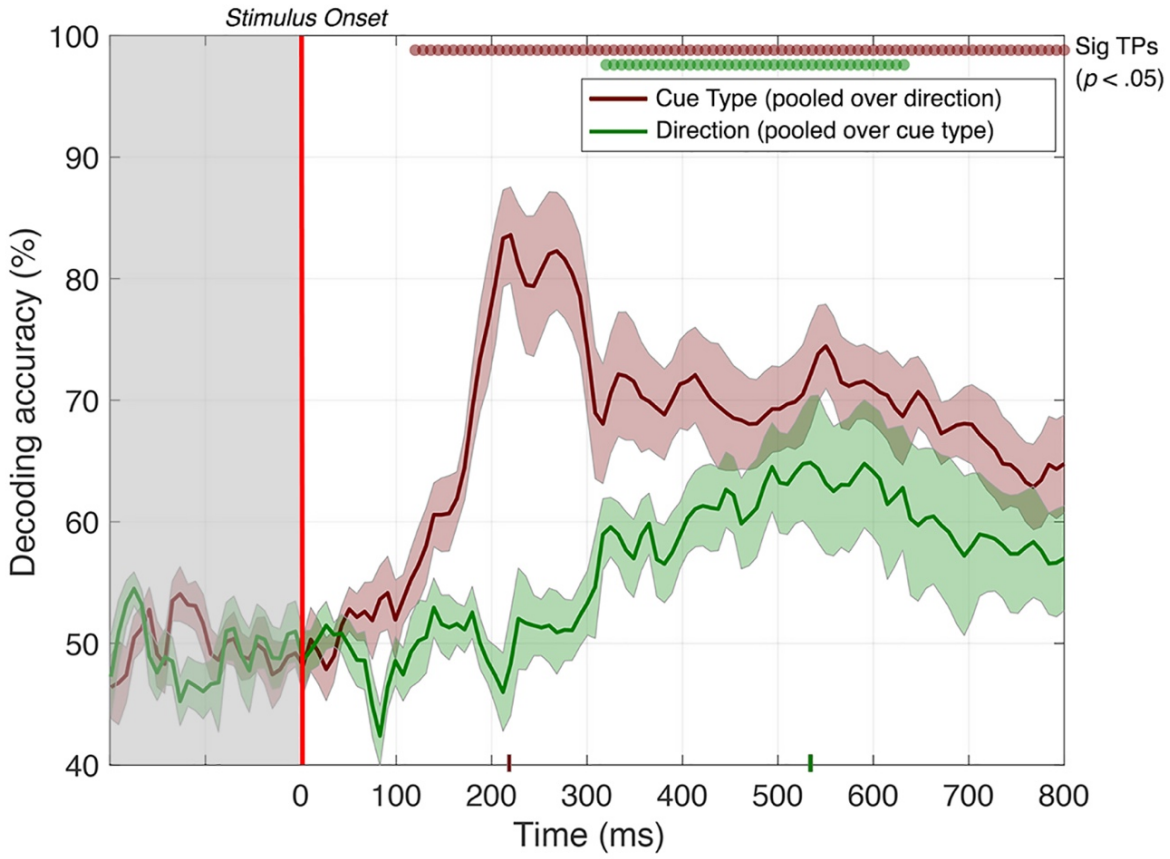

Figure 5 Decoding accuracy across 125 time points for cue type (pooled over direction) in red and direction (pooled over cue type) in green when all 64 electrodes are included as features. Cue type could be decoded from 120 ms - 800 ms after stimulus onset, while direction could be decoded from 320 – 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).

It is possible that the ability to decode between CD and IOVD cues is based on differential responses in visual cortex, and these responses might be inherited from retinal mechanisms. In Figure 6 we present a scalp montage of the support vector

(SV) weights at the time of peak decoding accuracy. These weights indicate the contribution of each electrode to decoding performance. The distribution of signal contributions appears to be complex and does not appear to be limited to occipital regions, however some interesting features are visible. In the CD vs IOVD (pooled direction) distribution (Figure 6A) there is a significant input from both visual cortex and a band of activity in more frontal regions, that we hypothesise may be driven by the frontal eye fields, perhaps due to vergence in the IOVD condition[27].

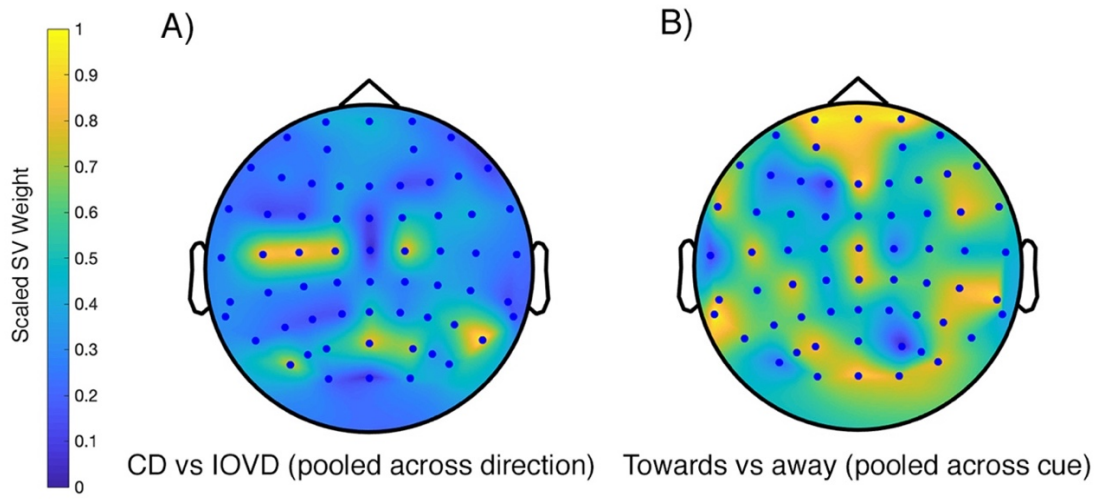

Figure 6 Standardised SV weights for CD vs IOVD (pooled across direction) and for towards vs away (pooled across MID cue) at their respective peak time point. In panel A) we present the SV weights at 224 ms for CD vs IOVD (pooled across direction) decoding, and in B) we present the SV weights at 536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

It is perhaps unsurprising that we are able to identify differences in the EEG response pattern generated by these network activations. 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, although the accuracy was lower than for cue type, peaking at 65% at 536 ms ($p < .05$). The 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, indicates that different neuronal populations might be recruited by CD and IOVD cues and by different motion directions. Alternatively, these cues could activate similar neuronal populations, however the generated responses might 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 similar time point, peaks early, and stays high, regardless of direction – an effect that could be driven by different neuronal responses to low-level monocular cues which we address in the cross-trained decoding analysis (see below).

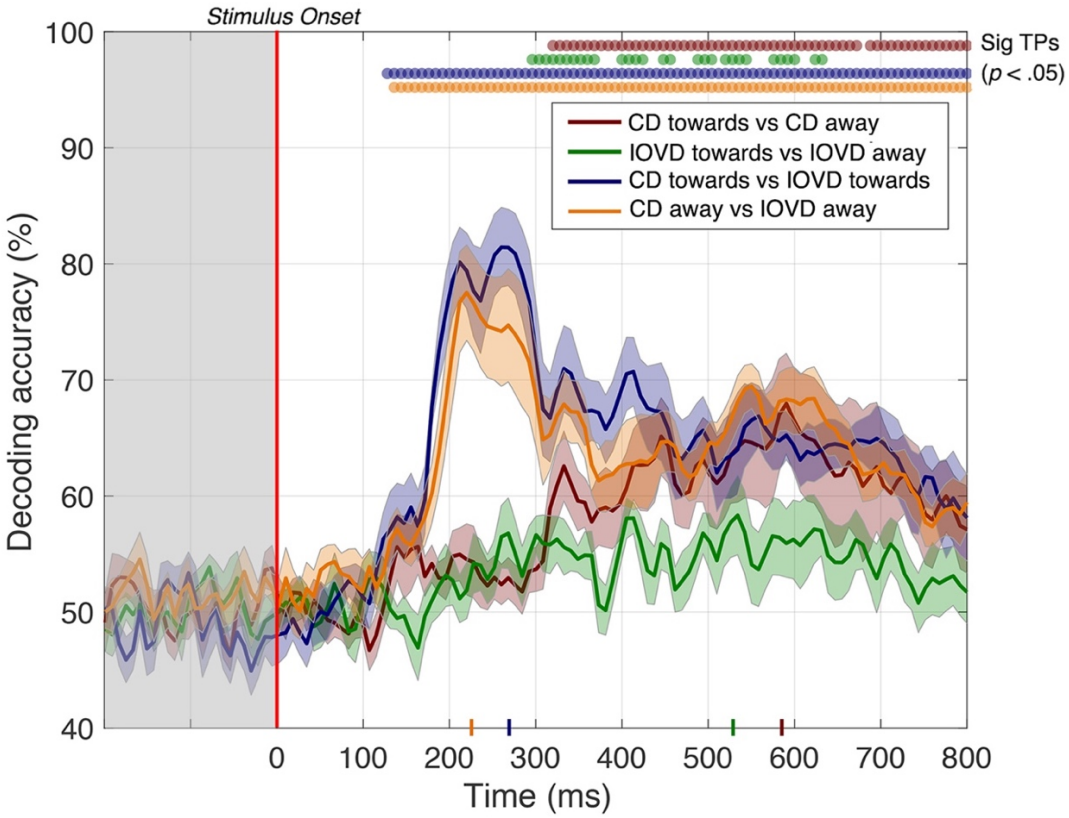

Figure 7 Pairwise decoding accuracy across 125 time points for four stimulus conditions. CD vs IOVD cue type (within motion direction) can first be decoded from 128 ms (towards) and 136 ms (away) after stimulus onset, while stimulus direction (within cue type) can first be decoded at 320 ms (CD) and 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. As presented in Figure 7, we could accurately decode motion direction *within* CD and IOVD cue conditions. CD towards vs CD 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 the electrode contributions for each of the four conditions, at their respective peak time of decoding, are illustrated 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 with an additional band of activity in more frontal regions that we suspect may be driven by the frontal eye fields[27]. Signal contributions for the within direction comparison (Figure 8A and B) are complex (again, perhaps due to the later time point of the peak), 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

that from a previous experiment that found occipital and superiorly located responses at a similar time point for CD motion stimuli[18].

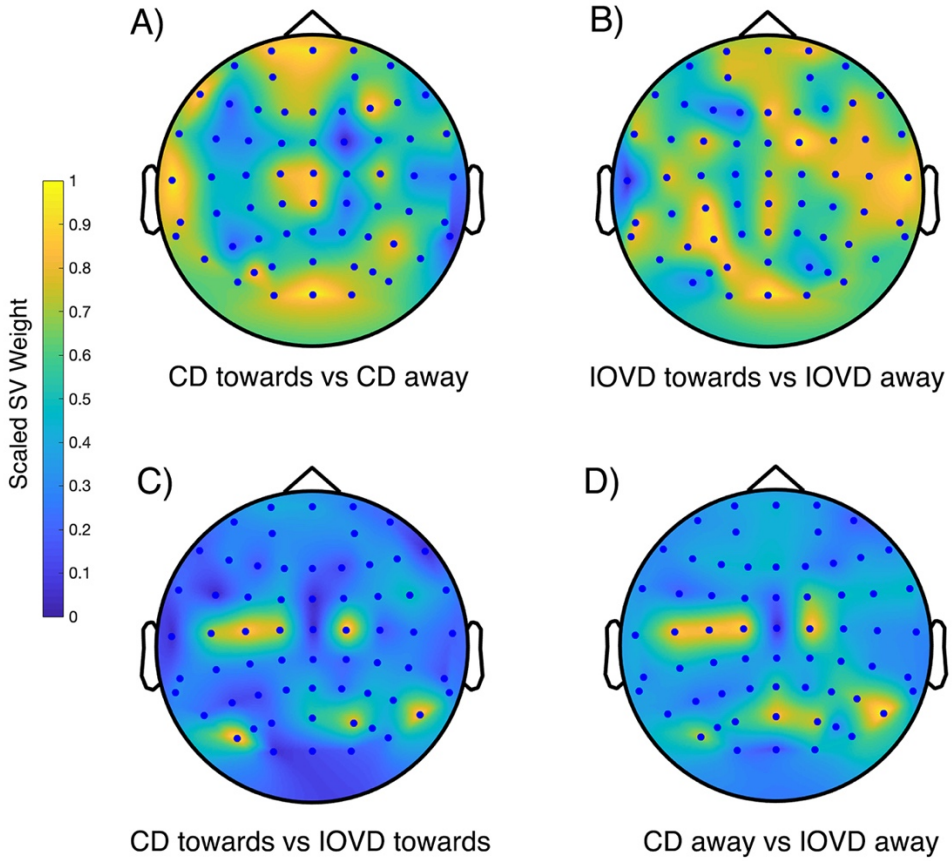

Figure 8 Distribution of standardised SV weights at their peak time point of decoding plotted onto a scalp montage for the four comparisons. In panel A) we present the SV weights for decoding CD towards vs CD away responses (592 ms), in panel B) for IOVD towards vs IOVD away responses (536 ms), in panel 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. Although different directions are very similar within both CD and IOVD cues at the retinal level, the ability to decode depth direction from EEG data suggests that these

responses could potentially come from an intermingled network of depth-direction sensitive neurons, or it may be that neurons elicit unique signals in response to towards and away stimuli that are different enough to be resolved using EEG.

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

One 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 CD towards vs CD away relies on static disparity information, rather than MID information. To address this, we ran a control experiment that tested whether we could decode static CD stimuli. 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 suggests that decoder performance is relying 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 CD stimuli can be decoded, this static information must have some contribution to MID CD decoding performance.

Our data (Figure 9) show that there was a short set of time points where the decoder could distinguish between near and far static CD stimuli. Irrespective of keyboard responses, decoding was possible between 280 – 366 ms (Figure 9, orange line), and 288 – 312 ms. 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).

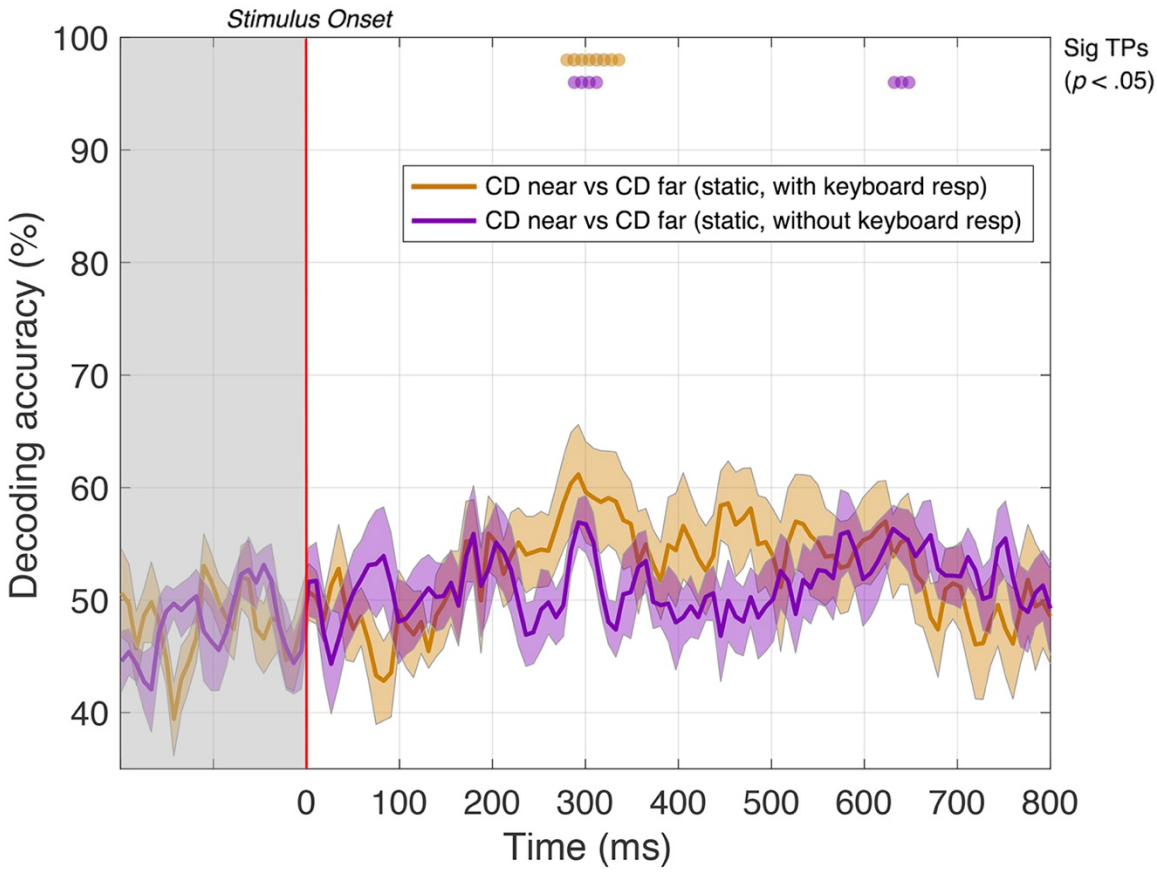

Figure 9 Pairwise decoding accuracy across 125 time points for static CD stimuli that are presented either near or far in depth. The orange line represents trials when participants were required to respond, via keyboard, whether they perceived the stimulus as either near or far in depth. The purple line 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 CD stimuli located near or far in depth for a short 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 motion information that was not present in these static CD 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 response, 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. These results suggest that stimulus decoding performance is likely to be driven by stimulus activity, rather than differences in signal from the motor cortex in response to keyboard button press.

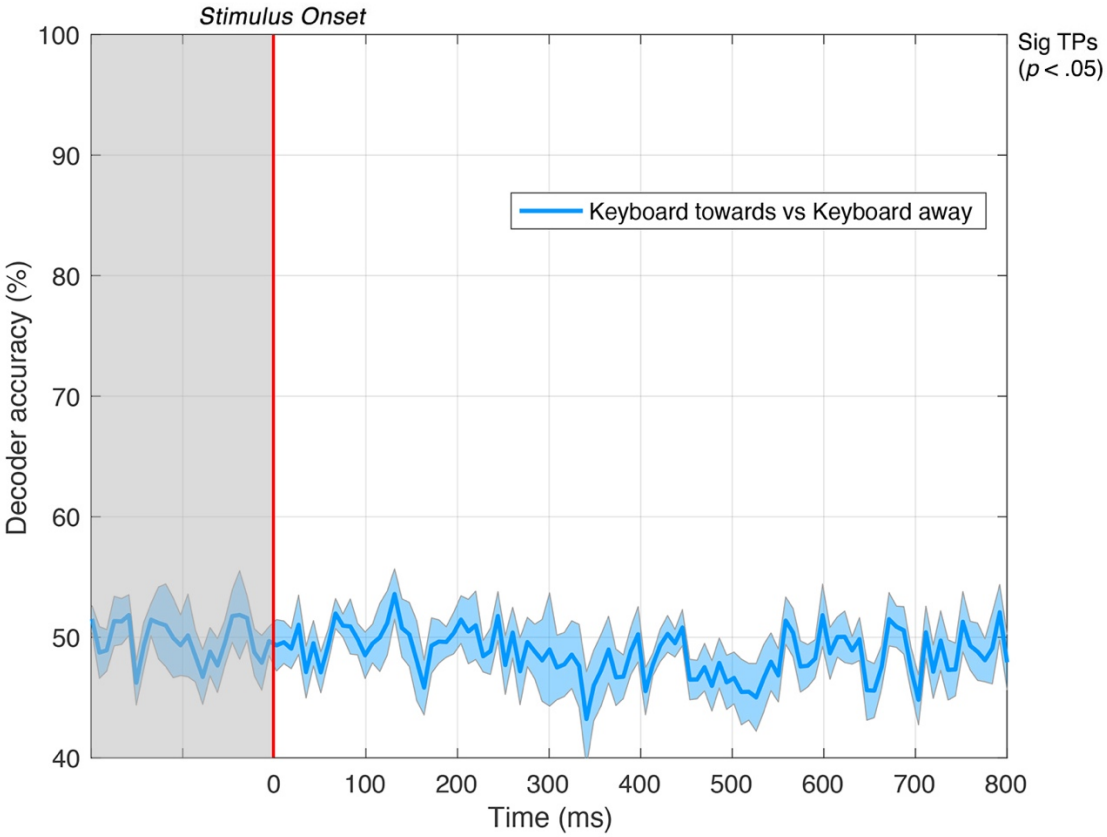

Figure 10 Pairwise decoding accuracy across 125 time points for keyboard response. Keyboard response (either pressing 2 on the keypad to indicate towards or 8 on the keypad to indicate away) cannot be accurately decoded: accuracy is at chance across all time points. Shaded error bars represent ±1SE of the bootstrapped mean (1000 iterations).

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

Next, we examined whether our decoder performance relied on unique CD and IOVD signals that could, potentially, originate from distinct neuronal populations. Here, we trained the decoder using CD towards vs CD away responses, and then tested how well it could decode IOVD towards vs IOVD away responses. If these signals are

truly distinct from each other, one might expect accuracy to fall around 50% chance as a model 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 at any time points, this might suggest the contribution of a mechanism that is agnostic to CD and IOVD cues, or some cross over between CD and IOVD pathways.

Remarkably, we found that we could decode motion direction from the cross-trained classifier for significant periods after ~500 ms. Specifically, IOVD direction could first be decoded at 512 ms after stimulus onset (56% accuracy) and was intermittently significant through to 800 ms (see Figure 11, purple line, $p < .05$). A similar result was found for CD direction - 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. One possibility is that the earlier classification performance we see in our original CD vs IOVD decoding analysis is due to low-level stimulus differences in the CD and IOVD cues, while later performance might depend upon a mechanism that is agnostic to these low-level features.

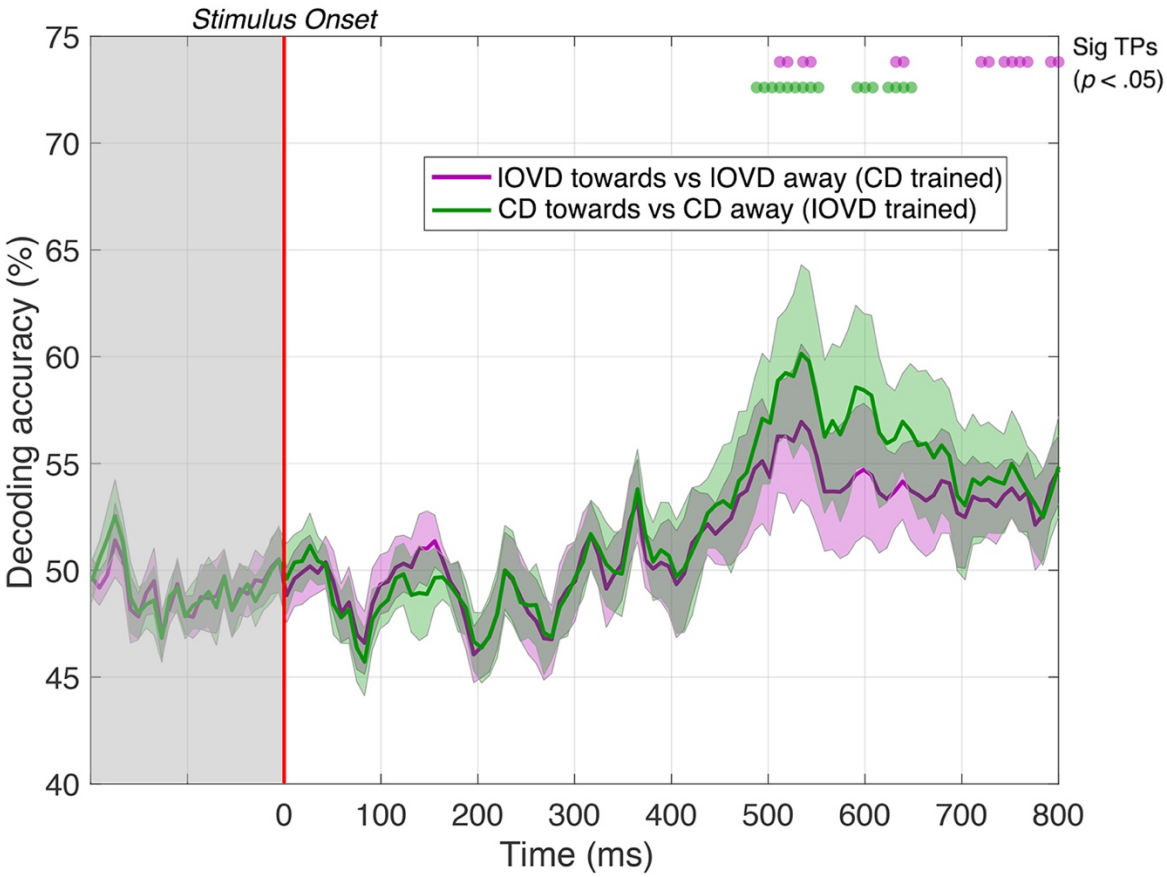

Figure 11 Pairwise decoding accuracy of a cross-trained decoder across 125 time points, comparing between CD and IOVD directions. The purple line represents decoding accuracy for IOVD towards vs IOVD away after decoder is trained using CD direction data. The green line represents decoding accuracy for 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 neuronal responses measured at the scalp. These signals can be decoded well in advance of behavioural reaction times for identifying MID cues, suggesting they reflect processes supporting MID perception. 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 stimulus-driven rather than driven by motor response. 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 must drive independent retinal computations and engage systems with different spatial and temporal resolutions[15,28,29]. Accurate decoding between CD and IOVD cues suggests that the neuronal 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 moving in the same direction in depth, indicating that the high performance of the decoder was indeed driven by differences in CD and IOVD cues, rather than motion direction. These findings are consistent with the hypothesis that CD and IOVD are processed through

two computationally distinct mechanisms and may be further processed by different neuronal pathways that could drive independent large-scale networks in the visual cortex[2,6]. Decoding accuracy was high, sitting around 84% at its peak, but this is unsurprising. 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.

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[30–32]. 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[6,9]. At the retinal level, the stimuli for the two motion directions are very similar. Decoding 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 different neuronal populations are intermingled within individual visual areas and the ability to resolve these responses using EEG is intriguing. A common suggestion 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 cues drive similar neuronal 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 3D-motion direction sensitive 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 CD cues

To examine the contribution of static disparity information to MID CD decoding performance, we ran a second experiment that used static CD stimuli with an identical distribution of starting positions, durations, and dot update rates to our original experiment. However, stereomotion velocity was set to zero – effectively holding our stimuli at one of two depth positions; near or far. It was possible to decode static CD information from the EEG signal at a few time points around 300 ms. Here, 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 response. One possibility is that transient static depth information contributes to the early phase of MID CD decoding (i.e. the time point where decoding was possible for both

experiments, around 300 ms) however later stages of decoding (beyond 300 ms) appear to be due solely to the smooth motion information that was confined to our MID stimuli. It is important to note that the mean disparity across the MID CD stimuli was lower than that used in the static CD stimuli, thus, the contribution of absolute depth to our MID CD decoding performance must therefore be *less* than that observed in the static CD decoding performance.

Cross trained decoder

Surprisingly, 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. This finding has two implications. First, it strongly suggests that the early classification accuracies we see in our original CD vs IOVD analysis are due to low-level stimulus differences between CD and IOVD stimuli. Second, the cross-trained decoder's ability to decode at later stages of the EEG time course suggests that these signals may originate from some mechanism that is agnostic to these low-level features. One possibility could be that these later responses originate from a cortical region that is involved in general stereomotion processing, rather than processing specific CD or IOVD mechanisms[11].

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 elicit 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[33–36]. 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 press. This has further implications, as we are effectively decoding the *percept* of 3D motion. Thus, for MID cues, it is not possible to decode the percept of motion direction. Instead, signals attributed to directional MID stimuli are driven by the input stimulus itself rather than the observers' percept.

Conclusion

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. Our findings are consistent 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 sensitive 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 performance we see in MID CD decoding. Finally, we also found that early stages of the EEG signal may rely upon low-level cue properties, while later stages of the signal may be driven by a more central, amodal motion-in-depth computation.

References

1. Cumming, B. G. & Parker, A. J. Binocular mechanisms for detecting motion-in-depth. *Vision research* **34**, 483–495 (1994).
2. Harris, J. M., Nefs, H. T. & Grafton, C. E. Binocular vision and motion-in-depth. *Spatial vision* **21**, 531–547 (2008).
3. Rashbass, C. & Westheimer, G. Independence of conjugate and disjunctive eye movements. *J Physiol* **159**, 361–364 (1961).
4. Regan, D. Binocular correlates of the direction of motion in depth. *Vision Res* **33**, 2359–2360 (1993).
5. Likova, L. T. & Tyler, C. W. Stereomotion processing in the human occipital cortex. *Neuroimage* **38**, 293–305 (2007).
6. Rokers, B., Cormack, L. K. & Huk, A. C. Disparity-and velocity-based signals for three-dimensional motion perception in human MT+. *Nature neuroscience* **12**, 1050–1055 (2009).
7. Czuba, T. B., Huk, A. C., Cormack, L. K. & Kohn, A. Area MT encodes three-dimensional motion. *J Neurosci* **34**, 15522–33 (2014).
8. Sanada, T. M. & DeAngelis, G. C. Neural representation of motion-in-depth in area MT. *J Neurosci* **34**, 15508–21 (2014).
9. Joo, S. J., Czuba, T. B., Cormack, L. K. & Huk, A. C. Separate perceptual and neural processing of velocity- and disparity-based 3D motion signals. *J Neurosci* **36**, 10791–10802 (2016).
10. Giesel, M., Wade, A. R., Bloj, M. & Harris, J. M. Investigating Human Visual Sensitivity to Binocular Motion-in-Depth for Anti- and De-Correlated Random-Dot Stimuli. *Vision* **2**, 41 (2018).

11. Kaestner, M. *et al.* Asymmetries between achromatic and chromatic extraction of 3D motion signals. *Proc Natl Acad Sci USA* **116**, 13631–13640 (2019).
12. Maloney, R. T. *et al.* Sensitivity to velocity- and disparity-based cues to motion-In-depth with and without spared stereopsis in binocular visual impairment. *Invest. Ophthalmol. Vis. Sci.* **59**, 4375–4383 (2018).
13. Brainard, D. H. The Psychophysics Toolbox. *Spatial Vision* **10**, 433–436 (1997).
14. Howard, I. P. & Rogers, B. J. *Seeing in depth*. **2**, (I. Porteous, 2002).
15. Shioiri, S., Nakajima, T., Kakehi, D. & Yaguchi, H. 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**, 1574–85 (2008).
16. Shioiri, S., Saisho, H. & Yaguchi, H. Motion in depth based on inter-ocular velocity differences. *Vision Res* **40**, 2565–72 (2000).
17. Baker, D. H., Kaestner, M. & Gouws, A. D. Measurement of crosstalk in stereoscopic display systems used for vision research. *J Vis* **16**, 14 (2016).
18. Cottereau, B. R., Ales, J. M. & Norcia, A. M. The evolution of a disparity decision in human visual cortex. *NeuroImage* **92**, 193–206 (2014).
19. Cottereau, B. R., McKee, S. P., Ales, J. M. & Norcia, A. M. Disparity-tuned population responses from human visual cortex. *J. Neurosci.* **31**, 954–965 (2011).
20. Cottereau, B. R., McKee, S. P. & Norcia, A. M. Dynamics and cortical distribution of neural responses to 2D and 3D motion in human. *J. Neurophysiol.* (2013). doi:10.1152/jn.00549.2013
21. Chang, C. & Lin, C. LIBSVM -- A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology* **2**, 1–27 (2011).

22. Boser, B., Guyon, I. & Vapnik, V. A Training Algorithm for Optimal Margin Classifiers. in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory* 144–152 (ACM Press, 1992). doi:10.1145/130385.130401
23. Cortes, C. & Vapnik, V. Support-Vector Networks. *Machine Learning* **20**, 273–297 (1995).
24. Wang, L. *Support Vector Machines: Theory and Applications (Studies in Fuzziness and Soft Computing)*. (Springer-Verlag, 2018).
25. Grootswagers, T., Wardle, S. G. & Carlson, T. A. Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience* **29**, 677–697 (2017).
26. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* **164**, 177–190 (2007).
27. Heinen, S. J., Rowland, J., Lee, B.-T. & Wade, A. R. An Oculomotor Decision Process Revealed by Functional Magnetic Resonance Imaging. *Journal of Neuroscience* **26**, 13515–13522 (2006).
28. Czuba, T. B., Rokers, B., Huk, A. C. & Cormack, L. K. Speed and eccentricity tuning reveal a central role for the velocity-based cue to 3D visual motion. *J Neurophysiol* **104**, 2886–99 (2010).
29. Wardle, S. G. & Alais, D. Evidence for speed sensitivity to motion in depth from binocular cues. *J Vis* **13**, (2013).
30. Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology* **148**, 574–591 (1959).

31. Hubel, D. H. & Wiesel, T. N. Receptive fields and the functional architecture of monkey striate cortex. *Journal of Physiology* **195**, 215–243 (1968).

32. Born, R. T. & Bradley, D. C. Structure and function of visual area MT. *Annual Review of Neuroscience* **28**, 157–189 (2005).

33. Bertamini, M., Rampone, G., Oulton, J., Tatlidil, S. & Makin, A. D. J. Sustained response to symmetry in extrastriate areas after stimulus offset: An EEG study. *Sci Rep* **9**, 4401 (2019).

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

35. Mullinger, K. J., Mayhew, S. D., Bagshaw, A. P., Bowtell, R. & Francis, S. T. Poststimulus undershoots in cerebral blood flow and BOLD fMRI responses are modulated by poststimulus neuronal activity. *Proceedings of the National Academy of Sciences* **110**, 13636–13641 (2013).

36. Stevenson, C. M., Brookes, M. J. & Morris, P. G. β-Band correlates of the fMRI BOLD response. *Hum. Brain Mapp.* **32**, 182–197 (2011).

Data Availability Statement

All data analysed during the current study may be requested from the authors and will be made available on OSF prior to publication.

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.

Author Contributions

A.R.W and R.T.M designed the experiment, F.G.S and M.M.H collected the data, M.M.H and A.R.W. analysed the data, M.M.H wrote the manuscript, and M.M.H, R.T.M, and A.R.W revised the manuscript. M.M.H and F.G.S contributed equally to this paper.

Competing interests

The authors declare no competing interests.