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

Authors and Affiliations: 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

* Authors contributed equally

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 CD and IOVD are processed by distinct neuronal mechanisms. Here, we decode brain patterns of electroencephalogram (EEG) signals generated in response to CD- and IOVD-isolating random dot stereogram stimuli. We ask *whether*, and *when*, we can accurately decode CD and IOVD cues from the pattern of EEG signals that they generate. We then ask how the time course of these signals relates to their perception. Generally, we found that we could decode signals generated from different stimulus cues (CD or IOVD) moving in different 3D-depth directions (towards or away). These signals could be decoded well before participants behaviourally identified the cues. Although participants were faster and more accurate in identifying CD cues, the decoder could distinguish IOVD signals earlier, and decoding accuracy peaked earlier, when compared to CD cues. Overall, these data are consistent with the hypothesis that CD and IOVD are processed through distinct neuronal mechanisms and suggest that the IOVD mechanism may have a faster temporal tuning.

Introduction

To navigate the visual environment successfully, the human 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 neuroimaging (fMRI) studies have similarly identified that such cues reliably activate human homologue MT+ and nearby regions (Likova & Tyler, 2007; Rokers, Cormack, & Huk, 2009). However, psychophysical measurements of CD and IOVD suggest that these mechanisms have different speed sensitivities, adapt independently, and engage systems with different spatial resolutions, with IOVD having a relatively fast temporal tuning when compared to CD (Czuba, Rokers, Huk, & Cormack, 2010; Joo, Czuba, Cormack, & Huk, 2016; Shioiri, Nakajima, Kakehi, & Yaguchi, 2008; Wardle & Alais, 2013). 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. We then ask how the timings of these signals relate to each other and to the perception of MID. 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*, and *when*, a multivariate pattern classifier could use the overall pattern of EEG signals to accurately decode stimulus cue type (CD or IOVD) and depth (towards or away). The ability to decode these cues would suggests that the evoked EEG responses are relatively independent from each other, providing support for the hypothesis that CD and IOVD processes are carried by two distinct mechanisms. Further, the high temporal resolution of EEG allows for the assessment of the temporal dynamics of these two potentially independent mechanisms at a millisecond scale.

Overall, the SVM could decode responses to CD and IOVD cues relatively accurately. 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.

Finally, although participants were faster and more accurate to identifying CD cues, IOVD cues could be accurately decoded, and peaked, relatively earlier than CD cues.

Methods

Participants

Twelve healthy subjects (mean age 24 years, 8 males) were recruited from the University of York. All participants had normal or corrected-to-normal vision. 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) were tested, and their data were included in the analysis. Each participant completed a 1-hour EEG session. All participants 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

The stimuli were very similar to those described 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 (σ = .005°). The dots were positioned at a density of 1 dot per 1° of visual angle on a mean grey background and had a Michaelson 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 monitor refresh (16.6 ms). Thus, CD stimuli consisted of temporally uncorrelated dynamic random dot stereograms, 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 Fig. 1A). CD stimuli changed unidirectionally in disparity (i.e. the stimulus moved either towards or away in depth). The shift in disparity began at the far point and finished at the near point for towards motion, and the opposite points for away motion. The maximum disparity was 16 arcmin (monocular) or 32 arcmin (binocular): this is the maximum limit for typical observers before image function fails and results in diplopia (Howard & Rogers, 2002).

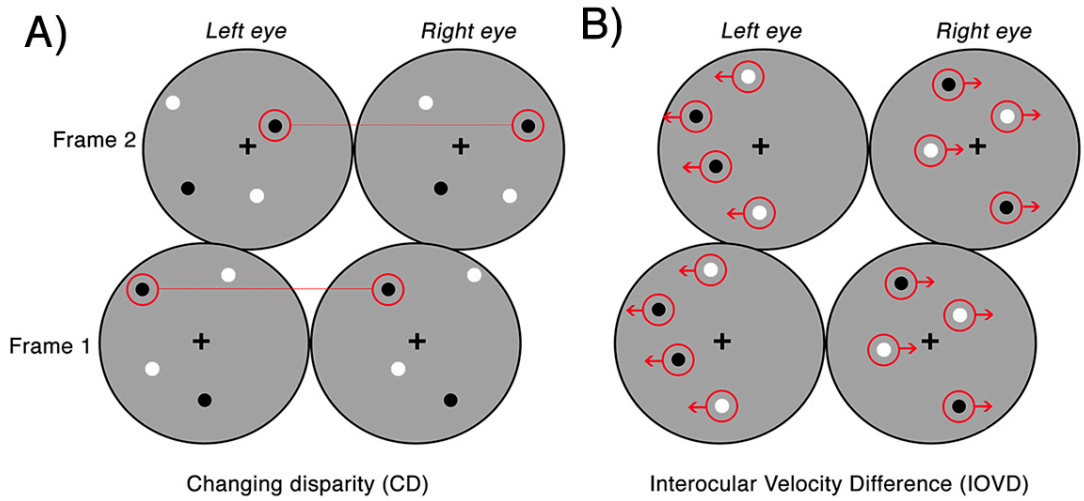


Figure 1 Examples of the CD and IOVD stimuli in two consecutive frames. For CD, the dots are positioned randomly and paired between the left and right eyes, and only differ in their lateral shift across each frame. For IOVD, the dots laterally shift through time and there is no pairing of dots between 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 Fig. 1B). 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 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 dots wide. The strips alternated across the eyes to ensure that they did not coincide on the two retinae. 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°. The edge of the annulus was smoothed with a cosine ramp 0.5° in size. 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 four 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.

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 (Lameris Ootech, NL). During the EEG cap set-up, participants were presented with a demonstration of the each of the four stimulus conditions on the VIEWpixx display. 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 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 210 stimulus trials. Within each block the four stimulus conditions (CD towards, CD away, IOVD towards, and IOVD away) were randomized 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 (thus, there was one cycle of the (4 Hz) sawtooth wave modulation on each trial) and was followed by an inter-trial interval. The length of the 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 across. 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.

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 1KHz 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 overall noise levels in the eye channels. These signals were amplified 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.

EEG pre-processing and bootstrapping

For each participant, EEG data were binned into 1000 ms epochs at -200 ms to 800 ms, relative to stimulus onset at 0 ms (see Fig. 2 for examples of EEG epochs). 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 Hz so that decoding would occur every 8 ms. Thus, we retained a sampling pool of 1260 EEG epochs (210 trials x 6 blocks) that included each of the four stimulus conditions (315 samples of each condition). From this sampling pool, we bootstrapped the data of 10 epochs from each condition. These 10 epochs were averaged together to create a mean

bootstrapped epoch for each condition. This process was repeated 21 times so that there were 21 mean bootstrapped EEG epochs, for each of the 64 electrodes, for each condition, and each participant. Data were then z-scored to standardize electrodes. This process was repeated for 1000 iterations of a bootstrapped vector machine (SVM) classifier, with each iteration using a new subset of sample epochs from the sampling pool to generate 21 unique mean bootstrapped epochs.

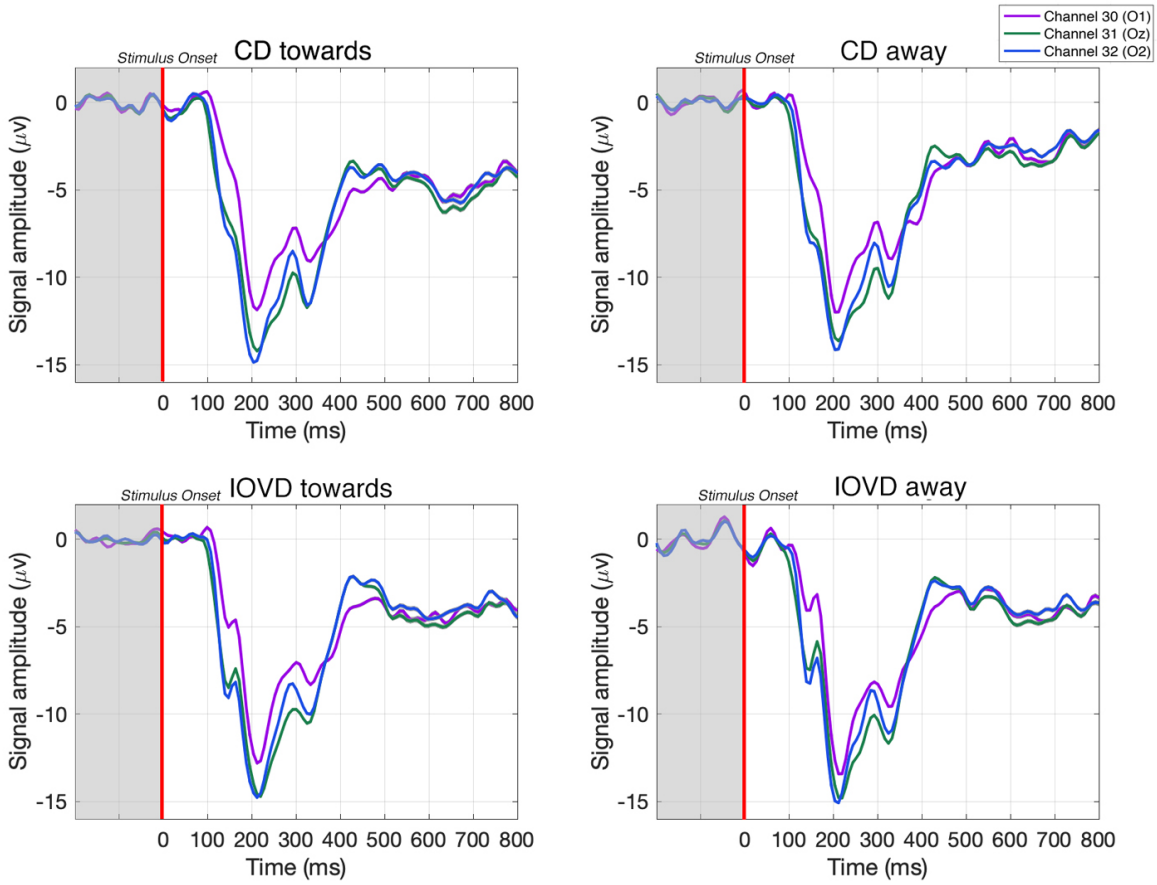


Figure 2 Examples of EEG waveforms from a single participant, taken from 3 occipital electrodes in response to each stimulus type. Each epoch is binned between -200 to 800 ms and the stimulus is presented between 0 – 250 ms.

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 before 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 the *all* 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 responses are pooled into keyboard-response conditions (with balanced stimulus types), rather than stimulus conditions, and cannot be accurately decoded, it would indicate that keyboard response does not influence decoding accuracy, and we can conclude that any differences in the pattern of EEG response are predominately driven by bottom-up MID cues.

We pooled our data into epochs that were associated with either a 'towards' or an 'away' button press response. This resulted in eight unequal pools of data, for each participant: four in which the direction response matched the stimulus, four in which it was erroneous. We then performed a classification analysis on button press using datasets with that had equal numbers of 'towards' and 'away' stimuli. Nine participants were included in this analysis, as one participant had 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, 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 proceeding 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 SVM 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 radial basis kernel (RBK) function 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 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) of the EEG epoch, using data from all 64 electrodes. EEG data were repeatedly bootstrapped (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 data samples and left out sample is the test data (i.e. the SVM is trained on the 40 epochs, and 1 epoch is kept as the test data) (Boser, Guyon, & Vapnik, 1992; Cortes & Vapnik, 1995; Wang, 2018). 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 (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 Fig. 1 and Supplementary Fig. 2 for decoding accuracies after shuffling labels).

In principle, time points at which the decoder performed at above-chance (as assessed statistically from the distribution of bootstrapped SVM classification accuracy) 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 (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 $p < .01$.

Decoding was first performed using six binary comparisons. First, we tested if we could decode CD and IOVD cue responses, pooled across direction. Next, we tested if could decode towards and away directions, pooled across MID cue type. We then asked if we could decode stimulus direction *within* MID cue type, and between MID cue type *within* stimulus direction; therefore, comparing between CD towards and CD away cues, IOVD towards and IOVD away cues, CD towards and IOVD towards cues, and CD away and IOVD away cues. In a supplementary analysis, we used a synthetic dataset to test if we could decode keyboard response.

Results

Behavioural responses

The overall percentage of correct responses and the mean reaction time to each of the four stimulus conditions are presented in Table 1. Participants were generally better at identifying CD cues (75.6% accuracy) when compared to IOVD cues (63.3% accuracy). Although there was little difference in proportion of correct responses towards and away conditions within IOVD cues (61.0% and 65.7%, receptively), participants were more accurate at identifying CD towards (84.5%) than CD away (66.8%) cues. In regard to reaction times, participants were faster to identify CD cues (648.2 ms) when compared to IOVD cues (675.2 ms), however there was little difference in reaction time when comparing between direction within CD and IOVD cue type.

Table 1 Behavioural responses for each participant. The proportion of correct responses and mean reaction time to each stimulus condition. For each participant, behavioural data were averaged across 315 trials per condition and 0-time was defined as the onset of the stimulus.

	Proportion of Responses correct				**Reaction time (ms)**			
	CD Towards	**CD Away**	**IOVD Towards**	**IOVD Away**	**CD Towards**	**CD Away**	**IOVD Towards**	**IOVD Away**
AW	0.918	0.963	0.703	.0791	614	582	636	613
BP	0.921	0.936	0.936	0.610	610	579	611	627
DL	0.893	0.145	0.269	0.755	697	663	741	786
FS	0.765	0.458	0.519	0.571	856	891	848	851
GM	0.917	0.665	0.732	0.297	580	683	623	646
IJ	0.698	0.342	0.511	0.527	510	533	540	538
IS	0.962	0.947	0.761	0.882	568	540	614	548
JS	0.642	0.413	0.380	0.627	676	682	693	701
MW	0.856	0.817	0.795	0.727	725	737	742	752
RM	0.883	0.996	0.483	0.785	642	596	744	668
	Averages							
	0.845	0.668	0.610	0.657	647.8	648.6	679.2	671.2
	0.756		0.633		648.2		675.2	
	Standard Deviation							
	0.242		0.186		97.23	108.7	90.65	98.89

EEG Results

Decoding between cue type and direction: CD and IOVD stimuli drive different neuronal populations

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. Most significantly, CD stimuli (deliberately) contained no coherent 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.

As expected, we were able to decode EEG responses to the different types of MID cue (irrespective of motion direction) at well above chance levels, suggesting that the stimuli drive different neuronal populations (Fig. 3, red line). Decoding occurred at chance until 144 ms after stimulus onset and continued at above chance levels across the remaining time course (Bonferroni corrected, $p < .01$). The accuracy of the decoder peaked at 78%, 272 ms after onset. The decoding performance for IOVD vs CD at 144 ms is the earliest difference that we see in this study.

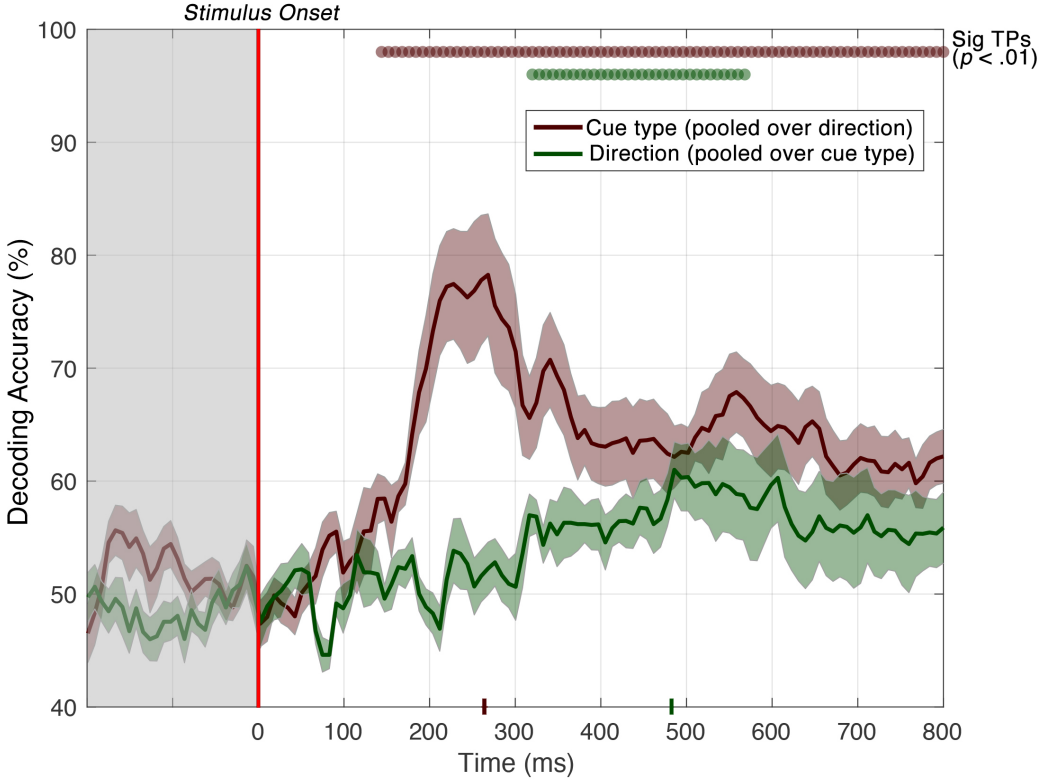


Figure 3 Decoding accuracy across 125 time points for cue type (pooled over direction) in red and direction (pooled over cue type) in green. Cue type could be decoded from 144 ms - 800 ms after stimulus onset, while direction could be decoded from 320 – 568 ms after stimulus onset. The coloured ticks 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 ($p < .01$). Shaded error bars represent ±1SE of the bootstrapped mean (1000 iterations).

It is possible that the ability to decode CD and IOVD cues is based on differential responses in early visual cortex, and these responses might be inherited from retinal mechanisms. Plotting the pattern of overall response for CD and IOVD at

the first point of possible decoding (Fig. 4.A and B) revealed subtle differences (Fig 4.C) in the response across the scalp. Similarly, we observed differences in the pattern of response at the time of peak decoding accuracy (Fig 4.F), with the largest (negative) responses occurring towards more posterior regions of the brain (represented by dark red in Fig 4.D and Fig 4.E). From these data, it appears that there is no obvious single source that is driving differences between CD and IOVD responses. Instead, it appears that the pattern of responses across the entire scalp contributes to decoding between CD and IOVD stimuli.

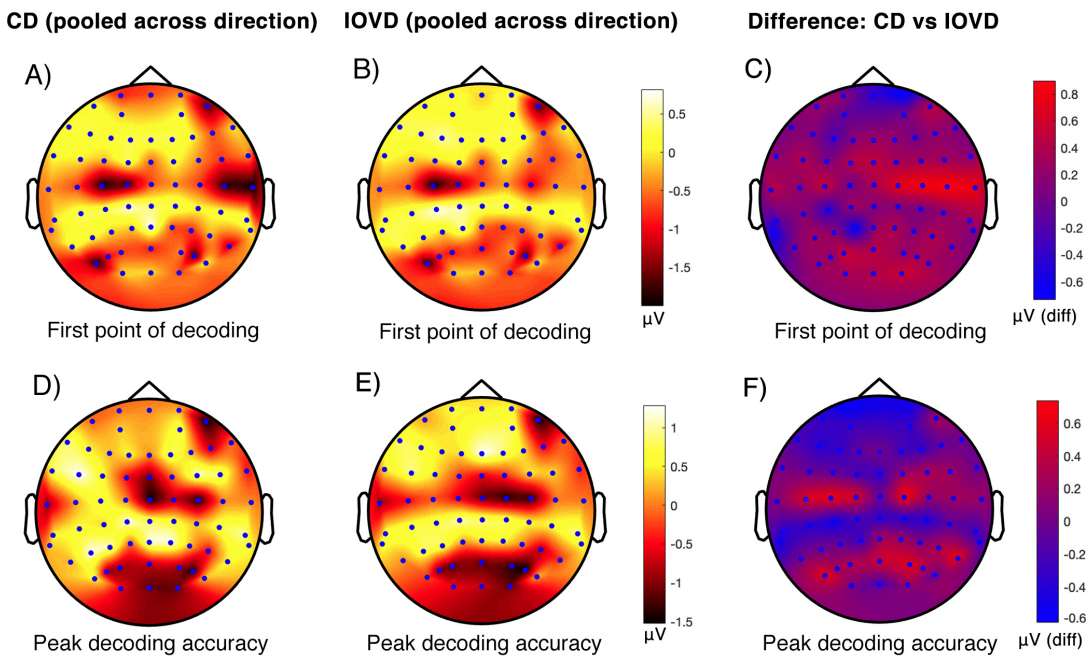


Figure 4 EEG response amplitudes to CD and IOVD (pooled across direction) and their differences visualised on a scalp montage. In panels A and B, we present the pattern of response at the first point of above-chance decoding accuracy (144 ms) between CD and IOVD responses (pooled across direction), and in panel C we present a visualisation of the difference in signal between these two conditions. In panels D-F, this is repeated for the pattern of response at the time of peak decoding accuracy (272 ms).

It is, perhaps, unsurprising that we are able to identify differences in the EEG response pattern generated by these network activations. However, our results suggest that we can also decode the *direction* of motion through depth (i.e. either towards or away) independent of cue type. As illustrated in Fig. 3 (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 568 ms after stimulus onset, although the accuracy was lower than for cue type, peaking at 61% at 488 ms.

Decoding within cue-type and direction

Decoding performance pooled across cue type (or motion direction) indicated that different neuronal populations might be recruited by CD and IOVD cues and by different motion directions. 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 motion directions *within* individual cue types. Our data (Fig. 5) show that we could do so at above-chance levels. Decoding between CD and IOVD stimuli moving towards in direction (Fig 5. blue line) was significant from 184 ms, and peaked at 74% at 272 ms, while the decoding between CD and IOVD stimuli moving away in direction (Fig 5. grey line) was significant from 141 ms and peaked at 73% at 224 ms. Performance in decoding between cue-type within motion direction peaks early (and stays high regardless of direction), consistent with the idea that they

represent qualitatively different inputs to the visual system that are processed by different anatomical pathways.

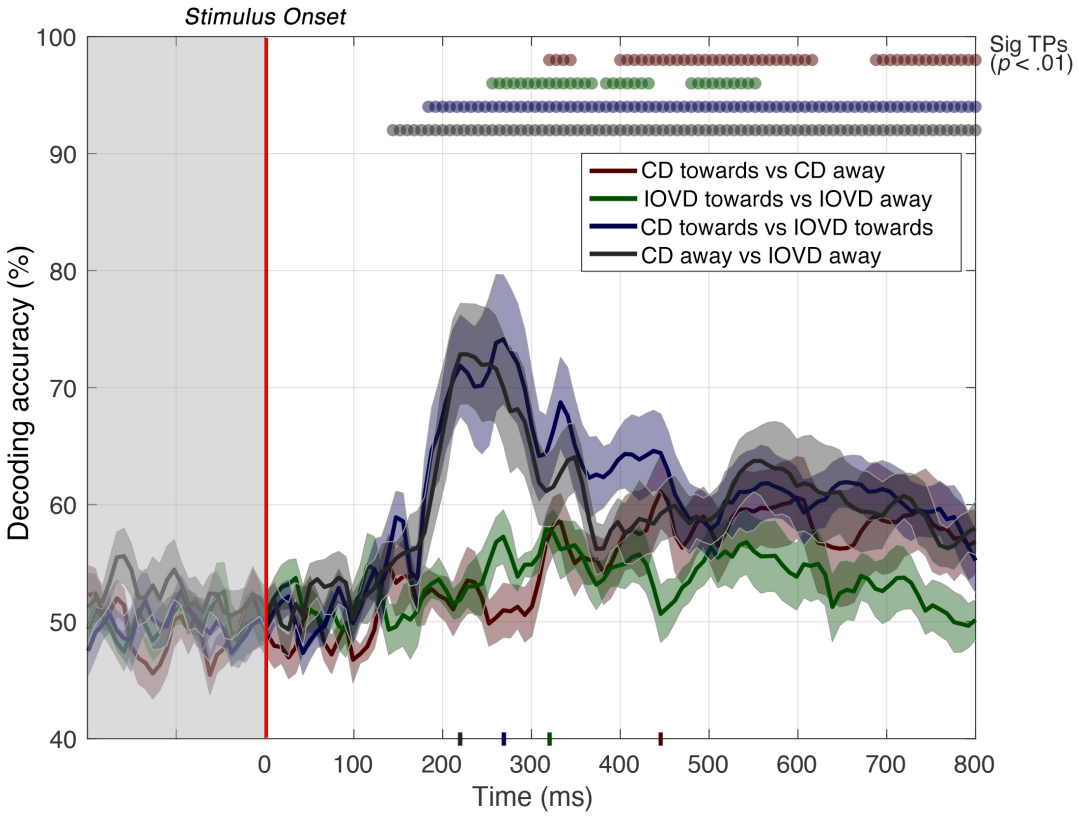


Figure 5 Pairwise decoding accuracy across 125 time points for four within cue and within direction conditions. Cue type (within direction) can be decoded from 144 ms (away) and 184 ms (towards) after stimulus onset, while stimulus direction (within each cue type) can be decoded at 320 ms (CD) and 256 ms (IOVD) after stimulus onset. Red, green, blue, and grey dots indicate time points when the Bonferroni-corrected t-tests were significant for each of the 4 conditions ($p < .01$) and the coloured ticks on the x-axis represent the time point of peak decoding performance for each condition. Shaded error bars represent ±1SE of the bootstrapped mean (1000 iterations).

Most surprisingly through, we are able to obtain above-chance performance in decoding the 3D motion *direction* of both CD and IOVD stimuli. As presented in Fig 5, we could accurately decode direction *within* CD or IOVD cue conditions. IOVD direction-polarity (Fig 5. green line) could be decoded from 256 ms post-stimulus onset, while CD direction-polarity (Fig 5. red line) could be decoded from 320 ms post-stimulus onset. Notably, the decoder could distinguish between IOVD direction-polarity 64 ms *before* CD direction-polarity. In addition, IOVD direction decoding accuracy (58% at 320 ms) peaked 128 ms *before* CD direction decoding accuracy (61% at 448 ms). Scalp montages illustrating the difference in EEG responses driven by towards and away CD and IOVD stimuli, at the first possible point and the peak point of decoding performance, are presented in Fig 6. The ability of EEG to resolve these responses to 3D-motion direction is intriguing. Although different motion directions are very similar for both CD and IOVD cues at the retinal level, the ability to decode depth direction from the data suggests that these responses may come from an intermingled network of depth-direction sensitive cells, and such responses can be resolved using EEG.

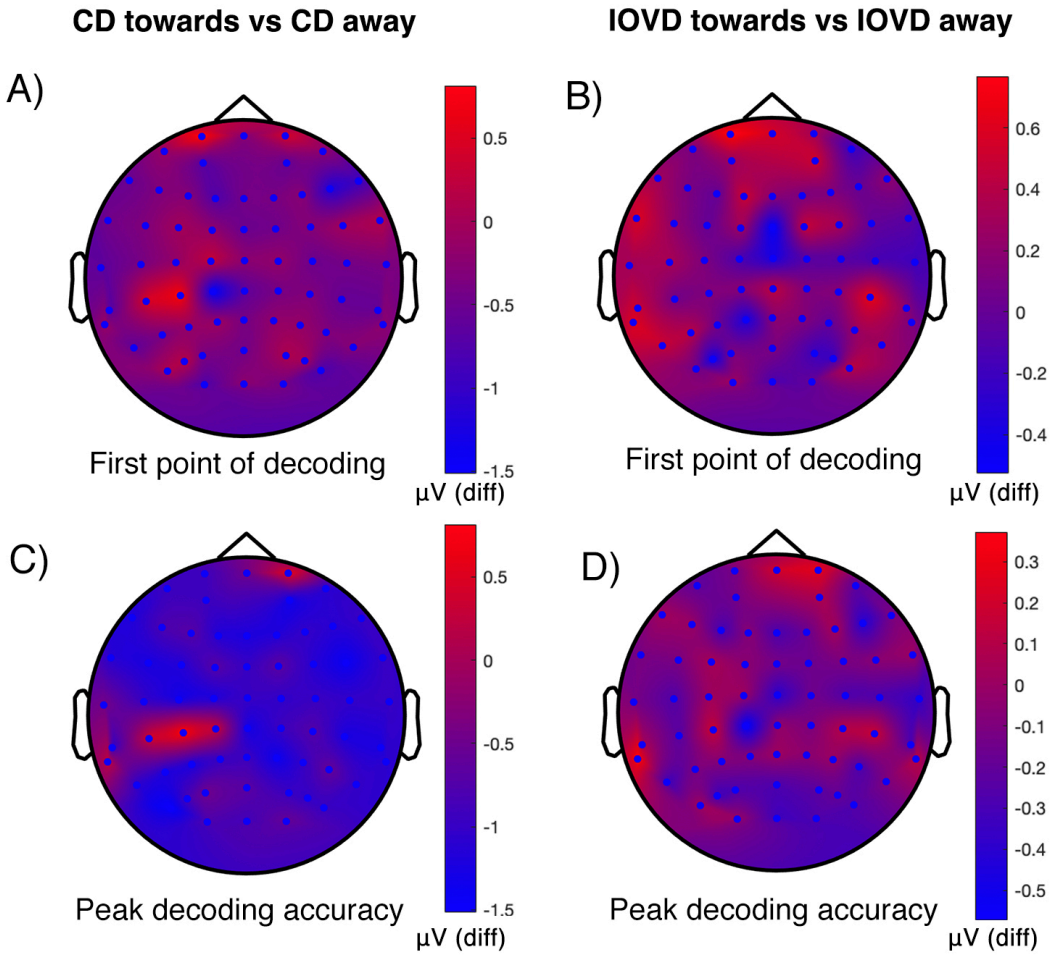


Figure 6 Differences in the pattern of EEG response when comparing between towards and away depth directions within CD or IOVD cue. In panel A and B, we present the difference in the overall pattern of response at the first point of above-chance decoding performance when comparing between towards and away responses for CD (320 ms) and IOVD (256 ms) cues, respectively. In panels C and D, we present the same at the time point of peak decoding accuracy for CD (448 ms) and IOVD (256 ms) cues.

Decoding keyboard responses

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 synthetic data that differed in how participants responded on the keyboard during the trial.

The results (Fig 7.) 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. Our stimulus decoding performance is therefore likely to be driven by bottom-up stimulus driven activity, rather than keyboard responses.

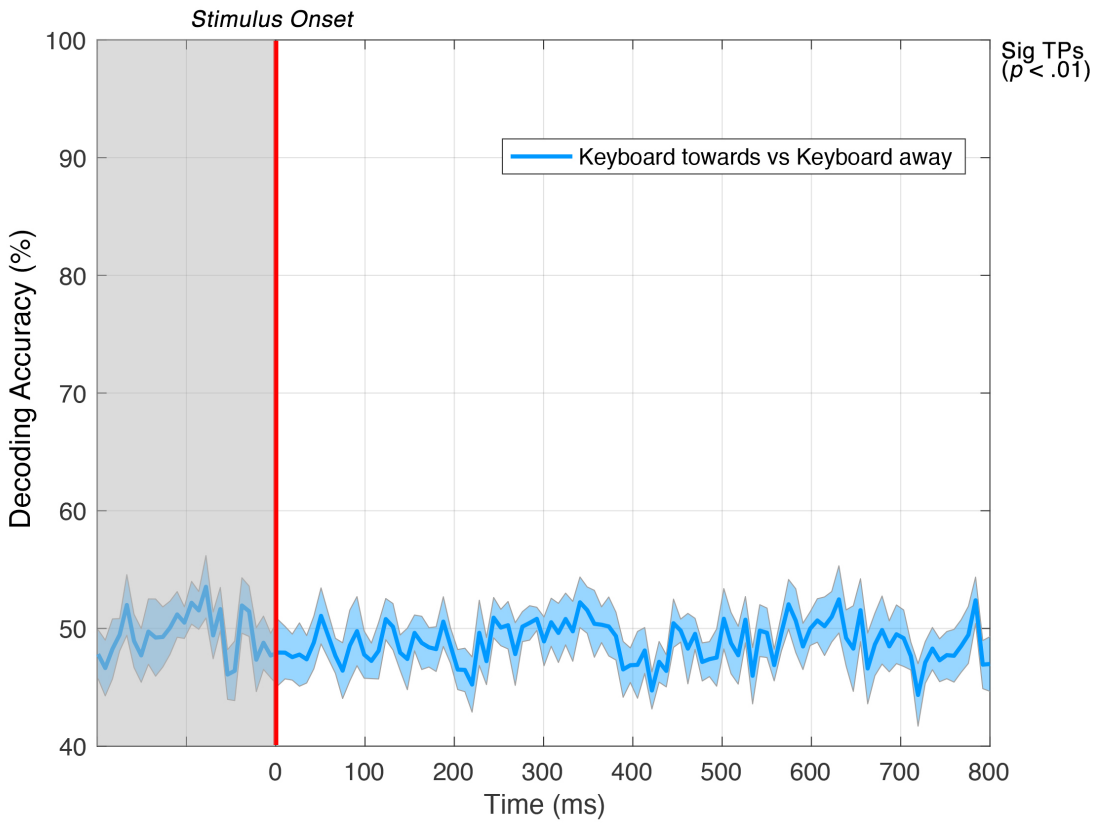


Figure 7 Pairwise decoding accuracy across 125 time points for keyboard response. Keyboard response (either pressing 2 on the keypad to indicate towards or 8 on the keypad to indicate away) cannot be accurately decoded: accuracy is at chance across all time points. Shaded error bars represent ±1SE of the bootstrapped mean (1000 iterations).

Discussion

We have examined whether we can decode EEG signals generated in response to CD and IOVD-isolating stimuli, moving either towards or away in depth. We then assessed how the timings of these signals relate to each other and to their perception. We find that cue type (CD and IOVD) and direction of 3D-motion (towards and away) can be decoded based on the distinct pattern of neuronal responses induced on the scalp. Further, IOVD cues can be decoded earlier, and decoding accuracy peaks earlier, when compared to CD cues. These signals can be decoded well in advance of behavioural reaction times for identifying MID cues, suggesting bottom-up processes are involved in the perception of MID. Finally, we could not decode EEG signals when partitioned into keyboard response conditions, confirming that decoding performance is likely to be stimulus-driven.

CD and IOVD mechanisms have independent retinal computations and speed sensitivities, and engage systems with different spatial resolutions (Czuba et al., 2010; Shioiri et al., 2008; Wardle & Alais, 2013). 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 at 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 direction. This is consistent with the hypothesis that CD and IOVD are processed through two computationally distinct mechanisms, with CD and IOVD cue types being processed by different neuronal pathways that may drive independent large-scale networks in the visual cortex (Harris et al., 2008; Rokers et

al., 2009). Decoding accuracy was high, sitting around 74%, and this is unsurprising. CD and IOVD mechanisms rely on fundamentally different properties that are intrinsic to such cues. For example, CD dots contain no motion energy and have a shorter lifetime, refreshing on every frame (i.e. every 16.6 ms), whereas IOVD dots travel short distances and have a relatively longer lifetime of 50 ms.

Perhaps surprisingly, the decoder could distinguish between signals generated from stimuli moving in different 3D-motion directions (either towards or away in direction polarity) within each cue type (CD or IOVD). 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 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 different motion directions are very similar for any given cue type. Decoding in the case of the CD stimulus must depend on neurons that can compute a temporal derivative from disparity selective neurons in primary visual cortex. Decoding in the case of IOVD 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. Overall, our data provides support for the existence of a spatial clustering 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). Alternatively, it is possible that different visual areas exhibit preferences for particular 3D motion directions and that these larger (cm-scale) neuronal biases underlie our findings.

EEG has a temporal resolution on the order of milliseconds, allowing one to assess temporal dynamics of CD and IOVD signals. The decoder could distinguish between IOVD direction-polarity 64 ms earlier than CD direction-polarity, and decoding accuracy for IOVD direction-polarity peaked 128 ms before CD direction-polarity. These differences are large, as it typically takes 40 ms for a neuronal signal to travel from the retina and be processed by the visual cortex (Celebrini, Thorpe, Trotter, & Imbert, 1993). Behaviourally, participants were generally slower and less accurate when identifying IOVD cues when compared to CD cues. Yet, IOVD cues could be classified earlier in the EEG time-course. Psychophysical evidence suggests differences in the temporal tuning of CD and IOVD mechanisms. Generally, the CD mechanism has a low-pass temporal tuning while the IOVD mechanism has a band-pass temporal tuning that is geared towards relatively higher temporal frequencies, which may be due to its initial computations being derived from monocular motion signals (Czuba et al., 2010; Joo et al., 2016; Shioiri et al., 2008; Wardle & Alais, 2013). While we are aware that the timing of decoder performance is not an absolute indication of information processing speed, our data provide support for the notion that the IOVD mechanism may have a relatively faster temporal tuning than CD (King & Dehaene, 2014). The preference of the IOVD system for rapid stimuli might reflect its role as part of a collision alert system, rather than a system supporting object

identification and location (Czuba et al., 2010; Maloney et al., 2016; Norcia & Tyler, 1984; Shioiri et al., 2008).

We have found that both MID cue (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. These 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 data suggesting the existence of 3D-motion direction sensitive neurons that may exist at a finer scale and are intermingled across the visual cortex. Finally, IOVD cues can be decoded and peak earlier than CD cues, reflecting a potentially faster tuning of the IOVD mechanism.

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.

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

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

Celebrini, S., Thorpe, S., Trotter, Y., & Imbert, M. (1993). Dynamics of orientation coding in area V1 of the awake primate. *Visual Neuroscience*, *10*(5), 811–825.

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., 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. *Journal of Neurophysiology*, *104*(5), 2886–2899. https://doi.org/10.1152/jn.00585.2009

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.

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

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-Newsom, Kirstie. 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*, *(in press)*.

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210.

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., 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

Maloney, R. T., Kaestner, M., Ansell, J., Bloj, M., Harris, J., & Wade, A. (2016). Mapping the temporal and neural properties of binocular mechanisms for motion-in-depth perception. *Perception*, *45*, 201.

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

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.

Shioiri, S., Nakajima, T., Kakehi, D., & Yaguchi, H. (2008). Differences in temporal frequency tuning between the two binocular mechanisms for seeing motion in depth. *Journal of the Optical Society of America A*, *25*(7), 1574. https://doi.org/10.1364/josaa.25.001574

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), 17–17. https://doi.org/10.1167/13.1.17

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.

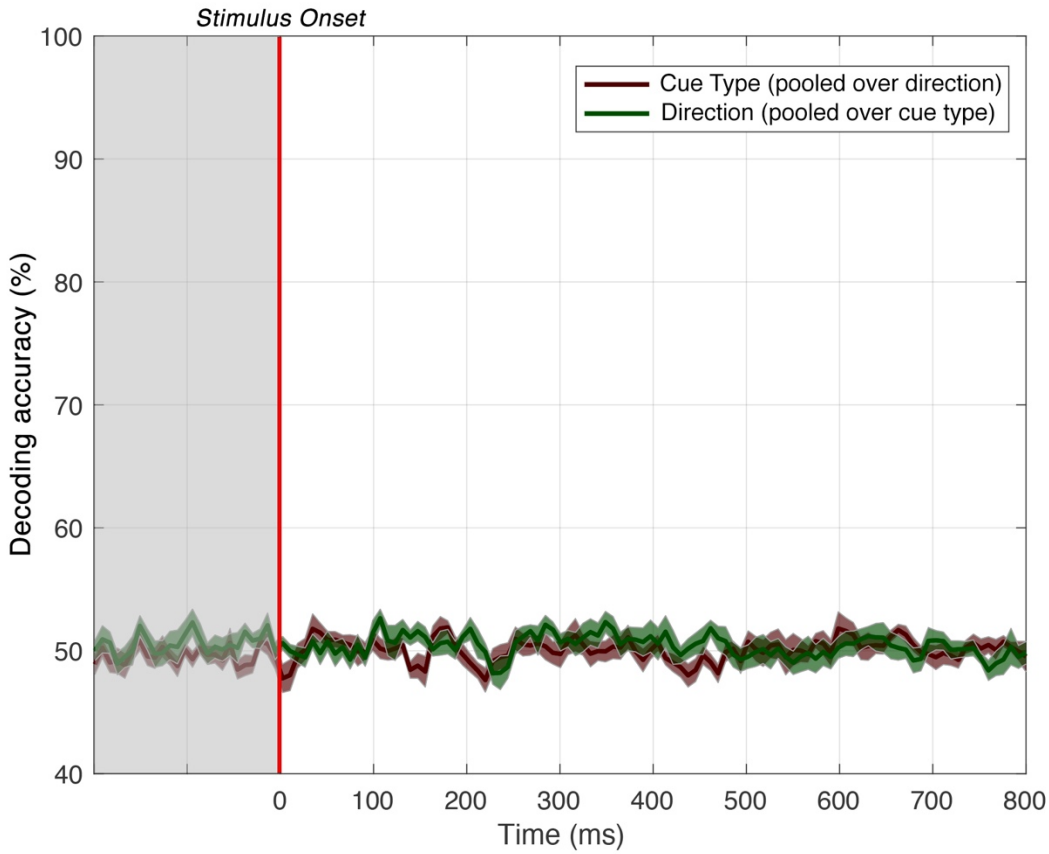
Author Contributions

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

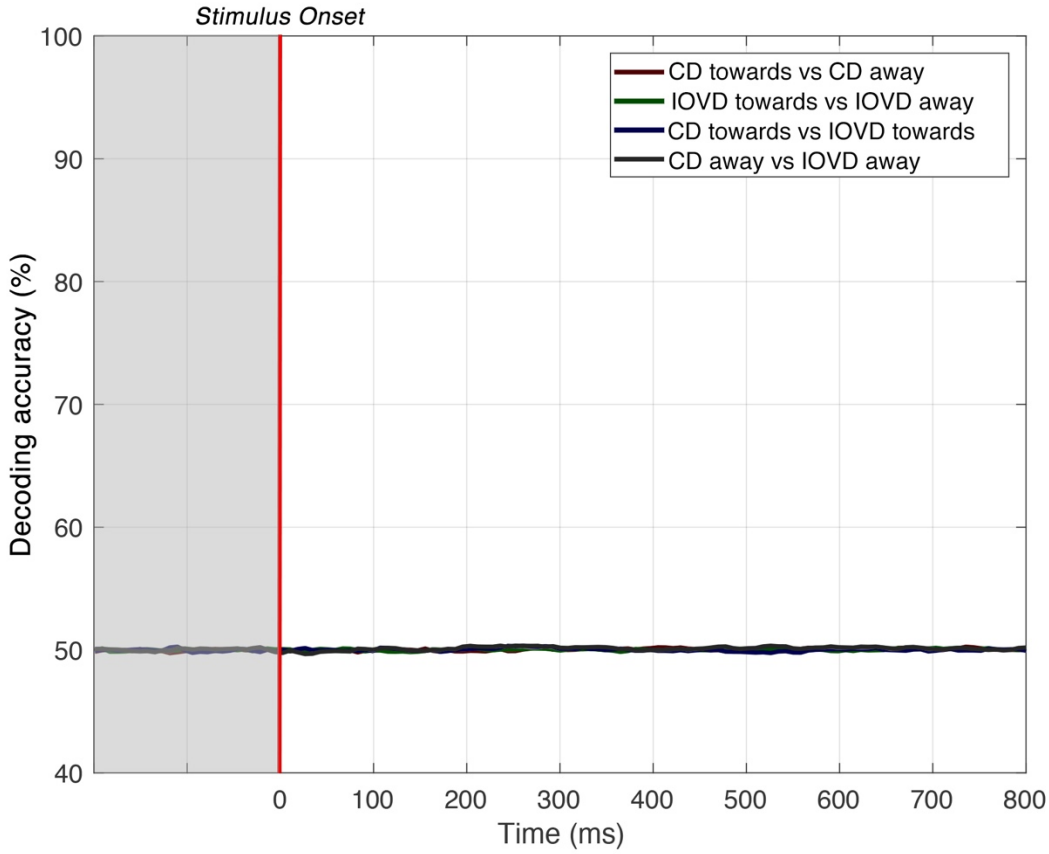
Competing interests

The authors declare no competing interests.

Supplementary Information



Supplementary Figure S1 Shuffled label pairwise decoding accuracies for cue type (pooled over direction) in red and direction (pooled over cue type) in green. After shuffling labels, the decoding accuracy of both comparisons fell around the 50% chance baseline, confirming that the SVM was working as expected.



Supplementary Figure S2 Shuffled label pairwise decoding accuracies for within cue and within direction conditions. After shuffling labels, accuracy for all four conditions was virtually identical, falling around the 50% chance baseline, confirming that the SVM was working as expected.