Disparity- and velocity- based signals for 3D motion perception in human MT+
Bas Rokers, Lawrence K. Cormack, and Alexander C. Huk

Supplementary Materials

Supplementary Figure 1. Effects of number of disparity-defined wedges on CD-motion responses. fMRI responses for all visual areas studied, as stimuli alternated between changing-disparity-defined (CD) 3D motion and spatiotemporally scrambled (STS) random dots (e.g., Experiment 2, Figure 2, blue bars). In the original Experiment 2, the CD motion display was split into 4 quadrants, with adjacent quadrants moving in opposite directions (see Figure 2a). We repeated the main CD-motion condition with similar displays that contained different numbers of wedges. If MT+ responses had been driven primarily by disparity-defined edges, one would expect little response to a full-plane stimulus that does not contain disparity-defined edges, and increasingly strong responses to stimuli that contained larger numbers of edges. MT+ responses modulated to all conditions including the 1-wedge/no-edge condition (i.e., all response amplitudes are >> 0%). Responses to displays that contained fewer (1, and 2) or more (8) wedges were not significantly systematically affected by the number of disparity-defined (w)edges ($F_{2,35} = 1.62, p = 0.305$). This pattern of responses confirms that the primary element driving MT+ responses was CD motion, and not other disparity- or depth-defined structure. Note that the 1-wedge condition (Likova and Tyler, 2007) is not directly comparable to the other conditions, because this full-field stimulus did not contain the rich pattern of relative disparities present in all other conditions (due to the absence of multiple simultaneous wedges), and may have produced stronger (unidirectional) drive to vergence. One out of three subjects did show a significant change in response with the number of wedges in LO ($F_{2,11} = 9.16, p = 0.007$), and two out of three showed this effect in V3A ($F_{2,11} = 12.81, p = 0.002$ and $F_{2,11} = 26.65, p = 0.0002$ respectively). This effect was primarily due to an increased response in the 8-wedge condition. These two areas might have carried a mix of CD-motion and motion- or disparity-defined edge signals.

Nature Neuroscience: doi:10.1038/nn.2343
Supplementary Figure 2. Direction-selective fMRI adaptation effects were evident in every subject in every scanning session (Experiment 4). Top row shows 3D motion fMRI adaptation results in MT+ for each subject and each scanning session (data average shown in main Figure 4b, left panel); bottom row shows results for 2D dichoptic control. Adaptation effects were evident in every 3D dichoptic session, for every subject (responses to opposite-direction probes were larger than to same-direction probes), consistent with the 3D dichoptic group average shown in Figure 4b. Differences were smaller and less regular in the 2D monocular experiments reflected by the smaller 2D monocular group average effects presented in the main manuscript. Values above horizontal axes in each subpanel indicate behavioral performance (% hit rate) in the attentional-control task during the two (color-coded) conditions (see Methods).
Supplementary Figure 3. Psychophysical evidence for 3D direction-selective adaptation. We observed a perceptual motion aftereffect for 3D motion using stimuli similar to those in our fMRI adaptation experiment. The figure shows proportion of times the probe stimulus was judged to move towards the observer in depth (y-axis), as a function of the direction and strength of the probe stimulus (x-axis, positive numbers indicate motion towards the observer; negative numbers indicate motion away and 0 indicates no net motion). The strength of probe motion in either direction was varied by changing the proportion of dots moving coherently away from or towards the observer (remaining dots followed random walks through depth). All other aspects of the stimulus and protocol were similar to those in the fMRI adaptation experiment. Red points show judgments after “away” adaptation, green points show judgments after “towards” adaptation, and white points show judgments without adaptation (baseline). “Towards” choices were much more likely after away adaptation than after towards adaptation (i.e., red curve is higher/left-shifted relative to green curve), demonstrating a direction-selective motion aftereffect through depth. Data shown for one observer (BR; n = 12 trials per point for each adaptation condition; 24 trials per point for the baseline condition, error bars depict ± 1 s.e.m.). We observed effects of similar magnitude in other observers. Logistic curves (black lines) were fitted to data points using a least squares algorithm.
Supplementary Figure 4. Relationship between MT and MST responses across experiments. We identified 2 subregions within MT+, putative MT and putative MST, based on a double-dissociation strategy involving retinotopic organization and ipsilateral responsivity (see Supplementary Fig. 5 for details). Responses in these 2 subregions were similar across all experiments: we did not discern any systematic or compelling patterns of distinction between MT and MST. (a) MT response amplitude plotted against MST response amplitudes from the first 3 (blocked design) experiments. Each point indicates a pair of MT and MST amplitudes from a single experimental run (i.e., every data point that went into the averages shown in main Figures 1-3 is plotted individually here). (b) MT versus MST responses for every trial from the adaptation (event-related) experiment (3D dichoptic trials only, collapsed over probe direction). Both scatterplots show an oriented ellipse with a positive slope near 1 (depicted by the dotted diagonal lines) indicating the relationship between the responses in these 2 small abutting subregions.
Supplementary Figure 5. Data analysis.

A. Typical flattened view of the phase-encoded retinotopic (polar angle) map in occipital cortex. Anatomical images were skull-stripped using FSL BET software (http://www.fmrib.ox.ac.uk/fsl). We used the mrVista/mrGray software suite (http://white.stanford.edu/software) for all subsequent anatomical processing. Retinotopically organized visual areas were defined by measuring the polar angle component of the cortical retinotopic map. We combined visual areas V3A and V3B (here referred to as “V3A”) and LO1 and LO2 (here referred to as “LO”) because the retinotopic maps did not allow for clear distinction of these adjacent areas for all subjects. MT+ was identified as a contiguous region of gray matter that responded more strongly to full-field dots moving in the plane of fixation than to a stationary pattern of dots in the plane of fixation, typically within or near the occipital continuation of the inferior temporal sulcus. Coherence, a measure of signal-to-noise, represents the correlation between the observed time-series and the best-fitting sinusoid at the fundamental of the stimulus frequency. We applied a coherence threshold of 0.3 (and a phase lag criterion of < 6 s) to an average of 10 MT+ localizer scans performed in a separate experimental session for each subject.

In a separate session, we distinguished putative human MT and MST by identifying MT as a subregion of MT+ that responded with a systematic phase progression to a retinotopic mapping stimulus, and MST as a distinct region that responded to ipsilateral stimulation (Huk et al., 2002). To map retinotopic responses, subjects viewed a 45 deg wedge of moving dots that slowly (24 s/revolution) rotated through the visual field. To identify ipsilateral responsiveness, subjects viewed a 45 deg wedge of dots that alternated between moving and stationary. The wedge was centered on the horizontal meridian, spanning an eccentricity range from 8.5-17.5 deg. In both displays, the dots followed sinusoidal paths with random phase, oscillating radially towards and away from the fixation point. We then performed standard coherence analysis for both types of stimuli, and hand-identified distinct regions with non-overlapping responses to the 2 conditions. We identified both subregions in each hemisphere and, as in prior reports, putative MT was posterior to putative MST (Huk et al., 2002). Some portions of MT+ were not assigned to either subregion, owing either to responses to both or neither of the retinotopy and ipsilateral stimuli. Because we typically observed rather similar patterns of response in the putative MT and MST subdivisions of the MT+ complex across our experiments, we predominantly report...
Motion reference scans were run at the beginning and end of each experimental scanning session to further restrict MT+ and all other ROIs (Experiments 1 through 3; Experiment 4 used a single reference scan at the beginning of each session). During the reference scans, subjects fixated a central point while the display alternated between expanding/contracting (12 s) and stationary (12 s) dots in the plane of fixation (presented in same location as the stimuli in all the main experiments). The averaged reference scan responses were used to exclude unresponsive voxels in the visual ROIs, i.e., gray matter regions that would have responded to visual field locations outside the stimulus apertures, or voxels that had too little overlap with gray matter. Voxels that did not exhibit strong coherence with the stimulus alternations ($r < 0.30$, phase lag $> 6$ s) were discarded from further analysis. Coherence thresholds ranging from 0.23 to 0.50 yielded similar results. If anything, stricter coherence thresholds yielded more pronounced effects.

B. Typical single session BOLD signal time-series during block alternation of a moving (12 s) and static (12 s) 3D stimulus in MT+. In the blocked-design protocol used in Experiments 1-3, the response time-series within each ROI was fit with a sinusoid at the same frequency of the stimulus alternation. The resulting vector (amplitude and phase) was projected onto a unit vector with the average phase of the two motion reference scans (described above) to obtain the univariate projected amplitudes for each scanning run (in units of % signal change).

We combined the resulting response amplitudes across sessions and performed standard statistical analyses (2-way or 3-way repeated measures ANOVA as appropriate, with subject as a random factor). We relied on the results of the ANOVAs to confirm the patterns of response evident in the bar graphs in Figures 1c, 2c, and 3c and followed up on significant effects with 2-tailed Student’s $t$-tests to further interpret the observed interactions. We note that similar conclusions are supported if the $t$-tests are replaced with bootstrapped, nonparametric confidence intervals and that largely identical trends were observed upon visual inspection of individual subject means and standard errors.

C. Typical single session event-related BOLD signal change to presentation of top-up adapter, and Same (yellow), Opposite (orange), or No (blue) probe in MT+. In the event-related adaptation protocol used in Experiment 4, the time series for each trial was extracted for each ROI as the 22.5 s (3 trials) following the start of each top-up adapter (the first and last two trials from each run were discarded, to ensure that each trial had the same number of preceding and following events). For the data analysis and figures reported in the main text (Figure 4b), we subtracted the mean time series for adapter-blank trials (blue) from the responses to adapter-probe (same/opposite) trials (yellow/orange). Standard errors of the mean for each time point were calculated over all trials. Adaptation effects appear small relative to the convolved BOLD response to adapter-blank trials, but reflected consistent responses to short (1 s) probe (same/ opposite) stimuli, that were significantly different in each individual session in each subject.

To estimate univariate response amplitudes, we used the analysis of Larsson et al.
(2006), projecting each adapter-probe trial onto the mean time series response (independent of condition), and normalizing by the amplitude of the mean response. Each projection results in a scalar value with units of % signal change that indicates whether that trial’s response was larger or smaller than the average response. We then sorted the trials by condition (e.g., same-direction versus opposite-direction), calculated the mean and standard error of these univariate response amplitudes and subjected these values to t-tests. Finally, to facilitate interpretation and comparison across visual areas, we converted these univariate amplitudes to a direction-selectivity index, calculated as the difference between the same and opposite responses, divided by their sum. We then bootstrapped confidence intervals, as shown in Figure 4c.

In all experiments, we chose to focus on making a large number of measurements in a small number of subjects (19 scanning sessions in each of 3 subjects, in addition to numerous pilot experiments). As in other methods that employ a similar data collection strategy (e.g., visual psychophysics, primate electrophysiology), we carefully inspected the pattern of means and standard errors for each subject in each experiment, and noted that each subject showed a pattern of results similar to those in the group-averaged data shown in the figures. Our strategy of making many measurements in 3 subjects does not allow us to make statistical inferences about the population as a whole, but does allow us to assess the consistency of the pattern of results within each subject. The regularity of results across subjects throughout all 4 experiments suggests that our inferences are likely to generalize to subjects with expertise in maintaining fixation and vergence while simultaneously performing challenging psychophysical tasks, and who have been shown in prior psychophysical studies to exhibit strong and reliable percepts of 3D motion (Rokers et al., 2008).
Supplementary Figure 6. Visual stimuli and 3D percepts.

A. **Experiment 1 - Still frame for the monocular horizontal condition.** In experiment 1 we investigated the contribution of motion orientation (horizontal versus vertical) and eye-of-stimulation (monocular versus dichoptic) to 3D motion. In all conditions, each dot followed a 1.6 Hz sinusoidal trajectory and the two dots in a pair always had opposite (anti-phase) motion. Dot velocity in each monocular half-image ranged between ±1.2 deg/s. The amplitude of the motion profile produced ±22 arcmin of binocular disparity. Each dot pair started at a random phase of its sinusoidal motion trajectory, so that there was no global motion resembling optic flow, and no strong drive on vergence. During block alternations, dot speeds were linearly ramped between 0 deg/s (stationary) and full speed over 1 s. Each dot was semi-transparent (rendered with an α transparency of 0.5) so that the luminance of each dot was half of the available range. In this way luminance would mix (instead of occlude) when two dots moved across the same space in one of the monocular images. We chose to present the motion with smooth and continuous oscillations of the dots because they supported strong percepts of motion through depth; the generally small responses we observed in V1 were most likely due to the resulting lack of contrast transients in these displays (the same holds for Experiment 3).

In all conditions dots alternated every 12 s between moving and stationary. Each scanning run contained 7 repeats of this 24 s period for a total of 168 s. For each of the 3 subjects, we repeated each of the 4 conditions 3 times across 3 scanning sessions.

B. **Experiment 2 - Still frame for the CD-motion condition.** In experiment 2 we investigated the contribution of changing disparity signals to 3D motion (CD-motion). In the main condition, each block contained the CD stimulus for the first 12 s, and then a spatiotemporally-scrambled (ST-scrambled) version of the display for the remaining 12 s. In the ST-scrambled stimulus, temporal frames were randomly drawn with replacement from the temporal sequence that made up the CD stimulus, and then each
(binocular) dot was assigned random x-y coordinates without regard to the wedge structure of the CD stimulus. This resulted in a stimulus that had the same overall low-level statistics (including disparity statistics) as the CD stimulus, but lacked the specific spatiotemporal structure that specified coherent motion though depth.

In three control conditions, each block contrasted different degraded versions of the CD stimulus (12 s) against the same ST-scrambled stimulus just described (12 s). In the first condition, we used a spatially-scrambled version, in which each frame was generated as for the CD-motion stimulus, and each dot was then randomly repositioned in the display (thus maintaining the overall temporal disparity structure but eliminating the wedges). In the second condition, we used a temporally-scrambled version, in which frames were randomly drawn with replacement from the temporal sequence that made up the CD stimulus (thus maintaining the spatial wedge structure but eliminating the smooth change in disparity over time). For the third, we used an anti-correlated version, which was identical to the CD stimulus except that one member of each binocular dot pair was flipped from white to black or vice versa. In this condition dots appeared to be located in or near the image plane at all times, and no clear CD-motion was perceived.

Each scanning run contained 7 repeats of the 24 s period for a total of 168 s. For each of the 3 subjects, we repeated the CD motion condition 12 times and each of the 3 other conditions (spatially-scrambled, temporally-scrambled, anti-correlated) 4 times across 4 scanning sessions.

C. **Experiment 3 - Still frame for the anticorrelated IOVD-biased condition.** In experiment 3, we investigated the contribution of the IOVD cue to 3D motion, using anti-correlation to disrupt the contribution of disparity-based signals. Furthermore, dot pairs oscillated on a horizontal sinusoidal trajectory with either the same phase in the two eyes (MWD) or opposite phase (MTD). In the 50% correlation condition, 50% of the dot pairs had the same, and 50% had opposite polarity contrast. At 25% correlation, 25% of the dot pairs had the same, and 75% had opposite polarity contrast.

All dots were assigned a random starting position along their sinusoidal motion trajectory, so that there was no global motion resembling optic flow and no strong drive on vergence that could affect eye position. Monocular dot velocity ranged between ±1.2 deg/s. The amplitude of the 1.6 Hz sinusoidal motion trajectory produced ±22 arcmin of binocular disparity, and the range of disparities in the MWD displays spanned the same range as in the MTD displays. As in Experiment 1, dot speeds were linearly ramped between stationary and fully moving at block alternations.

Stimuli had a 24 s period. Dot pairs moved during the first 12 s and the same dots were stationary during the second 12 s. Each scanning run contained 7 repeats of the 24 s period for a total of 168 s. For each of the 3 subjects, we repeated each of the 8 conditions (MTD 100, 50, 25 0% correlation, and MWD 100, 50, 25, 0%) 3 times across 3 scanning sessions.

D. **Experiment 4 - Still frame for the 3D adapt condition.** Experiment 4 employed a direction-selective adaptation protocol to investigate the selectivity of cortical areas to specific directions of 3D motion. In all adaptation experiments, 32 dots pairs were presented in 3.5 deg radius circular patches within the left and right eye half-images. All dots were assigned a random starting position and moved continuously towards or
away from the observer with a 1 deg/s monocular velocity until they reached the edge of a 3D volume defined by ±18 arcmin deg of disparity. Upon reaching the edge of the volume, dots were replotted on the opposite (near/far) side of the volume with a random image plane position. Dots were spaced to minimize false binocular matches; at 0 arcmin disparity they were at least 0.7 deg (in the image plane) away from any other dot. Because individual correlated dots moved continuously towards or away from the observer, the stimuli contained both IOVD and CD cues. The region outside the aperture contained a 1/f noise pattern at 0 arcmin disparity (i.e., in the fixation plane) to stabilize vergence.

E. Illustration of a typical 3D percept produced by the visual stimuli. To control attention across experimental conditions, observers performed a challenging visual dot-change task throughout all fMRI runs. At exponentially distributed random times a dot pair changed color or contrast and observers pressed a button when they detected the change. Average detection rates in the dot-change task for Experiments 1, 2, 3, and 4 were 64, 40, 62, and 69% respectively, confirming that the detection task was challenging and required continuous attention. Note that these are hit rates (a button press within 1 s of the event) during a continuous task, so that random and occasional button presses would have resulted in detection rates close to 0% (false alarms were extremely rare).

In Experiment 4 observers performed the dot color change detection task throughout all periods of stimulus presentation (adapter and probes). We did not observe differences in detection rates during the presentation of same versus opposite direction of motion probes, for either the 3D dichoptic ($t_{135} = 0.384$, $p = 0.702$) or the 2D monocular ($t_{142} = 1.430$, $p = 0.155$) adaptation experiments.

We found this dot-change task preferable over a variety of other standard tasks (e.g., speed discrimination or a dimming detection task at fixation), because it required attention to be deployed across the entire visual display, could be performed during both moving and stationary epochs, and was not substantially affected by the often large differences between 2D and 3D motion sensitivity (Tyler, 1971; Harris et al., 1998). The exponential distribution of detection events (resulting in a flat hazard function) further required subjects to evenly and continuously distribute their attention over time, instead of periodically performing regularly-spaced trials.

To maintain alertness, qualitatively strong stereoscopic percepts, and minimize head motion throughout the scanning sessions, subjects were given only short breaks between runs (~5-10 s), and were encouraged to continuously perform the task throughout the scanning session.
Supplementary references