Spatiotemporal integration of isolated binocular three-dimensional motion cues

Jake A. Whritner

Center for Perceptual Systems, Department of Psychology, The University of Texas at Austin, Austin, TX, USA ✉

Thaddeus B. Czuba

Center for Perceptual Systems, Department of Psychology, The University of Texas at Austin, Austin, TX, USA ✉

Lawrence K. Cormack

Center for Perceptual Systems, Department of Psychology, The University of Texas at Austin, Austin, TX, USA ✉

Alexander C. Huk

Center for Perceptual Systems, Departments of Neuroscience & Psychology, The University of Texas at Austin, Austin, TX, USA ✉

Two primary binocular cues—based on velocities seen by the two eyes or on temporal changes in binocular disparity—support the perception of three-dimensional (3D) motion. Although these cues support 3D motion perception in different perceptual tasks or regimes, stimulus cross-cue contamination and/or substantial differences in spatiotemporal structure have complicated interpretations. We introduce novel psychophysical stimuli which cleanly isolate the cues, based on a design introduced in oculomotor work (Sheliga, Quaia, FitzGibbon, & Cumming, 2016). We then use these stimuli to characterize and compare the temporal and spatial integration properties of velocity- and disparity-based mechanisms. On average, temporal integration of velocity-based cues progressed more than twice as quickly as disparity-based cues; performance in each pure-cue condition saturated at approximately 200 ms and approximately 500 ms, respectively. This temporal distinction suggests that disparity-based 3D direction judgments may include a post-sensory stage involving additional integration time in some observers, whereas velocity-based judgments are rapid and seem to be more purely sensory in nature. Thus, these two binocular mechanisms appear to support 3D motion perception with distinct temporal properties, reflecting differential mixtures of sensory and decision contributions. Spatial integration profiles for the two mechanisms were similar, and on the scale of receptive fields in area MT. Consistent with prior work, there were substantial individual differences, which we interpret as both sensory and cognitive variations across subjects, further clarifying the case for distinct sets of both cue-specific sensory and cognitive mechanisms. The pure-cue stimuli presented here lay the groundwork for further investigations of velocity- and disparity-based contributions to 3D motion perception.

Introduction

The primate visual system is known to extract two binocular cues to support the perception of motion through depth, one based on a binocular combination of monocular velocity signals and another based on binocular positional disparity signals (Czuba, Rokers, Huk, & Cormack, 2010; Nefs, O'Hare, & Harris, 2010). The use of the velocity-based cue takes retinal velocities as the core inputs to support an inference of three-dimensional (3D) direction: this is 3D motion perception based on retinal motion primitives per se.[1] The disparity-based cue, in contrast, is built from classical binocular disparities, and involves estimating how these disparities change over time: this is 3D motion perception based on temporal integration of a binocular position in depth estimate. Prior work has shown that the visual system contains separable mechanisms for encoding velocity- and disparity-based 3D direction information that are weighted differently in different regimes (Nefs et al., 2010): velocity-based information is relied on more for peripheral locations and medium-to-fast speeds, whereas disparity-based

information predominates for foveal/parafoveal locations and slow speeds (Brooks & Stone, 2004; Cumming & Parker, 1994; Czuba et al., 2010; Harris & Watamaniuk, 1995; Shioiri, Nakajima, Kakehi, & Yaguchi, 2008).

It is tempting to posit that functional dissociations between velocity- and disparity-based processing might reflect a difference in usefulness for tasks that emphasize a rapid response versus those that emphasize precision over speed. To consider extreme examples, catching a ball (or quickly ducking to avoid it) might be largely velocity-based, whereas the slow threading of a needle, fine surgery, and so on, might rely far more on disparity-based processing. These qualitative and anecdotal examples can be made quantitative by hypothesizing that velocity-based sensitivity would increase steeply over brief amounts of time, such that increases in viewing duration would be equivalent to increases in stimulus strength (aka, Bloch's law; Bloch, 1885). This regime of complete temporal integration—when measured behaviorally, often accomplished in less than 200 ms—can be loosely interpreted as consistent with the contribution of sensory neurons integrating signals over relatively brief time periods. Alternatively, disparity-based processing might additionally accumulate signals over a longer time period because the relevant tasks afford this luxury. Such a gradual dependence on viewing duration is expected from probability summation, in which the noisy sensory representation is accumulated by downstream mechanisms that benefit from multiple quasi-independent samples of the noisy sensory signal over time. Almost any form of post-sensory integration of noisy sensory signals will reflect sensitivity improvements with time, but shallower than for the ideal of linear, noiseless sensory integration (Palmer, Huk, & Shadlen, 2005; Watson, 1986).

However, these conjectures regarding distinct timescales and forms of integration between velocity-based and disparity-based processing have not been tested directly. Prior work has shown that 3D motion stimuli containing both velocity-based and disparity-based cues do support two phases of integration: an early, steep phase likely to reflect the contributions of sensory mechanisms, and a later, shallower phase likely to derive from more downstream integration (Katz, Hennig, Cormack, & Huk, 2015). But that work only compared the slopes of 3D integration with measures of standard two-dimensional (2D) motion integration, and did not dissociate the two binocular cues to 3D motion. We, therefore, performed direct measurements of the temporal integration of velocity and disparity-based mechanisms using purely cue-isolating stimuli. To do this, we created a novel psychophysics-friendly generalization of a spatio-temporal stimulus geometry previously used to examine vergence responses to the velocity-based cue (Sheliga et al., 2016). Their stimuli cleverly and completely isolated the two cues. Stimulus

designs for purely isolating the disparity-based cue have been described and used in many previous studies, but such stimuli often differed from stimuli containing velocity-based cues in many ways, which limits the clarity of such comparisons (Brooks & Stone, 2004; Czuba et al., 2010; Harris, Nefs, & Grafton, 2008; Sanada & DeAngelis, 2014). Isolating the velocity-based cue has proved more elusive, and all previous studies, including our own, have used stimuli that have merely minimized (in practice) but not eliminated (by design) the potential for disparity-based signal contamination. As far as we know, this psychophysical study is the first to use a geometrically pure velocity-based cue to 3D motion. Although some other attempts did use stimuli that removed local binocular disparities, possible matches on larger spatial scales were still possible; in the current work, no systematic disparities can be extracted by any mechanism, known or unknown (Joo, Greer, Cormack, & Huk, 2019; Maloney et al., 2018; Nefs et al., 2010; Rokers, Czuba, Cormack, & Huk, 2011).

We found that direction discrimination based on velocity signals saturated quickly (∼200 ms), whereas disparity-based direction discrimination evolved more slowly, over the time scale of a half-second or more (although interesting individual differences were also observed). Analogous measures of spatial integration yielded similar estimates for velocity and disparity processing, suggesting that a primary difference between the two cues is indeed the amount of time over which each is estimated.

Methods

Participants

Four experienced psychophysical observers (three authors, one naive, males, aged 27–51) participated in this study. All four subjects completed Experiment 1; two authors (JAW and ACH) and the naive observer completed Experiment 2. Participants were screened for good stereopsis and normal or corrected-to-normal vision. All observers participated with written informed consent and were treated according to the principles set forth in the Declaration of Helsinki of the World Medical Association. All procedures were approved by The University of Texas at Austin Institutional Review Board.

Experimental apparatus

Experiments were programmed using the Psychophysics Toolbox (Brainard, 1997) for MATLAB (MathWorks, Natick, MA). Stimuli were presented stereoscopically on a large-format projector display (PROPixx projector, 120 Hz per eye, 74.5 cm × 132.5 cm; VPixx, St. Bruno, Canada) outfitted with a

polarizing-preserving rear projection screen (Screen Tech ST-Pro-DCF acrylic glass screen; Hamburg, Germany) and frame-synced LCD circular polarizer (ProPixx 3D polarizer module by DepthQ) to provide stereoscopic presentation with high temporal precision and minimal crosstalk (ProPixx 'RB3D' interleaved stereo presentation sequencer; Kenny, 2020). Subjects wore passive 3D glasses (circular polarization). They were seated 57 cm in front of the screen with their head in a chin rest and their forehead resting against a stabilization bar.

Velocity-based stimulus

For the velocity-isolating condition, we modified the novel and clever oculomotor stimulus introduced by Sheliga et al. (2016). The general cue-isolating logic of the approach is based on presenting sinusoidal gratings that jumped in opposite directions in the two eyes: critically, the motion was generated by applying a discrete quarter-wave (90°) phase shift on each stimulus frame update in opposite directions in the two eyes. Stereoscopically, this results in an interocular phase difference of either 0° (in phase) or 180° (out of phase). Owing to the circular nature of phase, alternating between 0° and 180° interocular phase is ambiguous with respect to the direction of binocular disparity change, as stepping from 0 to 180° is equivalent to stepping from 0° to −180°. However, the 90° phase steps produce clean, unambiguous motion signals within each eye. This interocular phase logic is schematized in Figure 1.

To have a stimulus that could be used for visual psychophysics, we wanted to be able to measure how accuracy depended on other stimulus parameters, which required reducing the signal to noise of the motion (as opposed to the full-field, fully coherent motion in the original oculomotor stimulus). We therefore extended the interocular phase logic to a multiple-element stimulus comprising 16 binocular Gabor patches (each 0.5° wide (FWHM), 2 cycles/°, 25% Michelson contrast). Example frames of the stimulus presented to each eye are shown in Figure 2. Each Gabor element was randomly positioned around the chosen stimulus position (5° eccentric from fixation) in a standard normal distribution ($\sigma = 1°$). A spacing algorithm ensured all elements were nonoverlapping by imposing a minimum distance of 1° between all other gratings, as well as the central fixation target. Empirically, all element positions fell within a circle centered at 6.2° in eccentricity with a diameter of 11° (eccentricities ranged from 0.7° to 11.6°). Each element followed a phase updating rule such that each started with a randomized baseline phase that was matched in the two eyes. Then, on every update (4 frames, ~67 ms) each monocular element underwent a quarter wavelength shift in opposite directions between the two
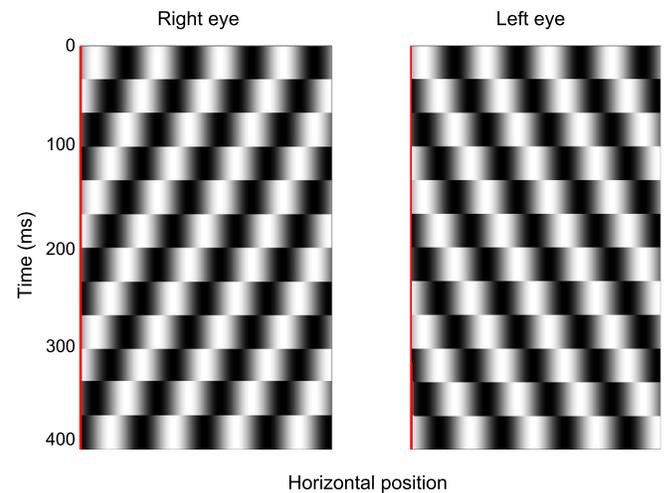
Figure 1. Schematic of the logic underlying the pure velocity-based stimulus. The left and right panels show a single luminance grating as a function of space (*x*-axis, "horizontal position") and time (*y*-axis, with time moving forward from top to bottom). A red reference line has been added at position zero in each eye to help visually clarify the phase changes. The two space–time half-images have an unambiguous velocity evident in the orientation of the displays (i.e., orientation in space–time is motion). Free-fusing these space–time images can be done to reveal that there is no coherent interocular disparity signal (i.e., there is not a coherent gradient of disparity across the vertical axis) owing to the interocular phase difference always being 0° or 180°. This interocular phase logic was first used by Sheliga et al. (2016) to purely isolate the velocity-based cue. In their displays, there was a single full-field grating; in ours, the sinusoidal pattern shown was windowed with a Gaussian contrast profile to create a small Gabor element, and multiple elements were presented within a particular region of the display, as described elsewhere in this article.

eyes. The resulting stimulus yields a compelling percept of 3D motion toward or away from the observer. The temporal frequency of the velocity-only gratings was 3.75 cycles/second·eye for a monocular angular speed of 1.875° per second per eye (°/second·eye), which was approximately matched to reported peak sensitivity for a velocity-based stimulus presented at eccentricities between 3° and 7° (Czuba et al., 2010). In future experiments, it would be desirable to measure speed sensitivity for these exact stimuli, but for the purposes of this article, we chose speeds that were near the previously measured peaks.

Velocity-based stimulus: Manipulating coherence

To measure how sensitivity changes as a function of stimulus viewing duration (or spatial extent) we needed to be able to measure changes in accuracy across a
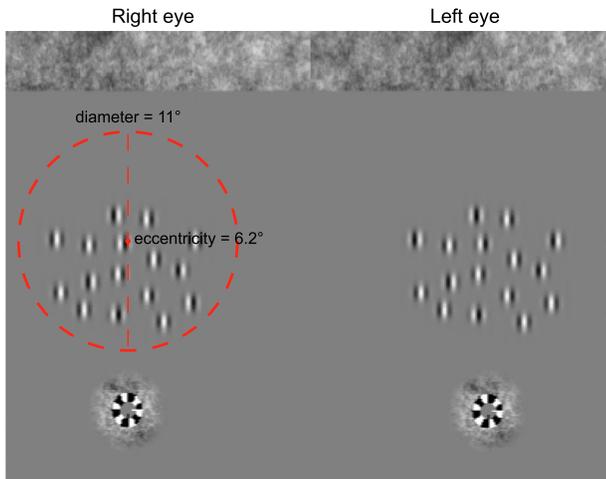
Figure 2. Example frames from the stimuli used for Experiment 1. Each Gabor patch was 0.5° wide (FWHM) and had a spatial frequency of 2 cycles/°. For visualization, the Michelson contrast has been increased from 25% (presented) to 100%. The patches were spaced such that there was a minimum distance of 1° between them. The red circle (not shown in the actual stimulus) illustrates the area within which patch positions were presented across all trials. The circle is centered at 6.2° eccentricity and is 11° in diameter. At the top of each frame, there is a rectangular pink noise texture as a zero disparity reference and boundary indicator. At the bottom of each frame is a fixation circle to aid in binocular alignment and provide a zero disparity reference near fixation, helping to minimize instability.

range of stimulus values. In pilot tests, we noticed that direction discrimination quickly saturated at perfect levels for the velocity-based stimulus, so we sought a way to effectively decrease the motion strength. To do this, we titrated the motion coherence of the stimulus as follows until we felt that we had enough "headroom" in performance to see any differences across conditions.

To implement the coherence manipulation, we designated a number of elements as signal every frame and assigned each one a distinct "date of birth" to ensure at least one informative Gabor phase step occurred every 17 ms. For example, in a subset of four elements, each would be assigned a date of birth indicating how far it was from completing the four-frame update. From one frame to the next, the element that was assigned the third DOB would complete the quarter-wavelength phase update on the next frame, whereas the rest of the elements were designated as noise over that stimulus update and would simply counter-phase flicker. Thus, our signal elements would provide an informative velocity drift, and the noise elements would provide neither an informative velocity component nor an unambiguous disparity signal. To prevent strategic identification of signal locations and/or perceptual "pop-out" effects, we

pseudo-randomly reassigning which subset of elements were designated as signal or noise on every frame. All subjects ran the velocity-based stimulus at the same coherence level (6%) for Experiments 1 and 2.

Disparity-based stimulus

Correspondingly, we developed a novel disparity-based stimulus that was matched with the velocity-based stimulus along as many stimulus dimensions as possible, while ensuring that the only coherent source of 3D direction information was from changes in binocular disparity—and not the interocular relationship between monocular velocities. This stimulus comprised an identical arrangement of 16 Gabor patches (0.5° wide (FWHM), 25% Michelson contrast, SF = 2 cycles/°) with the same spacing algorithm that was used to set patch positions and maintain a minimum distance between elements of the velocity-based stimulus (see Figure 2). Although matched along these dimensions, the disparity-based stimulus necessarily differed in the phase-update logic. To disrupt the interocular velocity cue, we randomly set the baseline Gabor phase on each frame while the interocular phase difference was gradually incremented, producing pure changes in binocular disparity with no net 3D direction from differences between monocular velocities (Figure 3). The speed of the disparity-based stimulus was also matched to previously measured speed sensitivity for disparity-based stimuli from Czuba et al. (2010), with a temporal frequency of 0.938 cycles/second·eye, resulting in a monocular angular speed of 0.469°/second·eye. We chose this value from the reported peak sensitivity in order to give participants the best chance at performing well. The disparity-based stimulus was full coherence for all subjects.

Experiment 1 procedure

We varied the stimulus duration and measured response accuracy during a 3D discrimination task. For the velocity-based condition, trials ranged in duration from 17 to 750 ms. Informed by pilot data, we decided to sample more from the briefest stimulus durations. The full vector of durations used for the velocity-based conditions was: [0.017, 0.033, 0.050, 0.067, 0.083, 0.100, 0.133, 0.167, 0.200, 0.233, 0.300, 0.367, 0.433, 0.550, 0.650, 0.750] (quantized by the 120 Hz refresh rate of the display). For the disparity condition, durations ranged from 67 to 1067 ms and were linearly spaced in 67-ms increments. The extended range of durations for the disparity-based condition was a result of pilot data indicating that some participants' performance continued to improve over long durations. Each trial began with a 500-ms settling period containing the
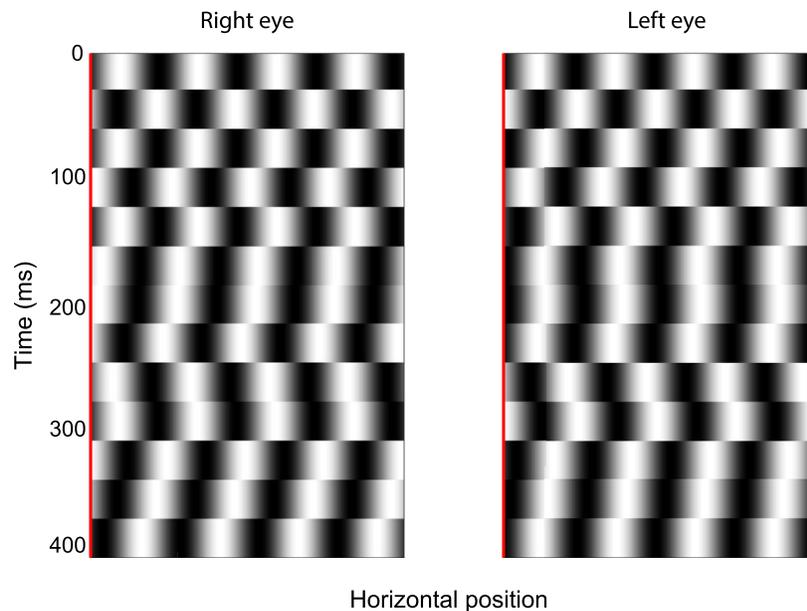
Figure 3. Schematic of the logic underlying the pure disparity-based stimulus. The axes are the same as in Figure 1, with the *x*-axis representing position and the *y*-axis representing time. A red reference line has again been added as an aid to clarify phase changes. In this case, shuffling the baseline phase every four frames (67 ms) results in a lack of a coherent velocity signal over time. Free-fusing, however, reveals a coherent gradient of disparity caused by gradually incrementing the interocular phase difference between gratings by 22.5° every four frames. As described elsewhere in this article, this logic was applied to multiple Gabor elements and presented using the same procedure as the velocity-based stimulus.

fixation and pink noise borders, which were included to aid fixational maintenance and alignment. On every trial, observers were then presented either a velocity- or disparity-based stimulus moving toward or away in depth for a randomly selected stimulus duration from the set above. Observers then reported the direction of the motion via a keypress. Audio feedback was provided to identify correct or incorrect responses. Following the response, there was a 500-ms intertrial-interval—with the fixation and pink noise borders remaining on screen—before the next stimulus was presented.

The experiment was completed in 10 runs for each condition (velocity- and disparity-based). Each run consisted of 320 pseudo-randomly interleaved trials (a total of 10 repetitions of each combination of 16 stimulus durations [sampled from 17 ms to 1066 ms] and two motion directions [towards and away]). This resulted in a total of 3200 trials per participant per condition. Figure 4 shows a visual depiction of a trial.

Spatial integration stimuli

The stimuli used in Experiment 2 followed the same cue-isolating algorithm as described in Experiment 1. A spatial manipulation was introduced by varying the angular subtense of the sector (α) that contained Gabor elements. We modified our spacing algorithm to sample
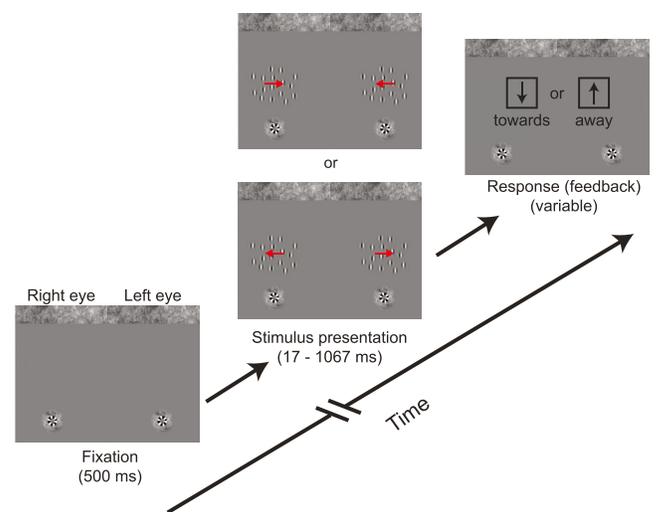
Figure 4. Depiction of a single trial for the temporal manipulation. The fixation and pink noise borders were presented for 500 ms prior to trial onset. The stimulus was presented for a variable duration ranging from 17 to 1067 ms. The participant was then required to respond towards or away via keypress (down arrow for toward, up arrow for away). Audio feedback was provided and then a 500 ms fixation period occurred before the next trial presentation.

possible element locations only within a limited angle of the sector for a given trial. This approach approximated the sector of varying angle used by Burr, Concetta
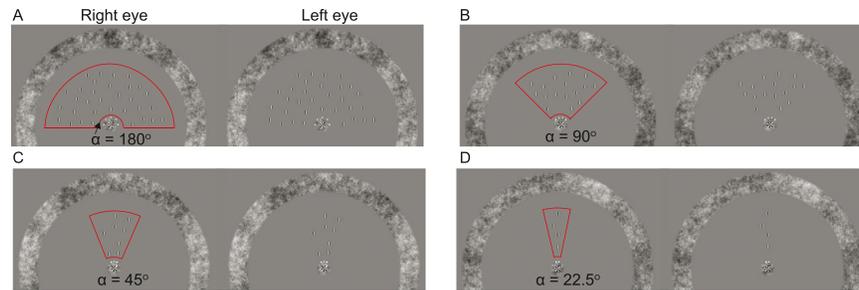
Figure 5. Left/right eye image pairs from example frames from Experiment 2. Each panel shows a different sector size: (A) $\alpha = 180°$; (B) $\alpha = 90°$; (C) $\alpha = 45°$; or (D) $\alpha = 22.5°$. Red lines depicting the sector size have been added as a visual aid and were not shown in the actual stimulus. Note the visual indicator for spatial location and angle size (white dots above the fixation), as well as the increased number of elements to fill the larger sectors.

Morrone, and Vaina (1998). Figure 5 shows stimulus frames for each of the four angles (illustrated by the red lines added to the image frames) at one of the spatial locations presented. To maintain a constant element density we increased the number of Gabor elements for larger-angle sectors. These values are as follows: 22.5° (4 Gabor elements), 45° (8 elements), 90° (16 elements), and 180° (32 elements). In addition to these stimulus sizes, we presented the stimulus randomly in one of four cardinal locations relative to fixation (right, left, up, and down) on every trial. To avoid spatial attention issues, we cued the subject to indicate the stimulus quadrant, as well as the size of the angle it would occupy. We did so by providing subjects with a series of small bright dots adjacent to fixation in the quadrant of presentation and taking up the same angle as the stimulus (see Figure 5). We also added a circular pink noise border to provide a zero-disparity reference surrounding the entire stimulus.

Experiment 2 procedure

The trial structure was identical to that of Experiment 1 (depicted in Figure 4). Participants viewed either velocity- or disparity-based stimuli and were asked to report whether the motion was moving toward or away in depth using a key press. For the velocity condition, stimulus duration was fixed to the 75% correct value of the fitted curve for each participant from Experiment 1 (subject 1: 157 ms, subject 2: 60 ms, subject 3: 400 ms). For the disparity condition, this value was not always straightforward to estimate, because maximal levels of performance for some observers remained below the 75% threshold. In these cases, the stimulus was presented for a duration of 500 ms (which achieved the desired range of performance between chance and perfect as we manipulated spatial extent of the stimulus). We presented four different angles of stimulus wedges, with the number of Gabor elements increasing

with the size of the sectors (Figure 5). As described elsewhere in this article, we also presented in four spatial locations around fixation. Each run consisted of five repetitions of toward and away presentations at each combination of angle and spatial location, resulting in 160 trials per run. Subjects completed 10 runs for 1,600 trials per condition.

Data analysis

The data and the MATLAB scripts used for analysis are available at: https://github.com/jwhritner/SpatioTemporalIntegration3D. We concatenated responses from all experimental runs for each subject and then computed the percent correct per stimulus duration for each condition. We fitted saturating exponential curves to the data where percent correct, *F*, after *t* milliseconds is expressed by:

$$F(t) = ae^{-x/\tau} + c \qquad (1)$$

where τ describes the time to saturation. Aggregate data were analysed in the same way, with percent correct computed on the concatenated vector of all trials per duration for each condition. The coherence implementation described elsewhere in this article for the velocity-based stimulus allowed us to avoid ceiling effects and fit exponential curves to these data. However, we were not able to perfectly match the maximal accuracy levels of the two cues. The fitting method described here is robust to the different levels of accuracy achieved.

We performed 10,000 bootstrap replications for both the individual and the aggregate data by pulling with replacement from the responses per duration and condition. Error bars represent 95% confidence intervals (CI) from these simulations. Separate simulations were performed to bootstrap the distribution of fitted τ values for the aggregate data.

Results

Experiment 1: Temporal integration

We measured accuracy in a 3D direction discrimination task as a function of viewing duration spanning time frames that captured both short time-scale integration done by sensory mechanisms and longer time-scale integration thought to be mediated by decision processes. We developed and used novel cue-isolating stimuli which allowed us to compare how velocity- and disparity-based cues for 3D motion direction might be integrated differently over time. Consistent with our hypothesis, we found that the velocity-based cue was integrated rapidly, with the majority of improvement in sensitivity occurring over the first 200 ms of stimulus presentation. The disparity-based cue was integrated more gradually, showing improvements up to 500 ms.

Figure 6 shows the main results, with 3D direction discrimination accuracy (percent correct, towards versus away; *y*-axis) as a function of viewing duration (*x*-axis) for the velocity- and disparity-based cues (orange and blue points and curves, respectively). Averaging over all trials from all subjects, visual inspection identifies clear differences in how accuracy

depends on duration between the two cue types. Sensitivity for the velocity-based cue saturates quickly (by 200 ms), whereas sensitivity for the disparity-based cue continues to improve gradually over time. To quantify the timescales of integration, we fit a saturating exponential function to the accuracy-versus-duration data, where the time constant τ serves as a standard univariate metric of time to saturation. This analysis confirmed what is apparent in the graph, with accuracy depending on the isolated velocity-based cue reflecting integration approximately twice as fast as that for the disparity-based cue (velocity: $\tau = 0.079$, 95% CI [.070, .090]; disparity: $\tau = 0.161$, 95% CI [.122, .200]). Error bars represent bootstrapped 95% CIs from 10,000 replications (with replacement) from the aggregate response data. There are, however, large individual differences.

Experiment 2: Spatial integration

Spatial integration for many forms of motion has been demonstrated psychophysically to occur over extended regions of the visual field. The extent of spatial summation has been shown to be dependent on the type and complexity of motion, and is indicative of the corresponding fundamental unit of detection in cortex (i.e., receptive field size; Anderson & Burr, 1987; Burr et al., 1998; Fredericksen, Verstraten, & Van De Grind, 1994). Consistent with the moderately large receptive fields measured in the middle temporal visual area (MT) (e.g., widths roughly equal to eccentricity in the visual field; Felleman & Kaas, 1984), psychophysical results have demonstrated that spatial summation for many forms of motion occur over correspondingly large areas of space (Komatsu & Wurtz, 1988; Mikami, Newsome, & Wurtz, 1986). Although electrophysiology studies have shown that the majority of neurons in macaque area MT exhibit directionally selective responses to binocular 3D motions (Czuba, Huk, Cormack, & Kohn, 2014; Sanada & DeAngelis, 2014), prior psychophysical efforts to examine spatial summation of 3D motion using binocular cue isolating stimuli have been complicated by an inability to truly isolate the velocity-based cue (e.g., Brooks & Stone, 2006). We, therefore, examined spatial integration of 3D direction discrimination by manipulating the size of the stimulus area of the cue-isolating stimuli introduced in Experiment 1. We hypothesized that spatial integration profiles of cue-isolating stimuli would exhibit similar asymptotic increases in discrimination performance with increasing stimulus size, plateauing as stimulus area approached that of MT receptive fields; owing to a common neural basis for both binocular cues (DeAngelis, Cumming, & Newsome, 1998; Joo et al., 2019; Sanada & DeAngelis, 2014).

Figure 6. A 3D direction discrimination performance as a function of the viewing duration for the aggregate data collected in Experiment 1. Performance for each of the two conditions is plotted as percent correct (*y*-axis) at 16 stimulus durations (*x*-axis). Each of the four subjects completed 3,200 trials per condition (200 trials per duration). Note that performance in the velocity-based condition (orange) saturates quickly while the disparity-based condition (blue) is more gradual. Error bars represent bootstrapped 95% CIs. The solid lines are fitted saturating exponential curves.

Figure 7. The 3D direction discrimination performance (percent correct, *y*-axis) as a function of the sector angle α (size, *x*-axis) containing Gabor elements. These data are the aggregate of all trials run by three subjects. Note that performance for both velocity- (orange) and disparity-based (blue) conditions increases with increasing α values. Error bars are bootstrapped 95% CIs.

Aggregate results for three subjects are plotted for each condition in Figure 7 as percent correct as a function of stimulus wedge angle. For both conditions, the results show a clear effect of angle, with percent correct increasing with larger angles and approaching saturation by 180°. These results are consistent with earlier findings suggesting that spatial receptive fields used for 3D direction discrimination are approximately MT-sized (i.e., bigger than an octant and smaller than a quadrant; Burr et al., 1998; Maunsell & van Essen, 1983; Van Essen, Maunsell, & Bixby, 1981). There is a small difference between cue conditions, with the velocity-based condition exhibiting a slightly steeper slope of improvement between 22.5° and 45.0°, but ultimately plateauing at the same accuracy as the disparity-based condition.

Individual differences enrich the interpretation of differences in velocity- and disparity-based integration

In addition to the robust effect found in the aggregate data, variation in overall performance between subjects was substantial, particularly in the disparity-based

condition. Figure 8 shows the data for each individual subject in Figures 8A to D. All subjects performed well overall on the velocity-based condition, all four subjects achieved asymptotic performance at or above 80% correct within 750 ms. Three of four subjects exhibited similarly fast integration profiles as seen in the aggregate; Subject 3 was a notable outlier.

There is greater between-subject variability in the disparity-based condition. Subjects 3 and 4 show higher sensitivity to the disparity-based cue, with performance saturating very quickly compared to Subjects 1 and 2. Subject 4's performance improves with duration at the same rate for both cues, and asymptotes to 100% for the disparity-based cue (interestingly, this subject is also an extremely experienced stereo observer with very good stereoacuity; Stevenson, Cormack, & Schor, 1989). These individual differences have been cited in earlier results showing variability in performance on 3D motion estimation tasks (Harris & Dean, 2003; Nefs et al., 2010; Rokers, Fulvio, Pillow, & Cooper, 2018).

Analyses of individual differences for spatial integration were mostly limited to differences in overall performance. Figure 9 shows the individual data plotted as percent correct as a function of the angle (α) of the sector containing stimulus gratings. All subjects show increased performance from 45° to 90° for the disparity-based condition (blue). For the velocity-based condition, Subject 3 does not see an increase in performance, whereas the other two subjects do.

There are striking individual differences in the two experiments presented here, both in terms of overall accuracy for different cues, and also in terms of the relationship between accuracy and stimulus duration. These results are consistent with other studies finding individual differences in 3D motion perception (Harris & Dean, 2003; Nefs et al., 2010; Rokers et al., 2018). We have minimized the possibility of variability across participants using a fixation that helps to ensure binocular alignment and prevent vergence drift. We cannot rule out the possibility that the individual differences here and reported by others result from sensory differences in sensitivity. We will discuss elsewhere in this article that there is another (not mutually exclusive) possibility that we believe could be playing a role: strategic integration differences.

Discussion

Using novel stimuli that completely isolated velocity- and disparity-based cues without large "nuisance" differences in spatiotemporal content (and appearance), we characterized the temporal and spatial integration profiles of velocity- versus disparity-based 3D motion perception. We did this by assessing how accuracy depended on either increased time (duration) or space
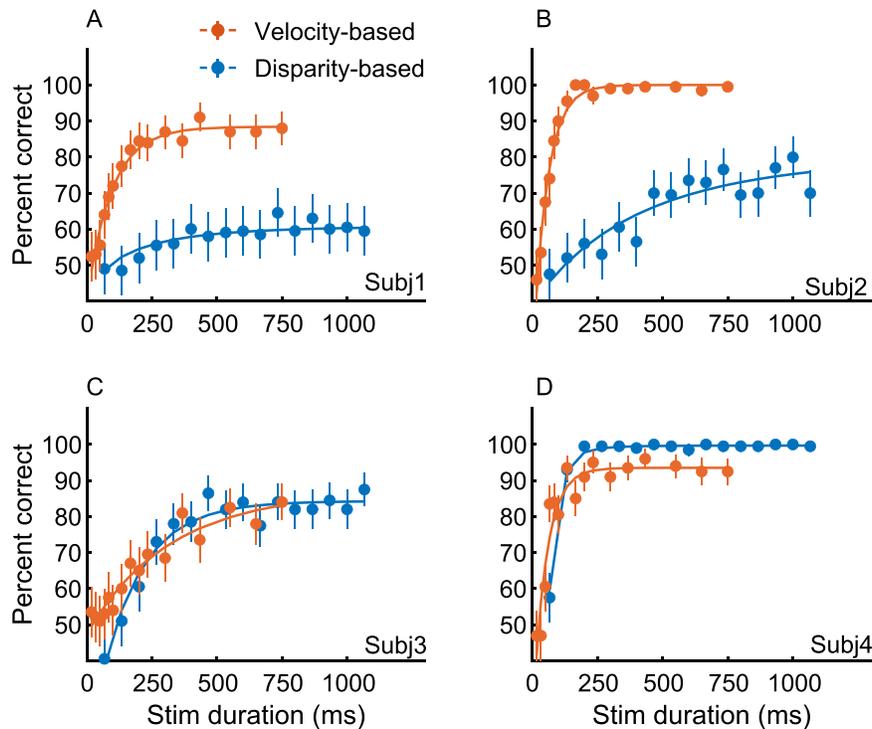
Figure 8. (A-D) Temporal integration data for four individual subjects. As in the aggregate above, the data are plotted as percent correct (*y*-axis) at 16 stimulus durations (*x*-axis). Note that despite some variation in overall performance, three of the four subjects show fast integration profiles for the velocity-based condition as seen in the aggregate. The disparity-based conditions shows more variability. Error bars are bootstrapped 95% CIs from each subject's data.
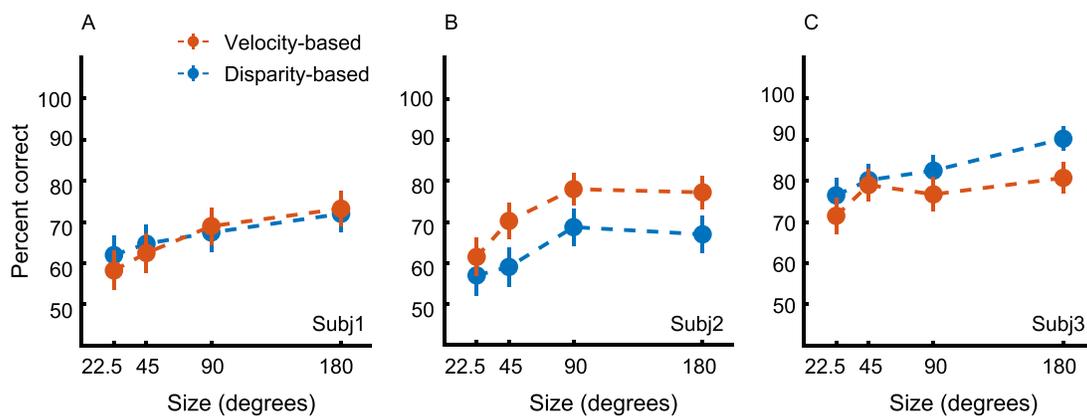
Figure 9. (A-C) Spatial integration data for three individual subjects. These are plotted as percent correct (*y*-axis) for each of the four stimulus angles (α) presented (size, *x*-axis). There are some differences in overall performance between subjects but the overall trends for both the velocity- and disparity-based conditions are similar, with performance increasing most as α increases from 45° to 90°. Error bars are bootstrapped 95% CIs from each subject's data.

(size) in a coarse 3D motion direction discrimination task (towards vs. away). Our hypothesis, supported by earlier results showing that these mechanisms have different speed sensitivity (Brooks & Stone, 2004; Czuba et al., 2010), was that velocity-based sensitivity would be dominated by a fast integration regime (consistent with classical sensory integration)—while the disparity-based mechanism would primarily reflect slower, statistical integration. In light of recent physiological evidence suggesting MT as a common site for the processing of velocity and disparity cues (Czuba et al., 2014; Joo, Czuba, Cormack, & Huk, 2016; Sanada & DeAngelis, 2014), we predicted that spatial integration would be similar for the two stimuli,

showing the steepest dependence on stimulus size up to the scale of MT receptive fields.

The data support both temporal and spatial integration hypotheses. Results presented in Experiment 1 indicate the velocity-based cue is integrated quickly, on a conventional time scale associated with the integration by motion-selective sensory neurons (<200 ms) (Bair & Movshon, 2004; Katz et al., 2015). In contrast, disparity-based information is integrated over a longer time frame for some observers, but less efficiently, consistent with a more deliberative, statistical accumulation of evidence. In Experiment 2, we show that spatial integration for both cues is similar, plateauing with sector angles greater than 90°. These data resemble those reported by Burr et al. (1998), who found spatial summation improved over larger areas, suggesting integration is performed by neurons with large receptive fields—such as those found in MT and the medial superior temporal area (MST). Our preliminary psychophysical results suggest, that performance on a depth motion discrimination task may not increase with stimulus areas greater than 180°. A full spatial experiment paired with electrophysiology would be an interesting follow-up and could help to reveal possible differences between binocular motion mechanisms that the coarse psychophysical manipulation could not.

The individual differences observed in these data may be due to differences in sensitivity. We wish to speculate that they may also be due to different integrative strategies (some of which may be compensation for sensory sensitivity). Our observers were motivated to perform well (and they were given feedback, so they were crudely aware of their performance after a few trials). Given this motivation, if they were not confident in a judgement based on "immediate" sensory information, then they tarried a bit to benefit from probability summation. Different subjects with different degrees of perceptual sensitivity to one or the other cue might further rely on more or less decision-stage (statistical) integration. Such a mixture of sensory sensitivity and compensatory decision-stage integration may have muddled the individual differences landscape in prior work. Our quantification of the temporal dependence of sensitivity allowed us to at least tentatively tease apart these two rather distinct contributions to overall performance. The subjective impressions of the participants were interesting. One observer found the disparity-based task to be easy, but subjectively had trouble seeing the velocity-based motion clearly despite his numerical performance being fine. Two other subjects had the opposite experience, finding the velocity-based stimuli to be subjectively salient and the disparity-based stimuli to be unclear. These experiences speak to the diversity in individual experiences of 3D motion. A larger scale study–enabled by the cue-isolating stimuli presented in this article–may be able to provide a more detailed, comprehensive model of these individual differences.

Together, these results suggest that 3D motion perception may depend not only on the cue(s) principally available or relied upon, but also on the temporal and spatial structure of the task at hand. In prior work, we have shown that full-cue 3D motion (i.e., containing both velocity- and disparity-based information) is integrated in multiple stages—with the early sensory stage persisting for about 150 ms before transitioning to a decision stage (Katz et al., 2015). Analyses of temporal integration profiles for cue-isolating stimuli suggest that subjects were likely relying on velocity-based cues present in the full-cue 3D stimulus to perform initial/rapid components of the 3D direction discrimination task. The psychophysical data provided here further support the interpretation that this component likely supports fast decisions related to the direction of 3D motion, such as whether or not an approaching object will collide with the head—which captures attention automatically (Lin, Murray, & Boynton, 2009) and requires an immediate response. Although a fast integration regime is beneficial for this kind of task, one could easily imagine other scenarios where timing is less important and where fine-grained depth estimates could benefit from more gradual or prolonged integration. In these circumstances, the disparity-based component could help to refine estimates over longer time periods, perhaps relying on feedback from higher visual and cognitive areas (although, obviously, one of our subjects was able to tap into the disparity signal very rapidly).

Conclusions

A major contribution of this article is introducing a class of stimuli suitable for psychophysics that isolate binocular 3D motion cues (both of which are represented in MT) while remaining matched along other principle stimulus dimensions. These stimuli make it possible to further characterize sensory- and decision-related contributions to 3D motion integration using a true cue-isolating paradigm. In Experiment 1, we measured temporal integration for 3D direction judgments and found that performance saturated quickly for the velocity-based cue and more gradually for the disparity-based cue (on average, but with interesting individual differences). Our results support the general hypothesis that there are separate underlying mechanisms for detecting and integrating these cues to perform 3D motion tasks. They also support the hypothesis that the velocity-based mechanism is relatively fast, and likely tapped during brief judgements of 3D motion direction and speed.

The disparity-based judgments were much slower in one-half of our observers, perhaps indicative of a higher decision stage at work. Experiment 2 provides an example of how these stimuli can be extended to test spatial integration as well. These data are roughly indicative of integration areas that are of the same scale as MT receptive fields. Future experiments paired with electrophysiology or an image-computable model of area MT could help to elucidate subtler spatial differences between the two mechanisms.

Keywords: binocular vision, 3D motion, psychophysics, spatiotemporal integration, motion perception

Acknowledgments

Corresponding author: Jake A. Whritner.
Email: jake.whritner@utexas.edu.
Address: Center for Perceptual Systems, Department of Psychology, The University of Texas at Austin, Austin, TX 78712, USA.

Footnote

[1]Throughout this article, we leave behind the commonly used acronyms IOVD and CD. The brain does not encode interocular velocity "differences" per se. Rather, it builds 3D tuning curves from a binocular combination of imbalanced monocular velocity signals that do not reflect velocity *differences* in the literal arithmetic sense. For this reason, we prefer to use "velocity- and disparity-based" as our core terminology, and hope others working on this topic will share this sensibility. See Bonnen et al. (2020) for more details.

References

Anderson, S. J., & Burr, D. C. (1987). Receptive field size of human motion detection units. *Vision Research, 27*(4), 621–635.

Bair, W., & Movshon, J. A. (2004). Adaptive temporal integration of motion in direction-selective neurons in Macaque visual cortex. *Journal of Neuroscience, 24*(33), 7305–7323.

Bloch, A. M. (1885). Experiences sur la vision. *Comptes Rendus de la Societé de Biologie, 37*, 493.

Bonnen, K. L., Czuba, T. B., Whritner, J. A., Kohn, A., Huk, A. C., & Cormack, L. K. (2020). Binocular viewing geometry shapes the neural representation of the dynamic three-dimensional environment. *Nature Neuroscience*.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436.

Brooks, K. R., & Stone, L. S. (2004). Stereomotion speed perception: contributions from both changing disparity and interocular velocity difference over a range of relative disparities. *Journal of Vision, 4*(12), 1061–1079.

Brooks, K. R., & Stone, L. S. (2006). Stereomotion suppression and the perception of speed: Accuracy and precision as a function of 3d trajectory. *Journal of Vision, 6*(11), 6–6.

Burr, D. C., Concetta Morrone, M., & Vaina, L. M. (1998). Large receptive fields for optic flow detection in humans. *Vision Research, 38*(12), 1731–1743.

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

Czuba, T. B., Huk, A. C., Cormack, L. K., & Kohn, A. (2014). Area MT encodes three-dimensional motion. *Journal of Neuroscience, 34*(47), 15522–15533.

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.

DeAngelis, G. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature, 394*(August), 677–680.

Felleman, D. J., & Kaas, J. H. (1984). Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. *Journal of Neurophysiology, 52*(3), 488–513.

Fredericksen, R., Verstraten, F., & Van De Grind, W. (1994). Spatial summation and its interaction with the temporal integration mechanism in human motion perception. *Vision Research, 34*(23), 3171–3188.

Harris, J. M., & Dean, P. J. A. (2003). Accuracy and precision of binocular 3-D motion perception. *Journal of Experimental Psychology: Human Perception and Performance, 29*(5), 869–881.

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

Harris, J. M., & Watamaniuk, S. N. J. (1995). Speed discrimination of motion-in-depth using binocular cues. *Vision Research, 35*(7), 885–896.

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. *Journal of Neuroscience, 36*(42), 10791–10802.

Joo, S. J., Greer, D. A., Cormack, L. K., & Huk, A. C. (2019). Eye-specific pattern-motion signals support the perception of three-dimensional motion. *Journal of Vision, 19*(4), 27.

Katz, L. N., Hennig, J. A., Cormack, L. K., & Huk, A. C. (2015). A distinct mechanism of temporal integration for motion through depth. *Journal of Neuroscience, 35*(28), 10212–10216.

Kenny, S. (2020). *Advanced vision research paradigms with the PROPixx high refresh rate projector* (V-SSS 2020) [Video file]. Retrieved from https://vpixx.com/vocal/advanced-vision-research-vss-2020/.

Komatsu, H., & Wurtz, R. H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *Journal of Neurophysiology, 60*(2), 580–603.

Lin, J. Y., Murray, S. O., & Boynton, G. M. (2009). Capture of attention to threatening stimuli without perceptual awareness. *Current Biology, 19*(13), 1118–1122.

Maloney, R. T., Kaestner, M., Bruce, A., Bloj, M., Harris, J. M., & Wade, A. R. (2018). Sensitivity to velocity- and disparity-based cues to motion-in-depth with and without spared stereopsis in binocular visual impairment. *Investigative Opthalmology & Visual Science, 59*(11), 4375.

Maunsell, J. H. R., & van Essen, D. C. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience, 3*(12), 2563–2586.

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. *Journal of Neurophysiology, 55*(6), 1328–1339.

Nefs, H. T., O'Hare, L., & Harris, J. M. (2010). Two independent mechanisms for motion-in-depth

perception: Evidence from individual differences. *Frontiers in Psychology, 1*(OCT), 1–8.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision, 5*(5), 1, https://doi.org/10.1167/5.5.1.

Rokers, B., Czuba, T. B., Cormack, L. K., & Huk, A. C. (2011). Motion processing with two eyes in three dimensions. *Journal of Vision, 11*(2), 1–19.

Rokers, B., Fulvio, J. M., Pillow, J., & Cooper, E. A. (2018). Systematic misperceptions of 3D motion explained by Bayesian inference. *Journal of Vision, 18*(3), 1–23.

Sanada, T. M., & DeAngelis, G. C. (2014). Neural representation of motion-in-depth in area MT. *Journal of Neuroscience, 34*(47), 15508–15521.

Sheliga, B. M., Quaia, C., FitzGibbon, E. J., & Cumming, B. G. (2016). Human short-latency ocular vergence responses produced by interocular velocity differences. *Journal of Vision, 16*(10), 11.

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, Optics, Image Science, and Vision, 25*(7), 1574–1585.

Stevenson, S. B., Cormack, L. K., & Schor, C. M. (1989). Hyperacuity, superresolution and gap resolution in human stereopsis. *Vision Research, 29*(11), 1597–1605.

Van Essen, D. C., Maunsell, J. H. R., & Bixby, J. L. (1981). The middle temporal visual area in the macaque: Myeloarchitecture, connections, functional properties and topographic organization. *Journal of Comparative Neurology, 199*(3), 293–326.

Watson, A. B. (1986). Temporal Sensitivity. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance 3D* (Vol. 1). New York, NY: John Wiley & Sons.