Feature-based attention promotes biological motion recognition

Sarah C. Tyler

Department of Cognitive Sciences,
Center for Cognitive Neuroscience, University of California,
Irvine, Irvine, CA, USA

Emily D. Grossman

Department of Cognitive Sciences,
Center for Cognitive Neuroscience, University of California,
Irvine, Irvine, CA, USA

Motion perception is important for visually segregating and identifying objects from their surroundings, but in some cases extracting motion cues can be taxing to the human attention system. We measured the strength of feature salience required for individuals to correctly judge three types of moving events: biological motion, coherent motion, and multiple object tracking. The motion animations were embedded within a larger Gabor grid and constructed such that motion was conveyed by a salient single-feature dimension (second order) or by alternating across equisalient feature dimensions (third order). In the single-feature displays, we found biological motion to require less difference in the Gabor features (relative to background) to yield equivalent task performance as the coherent motion or multiple object tracking. This main effect of feature magnitude may reflect the inherent salience of biological motion as a visual stimulus. In the alternating-feature displays, both the biological motion and coherent motion discriminations needed additional salience, as compared to the single-feature displays, to achieve threshold discrimination levels. Accuracy in the multiple object tracking task did not vary as a function of salience. Together, these findings demonstrate the effectiveness with which attention-based motion mechanisms operate in complex dynamic sequences and argue for a critical role of feature-based attention in promoting biological motion perception.

Keywords: motion—2D, biological motion, attention, perceptual organization


Introduction

Humans are naturally drawn to the rich visual kinematics of body movements, with the ability to intuit an abundance of social-communicative information from brief exposures to the actions of others. In vision science, this ability has historically been investigated using point-light biological motion animations in which actions are conveyed by the kinematics of the joints, portrayed as small dots on the body (Blake & Shiffrar, 2007; Johansson, 1973). Point-light animations are readily recognized by typical observers and offer the benefit of precise control over the superficial aspects of the body (faces, clothing, etc.) while retaining the essential features of actions.

What those essential features are and the mechanisms by which they are perceived, however, is poorly understood. Body kinematics are an obvious candidate for key features, born out by much empirical evidence. Biological motion is most effectively masked by dynamic noise matched for the space–time dynamics found in point-light sequences (Bertenthal & Pinto, 1994; Hiris, Humphrey, & Stout, 2005) and is more difficult to discriminate when the body parts that move most are eliminated (e.g., the extremities) or when the temporal dynamics are disrupted (Mather, Radford, & West, 1992; Pollick, Fidopiastis, & Braden, 2001). Biological motion perception also fails at isoluminance conditions that render local motion signals very difficult to perceive (Garcia & Grossman, 2008). Other studies have shown that gender discrimination of point-light actors is largely accounted for by the variance in kinematics across actors (Troje, 2002), and point-light animations that lack an underlying human body structure, but retain key dynamic features associated with biological motion, are spontaneously reported as biological (Casile & Giese, 2005).

Motion analysis, however, fails to account for other important behavioral findings, such as the orientation effect in biological motion recognition. Point-light sequences take longer to recognize and are much more difficult to discriminate when shown upside down, even though the kinematics are intact (Pavlova & Sokolov, 2000; Sumi, 1984). Moreover, point-light displays constructed with limited lifetime dots or with joint positions that are jittered from frame to frame are both readily discriminated, despite the distorted local motion trajectories (Beintema, Georg, & Lappe, 2006; Beintema & Lappe, 2002; Neri, Morrone, & Burr, 1998). In addition, interrupting the dynamics of...
biological motion with extended inter-frame intervals has little effect on the ability to discriminate the walking direction of the target, particularly if observers are shown long durations of the gait cycle (Thornton, Pinto, & Shiffrar, 1998). Evidence along these lines has led many researchers to conclude that biological motion recognition is an inherently form-based process, with actions constructed by temporally integrating sequences of stationary body postures (Lange, Georg, & Lappe, 2006).

Emerging evidence suggests that this debate is ill-posed, despite having sparked ambitious experimentation. Even with the inefficiencies apparent in biological motion recognition (e.g., Gold, Tadin, Cook, & Blake, 2008), studies show that observers shift flexibly between motion- and form-based strategies, depending on the nature of the potent information available. When only brief snapshots of actions are available, observers tend to rely more heavily on form-based information apparent in the displays, such as global body structure (Lu & Liu, 2006; Thirkettle, Benton, & Scott-Samuel, 2009). When viewing longer sequences, such as those seen in more naturalistic viewing conditions, observers rely more critically on dynamic features (Thurman, Giese, & Grossman, 2010; Thurman & Grossman, 2008). Thus, knowingly or not, observers are (wisely) able to extract the essential components of biological motion that are most salient in their current context.

This flexible strategy would appear to implicate a third important mechanism for biological motion perception that has been largely ignored in the ongoing debate: attention. This omission in the current literature is likely the consequence of relatively little direct investigation, with only a handful of studies directly targeting attentive mechanisms in biological motion perception. From these studies, however, we know that observers’ ability to discriminate biological motion is highly correlated to performance on Stroop tasks, linking recognition to individual differences in attentive selection (Chandrasekaran, Turner, Bulthoff, & Thornton, 2010). Biological motion recognition also suffers in dual-task paradigms and in conditions of visual crowding, evidence of capacity limitations, and limited acuity, respectively (Thornton, Rensink, & Shiffrar, 2002; Thornton & Vuong, 2004). Lastly, visual search through biological motion arrays proceeds serially at a cost of 100–200 ms per item, a quantitative estimate of the cost in analyzing features that are clearly not preattentive (Cavanagh, Labianca, & Thornton, 2001).

In this study, we examine the role of attention-based motion, or motion perceived via feature-based attention, in promoting biological motion perception. To do this, we created biological motion animations with local motion cues to which passive motion mechanisms are blind. To quantify the effectiveness of these animations in conveying biological motion, we compare the salience of local features required for equivalent performance when motion is apparent to low-level versus attentive mechanisms. To make this distinction clear, consider that at the earliest levels of analysis the visual system is equipped with an array of motion detectors sensitive to space–time shifts of edges and surfaces, and other visual features. These first- and second-order motion detectors are differentiated on the basis of the type of features compared (i.e., first-order motion is defined by changes in luminance, whereas second-order motion is defined by changes in texture, such as contrast, color, or disparity; e.g., Sperling, 1989) and are passive in the sense that they operate on relatively unprocessed inputs in early levels of analysis and are largely preattentive (Cavanagh & Mather, 1989). We know that biological motion perception is invariant to these low-level motion mechanisms. Traditional point-light sequences with dark dots against a light background are readily detected by first-order motion detectors, while second-order sequences that are invisible to first-order motion detectors are also readily recognized and discriminated (Ahlström, Blake, & Ahlström, 1997).

A third class of visual analyses (and now perhaps a fourth class; Blaser & Sperling, 2008) has been proposed to characterize a host of other motion phenomena that cannot be captured by these first- and second-order mechanisms. These third-order or attention-based motion analyses are argued to use active, top-down attentive selection of features and objects from background elements (Itti & Koch, 2001), can be driven by properties of the visual stimulus (bottom-up) or under guided control (Kastner & Ungerleider, 2000), and, critically, can promote perceived motion in visual displays with no low-level space–time position shifts (Lu & Sperling, 1995). It is worth noting that most instances of movement in natural vision likely reflect the combined workings of low-level motion detectors with these attentive mechanisms.

In these experiments, we seek to isolate the attentive demands of biological motion perception by constructing point-light sequences to which low-level motion mechanisms are blind. To achieve this, we embedded point-light animations within a stationary Gabor grid, with the joints of the actor depicted by highly salient tokens (relative to the background elements). Third-order or alternating-feature point-light animations were constructed by increasing the salience of the tokens across multiple dimensions of feature space (third-order or alternating-feature-defined motion; e.g., Blaser, Pylyshyn, & Holcombe, 2000; Figure 1B). Any perceived motion in this display is the result of perceptually binding the salient features across space, time, and feature dimensions. To ensure that all feature dimensions were readily and equally discriminated from the background elements, we carefully calibrated all four feature dimensions to be equisalient in preliminary measurements using single-feature versions of the displays. With these single-feature measurements as a metric, we were able to quantify salience required to perform the same tasks in the alternating-feature biological motion.
A. The three conditions

As a means for comparison, we obtained the same measurements for two additional tasks in embedded Gabor displays. First, to estimate the salience required to perceive simple local motion energy, without the demands of biological motion, subjects performed a coherent versus incoherent motion discrimination task on embedded random dot kinematograms (single-feature and alternating-feature versions). Second, to measure any inherent differences in token salience or figure–ground segregation between the two types of displays, subjects completed a multiple object tracking task using single-feature and alternating-feature versions (Pylyshyn & Storm, 1988).

We find that observers can recognize and discriminate the point-light displays depicted in alternating-feature space, evidence that the essential features of biological motion are readily analyzed by attentive mechanisms. Higher feature salience was required for these discriminations, however, than predicted from the single-feature measurements. We found a similar pattern for coherent motion task, suggesting that it is the difficulty in perceiving the dynamic features in biological motion that limits performance. Multiple object tracking, in contrast, did not vary as a function of feature salience (beyond a minimum threshold required for target detection). These

B. Single- and alternating-feature tasks

As a means for comparison, we obtained the same measurements for two additional tasks in embedded Gabor displays. First, to estimate the salience required to perceive simple local motion energy, without the demands of biological motion, subjects performed a coherent versus incoherent motion discrimination task on embedded random dot kinematograms (single-feature and alternating-feature versions). Second, to measure any inherent differences in token salience or figure–ground segregation between the two types of displays, subjects completed a multiple object tracking task using single-feature and alternating-feature versions (Pylyshyn & Storm, 1988).

We find that observers can recognize and discriminate the point-light displays depicted in alternating-feature space, evidence that the essential features of biological motion are readily analyzed by attentive mechanisms. Higher feature salience was required for these discriminations, however, than predicted from the single-feature measurements. We found a similar pattern for coherent motion task, suggesting that it is the difficulty in perceiving the dynamic features in biological motion that limits performance. Multiple object tracking, in contrast, did not vary as a function of feature salience (beyond a minimum threshold required for target detection). These
psychophysical results are the first to demonstrate that the biological motion from point-light sequences can be perceptually constructed on the basis of attentive tracking mechanisms alone, without passive motion analyses. These experiments also quantify the inherent salience of point-light sequences and argue for the importance of salience-driven feature-based attention in biological motion perception.

**Experiment 1**

In the first experiment, we compare the feature salience required for observers to perform three motion tasks: biological motion versus motion-matched scrambled sequences, coherent versus incoherent motion, and multiple object tracking. For each of these tasks, the moving sequence was embedded within a larger array of Gabors, with the target tokens differentiated from the background on the basis of increased feature salience for one of the Gabor features (luminance contrast, spatial frequency, orientation, and/or phase). In our first set of measurements, the four feature dimensions were carefully calibrated to be perceptually equivalent, with these equivalent salience levels then used to construct calibrated alternating-feature displays.

**Participants**

Thirty subjects (ages 18–40) with normal or corrected-to-normal vision participated in this experiment. Each subject participated in either the biological motion ($N = 9$), coherent motion ($N = 13$ for the 5%, with 10 of those participants also completing the 10% and 20% conditions), or multiple object tracking task ($N = 8$). All subjects gave informed, written consent approved by the University of California Irvine Institutional Review Board and received either monetary compensation or course credit for their participation in the experiment.

**Materials**

Stimuli for all three tasks were generated by varying features of stationary Gaussian-windowed sine-wave gratings, or Gabors (each 0.6 deg of visual angle), situated within a larger array of Gabors (16.2 deg; Figure 1). All Gabors were identical at the onset of each trial, with a Michelson contrast of 0.244, a spatial frequency of 4.71 cycles/deg, a drift speed of 4.2 cycles/s (3.1 cycles/s for the biological motion condition, due to the slower refresh rate; see below) and were vertically orientated. Target Gabors “carrying” the motion signals were differentiated from the background by increased salience, created by increasing or decreasing the magnitude of the luminance contrast, spatial frequency, tilt, or phase. The Gabors remained stationary within the grid throughout the animations, thus any perceived position shift of local target elements was induced by the sequential changes in relative salience (an apparent motion cue). Using this type of display, we created the following types of sequences.

Point-light biological motion animations depicted one of twenty-five unique human actions, including walking, kicking, and throwing, among others (Grossman & Blake, 1999; Johansson, 1973). Traditional point-light animations are constructed as dot displays with a small number of tokens (twelve in our animations, subtending approximately 7 deg of visual angle) capturing the movements of the joints and head of an actor. In these experiments, the tokens for the joints were replaced by target Gabors with increased salience relative to the background (Movie 1). Because the positions $(x, y)$ of the target Gabors were unchanged throughout the action sequence, the biological motion animations were effectively aliased over space as defined by the resolution of the grid. Nonetheless, even with the coarse resolution, naive subjects had no difficulty recognizing the human actions in highly salient conditions.

Motion-matched “scrambled” biological motion was created by randomizing the starting position of the joints in each biological motion sequence but left the body kinematics intact. These animations appear as meaningless clouds of dots, with some overall flow in common. Aside from the initial scrambling of the starting joint position, these animations were constructed identically to the biological sequences.

Coherent motion animations depicted a unidirectional motion signal in one of eight possible directions (0, 45, 90, 135, 180, 225, 270, 315 and 360 deg) at a speed of 6 deg/s (Movie 2). These animations were constructed by sequentially increasing the salience of adjacent Gabors in a subset of the array so as to induce apparent motion.
On any given frame, a subset of the total 729 Gabors was targeted resulting in approximately 5%, 10%, and 20% of the tokens carrying the directional signal. Each targeted location shifted one Gabor position over the course of 100 ms, at which time a new Gabor was randomly selected, creating a “limited lifetime” display. The lifetime phase was randomized across the targets to desynchronize the resetting of salient target positions and to prevent subjects from discriminating the displays on the basis of a single token.

Incoherent motion was created much like the coherent motion displays but with sequentially selected Gabors randomized across each of the eight possible nearest neighbors. Thus, the incoherent motion displays contained a mix of eight apparent motion directional signals, as opposed to one in the coherent displays. All the other parameters, including speed and “lifetime,” were identical.

Multiple object tracking displays were created by drifting eight tokens (at a speed of 6 deg/s) independently across the display (Movie 3). Three of the eight Gabors were identified by a blue outline for the 1 s prior to the onset of the motion sequence as the “to-be-tracked” tokens. The tokens were then set into motion for 1 s, colliding with each other and bouncing off of the sides of the grid. Immediately following the animation sequence, a blue outline surrounded the final resting position of a single Gabor (tracked or non-tracked), which indicated the Gabor on which the subject was instructed to make their judgment. Previous studies have shown that performance on the multiple object tracking task is mediated, in part, by the proportion of to-be-tracked tokens (out of the total possible) and their velocity and density across the array (which also affects the number of collisions between items; Franconeri, Jonathan, & Scimeca, 2010). In a control experiment (N = 4), we determined that tracking highly salient black tokens on a uniformly gray background (i.e., without the aperture grid) using the same speed and density parameters as in our aperture-grid experiment described above yields tracking performance near 80%.

All stimuli were displayed on a CRT monitor controlled by an Apple Macintosh computer equipped with Matlab 7.5 (Mathworks) and Psychtoolbox 3.0.8 (Brainard, 1997; Pelli, 1997). The biological motion sequences were displayed at a slightly slower screen refresh rate than the coherent motion and multiple object tracking stimuli (75 Hz versus 100 Hz) to most naturally replicate the speed at which the original actions were recorded.

Procedure

Each subject participated in only a single task (biological motion, multiple object tracking, or coherent motion). Subjects made a two-alternative forced choice (2AFC) discrimination on each biological motion trial, indicating if the target Gabors depicted an intact biological or a scrambled sequence. Subjects made a 2AFC discrimination on the coherent motion trials, indicating if the sequence depicted coherent motion (regardless of the direction) or incoherent motion (with the three coherence levels measured in separate blocks). Subjects made a 2AFC judgment on the multiple object tracking trials, indicating whether or not the cued Gabor was the final resting position of a to-be-tracked target. No feedback was provided on the subject performance on any task (Movie 4).

All subjects completed both single-feature and alternating-feature versions of the tasks. The goal of the single-feature task was to estimate equivalent perceptual salience for four key feature dimensions (luminance contrast, spatial frequency, orientation, or drift speed) of the Gabors. This was necessary so that each feature could be rendered with equivalent perceptual salience (adjusted for each observer).
Results

Figure 2 shows the magnitude of feature differences required, for each key feature, to yield 80% discrimination performance, as measured in the single-feature displays. Participants were able to discriminate the biological and scrambled motion, discriminate the coherent and incoherent motion (at all coherence levels), and successfully track target Gabors using any of the four features, albeit with differing levels of salience required for each feature. The adaptive staircase procedure was completed for each feature dimension in separate blocks to yield estimates of equivalent salience for each feature alone.

Subject performance was fitted with a Weibull psychometric function to yield (1) an estimate of 80% threshold performance and (2) the approximate slope surrounding that threshold (ΔT), estimated between 75% and 85% performance. Using this anchor point and range, we then estimated the magnitude for each feature that would induce the theoretical 100%, 120%, 140%, 160%, and 180% subject performance for each key feature alone (denoted as T + ΔT, T + 2ΔT, T + 3ΔT, T + 4ΔT, T + 5ΔT). These values were taken as the best estimates of feature magnitudes that would induce equivalent performance, across the four feature dimensions, at the specified accuracy levels.

Following these initial measurements using the single-feature displays, each subject then repeated the same alternating-feature display. In these displays, token salience was increased relative to the background in each of the four features sequentially, creating successive waves of salience across the feature maps over time (Figure 1B). The increase from background was calibrated to be equivalent across the four features, as measured from the single-feature experiments, so as to yield equivalent levels of perceived salience (or 80%–180% theoretical performance). In a typical trial, for example, the initial 50 ms of the animation may have defined target tokens on the basis of orientation (with all other feature dimensions being identical to the non-target Gabors), while the next 50 ms might have then defined the target tokens on the basis of spatial frequency (with the other features matched to the background tokens). In this way, each successive 200-ms interval cycled through the four key dimensions. The order of the four feature changes was randomized on each trial, and salience was sinusoidally ramped on and off to minimize any transients. The magnitudes of the feature changes were chosen to be perceptually equivalent across the four key features, based on the individual subject’s psychophysical estimates collected in the single-feature experiments. Using the method of constant stimuli, subjects completed a total of 258 trials (43 per estimated salience level). Feedback was not provided.

In the alternating-feature displays. To minimize the possibility that salience was signaled by non-specific visual transients that may accompany any abrupt feature change (e.g., a sudden increase in contrast), the magnitude of the feature differences were sinusoidally ramped on and off over the course of 50 ms in all displays. Equivalent salience was defined as the maximum difference of that feature, relative to background, that resulted in 80% threshold accuracy on the discrimination task for each individual. Therefore, subjects first completed the single-feature task for each of the key features prior to the alternating-feature task. The experiments proceeded as follows.

Subjects completed a double-interleaved 3–1 adaptive staircase (Levitt, 1971) that incrementally increased or decreased the difference between a single feature of the target and non-target Gabors, while keeping all other parameter values identical. Initial starting parameters for the staircased features were set to yield approximately 90–100% accuracy, as determined in pilot studies, and adjusted the relative strength of that single parameter based on subject performance. For example, three sequential correct trials on an orientation staircase resulted in the orientation of the target Gabors on the next trial to be more similar to that of the background Gabors, thus rendering them less salient. Each incorrect trial would result in the orientation adjusted to be more dissimilar from the background Gabors, rendering the targets more salient and the next trial slightly easier. The staircases terminated following fifteen reversals (on each staircase) in the trajectory of subject performance, yielding approximately 150–200 trials per subject.

The adaptive staircase procedure was completed for each key feature dimension in separate blocks to yield estimates of equivalent salience for each feature alone.

In the trajectory of subject performance, yielding approximately 150–200 trials per subject.

Feedback was not provided.
for targets defined by orientation. The high potency of some features as compared to others in these single-feature displays is not unexpected (Lu & Itti, 2005), with previous reports using visual search tasks showing, for example, that orientation is more inherently salient than luminance contrast (Nothdurft, 1992). Because salience is strongly dependent on context, the relative strengths of each feature in these tasks would be difficult to predict on the basis of prior studies using slightly different visual displays and tasks. To our knowledge, there are no previous studies measuring the relative salience of the four features we have targeted in the context of these three tasks.

An analysis of variance also revealed a main effect of task in the single-feature threshold estimates, such that the biological motion task required smaller feature differences as compared to the other two conditions \( F(4,188) = 5.08, p < 0.01 \). There are many ways to interpret this finding, given that each of these tasks imposes a unique set of cognitive demands on the observer. One interpretation is that this main effect simply reflects the intrinsic difficulty level associated with each task, with discriminating biological from scrambled motion being inherently easier than coherent motion discriminations or object tracking. Alternatively (or additionally), this main effect may reflect the preexisting, inherent salience of biological motion, which is often described as “interesting” and has greater social importance than the other two types of stimuli. Whichever the case, these measurements quantify the task differences in terms of feature magnitudes and, most importantly, underscore the need for calibrating salience, both across features and tasks, prior to measuring performance on the alternating-feature displays.

Figure 3 shows the mean performance on the same motion tasks for the alternating-feature displays. For most salience levels measured, observers were able to discriminate the biological motion in these displays with high accuracy, evidence that the essential features in point-light sequences are apparent when constructed via attentive mechanisms. Subject did, however, require slightly more salience (one standard deviation more) than estimated on the basis of the single-feature measurements (significant effect of salience: \( F(5,46) = 11.11, p < 0.01 \)). This required increase in salience may reflect the cost in constructing biological motion without the benefits of the low-level, passive motion detectors or a simply greater
difficulty attending to salient features across dimensions as compared to within.

To determine the effect of removing low-level motion analysis in the alternating-feature displays, we consider performance in the coherent motion task. We found a statistical effect of coherence level in the alternating-feature measurements \(F(2, 171) = 13.28, p < 0.01\), with 10% coherence yielding overall the best performance. That 10% coherence yielded better performance than 20% despite having a lower proportion of tokens moving coherently is likely due to the nature of embedding the limited lifetime kinematograms in the Gabor grid displays, with “figure” and “background” becoming increasingly difficult to segregate at higher coherence levels. Discrimination accuracy improved as function of salience \(F(5, 171) = 18.41, p < 0.01\), with the easiest condition (10% coherence) requiring just one standard deviation more salience for optimal performance on the alternating-feature displays as compared to the single-feature versions. These results quantify the costs, in terms of salience, of extracting local motion energy by attentive tracking versus with passive motion mechanisms.

To determine whether feature tracking is inherently more difficult in alternating-feature as compared to single-feature tasks, we consider performance on the multiple object tracking task. Subjects were just as successful tracking tokens defined by alternating features as by single features, with no benefit of increasing the salience levels of the token features \(F(5, 42) = 0.45, p = 0.8102\). This result is consistent with previous findings that observers are able to successfully track an object even when its features change throughout the trajectory (Blaser et al., 2000) and that tracking can use a position-based strategy that allows for changing surface features (Cavanagh & Alvarez, 2005; Pylyshyn, 2001a). Thus, the attention-limiting factors on multiple object tracking are not predicted by salience and likely depend on other task-related factors, such as working memory and spatial selection constraints (i.e., the number of items to be tracked and their spatial proximity; Fougnie & Marois, 2006; Franconeri et al., 2010; Pylyshyn & Storm, 1988). Together with the biological motion discrimination results, these findings dissociate the attention demands imposed by biological motion and tracking objects.

**Experiment 2**

Experiment 1 demonstrated a demand for increased salience in discriminating the alternating-feature biological or coherent motion displays but not in the multiple object tracking task. At most salience levels, however, subjects performed better on the alternating-feature biological and coherent motion discriminations, as compared to the multiple object tracking. This better overall performance suggests that the biological motion displays, which require a more complex visual analysis, place greater demands on attentional resources than the single-feature displays.
performance may reflect some important differences in stimulus configurations and task demands across these three conditions.

One consideration is the unique spatial selection and working memory demands imposed by the tracking task. Multiple object tracking requires subjects to split their attentional spotlight among three possible targets and rapidly track, or update, the current positions of those possible targets in working memory. In contrast, the biological and coherent motion displays could effectively be grouped into a single object or group, with no need to split attention among the individual tokens, no potential distracters from which to induce possible confusion with likely targets, and no working memory or position-tracking demands.

The three stimuli also all have differences in configurations that span different spatial scales. For example, the tokens in multiple object tracking displays drifted across the entire 16-deg stimulus window, whereas the target figure in the biological motion was constrained to a relatively small (7 deg) central region. It is an ongoing debate whether subjects use a global configuration strategy when discriminating biological from scrambled motion (Lange & Lappe, 2006; Lu & Liu, 2006) or attend to local features of the body (Casile & Giese, 2005; Thurman et al., 2010; Troje & Westhoff, 2006), but both types of features exist at a smaller spatial scale than dispersed objects in the tracking task. One consequence of this configural difference is the need to split attention across a much larger field of view in the multiple object tracking task as compared to the biological motion condition. Likewise, although the coherent motion tokens were distributed, a subject could theoretically discriminate the displays by attending to a relatively small, focal region in the center of the display. Thus, one possible explanation for performance differences across the alternating-feature versions of these three tasks is the usefulness of salience in promoting more localized features, which could benefit the biological motion and coherent motion tasks more than multiple object tracking.

Therefore, in a second set of experiments, we sought to determine whether having tokens embedded within a single-object configuration constrained to a smaller region of space would improve performance on the alternating-feature multiple object tracking task. To achieve this, we measured tracking performance when subjects attended to individual joints in the point-light sequences, which had the benefits of matching the biological motion stimulus configuration but imposing the multiple object tracking task demands.

In pilot measurements for this study, we found that subjects were unable to track even a single token in the point-light figure at the stimulus size used in Experiment 1 (7 deg, centrally presented). That tracking targets moving within a more confined region of space is more difficult should not be entirely surprising given limited spatial acuity of attention (Intriligator & Cavanagh, 2001) and the increased density (and subsequently the number of collisions between targets and distracters, both of which make tracking much more difficult) of the small point-light display (Franconeri et al., 2010). Alternatively, forcing subjects to individuate the point lights, which readily group perceptually, may have impaired tracking performance by inducing attentive competition between target and distracter tokens that are integrated into a single object (Scholl, Pylyshyn, & Feldman, 2001). Either way, tracking local features on these small point-light animations was beyond the ability of our subjects, even though the sequences were readily recognized as biological.

We found, however, that increasing the stimulus size of the biological figure to span the same spatial extent as the multiple object tracking, at its maximum, eliminated these tracking difficulties. Therefore, in these experiments, subjects tracked three joints on 14-deg point-light sequences in single-feature and alternating-feature versions. Because of the inherent higher salience of biological motion (as evidenced by the main effect of task in Experiment 1), we also measured tracking performance on non-biological scrambled sequences, which have the same distribution of motion trajectories and are constrained to the same stimulus window as the biological sequences but do not benefit from the perceptual organization of the dots into a coherent figure.

Participants

Eight subjects (ages 21–33) with normal or corrected-to-normal vision participated in this experiment. Subjects completed either the biological motion multiple object tracking (N = 4) or scrambled motion multiple object tracking (N = 4). All subjects gave informed, written consent approved by the University of California Irvine Institutional Review Board and received monetary compensation for their participation in the experiment.

Materials

Biological motion multiple object tracking stimuli (Movie 5) were constructed identical to the biological motion animations described in Experiment 1, with the exception that the size of the biological figure was increased by a factor of two (14 deg). This change was implemented on the basis of pilot studies that determined that subjects could not successfully track even a single individual token within the 7-deg biological figure. As in the multiple object tracking task in Experiment 1, three of the twelve Gabors in the biological figure were identified by a blue outline for 1 s prior to the onset of the animation as the “to-be-tracked” tokens. The biological sequence then proceeded to move naturally for 1 s, immediately
followed by the final resting position of a single Gabor (tracked or non-tracked) being outlined in blue. This indicated the Gabor for which the subject was instructed to make their judgment.

Scrambled multiple object tracking stimuli (Movie 5) were constructed identical to the motion-matched “scrambled” actions in Experiment 1, with the overall configuration size increased by a factor of two, as in the multiple object tracking biological displays. The tracking paradigm was otherwise identical to that of the biological motion tracking experimental condition.

Results

As in Experiment 1, we found that varying magnitude differences were required for each feature to generate equivalent tracking performance (80% accuracy) in the single-feature displays (Figure 4A). The greatest magnitude differences were required for targets defined by luminance contrast, and the lowest magnitude differences were required for targets defined by orientation in both stimulus configurations (biological: $F(3,15) = 20.31, p < 0.01$; scrambled: $F(3,15) = 5.037, p = 0.018$). We found no main effect of task (biological or scrambled) in the feature magnitudes required for threshold tracking accuracy ($F(1, 31) = 2.09, p = 0.16$), indicating that salience of targets with natural trajectories such as these are not improved by an imposed, known structure (in this case, the human body).

Performance on the alternating-feature displays is shown in Figure 4B. Analysis of variance revealed no significant effect of salience in these alternating-feature tracking configurations (biological: $F(5, 23) = 0.43, p = 0.82$; scrambled: $F(5, 23) = 0.42, p = 0.83$). Tracking accuracy on both the biological and scrambled tasks were well predicted by the single-feature displays, with no added benefit of increased feature salience, just as was found in the multiple object tracking measurements from Experiment 1. Thus, task, but not stimulus configuration, appears to play a key role in the effect of salience on attentive selection. Salience provides a cue to important locations in the visual scene, and dynamic shifts of these locations can be cued just as easily within feature dimensions as across features.

Discussion

Attention mechanisms serve to unify our visual perception, from the earliest stages of selecting parts of the visual scene for further analysis to maintaining and manipulating information about visual objects. We measured the extent to which feature-based attention promotes perception of dynamic events and in particular biological motion. We find that observers can recognize biological motion in point-light animations constructed so as to be effectively invisible to low-level, passive motion mechanisms, evidence that observers can perceive biological motion via attentive mechanisms alone. Recognizing the attention-based biological motion, however, required more salience as compared to the low-level depictions (single-feature displays). We believe that this demand for additional salience reflects the inherently dynamic nature of essential features in biological motion, which are more difficult to extract when low-level, passive motion mechanisms are not engaged.

We base this conclusion on a comparison with coherent motion discriminations. In the alternating-feature displays, coherent motion suffered from the same accuracy cost as biological motion, with both tasks requiring approximately one standard deviation higher salience levels than Movie 5. Multiple object tracking for the biological and scrambled configurations. Salience levels for the key features are adjusted to yield approximately 180% accuracy, as estimated from the pooled group mean in the single-feature displays.
predicted by the single-feature measurements to reach threshold performance. Thus, motion energy is more difficult to extract across feature dimensions as compared to within. That the pattern in performance was found for biological motion discriminations suggests that it was the analysis of dynamic biological motion features that suffered in our alternating-feature displays.

We do not attribute our findings to increased difficulty in perceiving the target tokens relative to the background in the alternating-feature displays. Accuracy on the multiple object tracking task was equivalent across both the single-feature and alternating-feature versions. Thus, subjects were able to discern the tokens equally well whether they were depicted by a single feature or across multiple features. This was true whether subjects tracked tokens that moved independently (Experiment 1), tokens that were constrained to the structure of the human body (biological motion tracking), or tokens that had the same kinematics as biological motion (scrambled tracking). For all three tasks, accurate tracking could be achieved just as easily within single-feature space as compared to across feature dimensions.

We also do not attribute our findings in the alternating-feature conditions to generalized task difficulty, which was equated across the three tasks. The initial single-feature measurements yielded a significant main effect of task, with single-feature biological motion requiring smaller feature differences than the other two tasks to yield the same level of accuracy. This main effect could reflect different task effect sizes when viewing conditions are not controlled and/or more inherent salience in biological motion (i.e., is more “interesting”) than the other stimuli. The single-feature salience measures provide a means to quantify, and then calibrate, those differences. Any inherent differences in task difficulty would have been eliminated in using the staircase procedure.

Instead, we believe that our findings reflect the importance of attentive mechanisms in promoting biological motion recognition. In particular, our findings implicate feature-based attention in constructing biological motion, which is the mechanism believed to perceptually compute velocity from alternating-feature displays (Lu & Sperling, 1995). Our study contributes to a growing literature documenting the link between active, attentive mechanisms and biological motion recognition (for review, see Thompson & Parasuraman, in press). These studies show that biological motion recognition is limited

![Figure 4](http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932789/)
by the spatial and temporal boundaries of selective attention (Cavanagh et al., 2001; Thornton et al., 1998, 2002; Thornton & Vuong, 2004) and is highly correlated to individual subject's ability to selectively attend to features in non-biological tasks (Chandrasekaran et al., 2010). Interestingly, biological motion perception also appears to engage reflexive attentive orienting, with subjects able to more quickly and more accurately report targets in the implied attended field of point-light walkers (Bosbach, Prinz, & Kerzel, 2004; Shi, Weng, He, & Jiang, 2010). The automatic orienting effect of biological motion appears to be driven primarily by local motion features over global body motion (i.e., walking direction) in the action kinematics (Hirai, Saunders, & Troje, 2011), which may reflect a higher priority in visually analyzing these key dynamic features. However, because reflexive orienting only occurs when the point lights are in their intact, canonical configuration, engaging this attentive mechanism is likely to be a consequence of perceptually organizing biological motion rather than promoting the construction itself.

One important consideration for our experiments is the unique set of demands each task imposes on specific attentive mechanisms. Presumably, discriminating biological from motion-matched scrambled animations requires detecting some diagnostic feature(s) in the intact point-light sequences. The spatial scale of those features is currently a matter of debate. While some researchers argue that biological motion is inherently a global, form-based analysis (Lange et al., 2006), others have argued for the importance of local, dynamic features. For example, researchers have shown that observers can discriminate the facing direction of biological motion from the velocity patterns of the feet alone (Troje & Westhoff, 2006). In all likelihood, the most helpful key features depend on the information available and task demands (Thirkettle et al., 2009), with some local features being better than others (Thurman et al., 2010).

How global versus local key features interact with passive and attentive motion mechanisms is an open question. Because we found similar patterns in the alternating-feature measures across the biological and coherent motion tasks, we believe that our results more strongly implicate the selection of local velocity features. This interpretation dovetails with recently published findings of local motion features driving automatic orienting from biological motion (Hirai et al., 2011). However, it would be reasonable to suggest that observers might switch to a more global strategy in conditions that would promote a more holistic analysis. For example, we know that attentive selection has limitations in spatial acuity that is much more coarse than visual acuity (e.g., He, Cavanagh, & Intriligator, 1996), and thus, one could hypothesize that shrinking the biological target may promote a reliance on more global features. Likewise, increasing the size of the point-light figure, like was done in our Experiment 2, could also encourage the selection of more local features. The extent to which these shifts in strategy would result in increased (or decreased) demands on target saliency is not entirely clear.

As a means for comparison, however, we consider the coherent motion task, in which subjects were required to pool motion energy across some spatial window. Because the coherent motion displays used in these experiments imposed a limited lifetime on the token trajectories, accurate performance could not be achieved by tracking individual objects and instead required pooling a global velocity signal. In the single-feature displays, this could be achieved through the summed activity of an array of velocity-tuned visual filters, computing the space–time shifts of the textured displays (Petersik, 1995; Sperling, 1989). In the alternating-feature versions, however, attentive mechanisms were required to bind the motion energy across feature space (Lu & Sperling, 1995). In theory, this binding could have been restricted to small, focal window, if sufficient motion energy could be extracted from within a relatively local spatial extent. This would be a reasonable strategy for optimizing performance, as some researchers have argued that motion perception improves for small visual displays as compared to larger ones (Tadin, Lappin, Gilroy, & Blake, 2003).

Multiple object tracking, however, incurs a set of demands on the attention system unique from the other two tasks. Despite the dynamic nature of the tokens in multiple object tracking, studies have shown that subjects can use a rapidly updating position-based indexing strategy that operates independent of the motion system (e.g., Pylyshyn, 2001b). This position-based strategy tracks the position of highly salient events in the visual scene but does not require tracking individual features of the tokens (for review, see Cavanagh & Alvarez, 2005). Attentive demands in this task, therefore, may reflect the active updating of the current position of multiple tokens. Note, too, that this requires attentive selection being split among the tokens and with the result that performance is impacted by the number of objects that must be retained in working memory (e.g., Fougnie & Marois, 2006; Tombu & Seiffert, 2008). Our measurements showed that this host of capacity-limited attentive demands is not improved by increased target salience.

Thus, the biological and coherent motion tasks imposed demands on feature-based attentive mechanisms that are not required for object tracking. One may consider, then, the role of feature salience in promoting selection. The displays in these experiments exploited bottom-up salience or differences in object features relative to the background context. We know from previous research that observers are able to use feature salience in visual search tasks with stationary targets (Wolfe, Cave, & Franzel, 1989), to compute velocity (Lu & Sperling, 1995), or to track the identity of multiple objects that coexist in the same location, even when salient features alternate across feature dimensions (Blaser et al., 2000). Here, we demonstrate that feature-based attention also promotes the integration of
dynamic point-light sequences into an intact biological motion percept.

One motivation for constructing the alternating-feature displays was to render ineffective any low-level, passive motion mechanisms and isolate the more active, attention-based motion system. The impact on local motion analysis is evident in the coherent motion measurements, for which observers were clearly able to extract velocity information using this active, attentive system, consistent with previous research using similar manipulations (Lu & Sperling, 2001). We also found that coherent motion discriminations required more salience than predicted from displays that could be captured from passive, low-level motion mechanisms alone. That we found the same required added salience for the alternating-feature biological motion as the coherent motion suggests that the essential features of biological motion being extracted by the observer are inherently dynamic.

While orthogonal to the motivation of these experiments, it is worth noting that the biological and scrambled motion tracking measurements speak to a discussion in the object tracking literature on the role of stimulus configuration. Some researchers have suggested object tracking to benefit from integrating the tokens into a single non-rigid object (Yantis, 1992), while others have argued for the importance of factors that promote the individuation of independent tokens (Franconeri et al., 2010; Ko & Seiffert, 2009). We found no impact of stimulus configuration on tracking performance in our alternating-feature depictions, even when the tokens were organized by the highly salient structure of the human body.

As a final point, the importance of attentive demands in biological motion perception has long been reflected in the patient literature, in which studies have overwhelmingly reported that individuals with difficulties perceptually constructing point-light sequences suffer from damage to brain structures important for attention, not form or motion perception (Cowey & Vaina, 2000; Schenk & Zihl, 1997; Vaina, Lemay, Bienfang, Choi, & Nakayama, 1990). Patients with damage to the temporal parietal junction are less likely to recognize point-light animations as depicting actions (Pavlova, Staudt, Sokolov, Birbaumer, & Krageloh-Mann, 2003) and require more time or make more errors in discriminating the sequences (Battelli, Cavanagh, & Thornton, 2003). The parietal lobe is heavily implicated in the role of target selection and orienting attention, not only for biological motion but for spatial and temporal attention more generally (e.g., Kastner & Ungerleider, 2000; Posner & Petersen, 1990). The importance of domain-general attentive processes in biological motion has been recently documented in EEG recordings over parietal cortex, in which the neural signature of attentive selection is apparent approximately 200 ms prior to any evidence of neural selectivity for biological motion (Safford, Hussey, Parasuraman, & Thompson, 2010).

In total, the findings from these experiments together with the neuropsychological literature underscore the importance of including attentive mechanisms, and in particular feature-based attentive selection, in contemporary theories of biological motion recognition. It is likely that with a stimulus as rich in structure and information as body movements, observers employ an array of cognitive strategies to extract essential visual cues. That psycho-physical experiments have demonstrated this flexibility depending on the external limits of the stimulus and the nature of the task suggests that an important consideration for understanding biological motion recognition is the nature of the cognitive operations imposed on these visual inputs rather than the shape of the visual inputs themselves.

Acknowledgments

This work was supported by NSF BCS0748314. We would like to thank Samhita Dasgupta, Javier Garcia, and Steven Thurman for their helpful comments on previous versions of this manuscript.

Commercial relationships: none.
Corresponding author: Emily D. Grossman.
Email: grossman@uci.edu.
Address: 3151 Social Science Plaza, University of California Irvine, Irvine, CA 92697-5100, USA.

References


