Perception of biological motion from limited-lifetime stimuli

J. A. BEINTEMA
Ruhr-Universität Bochum, Bochum, Germany
and Universiteit Utrecht, Utrecht, The Netherlands

and

K. GEORG and M. LAPPE
Ruhr-Universität Bochum, Bochum, Germany
and Westfälische Wilhelms-Universität, Münster, Germany

The visual perception of human movement from sparse point-light walkers is often believed to rely on local motion analysis. We investigated the role of local motion in the perception of human walking, viewed from the side, in different tasks. The motion signal was manipulated by varying point lifetime. We found the task of coherence discrimination, commonly used in biological motion studies, to be inappropriate for testing the role of motion. A task requiring temporal information showed a strong performance drop when fewer points were used or when the image sequence was sampled and displayed at a reduced frame rate. Irrespective of the frame rate, performance did not vary with point lifetime. We concluded that local motion is not required for the perception of tested biological movements, suggesting that the analysis of biological motion does not benefit from examining local motion. The reliance of perception on the number of displayed points and frames supports the idea that biological motion is perceived from a sequence of spatiotemporally sampled forms.

The role of visual movement patterns in the perception of moving biological organisms was well demonstrated by the introduction of point-light stimuli (Johansson, 1973). From only a few moving point lights, attached to the joints of an otherwise invisible moving actor, people readily perceive the underlying human figure, categorize the displayed action after viewing it for only fractions of a second (Johansson, 1976), and can even perform subtle tasks such as gender recognition (Kozlowski & Cutting, 1977). This broad range of abilities, jointly referred to as *biological motion perception*, is remarkable given the sparse distribution of points and the many degrees of freedom involved.

Because a static frame of point lights is insufficient to induce spontaneous recognition of the underlying figure in naive observers (Johansson, 1973), many studies have assumed that biological motion is perceived on the basis of local motion signals. In the classic point-light display, each image frame provides a new motion vector (velocity and direction) for each point, the displacement vector to the corresponding point in the previous frame. Three or more sequential frames will provide motion trajectories. Indeed, a number of theories make use of motion vectors to explain the perception of biological motion. From a theoretical point of view, the sparse point representation of the limb structure (two dots per limb segment) provides insufficient information to unambiguously recover the motion and structure from pure motion vectors alone. One way around this issue is to assume pairwise rigidity, that adjoining limbs share the same rotation axis (Hoffman & Flinchbaugh, 1982; Webb & Aggarwal, 1982). Another approach is vector analysis, as proposed by Johansson (1973) and later by Cutting (1981). The latter approaches assume that the visual system employs a set of vector-based rules that help perceive biological motion from a hierarchical sum of simpler motion components. For instance, Johansson proposed that common motion vectors could be used to identify the global coordinate frame of the translating shoulder and hips. After vector subtraction, other vector-based rules could be applied to identify the pendular motions of the limbs.

Yet, motion signals from individual points are inherently linked to the anatomy of connected limb segments, with each point's position fitting a spatial structure that deforms over time. Form information in one frame may be weak, but the information could become substantial when integrated over sequential frames, because the possible positions in each new frame are not random but instead are restricted by the underlying structure and pose of the body. Unquestionable, valuable information for recognition of body movement is contained in changes in

the images during the sequence. Does biological motion perception, though, require motion detected at the level of points, or would it suffice to capture changes at the global structure level, such as the changes in limb orientation or body shape?

Brain studies in humans and animals support a functional specialization in human visual cortex for processing biological motion. Electrophysiological evidence for selectivity to biological motion has been found in area STP of the superior temporal cortex (Oram & Perrett, 1994). In human fMRI studies, area STS has also been found to be selectively activated by biological motion (Bonda, Petrides, Ostry, & Evans, 1996). In this area, the fMRI blood oxygenation level dependent signal varied with humans' ability to recognize the human form when a stimulus was presented upside down (Grossman & Blake, 2001), and it increased when the task was to recognize biological motion rather than to detect global motion direction (Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001). Interestingly, area STP receives projections from two major pathways through the visual cortex (Ungerleider & Mishkin, 1982): a ventral stream that is believed to be involved with processing form, and a dorsal stream that is involved in processing motion. The latter pathway leads via motion-sensitive areas MT (human V5) and MST (sensitive to complex motion) to more parietal areas (Maunsell & Newsome, 1987). Curiously, patients with deficits in their motion-sensitive areas are still able to perceive biological motion stimuli (McLeod, Dittrich, Driver, Perrett, & Zihl, 1996; Vaina, Lemay, Bienfang, Choi, & Nakayama, 1990). Moreover, a recent study showed that MST cells, which respond well to optic flow, do not respond specifically to biological motion (Dannenberg, Krekelberg, & Bremmer, 2002). This finding raises the question of what role local motion signals play in the perception of biological motion.

The contribution of form has been shown in two studies from Maggie Shiffrar's group. Orientation cues that resemble a human form help observers to perceive biomechanically consistent paths of apparent motion (Heptulla Chatterjee, Freyd, & Shiffrar, 1996) or to interpret biological motion displays (Shiffrar, Lichtey, & Heptulla Chatterjee, 1997). Because these studies used line drawings or full-body photographs, both conveying local and global form cues, they still leave open the question of whether global form information plays a role in point-light displays. Other studies that did employ point-light displays have supported an analysis of biological motion at a level more global than that of individual points or subconfigurations. For instance, scrambling the positions of the moving point lights quite drastically disrupts the perception of recognizable form (Cutting, 1981). Studies in which the point-light walker was masked by scrambled point-light walker elements have shown that detection or direction discrimination is reduced but still possible when local analysis of the stimulus is rendered ambiguous by the duplicated motions (Bertenthal & Pinto, 1994; Cutting, Moore, & Morrison, 1988; Pinto & Shiffrar, 1999). Also, the global percept of form in biological motion

can overrule stereoscopic-defined (Bülthoff, Bülthoff, & Sinha, 1998) or motion-defined (Sinha & Poggio, 1996) depth cues. These studies show that global form analysis plays an important role in the perception of biological motion but do not explicitly address the question of whether local motion signals play a role.

The global dynamics of stimulus motion are also important. Using walkers on hands or feet displayed upright or upside down, Shipley (2003) showed that unfamiliarity of the dynamic relations is an important factor in the well-known reduced recognition associated with upside-down displays. Troje (2002) separated movement dynamics and body structure from a large set of walkers using linear methods from statistics and pattern recognition. He showed that movement dynamics are more important than body structure for gender recognition from biological motion. These studies clearly demonstrated the importance of the global dynamics of body and limb motion, but because global dynamics could be transported either by local dot motion or by global changes to the configuration of the walker's posture, the findings still do not clarify the role of local motion signals.

Studies in which the local motion content in the display has been directly manipulated have so far yielded unclear, or at least debatable, results. A first study, which favored a contribution of local motion signals, showed that adding a temporal delay between subsequent frames or reversing dot contrast impaired the discriminability of the coherence or direction of a walker masked by noisy background (Mather, Radford, & West, 1992). Another study that tested reversal of dot contrast found no evidence for low-level motion contributions (Ahlström, Blake, & Ahlström, 1997). Later, Thornton, Pinto, and Shiffrar (1998) showed that interstimulus intervals (ISIs) only had an effect with short stimulus durations. In addition, they found that ISI was an influence when cognitive/high-level analysis of the stimulus was impaired by a mask that duplicated the motion of the walker. However, various factors complicate the interpretation of these experiments on the role of motion. First of all, only studies conducted with a masking background have supported a role of motion (Mather et al., 1992; Thornton et al., 1998), which raises the possibility that the sole role of motion information may be to segregate the figure from the background. Second, in these studies it was assumed that local motion processes are disrupted by introducing ISIs longer than 50 msec (Baker & Braddick, 1985) or by reversing dot contrast randomly in subsequent frames (Chubb & Sperling, 1989). These manipulations, however, may not only disturb low-level motion detection, but also influence other processes, such as form detection. Third, delaying stimulus frames changes the temporal sequence, so that either the sequence is undersampled, resulting in jerky stimuli, or the stimulus is presented at unnatural speed. Such manipulations may well have an effect on higher level processes.

Recently, we introduced a novel way to study the contribution of local motion signals in biological motion perception (Beintema & Lappe, 2002). The local motion signal was manipulated by limiting the lifetime of points. To

this end, the points can be situated anywhere on the four limbs (arms and legs), with each point jumping to a new random location after a certain number of image frames. Such jumps remove the local motion information (motion vector and trajectory information) carried by each point, without altering the temporal sampling of the sequence. When points are reallocated with each single frame (Figure 1), all motion signals consistent with the limb's movement are eliminated. Because the information is carried only by the sequence of point positions, we refer to this stimulus as the *sequential position walker*.

Using the technique of setting a limited lifetime for points, we found that naive observers recognized human movement from the display of a sequential position walker as easily as from classic point-light displays (Beintema & Lappe, 2002). Moreover, addition of local motion signals through increases in point lifetime only improved detection performance for a walker against a noisy background of flickering points. The latter result suggests that local motion signals play a role in the segregation of figure from background.

In our previous study, we investigated the role of local motion in biological motion perception using a single-frame duration, mainly with naive subjects. Here, we take a parametric approach by varying lifetime or number of points at different frame durations. Moreover, most studies on the role of motion, including Beintema and Lappe (2002), have employed tasks such as discriminating the direction that the figure is walking (facing left or right) or its *coherence* (i.e., whether the upper and lower body parts are oriented in the same direction [coherent] or in opposite directions [incoherent]). Coherence discrimination requires more spatial integration than does the direction task (Mather et al., 1992) but still does not require a temporal analysis of the stimulus. To assess the role of motion using a more appropriate task, we let subjects discriminate between a figure walking forward (normal sequence) or backward (reversed sequence).

First, we studied the perception of biological motion from the sequential position walker (single-frame lifetime) for different numbers of points and frame durations (Experiment 1). Second, we investigated the appropriateness of the coherence task through a comparison with the forward/backward task (Experiment 2). Third, we carried out a control experiment for the effectiveness of the limited-lifetime technique in removing motion signals (Experiment 3). Then, we used the forward/backward task to examine whether the addition of local motion signals aids the perception of biological motion (Experiment 4). Finally, we repeated this experiment to see whether the results would hold as well for natural stimuli (Experiment 5).

GENERAL METHOD

Setup and Stimuli

Stimuli were displayed on a CRT monitor (40 × 30 cm, 1,280 × 1,024 pixels) at a vertical refresh rate of 96 Hz, and consisted of white points (5 pixels wide) on a dark background. In most experiments, the subjects viewed the stimuli from a distance of 45 cm in a darkened room. The head was supported by a chinrest.

The sequence of point-light images in Experiments 1–4 was derived from an algorithm by Cutting (1978). This algorithm generates the joint positions of a human, viewed from the side, walking on a treadmill at 0.625 cycles/sec. Experiment 5 used stimuli captured from real walking humans; these stimuli are explained in detail in the description of that experiment. The stimulus figures in all experiments were positioned in the middle of the screen so that the center of the hips did not move horizontally. The width and height of the stimuli subtended 5º × 11º visual angle, unless stated otherwise. The starting phase in the step cycle was randomized from trial to trial.

The local motion signal was varied by removing points after 1, 2, 4, or 8 frames and by replacing them at random locations on the underlying stick figure. The possible positions were distributed uniformly across the eight limb segments, with each segment defined by the line connecting the joints. The initial position of points on the walker was determined at random. Except for the single-frame lifetime, points were refreshed asynchronously in time to prevent a simultaneous refresh of all the points. To keep a constant number of points displayed per frame, initially a fraction of points had a shorter lifetime. The number of points displayed per frame could be one, two, four, or eight. Frame duration was varied in multiples of the monitor's refresh duration (from 10.4 to 208 msec). For the frame duration of 52 msec (five monitor refreshes), the sequence contained 31 frames per step cycle. To maintain a natural walking speed (0.625 cycles/sec), we adjusted the number of displayed frames within the step cycle.

Procedure

Each trial started with a 0.5-sec interval during which the screen was dark, except for a central white point. The sequence was then presented for a fixed duration (1.6 sec in the first experiment, 2.0 sec in later experiments), after which the stimulus disappeared. The subjects indicated their choice by pressing one of two possible mouse buttons. The next trial started upon the button response. Each experiment was preceded by a dozen practice trials without feedback. In the coherence task, half of the trials contained stimuli in which the lower limbs of the body faced in the opposite direction from the upper limbs. In the forward/backward task, half of the trials were presented with the sequence of frames in reversed order. In all experiments, the direction in which the figure was facing and walking (left or right) was balanced over trials.

Data Analysis

Performance in the coherence/incoherence or forward/backward task was quantified by the mean correct rate (average fraction of
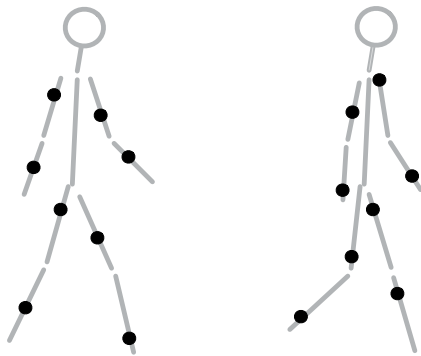

Figure 1. The sequential position walker (lifetime = 1 frame), in which each point jumps to a different random location with each subsequent frame.

correct responses for the two stimulus alternatives p_A and p_B). Statistical inferences were based on d' (the sum of z-transformed p_A and p_B values), because d' provides an unbiased estimate of the discriminability and is more appropriate for linear inferences such as ANOVAs than are simple mean correct rates. The d' curves are very similar to the curves of mean correct rates, although correct rates of 0% or 100% yield infinite d' values. Because p values suffer from round-off errors on the order of $1/n$ due to the limited number of trials (n), we computed d' from p values that were clipped at $p_{min} = 1/n$ and $p_{max} = 1 - 1/n$.

Stimulus Analysis

The reallocation of points after each frame strongly reduces the spatial correspondence between the nearest points in successive frames. As can be seen in Figure 2, the jump of a point in the first frame (gray dot) to the nearest point in the next frame (black dot) has little in common with the displacement of the point expected had it not been reallocated (open circle). Generally, the jumps are much larger and have their largest component perpendicular to the expected motion vector. Nevertheless, some jumps are in the direction of limb motion, and spurious jumps may also have a small component in the direction of the limb motion. To quantify how much these jumps were consistent with the limb movements, we computed the statistics of the frame-to-frame jumps as a function of lifetime for our different experimental conditions.

To estimate how much the jumps were consistent with the movement of the limb, we compared each jump with its expected real limb-motion vector. For the sequential position walker displayed with a 50-msec frame duration and eight points, the median jump pointed away from the true limb motion vector at an angle of about 65º and had a median amplitude roughly five times larger than the expected displacement during true limb motion. Only 0.8% of the jumps lay within 10% (Weber's fraction) of the expected motion vector. This fraction of jumps coherent with respect to the expected limb movement would increase with the number of displayed points or the frame duration, because in both cases the mean jump along the limbs decreases relative to the jump in the direction of limb motion. Even for eight points at the 200-msec frame duration, however, we

maximally found 1.5% of the jumps to be coherent with the limb movement. In comparison, the lowest fraction of coherently moving points that can drive MT neurons in motion coherence tasks (an arguably much simpler task) lies around 5% (see, e.g., Newsome, Britten, & Movshon, 1989). Thus, even when frame-to-frame jumps could elicit motion responses, at a single-frame point lifetime we consider them to be spurious noise. Only when point lifetime increases, as shown in Figure 3, does the fraction of locally consistent motion change substantially, up to 50%, 75%, and 87.5% for the two-, four-, and eight-frame lifetimes, respectively.

EXPERIMENT 1
Role of Form in Biological Motion Without Local Motion

To investigate the perception of biological motion for a stimulus with minimized local motion, we presented the sequential position walker for different numbers of displayed points. The present experiment differs in two important respects from an earlier test of the role of local motion in biological motion that also employed a limited-lifetime walker with different numbers of points (Beintema & Lappe, 2002). First of all, our previous stimuli were sampled and presented at one frequency of 19 Hz, with each frame lasting 52 msec. This frame duration had been chosen because it was long enough to be in the range (>50 msec) in which the "short-range" processing of apparent motion is reported to break down (Baker & Braddick, 1985), and short enough for the walker's movements not to be perceptually jerky. In the present experiment, we wished to see whether our earlier conclusions regarding the role of local motion hold for other frame durations as well. Second, we previously investigated how well subjects can discriminate between figures walking (and facing) toward the left or the right. Such a direction task, however, does not require the spatial integration of visual information, since discrimination of the walker's direction is already possible from a single point positioned on a moving ankle (Mather et al., 1992). In the present experiment, we employed the task of judging coherence, a task that at least requires spatial integration over lower and upper body parts.

Method

The sequential position walker was presented at seven different frame durations (10–200 msec), and the number of points displayed per frame was one, two, four, or eight. For each possible combination, eight coherent and eight incoherent walkers were presented, and all trials were randomized. The subjects (authors J.A.B. and K.G.) completed two sessions.

Results

Figure 4A shows that the fraction of correct responses, averaged over the 2 subjects and the two types of walkers (coherent and incoherent), was strongly reduced with longer frame durations and fewer points per frame.

Discussion

The decreased performance with fewer displayed points confirms our previous findings for direction discrimination at the 52-msec frame duration (Beintema & Lappe,
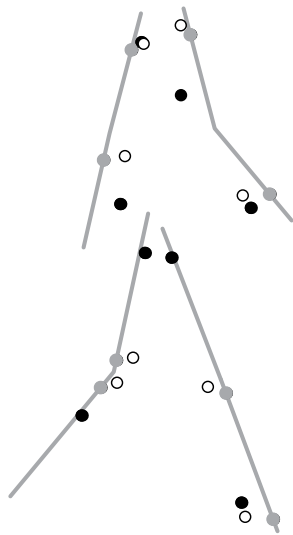

Figure 2. Two frames of the sequential position walker with a one-frame lifetime for eight points (gray dots) that are reallocated in the subsequent frame (black dots). The open dots indicate the expected point positions had each point moved according to the limb movement. The gray lines indicate the position of limb segments in the first frame.
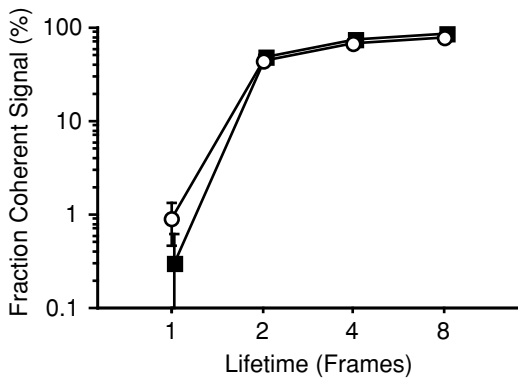
Figure 3. Fraction of motion vectors that lie within 10% of the expected motion vectors, as a function of lifetime. Each fraction is based on 2,400 nearest point displacements obtained from frame-to-frame analysis over multiple step cycles. Each point and its error bar represent the average fraction and 95% confidence interval over a range of experimental conditions tested. The solid symbols represent average fractions over frame durations of 10, 20, 50, 110, 150, and 200 msec for eight points per frame. The open symbols represent average fractions over number of points per frame (one, two, four, and eight) at a 50-msec frame duration.

2002), suggesting that spatially sampled form plays an important role in biological motion perception. Interestingly, frame duration also strongly reduced performance, indicating that temporal sampling of the sequence is important in biological motion perception as well. Specifically, two factors may underlie the decrease in performance with frame duration. First of all, since we kept the simulated walking speed constant, longer frame durations resulted in increasing phase differences between the sampled poses, creating a jerky stimulus. Second, since the total duration was kept constant, fewer poses were displayed in total. Thus, performance was strongly related to the number of presented frames.

Curiously, however, when eight points were displayed, an average of one point per limb segment, subjects could still discriminate coherence well above chance level at long frame durations. This result seems remarkable, because (for instance) at the 200-msec duration only eight different frames were presented in total. This suggests that the coherence task demanded little sequence information. We conducted the following experiment to investigate the importance of pure form information and to check performance in a task that requires temporal integration of information.

EXPERIMENT 2
Influence of Task in Biological Motion Perception

Experiment 1 showed that the spatial configuration of points in a few selected image frames already provides sufficient information to discriminate the walker's coherence. This suggests that the coherence task may not require temporal integration over different poses. To test this hypothesis, we repeated the coherence task, but this time using a stimulus in which points were reallocated on a static walker (walking speed = 0 cycles/sec). The phase was chosen so that the walker was presented with its feet maximally apart.

An analysis of human posture over time is ensured in tasks in which the order of displayed frames is relevant. One such task is discrimination of normal from frame-scrambled walking sequences (Neri, Morrone, & Burr, 1998). However, one could argue that a frame-scrambled walker appears qualitatively different because of its more diffuse temporal-averaged contour. Such a cue is eliminated when the task is to discriminate between sequences played in a normal or reversed order, because then the sequences contain the same spatiotemporally averaged form. Thus, in a second task, we presented the frames of
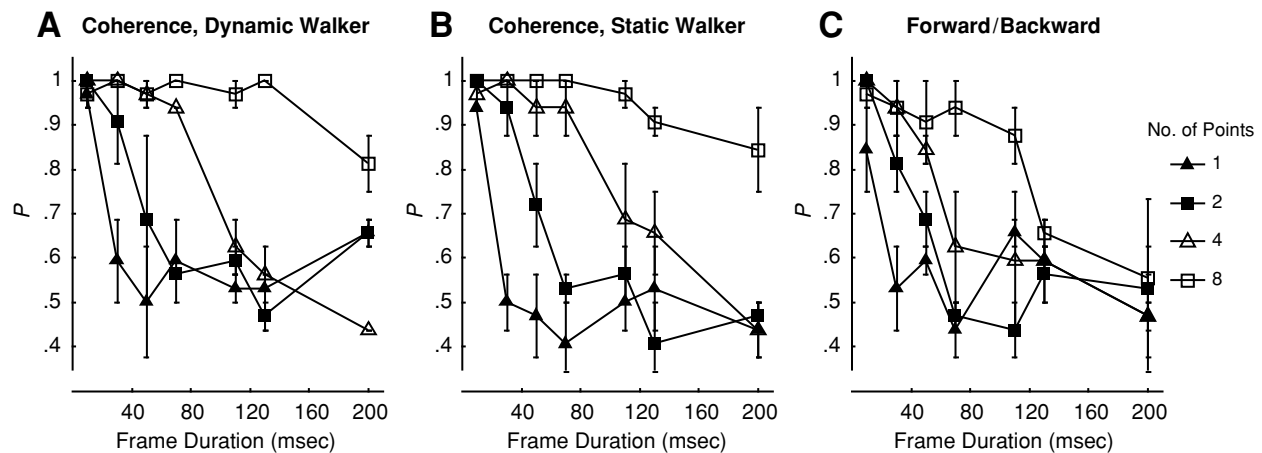

Figure 4. Proportion of correct responses as a function of frame duration split by the number of points per frame for the sequential position walker in three experiments. (A) Coherence discrimination for a dynamic walker (Experiment 1). (B) Coherence discrimination for a static walker (Experiment 2). (C) Forward/backward discrimination (Experiment 2). Points and error bars show means ± standard errors calculated across stimulus alternatives (coherent/incoherent or forward/backward), sessions, and 2 subjects.

the sequential position walker in normal or reversed order. The subjects (authors J.A.B. and K.G.) were to indicate whether the figure was walking forward or backward, and the stimulus duration was 2.0 sec.

Results

Performance with static walkers (Figure 4B) was similar to that with dynamic walkers (Figure 4A). An ANOVA on d', with number of points and frame duration as significant factors, showed that the task was not a significant factor [$F(1,56) = 2.4$, $p = .13$]. This shows that judging coherence does not require information about limb movement over time.

In contrast, the data for the forward/backward discrimination task (Figure 4C) do show reduced performance in comparison with the coherence task (Figure 4A). In this case, task was a significant factor [$F(1,56) = 9.8$, $p = .002$] in an ANOVA on d' with task, number of points, and frame duration as factors. Specifically, we found significant differences for the four-point stimuli presented at frame durations of 50 and 70 msec and for eight-point stimuli with frame durations of 130 and 200 msec. Most clearly, for eight-point stimuli the coherence task showed correct rates above .8, whereas the forward/backward task showed rates that dropped below .6 at the 200-msec frame duration. This suggests that the forward/backward task does require temporal information.

Discussion

In theory, discriminating coherent and incoherent walkers can be based on static form cues, so it is questionable whether such a task could profit from movement information in biological motion stimuli. Indeed, we found no performance difference between dynamic walkers and walkers that contained static information only. Since the performance for the dynamic walker was not significantly different, degrading in a way similar to performance with the static walker with frame duration and number of points, it seems plausible that even in dynamic walker conditions subjects relied on static form information. For instance, subjects may have used just a few salient form frames when eight points were displayed per frame. For fewer displayed points, subjects may still have relied on static form by integrating form over several frames. Indeed, since points were reallocated after each frame, over time more of the underlying stick figure was drawn out. Thus, the drop in performance with longer frame duration could be explained by reduced spatiotemporally averaged form.

For the forward/backward task, we did find a drop in performance with longer frame durations even for the eight-point walkers, demonstrating that this task does require information that depends critically on the temporal sampling of the sequence. Note, however, that this drop in performance was observed only for four and eight points per frame. For fewer points per frame, perceptual degradations may have gone unnoticed because for these stimuli, performance was already nearer chance level. We conclude that judging the coherence of a walker, or other tasks (such as judging walker direction) that can be based on form information, are not appropriate for testing the role of motion for the perception of biological motion.

EXPERIMENT 3
Effect of Intermitting Blank Intervals

Our results so far indicate that local motion signals are not necessary for the perception of biological motion. The evidence for this, however, hinges on our assumption that the limited-lifetime paradigm eliminated the presence of valid motion cues in the stimulus. Earlier investigations in support of a role for low-level motion processing in biological motion perception have used another technique to remove local motion cues (Mather et al., 1992; Thornton et al., 1998): In these studies, low-level processing was assumed to be disrupted if a long enough ISI was presented between frames, during which time no points were visible (Baker & Braddick, 1985).

If intermitting intervals are indeed effective in eliminating low-level motion processing, we may apply this paradigm as a control: If the stimulus with a single-frame lifetime still contains valid motion cues, and if low-level motion processing does play a role in biological motion perception without a background, then the insertion of blank intervals between frames should reduce performance specifically at 50-msec or longer ISIs.

Method

The forward/backward experiment with eight points per frame was repeated with and without an ISI. Without an ISI, each frame was visible until the next frame appeared. With an ISI, each frame was displayed for 20 msec, followed by an interval during which the screen was black until the next frame appeared. Because the stimulus with ISI was perceptually less bright because of its shorter illumination duration, which might have impaired the discrimination, we presented the stimuli at normal luminance as well as at half luminance (50% and 100%). Perceptually, at the 50-msec frame duration we found the stimulus with ISI at 100% luminance to be about equally as bright as the stimulus without ISI at 50% luminance. The experiment was conducted in a fully darkened room, with a viewing distance of 55 cm instead of 45 cm, and the walker subtended 7° × 12°. Four subjects participated (authors J.A.B. and M.L. and 2 naive subjects, J.L. and P.S.).

Results

Plotted in Figure 5 are mean correct rates as a function of frame duration for the two levels of brightness (50% and 100%). Note that frame duration here represents the sum of light duration (20 msec) and ISI. Clearly, no overall decrease in performance occurred when black intervals were intermitted (Figures 5A and 5B). Neither did we find an overall difference in the influence of ISI for conditions that were perceptually matched in brightness (compare no ISI at 50% luminance [triangles in 5A] with ISI at 100% luminance [squares in 5B]). Discriminability was somewhat reduced by ISI for the 150-msec frame duration, but across the brightness-matched conditions ISI did not yield a significant effect ($t = 1.3$, $p = .3$, in a t test on d' with six degrees of freedom).
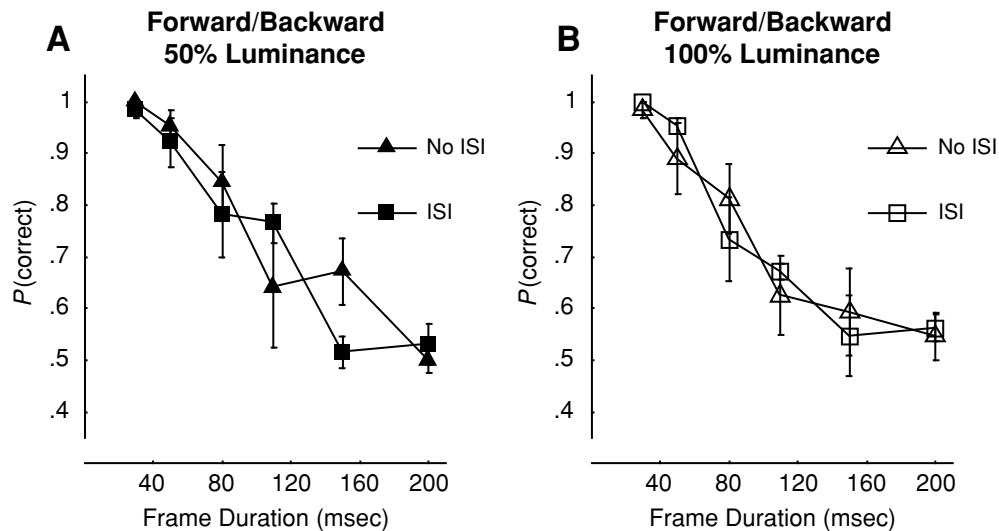
Figure 5. Proportion of correct responses in a forward/backward identification task (Experiment 3) as a function of frame duration, with and without an interstimulus interval (ISI), for 50% luminance (A) and 100% luminance (B). Points and error bars show means ± standard errors from all 4 subjects.

Discussion

We found that intermitting intervals did not reduce performance for a stimulus with a single-frame lifetime. Had we found an effect, we would have had to conclude that biological motion perception is sensitive to local motion signals and that our paradigm does not effectively remove these local motion cues. These results are therefore consistent with our findings that local motion signals are not required for biological motion perception.

EXPERIMENT 4
Effect of Lifetime as a Function of Frame Duration

The experiments above show that in the absence of local motion signals, human performance is well above chance level in a discrimination task that requires biological motion perception, provided that sufficiently many points are shown per frame. The performance, however, is reduced with frame duration. In this experiment, we examined whether this reduction as a function of frame duration still occurs when local motion is increased. We have looked into the effect of lifetime previously (Beintema & Lappe, 2002) and found no improvement with longer lifetime. However, that test was only performed with a frame duration of 50 msec. Moreover, we used a direction discrimination task that, as we showed for the coherence task (see Experiment 2), can be performed using purely static form information, which may have concealed an effect of adding local motion. Thus, here we reinvestigated the contribution of local motion at a range of frame durations and employed the forward/backward discrimination task.

Method

The stimuli and conditions were identical to those of the forward/backward task in Experiment 2. Eight points per frame were presented, and point lifetime was varied (one, two, four, or eight frames) at a range of frame durations (10–200 msec). Six subjects participated (authors J.A.B., K.G., and M.L. and 3 naive observers) in one session apiece.

Results

As in Experiment 2, we found that performance degraded with longer frame duration. If motion signals played a role, we would expect performance to increase with longer lifetime. But, as illustrated in Figure 6, we found no effect of lifetime on the mean correct rate at any frame duration. At the 110-msec frame duration, the data do suggest a small positive effect of lifetime, but this effect was not significant [$F(3,20) = 0.8$, $p = .48$, in an ANOVA on d' pooled over 6 subjects].

Discussion

Thus, in support of our previous conclusions from direction discrimination and recognition experiments carried out with a 50-msec frame duration (Beintema & Lappe 2002), we found no evidence for a contribution of local motion signals in the perception of biological motion. However, so far, we have only tested perception with synthetic stimuli derived from an algorithm by Cutting (1978). This synthetic walker might not be representative of natural walking movements, which might contain more detailed motion information. We therefore carried out a control experiment using natural stimuli.

EXPERIMENT 5
Effect of Lifetime for Real Motion Stimuli

Possibly our choice of an artificially created stimulus (adapted from Cutting, 1978), as used in the previous experiments, rendered the contribution of local motion less powerful than it is with natural walking stimuli. More-
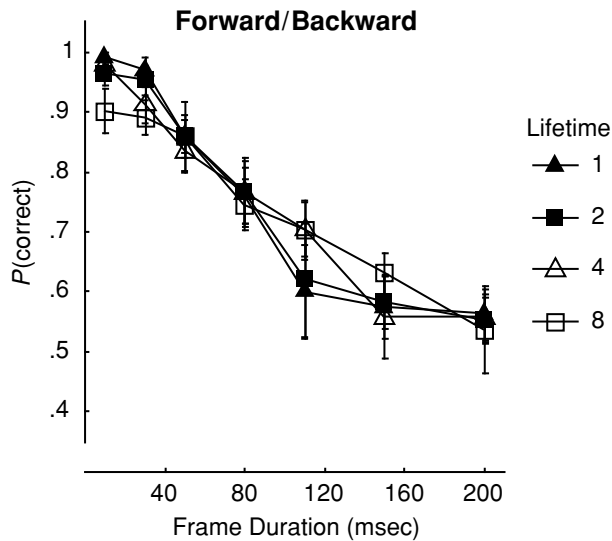
Forward/Backward

Figure 6. Proportion of correct responses in a forward/backward identification task (Experiment 4) as function of frame duration for different point lifetimes (in frames). Points and error bars show means ± standard errors from all 6 subjects.

over, our subjects had extensive training with the artificial stimulus, which possibly allowed them to adopt strategies that may not be representative of biological motion perception. We therefore repeated the forward/backward task of Experiment 4 using natural stimuli with experienced and naive subjects.

Method

Using a MotionStar Wireless motion capture system (Ascension Technology, Burlington, VT), walking patterns were recorded from 4 female and 5 male walkers (students from the department at Ruhr-Universität Bochum, 25–28 years old, with normal weight and an ankle-to-shoulder height of 1.2–1.6 m). They were instructed to walk at a natural pace. The 3-D locations of sensors attached to the

ankle, knee, hip, wrist, elbow, and shoulder joints were registered at 86 frames/sec. The stimuli were smoothed, centered, and clipped to reduce measurement noise and to standardize the displays. Each stimulus was the side view of a single walking cycle, normalized to 0.625 cycles/sec. The average forward hip position was subtracted from the walker's position. Thus, the walker remained at the same location, much like a person walking on a treadmill. We presented the stimuli with four rather than eight points per frame, because pilot experiments revealed that performance for eight points per frame remained at ceiling even at the longer frame durations. Four experienced observers (the authors and subject M.d.L.) and 3 naive observers participated.

Results

As in Experiment 4, performance dropped at longer frame durations (see Figures 7A and 7B for experienced and nonexperienced subjects, respectively). We found no significant increase in d' with longer lifetime (unpaired t tests) for either group of observers. Instead, as in our previous work with the direction discrimination task (Beintema & Lappe, 2002), recognition performance occasionally decreased, particularly at short frame durations. Significant decreases occurred for the experienced subjects at a frame duration of 40 msec between the one-frame and the four- ($p = .025$) and eight- ($p = .045$) frame lifetimes, and for the naive subjects at frame durations of 10 msec (1 frame–4 frames, $p = .05$; 1 frame–8 frames, $p = .01$; 2 frames–8 frames, $p = .02$) and 20 msec (1 frame–8 frames, $p = .001$; 2 frames–8 frames, $p = .02$).

Discussion

Generally, with eight points per frame, detection of forward/backward walking was much easier with the natural stimuli than with the synthetic walker. Our subjective impression was that sway in the arms and legs seemed larger in the natural stimuli than in Cutting's walker. We therefore attribute the increased performance to more pronounced form and motion cues. To assess the role of lifetime, we
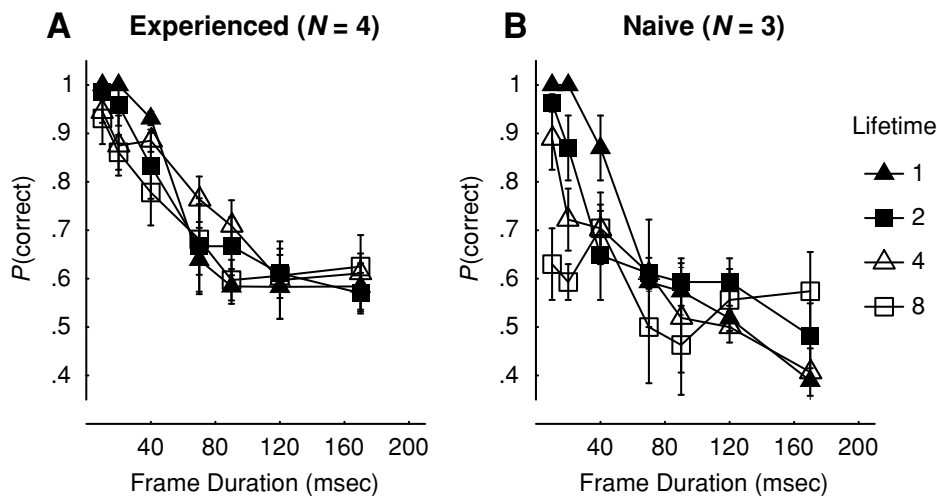

Figure 7. Proportion of correct responses in a forward/backward identification task with natural stimuli (Experiment 5) as a function of frame duration for different point lifetimes (in frames). Points and error bars show means ± standard errors for (A) experienced or (B) naive subjects.

showed the stimuli with only four dots per frame. We do not think that the smaller number of dots changed the interpretation of the data. In fact, the smaller number of dots reduced the amount of motion that (coincidentally) corresponded to the expected limb motion. Given less spurious motion, one would expect a clearer influence of lifetime if motion contributed to the percept. Instead, the effect of adding motion was small for the natural stimuli.

GENERAL DISCUSSION

Role of Local Motion

Concerning the role of motion, we found that performance remained well above chance level in a task that required motion integration. These data support previous work in which we concluded that local motion was not required for the perception of biological motion. In that study, we found that naive subjects recognized a walking figure from the sequential position walker just as often as from a classic point-light walker, in which points are continuously visible on the joints. Also, subjects performed coherence and direction discrimination tasks just as well as they did for classic point-light displays. Our present data confirm that these findings not only apply to the 50-msec frame duration previously tested, but they extend to other frame durations as well.

Concerning the role of form, we found that performance in coherence and forward/backward discrimination was reduced with fewer points per frame and with longer frame duration. As discussed in Experiments 1 and 2, at least two factors might underlie this reduction with longer frame duration. First of all, it may have been caused by temporal undersampling of the sequence, either by a decreased number of frames or by the increased phase differences between subsequent sampled poses. Second, as Experiment 2 demonstrated, the spatiotemporal sampling of form (i.e., the number of visible points within a time window) is also likely to play a role. Both factors predict that performance will increase with the number of points per frame and the number of frames. To see whether a simple multiplicative interaction between these two factors could explain the data, we plotted data from four experiments that used the Cutting stimulus against the number of points × number of frames (= 1.6/frame duration). As Figures 8A–8C show, this description of the data as a function of the total number of points per trial fits surprisingly well, because whether split by number of points, point lifetime, or frame duration, the data fall onto the same curve.

Although Experiments 1–3 support the view that local motion is not the basis of biological motion perception, does this mean that the addition of local motion does not
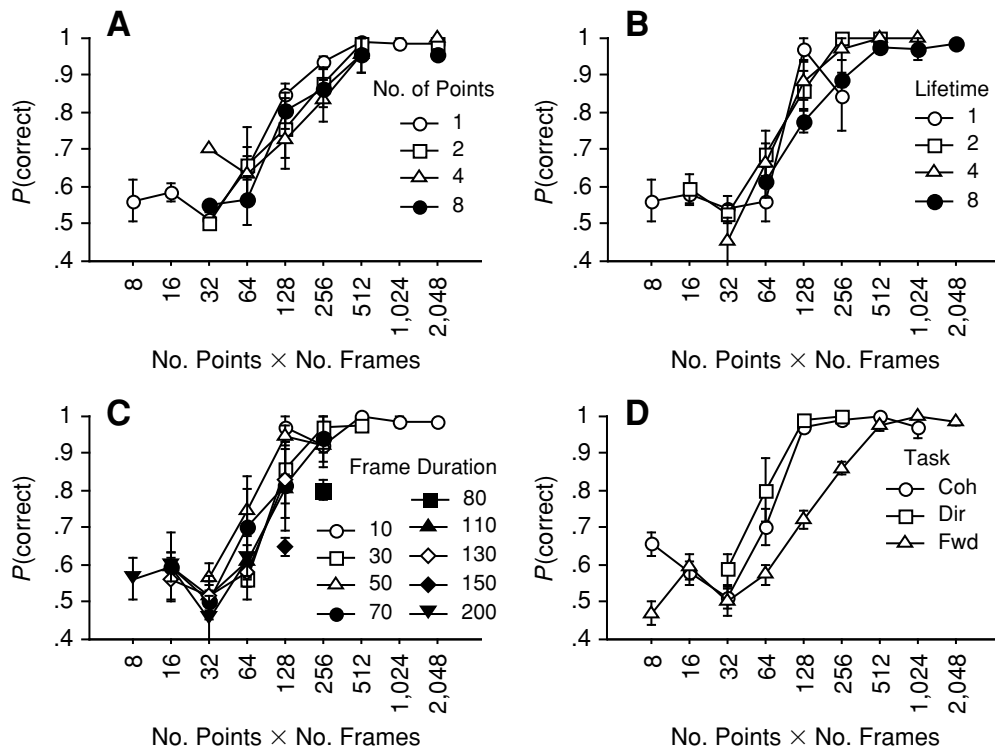

Figure 8. Proportion of correct responses as a function of total number of points per trial (number of points per frame × number of frames per trial). The data were binned in powers of 2. Each point and error bar represent the mean and standard error of data pooled over four experiments for subjects J.A.B. and K.G. We also included data collected in a direction discrimination task for different lifetimes and numbers of points (Beintema & Lappe, 2002). Each subplot presents the same data, split either by (A) number of points, (B) point lifetime, (C) frame duration, or (D) task.

aid perception? We found no significant effect of lifetime on performance (Experiment 4), but could the increased contribution of local motion have been masked by a decrease in sequential position information? We do not expect a longer lifetime to reduce performance because of an undersampling of the temporal sequence, because the number of different poses presented is unaltered by lifetime. However, an increase in lifetime does reduce the number of reallocations, thereby possibly degrading the spatiotemporally sampled form.

One possible factor that can account for the spatiotemporal averaging of form is visible persistence (see Coltheart, 1980, for a review). A visual stimulus displayed for a brief interval remains visible for a short period after its termination. This period of additional visibility is referred to as *visible persistence* and is estimated to be 100–150 msec for brief stimuli of about 10 msec. It is also known to decrease exponentially with exposure duration. Although it is unknown whether visible persistence also applies to the units that process biological motion, it might explain the slight decrease in performance observed with longer lifetime at the 10- and 30-msec frame durations (Figure 6). To estimate the effect of visible persistence on our data, we assumed that visible persistence causes points to remain visible for 100 msec but does not extend the visibility of points that are displayed longer than that. This effect would, for instance, cause the number of visible points at the 50-msec lifetime to be doubled with respect to the actual number of displayed points per frame. Figure 9 shows the data of Experiment 4 as a function of lifetime, with visible persistence either taken into account (9B) or not (9A).

Note that we split the data by the total number of points per trial because, as argued above (Figure 8), this factor not only takes into account the number of points, but also the effect of undersampling the temporal sequence when frame

duration is increased. Whereas Figure 9A shows no influence of lifetime, we do observe a small effect of lifetime when visible persistence is taken into account (Figure 9B). Still, an ANOVA on d' showed no significant effect of lifetime for any binned number of visible points per trial [e.g., at 128 visible points per trial, $F(3,38) = 1.6, p = .2$]. Thus, even when visible persistence is taken into account, our data still suggest that local motion plays a minor role.

The idea that local motion is not necessary for biological motion perception is consistent with findings by Ahlström et al. (1997), who concluded that contrast polarity is not an essential stimulus feature for matching in biological motion perception tasks. At first glance, our results seem at odds with those of Mather et al. (1992) and Thornton et al. (1998), who reported decreased discrimination performance when low-level motion processing was removed. Their stimuli, however, were embedded in a background of noise, so their task also required discrimination of the figure from background. When our limited-lifetime stimulus was masked by points that were randomly reallocated each frame, we found that the tolerance for noise strongly increased with point lifetime (Beintema & Lappe, 2002). Thus, their findings and ours may both be consistent with the idea that motion plays a role in biological motion perception by helping to distinguish figure from background.

Influence of Task, Stimulus, and Experience

The term *biological motion perception* refers to a wide range of features that may be perceived from moving point lights, such as the depth and connectivity of points; the presence of something biological; the identification of movement, intention, or mood; or identification of even more individual characteristics, such as gender or identity. It is quite possible that more than one mechanism is involved in biological motion perception. What features are
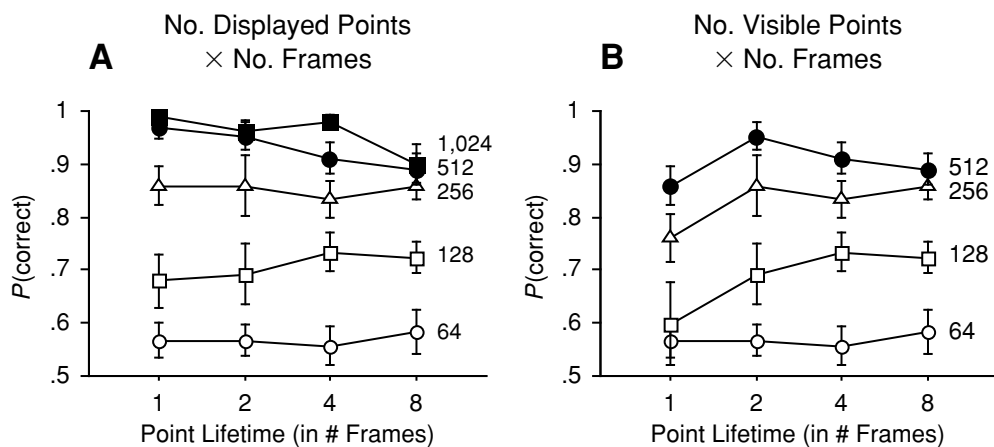

Figure 9. Proportion of correct responses as a function of point lifetime. The data were binned and split by the number of points per trial (in powers of 2) according to the number of (A) points displayed per frame or (B) visible points, once visible persistence is taken into account. Each data point and error bar represent the mean and standard error of data in the forward/backward task in Experiment 4. Note that in panel B, data for bins exceeding 512 visible points per trial have been left out, because they could not be compared across different lifetimes.

most informative might strongly depend on the task and stimulus (e.g., individual, viewpoint, type of movement, time of observation) and experience. Purely temporal information from single-point trajectories, for instance, can be sufficient to detect the direction of walking (Mather et al., 1992). Also, the enhanced recognition of individuals found when exaggerating temporal differences in arm movement displays of grasping and drinking out of a glass (Hill & Pollick, 2000) is likely to have been based mostly on motion of the wrist.

Our main goal was to investigate at what level local motion contributes to the basic perception of biological motion. We restricted ourselves to investigating the perception of human walking viewed from the side, a stimulus condition that is most often cited in the literature and for which it is not a priori clear that local motion does *not* play a role. We do not wish to claim that our results generalize to all other types of movement or viewpoints, which would be beyond the goal of this study; such generalization would certainly be valuable to look into, though. We varied the information content of the stimulus by comparing performance for synthetic and naturally captured stimuli (Experiments 4 and 5). It does seem possible in more natural stimuli to benefit from local motion in the stimulus, even if this benefit is very limited and available only for experienced subjects. This emphasizes that testing naive subjects on their initial percept is most appropriate, since such testing avoids effects of learning a strategy to rely on a single cue (Beintema & Lappe, 2002).

The results from Experiment 2 clearly show that forward/backward walking discrimination is a more appropriate task for studying the role of local motion in biological motion perception than discrimination of a walker's coherence, which can be performed purely on the basis of static-form information. To compare the results obtained for our different tasks, in Figure 8D we plotted the results for the coherence and forward/backward tasks together with previous results on direction discrimination (Beintema & Lappe, 2002). Indeed, the figure reveals clear differences in performance between different tasks. Discriminating between forward and backward walking is most difficult, followed by coherence and then by direction discrimination. Our comparison between the coherence task and the forward/backward task (Experiment 2), however, was based solely on data from 2 trained, non-naive subjects, who might have been able to attend to features that naive subjects normally would not. However, the comparison with the direction task, which was based on a larger data set from 4 subjects (Figure 8D), confirms the trend that the direction task is much easier than the forward/backward task. These differences illustrate that one must be cautious in interpreting data from only one task, since each task tests only one aspect of biological motion perception.

Models

Our results have implications for models on biological motion processing (Hoffman & Flinchbaugh, 1982; Rashid, 1980; Webb & Aggarwal, 1982) that rely on local motion input. We suggest another direction for modeling the analysis of biological motion, one that puts form before motion. Indeed, a number of form-based models have been proposed in the computer vision literature (e.g., O'Rourke & Badler, 1980; for a review, see Aggarwal & Cai, 1999). Our data suggest that perception is derived from an integration of position signals over space and time. This process must take into account the changes in form over time. We have speculated that the brain might accomplish this with dynamic form templates that, over time, accumulate the evidence for a specific sequence of postures (Beintema & Lappe, 2002). Such form templates would have to be selective to a pose and dynamically update its form over time. Preliminary results from a modeling study (Lange, Georg, & Lappe, 2002) show that such analysis of a sequence of forms can replicate our psychophysical data remarkably well. Giese and Poggio (2003) proposed a neural model for processing biological motion that could possibly challenge our explanation of the data in terms of a form-based analysis. Their model allows visual processing in two cortical streams, one along the form pathway, the other along the motion pathway. In their model, point-light stimuli only activate the end stages along the motion-based pathway, which suggests that the sequential position walker is not processed by form-based modules. Strictly speaking, however, their motion-based variant also does not rely solely on motion vector information, because the motion sensors are only roughly position invariant, so positional information conveyed by those motion sensors may still be the input basis for later sequence matching. It would therefore be interesting to see whether such a motion-based approach can replicate our data from the sequential position walker stimuli as well, specifically the dependency on number of points and frame duration and the insensitivity to lifetime.

Conclusion

Analyzing the contribution of local motion information at different frame durations, we found that identification of biological movements is possible on the basis of positional information alone and that adding local motion signals does not improve subjects' performance. In addition, we found that form-based tasks like coherence discrimination are not suitable for investigating the role of motion in biological motion perception. We suggest that dynamic form is the basis for biological motion perception and that local motion detectors may play a role that is nonspecific to biological motion analysis, such as segregating figure from background or perceiving depth structure.

REFERENCES

Aggarwal, J. K., & Cai, Q. (1999). Human motion analysis: A review. *Computer Vision & Image Understanding*, **73**, 428-440.

Ahlström, V., Blake, R., & Ahlström, U. (1997). Perception of biological motion. *Perception*, **26**, 1539-1548.

Baker, C. L., Jr., & Braddick, O. J. (1985). Temporal properties of the short-range process in apparent motion. *Perception*, **14**, 181-192.

Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences*, **99**, 5661-5663.

BERTENTHAL, B. I. J., & PINTO, J. (1994). Global processing of biological motion. *Psychological Science*, **5**, 221-225.

BONDA, E., PETRIDES, M., OSTRY, D., & EVANS, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience*, **16**, 3737-3744.

BÜLTHOFF, I., BÜLTHOFF, H., & SINHA, P. (1998). Top-down influences on stereoscopic depth-perception. *Nature Neuroscience*, **1**, 254-257.

CHUBB, C., & SPERLING, G. (1989). Two motion perception mechanisms revealed through distance-driven reversal of apparent motion. *Proceedings of the National Academy of Sciences*, **86**, 2985-2989.

COLTHEART, M. (1980). The persistence of vision. *Philosophical Transactions of the Royal Society of London: Series B*, **290**, 57-69.

CUTTING, J. E. (1978). A program to generate synthetic walkers as dynamic point-light displays. *Behavior Research Methods & Instrumentation*, **10**, 91-94.

CUTTING, J. E. (1981). Coding theory adapted to gait perception. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 71-87.

CUTTING, J. E., MOORE, C., & MORRISON, R. (1988). Masking the motions of human gait. *Perception & Psychophysics*, **44**, 339-347.

DANNENBERG, S., KREKELBERG, B., & BREMMER, F. (2002, November). *Biological motion processing in the macaque—Electrophysiology*. Paper presented at the meeting of the Society for Neuroscience, Orlando, FL.

GIESE, M. A., & POGGIO, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, **4**, 179-192.

GROSSMAN, E. D., & BLAKE, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research*, **41**, 1475-1482.

HEPTULLA CHATTERJEE, S., FREYD, J. J., & SHIFFRAR, M. (1996). Configural processing in the perception of apparent biological motion. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 916-929.

HILL, H., & POLLICK, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Sciences*, **11**, 223-228.

HOFFMAN, D. D., & FLINCHBAUGH, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics*, **42**, 195-204.

JOHANSSON, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, **14**, 201-211.

JOHANSSON, G. (1976). Spatio-temporal differentiation and integration in visual-motor perception. *Psychological Research*, **38**, 379-393.

KOZLOWSKI, L. T., & CUTTING, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, **21**, 575-580.

LANGE, J., GEORG, K., & LAPPE, M. (2002). Ideal-observer-model and psychophysical studies on the role of form information in biological motion perception. In R. P. Würtz & M. Lappe (Eds.), *Dynamic perception* (pp. 109-115). St. Augustin, Germany: Infix.

MATHER, G., RADFORD, K., & WEST, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society of London: Series B*, **249**, 149-155.

MAUNSELL, J. H. R., & NEWSOME, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, **10**, 363-402.

MCLEOD, P., DITTRICH, W., DRIVER, J., PERRETT, D., & ZIHL, J. (1996). Preserved and impaired detection of structure from motion by a "motion-blind" patient. *Visual Cognition*, **3**, 363-391.

NERI, P., MORRONE, M. C., & BURR, D. C. (1998). Seeing biological motion. *Nature*, **395**, 894-896.

NEWSOME, W. T., BRITTEN, K. H., & MOVSHON, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, **341**, 52-54.

ORAM, M. W., & PERRETT, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *Journal of Cognitive Neuroscience*, **6**, 99-116.

O'ROURKE, J., & BADLER, N. (1980). Model-based image analysis of human motion using constraint propagation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **2**, 522-536.

PINTO, J., & SHIFFRAR, M. (1999). Subconfigurations of the human form in the perception of biological motion displays. *Acta Psychologica*, **102**, 293-318.

RASHID, R. F. (1980). Towards a system for the interpretation of moving light displays. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **6**, 574-581.

SHIFFRAR, M., LICHTEY, L., & HEPTULLA CHATTERJEE, S. (1997). The perception of biological motion across apertures. *Perception & Psychophysics*, **59**, 51-59.

SHIPLEY, T. F. (2003). The effect of object and event orientation on perception of biological motion. *Psychological Science*, **14**, 377-380.

SINHA, P., & POGGIO, T. (1996). Role of learning in three-dimensional form perception. *Nature*, **384**, 460-463.

THORNTON, I. M., PINTO, J., & SHIFFRAR, M. (1998). The visual perception of human locomotion. *Cognitive Neuropsychology*, **15**, 535-552.

TROJE, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, **2**, 371-387.

UNGERLEIDER, L. G., & MISHKIN, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press.

VAINA, L. M., LEMAY, M., BIENFANG, D. C., CHOI, A. Y., & NAKAYAMA, K. (1990). Intact "biological motion" and "structure from motion" perception in a patient with impaired motion mechanisms: A case study. *Visual Neuroscience*, **5**, 353-369.

VAINA, L. M., SOLOMON, J., CHOWDHURY, S., SINHA, P., & BELLIVEAU, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences*, **98**, 11656-11661.

WEBB, J. A., & AGGARWAL, J. K. (1982). Structure from motion of rigid and jointed objects. *Computer Vision & Image Understanding*, **19**, 107-130.