Masking the motions of human gait

JAMES E. CUTTING, CASSANDRA MOORE, and ROGER MORRISON
Cornell University, Ithaca, New York

In three experiments we tried to mask the motions of human gait. We represented human walkers as a set of 11 computer-generated elements on a display monitor, moving as a nested hierarchy of motions that mimicked the motions of the head and major joints. The walker was seen in sagittal view, facing either right or left and walking as if on a treadmill. On the walker was superimposed a simultaneous mask composed of elements with the same brightness, shape, and subtense as those of the walker. We varied the mask parameters—particularly the number of elements and style of motion—to discern what masks best camouflaged the walker's direction. In general, many kinds of masks impeded viewer performance at durations of 200 msec, but only relatively complex masks continued to impede performance to 400 msec and beyond. Four results stand out concerning the concurrent perceptual organization of target and mask. First, if the mask is easily divided into groups by its motion parameters, viewer performance with respect to the stimulus is generally impeded by increasing the number of groups in the mask. Second, the most successful masks are those composed of scrambled parts of walkers. Third, given a sufficient number of scrambled-walker elements, viewer performance does not improve above chance even at 800 msec. And fourth, this lack of improvement appears to be confined to scrambled-walker masks that share the particular gait parameters of the walker target.

Those who study the perception of certain types of events often use the term *biological motion* for patterns of movement generated by living forms. By far the most frequently studied biological motion is human gait (e.g., by Bertenthal, Proffitt, & Cutting, 1984; Bertenthal, Proffitt, & Kramer, 1987; Cutting & Kozlowski, 1977; Cutting, Proffitt, & Kozlowski, 1978; Hoffman & Flinchbaugh, 1982; Johansson, 1973, 1975; Proffitt, Bertenthal, & Roberts, 1984; see also Cutting, 1982). In these studies, "motion patterns" consist of the movements of small patches of light, typically less than a dozen, that are mounted on individuals as they perform certain acts, particularly as they walk. The ability of viewers to organize the array into a coherent percept is interesting and important because, regardless of one's theoretical perspective, it is not obvious how such a complex set of motions should be analyzed.

Motion studies show that human gait is discerned rapidly. Johansson (1976) reported that naive viewers can perceive the human form and identify its action (doing jumping jacks, pushups, etc.) in displays with durations of only 200 msec, corresponding to about five frames of film sequence. This is impressive because naive subjects rarely perceive the human form as a static stimulus (Kozlowski & Cutting, 1977). The small amount of motion information provided by a few subsequent frames is enough for the human visual system to organize the elements into a coherent structure (see also Doner, Lappin, & Perfetto, 1984; Lappin, Doner, & Kottas, 1980).

What is not known about the motions of human gait is whether they can be masked, and if so by what. The goal of these experiments was to find out the parameters necessary for such masking. Since time-courses of forward and backward masking do not allow easy adaptation to the study of masking of moving stimuli, a simultaneous mask was used.

The present set of experiments examined the effect, on perception, of various masks placed on the image of a walker represented by elements mounted on major joints and the head. The literature consistently calls this stimulus a *point-light walker*. The walker itself always remained stationary at the center of the display screen, as if walking on a treadmill. Pilot research quickly demonstrated that elements with shape, brightness, or subtense different from those of the walker provided no masking at all. Thus, in the present experiments the walker was camouflaged by masks consisting of elements equivalent to those composing the walker.

Pilot work also revealed that judgments of presence or absence of a walker within any mask that we could compute were far too easy. Instead, then, we asked viewers to determine which way the walker faced, right or left? This task requires some reasonably fine-grained organization of the structure of the walker, since arms and legs move in both directions, regardless of which way the treadmill walker faces. In the first two experiments, we carefully varied the phase of the walker within the step cycle so that neither static nor dynamic cues within any given region on the display would dictate which way the walker faced. Durations of the stimuli were typically brief enough so that no position on the display monitor could be scrutinized exclusively if the viewer was to perform well on the task.

Experiment 1 was initially conducted as an exploration of what is now called "structural information theory" (Buffart, Leeuwenberg, & Restle, 1983; Cutting, 1981; Leeuwenberg, 1971; Restle, 1979). Our idea was that a mask that required more parameters to specify its movement would be more effective in precluding the identification of the walker's direction. This turned out not to be true. But another aspect of information load—those parameters shared with the target stimulus—did prove to be important in later experiments. Regardless, Experiment 1 served as a guide for further mask selection. We used many different types of masks in this study, all with 22 elements in them. Our results allowed us to hone our stable of masks in Experiment 2, but we also found it necessary to increase the number of elements in them to 55. Experiment 3 replicated and extended some of the findings of Experiments 1 and 2.

EXPERIMENT 1:
ON TRYING TO MASK POINT-LIGHT WALKERS WITH A FEW ELEMENTS

Method

Viewers and equipment. Thirty-two viewers participated individually: 20 in Condition 1 and 12 in Condition 2. All were members of the Cornell University community, 19 to 25 years in age, and with normal or corrected-to-normal vision. Each was paid $4. They sat in a dimly lit room 1 m from a Hewlett-Packard (HP) 1350S vector-plotting graphics monitor with a P31 phosphor. Stimuli were driven by an HP1000L series minicomputer.

Stimuli. The basic walker stimulus consisted of 11 elements, moving in such a way as to simulate the motions of a person walking. Positions of elements simulated the movements of all major joints (both elbows, wrists, knees, and ankles, and one shoulder and hip) and the head (Cutting, 1978b); one wrist, one elbow, one ankle, one knee, and the visible hip were appropriately and periodically occluded during the step cycle. Motion was generated through presentation of as many as 12 successive frames completing one step. In Condition 1, each frame was presented for 67 msec (Condition 2 will be discussed later). As stated above, the walker did not move across the screen but was stationary at midscreen as if walking on a treadmill. The walker subtended a height of 4° visual angle and a width of 1.5°. Each element subtended 0.2° and consisted of a blurred @ symbol. The walker could face either right or left, and this stimulus was all that was presented on *no-mask* trials.

In all other types of trials, 22 other elements were added to the walker and formed what we call a mask. Initial positions of these were randomly determined within a rectangular area 6° × 3° centered on the walker. Nine different types of masks were used in this study; later, for purposes of discussion, these will be condensed to five. Static target/mask relations are indicated in Figure 1.

In the *static* mask, the additional elements did not move during the stimulus trial. While the walker trod, the masking elements remained still. In the *dynamic* mask, different positions for each element were generated randomly for each frame. This prevented them from being perceived as having any coherent motion. In the *linear* mask, the masking elements moved as a unit diagonally across the display space, either downwardly to the right (45°) or downwardly to the left (135°), wrapping around (as if on the surface of a torus) as they disappeared from the masking area. The velocity of these elements was about 2°/sec, which resulted in the phenomenal impression of a rigid plane covered with elements being moved over the stimulus and beneath a rectangular aperture.

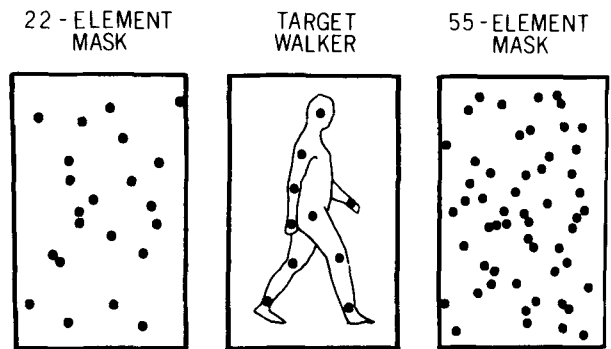22 - ELEMENT MASK   TARGET WALKER   55 - ELEMENT MASK

Figure 1. An indication of the point-light walker facing right, with masks of 22 and 55 elements. Experiments 1 and 3 used 22 masking elements, Experiments 2 and 3 used 55, and Experiment 3 also used 33 and 44. The rectangular outlines delimit the area within which masking elements could appear. In the experiments, masks were superimposed on the walker, and of course no outline of the walker appeared.

In two masks, the elements moved in circular paths. In the *synchronous circles* mask all of the distractors moved in phase; in the *asynchronous circles* mask the elements were divided into two groups, 120° out of phase. In both cases, the diameters of all circular paths were equal to the upper leg length of the walker, and all paths completed a circle in 24 frames. There were also two masks in which the distractors moved along elliptical paths, chosen because the treadmill walker's shoulders and hips traverse ellipses. These ellipses had a minor vertical axis equal to two-thirds the diameter of the circles, and a horizontal radius 1.5 times the diameter. Again they completed a full revolution in 24 frames. The *synchronous ellipse* and *asynchronous ellipse* masks were analogous to the two circular masks.

The final two masks had the same motions as the walker, but in a scrambled spatial arrangement. In the *synchronous scrambled walker* mask, all homologous walkers' parts moved in phase; for example, all three right-ankle elements moved together but in different locations on the screen, and so forth. In the *asynchronous scrambled walker* mask, the phases of the elements representing the same joint were shifted by 120°; that is, if the target-walker elements were considered referents at 0°, one scrambled walker was at 120°, and the other was at 240°. For circular, ellipsoidal, and scrambled-walker masks (as with the linear masks), when a dot reached the edge of the masking area, it wrapped around to the opposite side. One additional stimulus was created for a familiarization procedure, with masking elements moving in a linear fashion, but horizontally rather than diagonally.

Each mask/target combination was generated in four different tokens. The stimulus walker could face either right or left, and the mask could move either one direction (right, or clockwise) or the other (left, or counterclockwise). Of course for no-mask, static, and dynamic patterns, there was no mask direction, so these stimuli varied only with respect to the direction in which the walker faced. Two different tokens were generated for each right and left version for these masks.

Because each stimulus took about 40 sec to generate, all were precomputed and stored as data files. A total of 38 experimental stimuli of 12 frames each were stored: two that contained no mask, four each of the static, dynamic, and linear masks, and four each of the circle, ellipse, and scrambled-walker masks in both synchronous and asynchronous form.

Familiarization. Participants were told that they would view complex stimuli on a computer-driven display, all with elements that mimicked the motions of a person walking. A series of six blocks
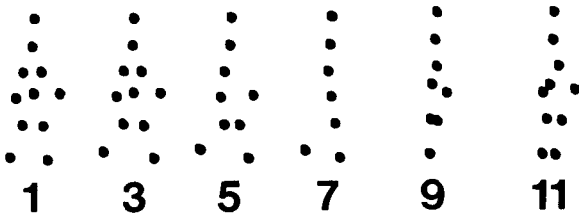
Figure 2. Sample frames taken out of the step-cycle for right-facing walkers.

of trials were then run to familiarize them with the task and stimuli. All of these trials, as well as those in the three experiments, began with a warning tone and a trial number that stayed on the screen for 1 sec. After the stimulus was then displayed, the screen remained blank until the viewer responded; then the warning tone for the next trial was sounded. The task was thus self-paced and response times were not measured.

The first familiarization block presented walkers with no masks, facing right and left. These stimuli lasted 72 frames, or about 4.8 sec, during which the walker completed six steps. The second practice block used the horizontal linear mask, and had the same duration as the no-mask trials. The remaining four practice blocks used the same two stimuli as the first two blocks, but with progressively shorter durations. The third and fourth presented 12 frames (804 msec) of the no-mask and horizontal linear masks, respectively. The fifth and sixth blocks presented three frames (201 msec) of these stimuli. The number of trials in each practice block was not fixed. Instead, each viewer continued until achieving eight consecutive correct responses. This served as a check on the viewer's ability to perform the task. Most were able to reach this criterion in less than 30 trials in Experiments 1 and 3, although not in Experiment 2, where more complex masks were used; when they did not, the familiarization sequence was truncated.

Experimental procedure. Each viewer then participated in 392 randomly presented trials over the course of about 40 min: (10 types of stimuli, the no-mask stimulus plus 9 mask types) × (3 durations of presentation) × (3 parts of the step cycle) × (4 tokens, or right and left combinations) + 32 practice trials of varied sorts that began the sequence.

Condition 1. The stimulus/mask durations were 67, 201, and 804 msec, consisting of 1, 3, and 12 stimulus frames, respectively. The 12-frame stimuli spanned half a step cycle (one step) and were not sampled in any way—that is, Frames 1-12 were always presented. As noted above, the 1- and 3-frame stimuli sampled different parts of the step cycle. In Frames 1, 2, and 3 of any stimulus/mask the walker's arms and legs were most outstretched; in Frames 7, 8, and 9 they were most aligned. Sample frames for the left- and right-facing walkers are shown in Figure 2. The 1-frame stimuli were either the 2nd, 6th, or 10th frame of the stored stimulus; the 3-frame stimuli contained Frames 1-3, 5-7, or 9-11. This manipulation controlled for the position of the walker's limbs, and made it impossible for the viewer to monitor a single location on the display and achieve good performance. Twelve-frame stimuli were presented as often as 3- and 1-frame stimuli.

Condition 2. This condition was identical to Condition 1 except that the durations of the stimuli were made equal. That is, the 3- and 1-frame stimuli were increased proportionally so that they had the same duration as the 12-frame stimuli. The purpose of this manipulation was to try to separate possible effects of duration from the number of different frames presented to the viewer.

Results

There were no differences of any kind between Conditions 1 and 2. This suggests that all effects are due to the

amount of visual information in the target/mask relations, not in the rate of information uptake. This result is consistent with Doner et al. (1984), who found that viewers' ability to detect a sphere with random dots on it in a field of similar noise dots was a function of the number of frames presented, not stimulus duration. In all further analyses the data from the two conditions were pooled.

Preliminary analyses also revealed no differences among several types of masks, so these were also pooled to aid discussion; in addition, several others were pooled for conceptual reasons. The pairs of masks with similar results were the synchronous circles and ellipses, the asynchronous circles and ellipses, and the synchronous and asynchronous scrambled walkers; here, however, all circles and ellipses were pooled for initial discussion. Because of pooling and the differences in numbers of responses per mask category, analyses of variance were performed on individual viewers' percent correct responses. In the discussion of masking effects, the terms used for the resulting six types of stimuli are: no mask, linear mask, dynamic mask, circular mask (for both ellipses and circles, both synchronous and asynchronous), and scrambled-walker mask (for both synchronous and asynchronous scrambled walkers).

As expected, there were reliable effects of number of frames $[F(2,62) = 352]$, mask type $[F(5,155) = 98.5]$, and their interaction $[F(10,310) = 16.7;$ all $ps < .001]$. Since our main interest was in the interaction of mask type and number of frames, separate analyses were then performed and comparisons made using a Scheffé test. The confidence interval $(\alpha = .01)$ was 10.0% in all three frame conditions. Overall results are shown in the three panels of Figure 3.

In one frame, of course, there is no motion. Nevertheless, viewer performance given the no-mask stimulus was quite good, just above 90%. This means that static configural information in the walker at different parts of the step cycle is sufficient for observers to detect the direction the walker faces. Such information, of course, would also be available in all other stimuli, but since there is also no mask motion one would expect that in judging walker direction, performance in response to all masks should be equal. Moreover, to the extent that the locations of the masking elements concealed this configural information, performance in judging the direction a walker faces should be poor, if not at chance level. Indeed, performance across the six masks was 59%.

In three frames, motion appears, but the pattern of results is little different. Viewer performance under the no-mask condition improved to 95% and remained reliably above that of all masking conditions, in which performance improved to an average of 75%. There were no clear-cut differences among mask types. Most surprising to us was that the static mask was as effective as other masks.

In 12 frames, however, masks yielded interesting differences. The static and linear masks became ineffective, with viewer performance not reliably different from that in the no-mask condition. Circular, dynamic, and
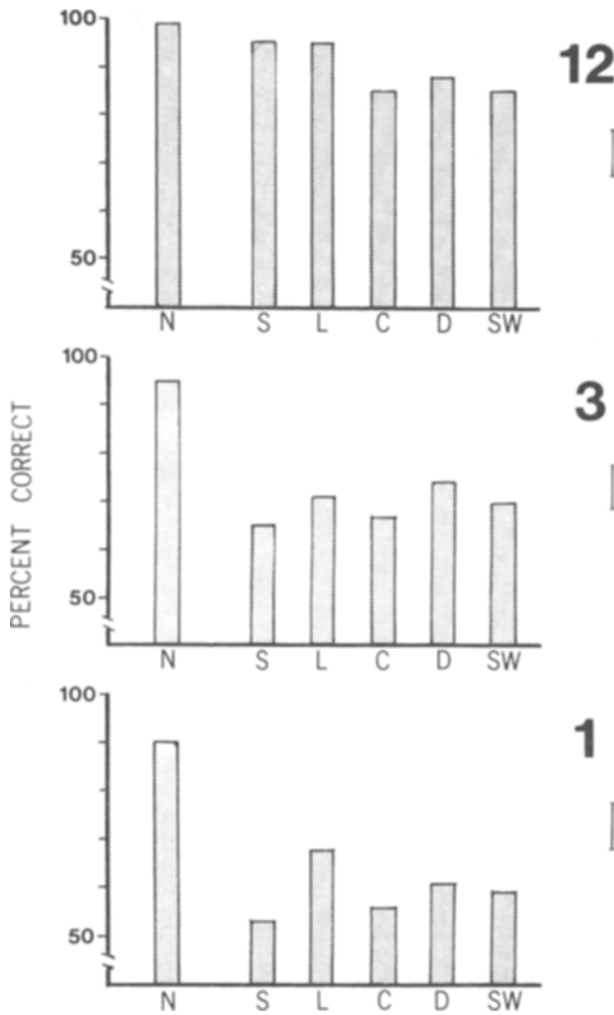
Figure 3. Results of Experiment 1 with 22 masking elements for 1-, 3-, and 12-frame masks of various types. N = no mask, S = static mask, L = linearly translating mask, C = circularly and ellipsoidally translating mask, D = dynamic noise mask, and SW = scrambled-walker mask. Sheffé confidence intervals (α = .01) are given for each frame condition.

scrambled-walker masks, on the other hand, yielded performance reliably worse.

One interaction, not shown in Figure 3, is that asynchronous circular and ellipsoidal masks were more effective than synchronous circular and ellipsoidal masks [$F(1,31) = 12.4, p < .002$], as shown in Table 1. This effect occurred for both 3- and 12-frame stimuli ($ps < .05$), and is the only result generally consistent with structural information theory (Buffart et al., 1983). That is, the coding scheme for an asynchronous mask of whatever type must have more parameters than that for a synchronous mask, and one might predict that the information load of the mask would impede identification of the stimulus.[1] The synchronous/asynchronous differ-

ence provoked one aspect of Experiment 2, in which we investigated the possible perceptual organization of masks as part of picking out the direction of the target walker.

Four other effects should be considered. First, performance with respect to all mask types improved with stimulus duration (all $ps < .01$), a result pertinent to Experiment 2. Second, a reliable difference was obtained in terms of what part of the step cycle was sampled on a stimulus trial [$F(2,62) = 56.6, p < .001$]. Judgments of walker direction were easiest in the trials that presented arms and legs most outstretched (75%), more difficult in those with intermediate extensions (68%), and most difficult on those with arms and legs held most closely to the body (59%). This effect was similar and reliable for 1- and 3-frame stimuli. Third, viewers appeared to have a bias for walkers facing right, where performance was 77%, over those facing left, where performance was 67% [$F(1,31) = 8.5, p < .001$]. This may be due to a propensity to respond with the dominant (right) hand in situations of uncertainty. And fourth, there were several uninterpretable interactions that involved the different tokens and cycle parts for each stimulus, a result found also in Experiments 2 and 3.

Discussion

Two types of simultaneous masking effects can be discerned. First, configural masking effects are substantial in static (1 frame) displays: Performance in single-frame trials with simultaneous masks was substantially below that for no masks (59% vs. 90%, respectively). Second, when motion is added to the display, performance improves for all stimuli, particularly those that are masked—to 75% and 89%, for 3- and 12-frame masked stimuli, respectively. In addition, given the near-identity of results in Conditions 1 and 2, it is the motion (or change) in the display, not the duration, that is important.

Reliable differences in performance with respect to different masks appeared only when stimuli had 12 frames, not 3. Thus, at least for masks of 22 elements, motion properties of the mask have few effects that accrue within 3 frames. The exception is that at 3 frames (as well as 12) there was a reliable effect of grouping (synchrony vs. asynchrony) in the circular and ellipsoidal masks. Thus there is ample evidence that perceptual or-

Table 1
Percent Correct Performance as a Function of the
Number of Element Groups in the Mask

	Element Groups		
	1	2	5
Experiment 1	84	68	
Experiment 2	77	73	69

Note—In Experiment 1, synchronous and asynchronous circularly and ellipsoidally translating masks had 22 masking elements; in Experiment 2, linearly and circularly translating masks had 55 masking elements. Data are for all motion conditions: means across 3 and 12 frames in Experiment 1 and across 3, 6, and 12 frames for Experiment 2.

ganization occurs for the target over 3 frames and within 200 msec (configural information is available in a single frame), but somewhat less that it occurs for the mask within 3 frames regardless of duration.

We think that the most important effect of this study, however, is that it is quite difficult to mask the motions of a point-light walker at all. Since little masking, and even less differential masking, occurred with 22 masking elements, we next tried 55.

EXPERIMENT 2: MASKING POINT-LIGHT WALKERS WITH MORE ELEMENTS

Method

Twelve different viewers from the Cornell University community were paid $5.00 to participate in a similar experiment lasting about 1 h. The same apparatus was used and similar mask/target stimuli were generated, but this time with 55 masking elements rather than 22. The no-mask stimulus again had only the 11 elements of the walker.

Five types of simultaneous masks were generated: linear, circular, dynamic, and scrambled-walker, as before, and a fifth mask that combined aspects of all four, which we call a *mixed* mask. No ellipsoidal or static masks were used. Dynamic and scrambled-walker masks were generated as before. Linear and circular masks were generated in three types. Masking elements could move uniformly as a single group, in two groups in different directions or phases, or in five different groups of directions or phases. The rationale for this manipulation stemmed from the effect, in Experiment 1, of synchronous versus asynchronous circular and ellipsoidal masks, where the latter proved much more effective.

We thought that if perceptual grouping occurred for the mask, it would be worthwhile to investigate the number of groups that the mask might be segregated into. One-group linear masks had their elements move either to the right (144°) or left (216°); two-group mask elements moved in both directions at these angles (22 one way, 33 the other, counterbalanced across tokens); and five-group masks had 11 elements move vertically (0°) and 11 each move at 72°, 144°, 216°, and 288°. All 55 circular mask elements moved either clockwise or counterclockwise. The one-group masking elements moved as a group, each starting at 12 o'clock (or 0°); two-group masks moved 180° out of phase (22 in one group, 33 in the other, counterbalanced across tokens); and five-group masks moved in equal 11-element groups, at phases of 0°, 72°, 144°, 216°, and 288°. Mixed masks had 11 elements moving linearly to the right (at 144°), 11 linearly to the left (216°), 11 moving circularly, 11 that were dynamic (changing position randomly across frames), and 11 that were scrambled-walker parts.

Rather than present 1, 3, and 12 frames as in Experiment 1, we eliminated the single-frame condition and explored the region between 3 and 12 frames. Thus, three different durations were employed: 201 msec (3 frames), 402 msec (6 frames), and 804 msec (12 frames). No condition was added here that equated durations, unlike in Experiment 1. Cycle parts were sampled as before, with Frames 1-3, 5-7, and 9-11 for the 3-frame stimuli, and Frames 1-6, 4-9, and 7-12 for the 6-frame stimuli. Twelve-frame trials had no cycle-part manipulation, but were represented equally often. Four tokens were generated for all mask types, with walker direction counterbalanced with mask movement.

The procedure was as before, a similar familiarization sequence followed by 396 experimental trials (the equivalent of 11 mask types: no-mask represented twice as often as the others, 3 linear, 3 circu-

lar, 1 dynamic, 1 scrambled walker, and 1 mixed) × (3 durations) × (3 cycle parts) × (4 tokens).

Results and Discussion

Again, there were reliable effects of duration [$F(2,22)$ = 21.5], mask type [$F(5,55)$ = 66.4], and their interaction [$F(10,110)$ = 4.39; all ps < .001]. And again, our main interest is viewer performance at different stimulus durations for different masks. As before, analyses of variance were performed on the percent correct performance for all mask types. Mean values are shown in Figure 4, and most results will be discussed in terms of Scheffé test confidence intervals (α = .01), which average 14.9% across frame conditions.
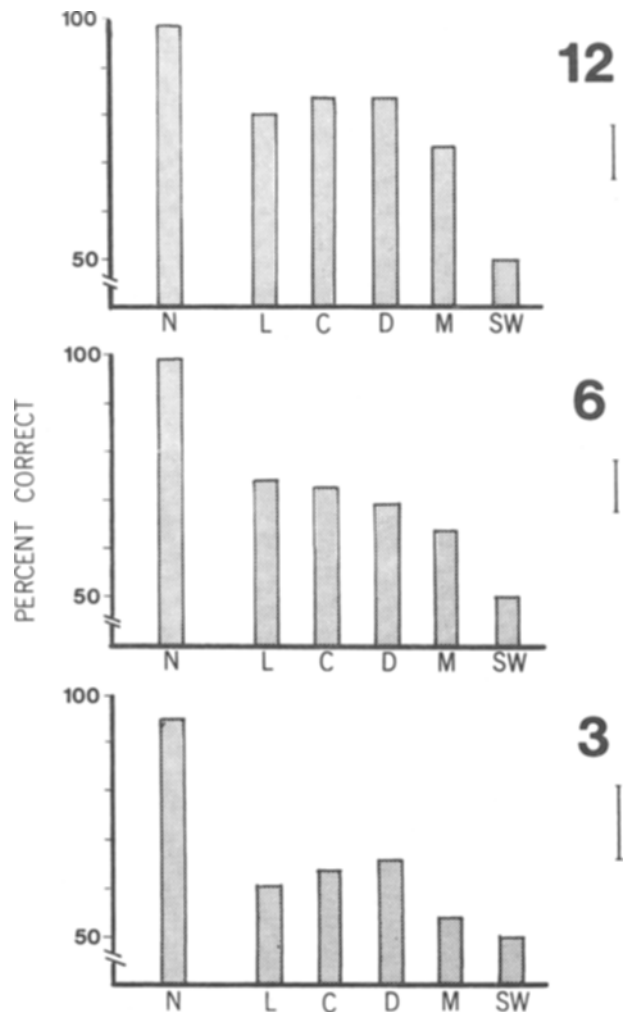

Figure 4. Results of Experiment 2 with 55 masking elements for 3-, 6-, and 12-frame masks of various types. N = no mask, L = linearly translating mask, C = circularly translating mask, D = dynamic noise mask, M = mixed mask, and SW = scrambled-walker mask. Scheffé confidence intervals (α = .01) are given for each frame condition.

In general, performance was worse than in Experiment 1 for all mask types. This effect is surely due to the increased number of masking elements, a notion explored in Experiment 3. At three frames, performance was already near ceiling for no-mask stimuli (95%); reliably more difficult to discern were the walkers covered with linear, circular, and dynamic masks (at 62%, 63%, and 68%); and reliably below these were walkers covered with scrambled-walker parts and mixed masks (at 53% and 50%).

At six frames, performance on unmasked stimuli was nearly perfect (99%); performance with linear, circular, dynamic, and mixed masks was reliably worse (at 74%, 73%, 69%, and 64%); and performance with scrambled-walker masks was still at chance (50%).

Twelve-frame stimuli revealed a similar pattern: near-perfect performance for no masks (98%); somewhat worse performance for linear, circular, dynamic, and mixed masks (82%, 84%, 84%, and 74%); and chance performance (49%) with scrambled-walker masks.

Another way to express these results is to say that there was reliable improvement in determining walker direction with increased duration for all stimulus types (all $ps < .02$), except the no-mask stimuli ($p = .07$) and the scrambled-walker stimuli ($p = .75$). Performance on these was at ceiling and floor, respectively, regardless of stimulus duration. We consider the main result of this study to consist of the lack of improvement with increased stimulus duration for scrambled-walker masks.

In addition, there was a reliable effect of number of masking-element groups (1 vs. 2 vs. 5) for the linear and circular masks, $[F(2,22) = 6.15, p < .008]$. Here the effect was reliable at 3 and 6 frames ($ps < .05$), but unlike in Experiment 1, not at 12 ($p = .37$). Table 1 shows performance by number of masking-element groups. This result, in conjunction with the result in Experiment 1 for synchronous versus asynchronous circles and ellipses, suggests that grouping principles occur for the mask as well as the target, and do so within the first 200 msec. Thus, when a viewer tries to discern walker direction, a unitary mask (one that forms a single group) is less effective than one that forms two groups, which is in turn less effective than one forming five.

Finally, there were reliable effects for the cycle part selected in the 3- and 6-frame stimuli $[F(2,22) = 7.66, p < .003]$, and for the direction that the walker faced $[F(1,11) = 8.03, p < .02]$. The magnitudes of these were like those in Experiment 1. And, as before, there were several uninterpretable interactions involving cycle parts, walker directions, and tokens.

Given these results, two ideas seemed worth pursuing. First, we investigated the effects of masks with 22, 33, 44, and 55 elements, making sure that performance differences in Experiments 1 and 2 were due to the number of elements in the mask. Second, we explored aspects of the scrambled-walker masks. Were their effects due to the identity of mask parameters with the walker targets, or were they due to a general similarity of vector paths?

EXPERIMENT 3: MASKING POINT-LIGHT WALKERS WITH VARIED NUMBERS OF ELEMENTS

Method

Eight observers from the Cornell University community participated; 4 of them had participated in Experiment 2. Each was paid $5. The same apparatus and procedure were used as before.

Four general types of mask stimuli were generated: one with circular masks and three with various types of scrambled walkers. The first of the latter type was like that of Experiments 1 and 2, with walker parts having the identical vector paths of the walker itself. In the second, masking elements were generated from a walker that took steps twice the size of the normal walker; and in the third, step size was one-half normal. This type of manipulation changes the motions of all elements (Cutting, 1978a), but particularly noticeable are the swings of the upper arm and upper leg: swings of the "normal" walkers are 19° and 27°, respectively, for the femur at the hip and the humerus at the shoulder; swings for the double-step walkers are 45° and 53°; and those for the half-step walkers, 9° and 13°. Elsewhere, extents of motions in double- and half-step walkers also vary roughly according to the ratios of step sizes. Also, since step-cycle time was the same in the three simulations, the instantaneous velocities in the mask vary accordingly. In general, the velocities are doubled for the double-step walker and halved for the half-step walker. These normal-, double-, and half-step walkers were then multiply scrambled to form the masks; all target walkers were as before, "normal" in their step size.

In addition, all types of masks varied in number of elements: 22, 33, 44, or 55. Phase was manipulated for the circular masks, with two groups of 11 elements 180° out of phase, three groups at 120°, four groups at 90°, and five groups at 72°. Phase was not varied in the scrambled-walker masks, since such manipulation produced no effect in Experiment 1.

More durations were sampled in this study, with 3, 6, 9, and 12 frames used. This time we decided against varying the part of the cycle sampled, so all trials began with Frame 1 (where arms and legs were nearly aligned). All types of masks were computed in four tokens, two each for walkers facing left and right. Thus, viewers participated in a random ordering of 320 trials: 5 mask types (no mask, circular, and three scrambled-walker types) × 4 durations × 4 masking element numbers (22, 33, 44, and 55 elements, with no-mask trials represented as often as the others) × 4 tokens.

Results and Discussion

Some results were as expected. The number of masking elements mattered $[F(3,21) = 5.74, p < .005]$, with viewer performance at 84%, 80%, 75%, and 74%, respectively, with 22, 33, 44, and 55 masking elements. In addition, duration also affected performance $[F(3,21) = 18.2, p < .001]$, but not in the predicted way. Performance was 65%, 80%, 83%, and 83% at 3, 6, 9, and 12 frames, respectively, with no reliable differences among mask/target stimuli longer than 201 msec. This leveling off of performance is almost surely due to the fact that, since all stimuli began with the first frame and since the sixth frame generally presented the walker with arms and legs most outstretched, little new information accrued after that.

More interesting, however, are the patterns of relations among types of stimuli. First, there were no errors whatsoever for no-mask stimuli at any duration, so these were

Table 2
Percent Correct Performance in Experiment 3 as a Function of
Mask Type and Number of Masking Elements

Mask Type	Masking Elements			
	22	33	44	55
Circular	80	84	71	80
Scrambled-Walker				
Normal-Step	85	78	75	54
Double-Step	87	73	75	80
Half-Step	84	84	78	80

removed from further analyses. Second, although there was no difference in overall performance among mask types [$F(3,21) = 1.78, p = .18$], there were two interesting interactions with mask type that we think are the main results of this study: one with number of masking elements [$F(9,63) = 3.31, p < .002$] and the other with both number and duration [$F(27,189) = 1.58, p < .05$].

The first interaction is shown in Table 2. Across the varying numbers of masking elements there was *no* reliable effect on viewer performance in judging walker direction under circular masks, or double-step or half-step scrambled-walker masks [all $Fs(3,21) < 2.35$, all $ps > .10$], but there was a reliable effect for normal-step scrambled-walker masks [$F(3,21) = 9.9, p < .001$]. This surprising result suggests, among other things, that it is the exact style of motions in the normal-step scrambled-walker mask that camouflages the walker's direction, not a general pattern of pendular motions.

The second interaction, which is more complex, suggests the same conclusion. It centers on a replication of a result from Experiment 2. First, performance with respect to all masks—circular and scrambled-walker—improved with duration [all $Fs(3,21) > 3.7$, all $ps < .03$]. But only the normal-step scrambled-walker masks provided an interaction of duration and the number of masking elements [$F(3,21) = 9.86, p < .001$], and here there was general improvement in judging the direction a walker faced with all numbers of masking elements—except those with 55, where performance hovered between 46% and 59%, as in Experiment 2.

This last result, like the previous interaction, suggests that the visual system is particularly sensitive to the exact vector paths of the masking elements, not just their general pendular swing. When sufficient numbers of masking elements with the exact vectoral parameters as the walker are superimposed on it, walker direction simply cannot be determined; when those parameters are violated, performance improves with duration as one would expect.

GENERAL DISCUSSION

We will consider first three theoretical issues about perceptual organization and motion, then turn to a concluding discussion and summary.

Masking of Wholes or Parts?

One question seems to arise inherently in any consideration of the masking of a stimulus as complex as a human walker: Does the masking cover the whole or just parts? The answer, we believe, is parts. But the parts are not smaller than an arm or a leg, at least not for Experiments 1 and 2. Our rationale for this conclusion is threefold:

First, knowledge plays an important role in the recognition even of unmasked walker displays. For example, only about 85% of naive viewers see a human form when presented with a treadmill walker identical to those used here (Proffitt et al., 1984). Thus, the stimulus cannot really be said to organize itself spontaneously and fully for a naive viewer, at least not as a human form. All viewers, however, report seeing bendable parts moving against one another—or arms and legs.

Second, when the viewer knows that a walker is to be presented, the entire stimulus need not be present for recognition to occur. Kozlowski and Cutting (1977, 1978) and Cutting (1978a) showed that a few lights, such as those on the arms or the legs, are sufficient even to determine the gender of the walker. If gender can be determined from walker fragments, surely the presence of a walker can also be determined from fragments if one knows a walker may appear. Thus, there is no reason for a knowledgeable viewer ever to look for more than an arm or a leg.

Third, throughout the familiarization and experimental procedure, the task we set for our viewers was to recognize the direction faced by an increasingly well-known and well-scrutinized walker stimulus. The viewer was free to adopt whatever strategy he or she might choose. When asked afterwards about how they did the task, most viewers claimed to have looked near the center of the display for information about arms or legs. They were aware that no single location would give away the answer, due to variation in phase, but a bent arm or a bent leg would suffice. This means that they had to organize perceptually at least three elements—a shoulder, elbow, and wrist, or a hip, knee, and ankle—subtending a region of about 1.5° in width by 2° in height, regardless of whether they were looking for an arm or a leg. An average of 1 out of every 6 masking elements would also lie in this region in a given frame, or about 3 to 9, depending on the experiment.

. Thus, we claim that the task the viewers were performing included at least two parts: The first was something of a straight filtering task that ignored about 5/6 of the display area; the second was an organizational task in which pendular relations among triads of elements were to be discerned and separated from a number of distractor elements.

In Experiment 3 something a bit different could occur. Since all stimuli, regardless of duration, began with Frame 1 (with arms and legs nearly aligned with the body), viewers could scrutinize a quite local area of the

display, perhaps attending to the presence or absence of a single element. What the results of this experiment show is that even the most highly informed viewers make substantial errors due to the locations of distractor elements in the mask.

Thus we claim that our results indicate how the patterns within various masks can impede performance for a knowledgeable observer, one who may even know exactly where to look.

Minimality and Structural Information Theory

The ability of the visual system to discover underlying structure in a visual pattern is often discussed in terms of a simplicity, or a minimum, principle (Cutting & Proffitt, 1982; Hatfield & Epstein, 1985; Hochberg & McAlister, 1953; Leeuwenberg, 1971; Restle, 1979). Indeed, Cutting (1981) worked out an application of this idea to dynamic point-light displays of walkers. Here, following information theory (Chaitin, 1977; Kolmogorov, 1968), it is as if the visual system assembles the simplest possible program to interpret the pattern of motions. This process is thought to proceed by reducing the number of parameters needed for the description, as the description becomes the percept. Unfortunately, there is no general agreement on what parameters are to be included in such a process, and even when one might agree on them, it is clear that minimal relations are not always perceived (Cutting, 1987; Hochberg, 1986).

Nonetheless, there are two results from these experiments that would seem to support a minimum principle and structural information theory. In Experiments 1 and 2, there was an effect of grouping for masks, where masks with two and five groups impeded recognition of walker direction much more than did masks with one group. It is as if grouping of the target and mask co-occur, even as the number of element groups in the mask impedes the grouping of the target. The simpler the mask, that is, the fewer the number of element groups in it, the easier it is for the viewer to organize the target. Second, in Experiment 3, the masking effects of normal-step walkers were much more striking than those of double- and half-step scrambled walkers. Such a result suggests that appropriate information loads for a mask are best computed when using the particular parameter values of the target stimulus.

Form, Motion, and Masking

Treisman, Russell, and Green (1975) investigated masking and motion in a study of iconic memory. Krumhansl (1984) adapted their procedure in a study of the masking of form and motion. In her study, rectangular and ellipsoidal patterns of elements (forms) rotated or remained stationary during the course of a trial. These were simultaneously masked with a random pattern of elements, also rotating or stationary. The results showed that form and motion were processed independently.

In the present three experiments, however, there is ample evidence that the perception of the particular form of the walker was impeded by the particular motions of the mask. The reason for this difference seems clear. The configuration of the target in Krumhansl's (1984) study did not change over time, it simply rotated or remained stationary. In the present experiments, the configuration of the walker changed continuously while only the underlying form remained constant.

Of course, many other researchers have studied the perception of moving objects under noise. Typically, this research is couched in terms of the *correspondence problem*: the matching of elements across frames when many different matches are possible. Ullman (1979), for example, showed that elements mounted on two moving cylinders could be easily segregated, and Petersik (1979), Lappin et al. (1980), and Doner et al. (1984) showed that a sphere with elements mounted in it or on it could be easily recognized in a field of noise elements of the same type. Todd (1985) showed that viewers could discern objects in noise at signal-to-noise ratios of greater than 1-to-6. None of these objects, nor those of Treisman et al. (1975) and Krumhansl (1984), are as complex as walkers. In particular, all were rigid; walkers are not.

Some Conclusions

We think four results of these studies are most important. First, from Experiments 1 and 2, we have evidence that the grouping of elements in the mask affects viewer performance in identifying which way a walker faces. That is, when the elements of a mask are easily grouped into a single object, performance at judging a point-light walker's direction was very good after 800 msec, but less so when those elements form two or even five groups.

Second, the most successful masks were generally those composed of scrambled walkers. In Experiment 1, with 22 masking elements, this effect did not accrue, with scrambled walkers providing no more interference than dynamic or circular masks; but in Experiments 2 and 3, with more masking elements, differences did occur.

Third, scrambled-walker masks with sufficient numbers of masking elements (here 55) were so effective that performance in judging the walker's direction was at chance regardless of stimulus/mask duration, at least up to 800 msec in Experiments 2 and 3. Substantial release from masking did occur with fewer masking elements.

Fourth, and perhaps most surprising, two masking effects from Experiment 3 seem particular to the exact style of movement in the scrambled-walker masks. First, the number of elements in the mask affected viewer performance in judging target-walker direction only when the mask shared the exact motion parameters of the target. That is, increasing the number of scrambled walkers in the mask impeded performance only when the prescrambled walkers took the same sized steps. Second, the lack of improvement with stimulus/mask duration occurred only when the masking elements shared the same vector

paths as the target walker. These two interactions suggest that processes of perceptually organizing the walker target are quite sensitive to particular patterns of motion; they may be disrupted relatively easily by movement patterns identical to those in the walker, but not for those that in many ways are similar but not identical.

REFERENCES

BERTENTHAL, B. I., PROFFITT, D. R., & CUTTING, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions. *Journal of Experimental Child Psychology, 37*, 213-230.

BERTENTHAL, B. I., PROFFITT, D. R., & KRAMER, S. J. (1987). Perception of biomechanical motions by infants: Implementation of various processing constraints. *Journal of Experimental Psychology: Human Perception & Performance, 13*, 577-585.

BUFFART, H., LEEUWENBERG, E., & RESTLE, F. (1983). Analysis of ambiguity in visual pattern completion. *Journal of Experimental Psychology: Human Perception & Performance, 9*, 980-1000.

CHAITIN, S. J. (1977). Algorithmic information theory. *IBM Journal of Research & Development, 21*, 350-359.

CUTTING, J. E. (1978a). Generation of synthetic male and female walkers through the manipulations of a biomechanical invariant. *Perception, 75*, 393-405.

CUTTING, J. E. (1978b). 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*, 393-405.

CUTTING, J. E. (1982). Blowing in the wind: Perceiving structure in trees and bushes. *Cognition, 12*, 25-44.

CUTTING, J. E. (1987). Perception and information. *Annual Review of Psychology, 38*, 61-90.

CUTTING, J. E., & KOZLOWSKI, L. T. (1977). Recognizing friends by their walk. *Bulletin of the Psychonomic Society, 9*, 353-356.

CUTTING, J. E., & PROFFITT, D. R. (1982). The minimum principle and the perception of absolute, common, and relative motions. *Cognitive Psychology, 14*, 211-246.

CUTTING, J. E., PROFFITT, D. R., & KOZLOWSKI, L. T. (1978). A biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception & Performance, 4*, 357-372.

DONER, J., LAPPIN J. S., & PERFETTO, G. (1984). Detection of three-dimensional structure in moving optical patterns. *Journal of Experimental Psychology: Human Perception & Performance, 10*, 1-11.

HATFIELD, G., & EPSTEIN, W. (1985). The status of the minimum principle in the theoretical analysis of visual perception. *Psychological Bulletin, 97*, 155-186.

HOCHBERG, J. (1986). Representation of motion and space in video and cinematic displays. In R. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, pp. 1-64). New York: Wiley.

HOCHBERG, J., & MCALISTER, E. (1953). A quantitative approach to figural "goodness." *Journal of Experimental Psychology, 46*, 361-364.

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. (1975). Visual motion perception. *Scientific American, 232*(6), 76-88.

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

KOLMOGOROV, A. N. (1968). Logical basis for information theory and probability theory. *IEEE Transactions on Information Theory, IT-14*, 662-664.

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

KOZLOWSKI, L. T., & CUTTING, J. E. (1978). Recognizing the gender of walkers from point-lights mounted on ankles: Some second thoughts. *Perception & Psychophysics, 23*, 459.

KRUMHANSL, C. L. (1984). Independent processing of visual form and motion. *Perception, 13*, 535-546.

LAPPIN, J. S., DONER, J. F., & KOTTAS, B. L. (1980). Minimal conditions for the detection of structure and motion in three dimensions. *Science, 209*, 717-719.

LEEUWENBERG, E. (1971). A perceptual coding language for visual and auditory patterns. *American Journal of Psychology, 84*, 307-347.

PETERSIK, J. T. (1979). Three-dimensional object constancy: Coherence of a simulated rotating sphere in noise. *Perception & Psychophysics, 25*, 328-335.

PROFFITT, D. R., BERTENTHAL, B. I., & ROBERTS, R. J. (1984). The role of occlusion in reducing multistability in moving point-light displays. *Perception & Psychophysics, 36*, 315-323.

RESTLE, F. (1979). Coding theory of the perception of motion configurations. *Psychological Review, 86*, 1-24.

TODD, J. T. (1985). Perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception & Performance, 11*, 689-710.

TREISMAN, A. M., RUSSELL, R., & GREEN, J. (1975). Brief visual storage of shape and movement. In P. Rabbit & S. Dornic (Eds.), *Attention and performance V* (pp. 699-721). London: Academic Press.

ULLMAN, S. (1979.) *The interpretation of visual motion.* Cambridge, MA: MIT Press.

NOTE

1. Structural information theory (Buffart et al., 1983) generally fails elsewhere in this study. The information loads for dynamic and scrambled-walker masks considerably exceed those for static, linear, and circular masks. Consider the following analyses: Each element in the dynamic masks must have two new parameters (x and y position) for each frame, yielding 44, 132, and 528 parameters for 1-, 3-, and 12-frame masks, respectively. The scrambled-walker masks generally need 5 parameters per element (x and y position for a pivot, length of moment arm, extent of motion, and phase), regardless of number of frames presented. This yields roughly 110 parameters. The static mask needs only 44 parameters (x and y position of each element), the linear mask needs only 46 (the 44 of the static mask plus an x and y vector), and the circular mask needs only 47 (the 44, plus a radius and two phases; the elliptical mask needs an additional radius for the minor axis). Asynchronies add only 2 additional phase parameters. Thus, the five types of masks (dynamic, scrambled-walker, static, linear, and circular) at 3 frames had 132, 110, 44, 46, and 47 (up to about 50) parameters, respectively. Yet performances were 74%, 69%, 65%, 72%, and 67%, respectively. The correlation is modest, $r = .61$, but it should be negative. For 12-frame stimuli the result is little different, $r = -.21$. Omitting the dynamic mask, which one might feel inflates the information load, offers little more: for 3-frame stimuli, $r = -.01$; and for 12-frame stimuli, $r = -.49$.