

Research Report

GLOBAL PROCESSING OF BIOLOGICAL MOTIONS

Bennett I. Bertenthal and Jeannine Pinto

University of Virginia

Abstract—*The structure of the human form is quickly and unequivocally recognized from 10 to 13 points of light moving as if attached to the major joints and head of a person walking. Recent psychophysical and computational models of this process suggest that these displays are organized by low-level processing constraints that delimit the pair-wise connections of the point lights. In the current research, these low-level constraints were rendered uninformative by a masking paradigm. The results from four experiments converged to show that the perception of structure in a point-light walker display does not require the prior detection of individual features or local relations.*

Among the varieties of motion in the world, those associated with the articulated movements of the human figure are some of the most unique and complex. Consider, for example, the nested sequences of movements produced by a ballet dancer or a basketball player. The intricacy of these movements is in large part attributable to the skeletomuscular structure of the human form, which comprises a hierarchical nesting of jointed limb segments. Although the motion patterns produced by these jointed segments are unusually complex, the human visual system appears highly sensitive to this information.

Johansson (1973) was among the first researchers to investigate systematically this perceptual sensitivity for human motions. He developed an ingenious technique for depicting the motions of a person devoid of all featural information. A person was filmed walking in the dark with small lights attached to the major joints and head. Although the resulting display consisted of a mere 10 to 13 points of light moving against a dark

background (see Fig. 1, left panel), observers quickly recognized the moving array of point lights (dubbed biological motions) as depicting a person. More recent research has revealed that observers are capable of recognizing friends, the gender of a person, and even certain dispositional characteristics from these biological motions (Kozlowski & Cutting, 1977; MacArthur & Baron, 1983).

The extremely rapid recognition of these reduced displays as depicting the human form suggests that the correct grouping of the point lights is accomplished early in visual processing (Johansson, 1973). Consistent with this view, most recent psychophysical and computational models implicate low-level processing heuristics that delimit the pair-wise connections of the point lights. For example, moving point lights that correspond to connected joints, such as the elbow and wrist, are constrained by human anatomy to remain locally rigid. By detecting these locally rigid connections, and then combining them into larger groupings, it is possible to recover the correct connectivity pattern of the human form (Hoffman & Flinchbaugh, 1982; Webb & Aggarwal, 1982).

One important implication of these models is that the processing of biological motions begins with the detection of local relations, which are then hierarchically organized into a global form. Surprisingly, this prediction has never been put to a test even though recent evidence suggests that some objects—biological objects, in particular—are perceived holistically (Farah, 1992). Indeed, subjective reports by observers viewing biological motions suggest that they perceive an emergent human form, and not the individual features or local relations making up this form (Bertenthal, 1993).

The purpose of this research was to investigate whether the global structure of a moving point-light display depicting the human form can be perceived prior to the perception of the individual features and local relations the form comprises. One approach to studying this

problem is to interfere with the perception of the individual features of the form by generating additional moving point lights and adding them to the point-light display at random locations (see Pomerantz, 1981, for a more general discussion of this approach). If the form is perceived globally, then the perception of the human figure will be unimpaired, even though the perception of the individual features and local relations will be disrupted. In the current study, this technique was used to evaluate whether a point-light display is perceived when the individual features or local pair-wise relations are masked. We conducted a series of four experiments in which the detection of the individual elements or local relations was rendered uninformative by the addition of identical moving elements that were randomly distributed in the display.

GENERAL METHOD

The detection task used in all experiments was an adaptation of one introduced by Cutting, Moore, and Morrison (1988) in which a computer-simulated point-light walker display is simultaneously masked by additional moving point lights. In our implementation of this paradigm, the target display consisted of 11 point lights moving as if attached to the major joints and head of a person walking on a treadmill (i.e., the translatory component of motion was subtracted from the display; Bertenthal & Kramer, 1984). The masking elements were generated by creating multiple sets of the absolute motion vectors of the point-light target and scrambling their spatial locations. Although the relative motions of these point lights were not constrained as were the motions corresponding to the target, the masking elements were identical to the target elements with regard to size, luminance, and motion trajectories. Figure 1 illustrates how the stimulus display was created by superimposing the masking elements onto the target display.

Address correspondence to Bennett I. Bertenthal, Department of Psychology, Gilmer Hall, University of Virginia, Charlottesville, VA 22903.

Global Processing of Biological Motions

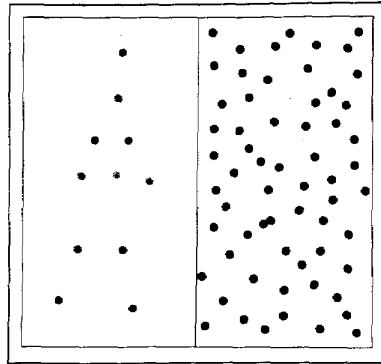


Fig. 1. Point-light walker display with and without masking elements. The left panel depicts a static frame from one of the moving point-light targets. (Note that the outline depicting the human form was not visible in the stimulus display.) The right panel depicts a static frame in which the point-light target (light circles) is embedded in an array of masking elements (dark circles). The absolute motions of the masking elements are generated using the same algorithm used to create the absolute motions of the point-light target. (Note that the target dots were the same luminance as the masking dots in the stimulus display shown to the observers.)

Stimulus displays, created with a 386 PC and a VGA graphics card, were presented sequentially with a 6.0-s inter-stimulus interval. Individual point lights subtended a visual angle of 0.25° . The target subtended a visual angle of 6.6° in height and 2.7° in width at its fullest extension, and the masking elements were distributed randomly in a region subtending $9.1^\circ \times 10.5^\circ$ on the computer screen. The target was designed to step through a complete gait cycle in 40 discrete frames at a rate of 20 frames/s. After each display was presented for 1.0 s (Experiment 1) or 2.0 s (Experiments 2, 3, and 4), observers responded by pressing a key on the keyboard. Detection performance was scored using d' as the dependent measure.¹

1. d' is a bias-free index of sensitivity to the presence of a stimulus. It is defined in terms of z , the inverse of the normal distribution: $d' = z(H) - z(F)$, where H is the proportion of correct detections of the stimulus and F is the proportion of mistaken detections.

EXPERIMENT 1: DETECTION OF POINT-LIGHT TARGETS

Stimuli and Procedure

This experiment was designed to test whether observers could detect the human form in a point-light display when the human form was masked by the superimposition of 66 additional moving point lights. Three observers judged whether the point-light target was present or absent on each trial. In the first condition, observers were presented with 50 trials in which the target was present and 50 trials in which the target was absent. On the latter trials, an additional 11 moving point lights were included, so that each display contained a total of 77 point lights. In the second condition, observers were presented with a target that was inverted on 50 trials and absent on 50 trials; the absolute motions of the masking elements were also inverted. Trials were blocked by condition and randomized within condition.

Results and Discussion

The results from this first experiment are presented in Table 1. Performance was significantly above chance when the point-light walker was presented upright, but not when it was inverted. This difference between the two conditions is statistically reliable, $F(1, 4) = 24.67$, $p < .01$. To fully appreciate the performance of the observers in the upright condition, it is useful to refer back to Figure 1 (right panel) to consider the apparent difficulty

of detecting the target when embedded in the masking elements. Nevertheless, correct detection was observed on 81% of the trials.

Given the brief duration and design of the stimulus displays, it is unlikely that the human form was detected from the absolute movements of the individual point lights. It is also unlikely that local relations in the target display (e.g., pairwise rigid relations between connected joints) were responsible for the high level of correct detection performance. These same relations were present in the inverted target condition, and yet performance was not significantly different from chance in this condition.

The finding that performance was not significantly different from chance in the inverted-target condition suggests that the perception of the human form is orientation-specific. As such, this finding is clearly contrary to the predictions of some computational theorists (e.g., Marr & Nishihara, 1978) who propose that three-dimensional shape descriptions are object-centered, and are thus not specific to the viewer's perspective of the human form. We return to this issue in the General Discussion.

EXPERIMENT 2: SINGLE ELEMENTS VERSUS TRIADS IN MASKING DISPLAYS

A second experiment was designed to investigate further whether detection of the targets in the previous experiment could be attributable to local configural relations (such as pair-wise rigidity) that were present in the targets and not in the

Table 1. Detection results as a function of experiment and condition

Experiment	Task	Condition	Mean d'
1	Presence/absence	Upright target	1.84*
		Inverted target	0.39
2	Presence/absence	Single elements	2.56*
		Limb triads	1.28*
		Random triads	1.49*
3	Directional judgment	Upright target	2.93*
		Inverted target	0.45
4	Presence/absence	Canonical target	2.47*
		Off-joint target	1.70*
		Phase-shifted target	1.16*

* $p < .05$.

masks. Although the strong form of this interpretation was already rendered invalid by the failure to find detection of the inverted targets (which contain the same configural relations as the upright targets), it remained possible that detection of local configural relations interacts with the orientation of the point-light displays. In order to evaluate this possibility more directly, we manipulated the masking elements so that the local configural relations in the target were no longer unique.

Stimuli and Procedure

Two new masking displays were created. One involved triads of elements that corresponded to each of the four limbs of the point-light target (limb triads). The other display involved triads of elements that corresponded to randomly selected joints on the point-light target (random triads). The triads were randomly distributed within the viewing region of the screen until a total of 72 masking elements was generated.

Eight observers were tested for detection of the point-light walker target in each of three conditions defined by the structure of the masking elements. The first condition involved the individual elements used in the previous experiment, the second condition involved the limb triads, and the third condition involved the random triads. All other aspects of the procedure were identical to that used in the first experiment.

Results and Discussion

The results from this experiment are presented in Table 1. An analysis of variance revealed that detection was more difficult when the masking elements were composed of triads than when the masking elements were composed of individual elements, $F(1, 22) = 10.78, p < .01$. Nevertheless, performance remained significantly above chance in all three conditions. It thus appears that performance in these tasks does not depend primarily on detecting local configural relations in the target that are not present in the masking elements.

EXPERIMENT 3: DETECTION OF DIRECTION OF TARGET

The next experiment was designed as a converging test for assessing whether

biological motions are perceived globally. It was inspired by recent neurophysiological findings showing that some cells in the upper bank of the superior temporal sulcus (areas TPO and PGa) in monkeys are selectively responsive to the perspective view of the human form (facing left or right) as well as its direction of movement (Perrett, Harries, Benson, Chitty, & Mistlin, 1990). This selectivity is quite impressive because the two perspective views of the human form are hardly distinguishable in terms of their parts: they differ almost exclusively at a global level. In view of this neurophysiological evidence, we predicted that observers would discriminate a point-light display facing left from one facing right, even though the constituent elements in both displays were virtually identical.

Stimuli and Procedure

The stimulus display was similar to the one used in Experiment 1, except that a target was present on every trial. The task was to judge whether the target was facing right or left. In Condition 1, targets were presented upright for 100 trials; in Condition 2, inverted targets were presented for 100 trials. Direction of targets (facing right or left) was counterbalanced and randomized within conditions.

Results and Discussion

Three observers were tested, and their results are presented in Table 1. These results are similar to those from Experiment 1: Detection of the direction of targets was significantly above chance for upright targets, but not for inverted targets. This difference as a function of orientation was significant, $F(1, 4) = 27.64, p < .01$. Once again, correct detection in the upright condition was very high, 84%.

The finding that a point-light walker appearing to face to the left was discriminated from one appearing to face to the right is clearly consistent with neurophysiological reports of selective responding for the perspective view of the body by cortical cells. More important, these findings confirm that the global form depicted by the biological motions is detected prior to local relations. If per-

formance was based on some local relation such as the direction of the angle formed by the point lights corresponding to the left leg, then directional judgments would not have varied with orientation. Apparently, observers perceive a global form that is orientation-specific as well as view-specific (see Verfaillie, 1993, for related results).

EXPERIMENT 4: SPATIAL AND TEMPORAL PERTURBATIONS OF THE TARGET

Thus far, it has been assumed that the visual system is uniquely sensitive to the global human form specified by the point-light walker display. Conceivably, all of the information present in a point-light display is not necessary for eliciting the responsiveness observed in the preceding experiments. This issue was investigated in Experiment 4 by presenting observers with three different point-light targets.

Stimuli and Procedure

The first stimulus target corresponded to the upright displays used in the preceding experiments, and is referred to as the canonical target (Fig. 2a). The second target involved spatially perturbing the point lights on the target so that they corresponded to locations between joints rather than locations on joints (Fig. 2b). The third target corresponded to a point-light walker display in which the point lights were temporally perturbed by randomly shifting their phases within the gait cycle (Fig. 2c). In contrast to the canonical display, in which all the point lights changed oscillation direction at the same moment in time, the point lights in this display all changed direction at different times. Previous developmental research with these three point-light displays suggests that both spatial and temporal perturbations reduce the perception of a coherent form, but that the effect for the temporal perturbation is much more pronounced (Bertenthal, 1993; Bertenthal & Pinto, 1993).

The three stimulus targets were embedded in the same masking display used in Experiment 1. Four observers were tested for detection of these targets.

Global Processing of Biological Motions

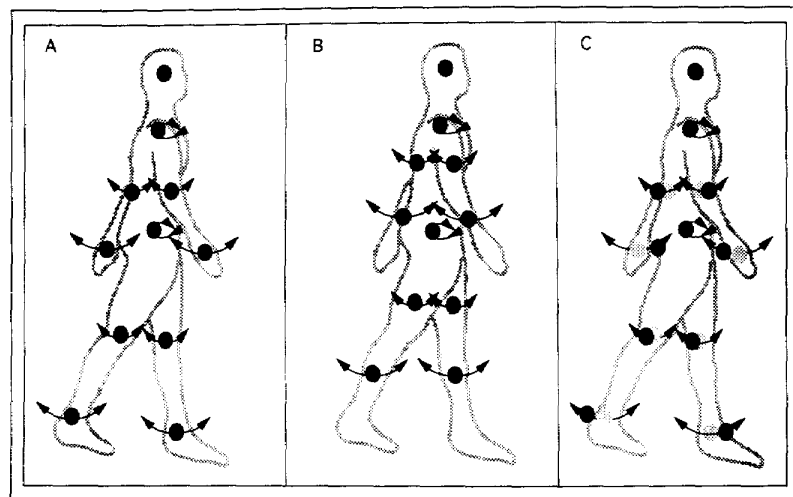


Fig. 2. Point-light stimulus displays used in Experiment 4. The canonical target (a) consisted of 11 point lights moving as if attached to the head, shoulder, hip, two wrists, two elbows, two ankles, and two knees of a person walking. The motion vectors drawn through each circle represent the perceived motions of the display. The off-joint target (b) consisted of 11 point lights moving as if attached to the head, shoulder, hip, and eight additional locations between the major joints of a person walking. The phase-shifted target (c) consisted of 11 point lights in which the temporal phase of each point light was shifted randomly relative to the gait cycle. Light circles are added to indicate the simultaneous in-phase locations of the point lights on the canonical target. Note that the outlines depicting the human form were not visible in the stimulus displays.

They were presented with 100 trials per stimulus condition (target present, 50 trials; target absent, 50 trials), for a total of 300 trials. Trials were blocked by condition and randomized within condition.

Results and Discussion

Detection of all three stimulus targets was significantly above chance (see Table 1). Nevertheless, performance differences as a function of target were observed. Detection of the off-joint display was poorer than detection of the canonical display, but this difference was non-significant. By contrast, detection of the phase-shifted display was significantly poorer than detection of the canonical display, $t(3) = 4.1, p < .05$. Apparently, the temporal phase relations of the point lights represent an especially important source of information used by the visual system to organize the display.

GENERAL DISCUSSION

These results present compelling evidence that the human form is perceived

as an emergent property in a biological-motion display. Discrimination performance remained above chance when the point-light targets were masked by individual elements or triads, but not when they were inverted. In the latter case, local relations were preserved, but that information was still not sufficient to ensure detection. These findings thus suggest that the perception of a global form specified by biological motions precedes the perception of the individual elements or the local relations, such as joint angles and pair-wise relative motions. Our results therefore stand in opposition to claims that the recognition of biological motions depends on local processing constraints, such as local rigidity or perceptual vector analysis (Hoffman & Flinchbaugh, 1982; Johansson, 1973; Webb & Aggarwal, 1982).

It is important to emphasize that it is not our intention to suggest that processing constraints based on local relations are irrelevant to the perception of biological motions. Indeed, our position is that multiple processing constraints are available for this purpose (Bertenthal &

Pinto, 1993). The current results suggest that the usefulness of these constraints varies as a function of the specific task. In much of the earlier research, the point-light displays shown to observers were restricted to 10 to 13 moving elements corresponding to the major joints and head of a person walking. Any algorithm that could connect these point lights correctly in their appropriate depth planes was sufficient for explaining the perception of structure from motion. We do not dispute the possibility that processing constraints based on local relations are sufficient for explaining these previous results. Conversely, local relations were rendered uninformative in the current study, yet detection performance was still above chance. We therefore conclude that local processing constraints are generally not necessary for detecting structure in point-light walker displays, although they may be sufficient under some well-defined conditions.

Let us now return to the finding that performance interacted with the orientation of the target. This result is consistent with results from previous studies suggesting that recognition of point-light walker displays is orientation-specific (Bertenthal, 1993; Pavlova, 1989; Sumi, 1984). Still, the results from the current study are somewhat surprising because our masking task did not require recognition of the point-light walker display; it was only necessary that observers judge whether or not a global structure, such as a hierarchical nesting of pendular motions, was present. These same structural relations are present regardless of orientation.

Currently, it is difficult to offer a definitive explanation for the orientation specificity of the results, but one plausible explanation is suggested by recent findings on the perception of reversible figures (Peterson & Gibson, 1991; Peterson, Harvey, & Weidenbacher, 1991). In this research, the authors show convincingly that stored representations of meaningful shapes contribute to the perceptual organization of ambiguous (or reversible) figures, especially when the figures are presented in the same orientation as the stored representations. Likewise, a point-light walker display is ambiguous, but it is perceived as a coherent shape as long as the stimulus display matches an orientation-specific rep-

resentation of the human form. These similarities between reversible figures and biological motions suggest that some stored orientation-specific representation of the human form could contribute to the perceptual organization of a point-light display in the same manner that stored representations constrain the perceived shape of reversible figures.

Additional evidence suggesting the contribution of some representational process is provided by a consideration of the task itself. Unlike the recognition paradigms used previously, the current detection task involved more than simply grouping the elements together. If observers were not instructed to search explicitly for a point-light target, then it would not have been detected. Successful performance thus depended on a guided search that had to be directed by some representation of the human form. (For similar arguments concerning higher level guidance of visual search, see Duncan & Humphreys, 1989; Wolfe, Cave, & Franzel, 1989).

Currently, the representation guiding this perceptual process is not well specified, but the results from Experiment 4 suggest that the perceptual process will tolerate some spatial perturbations in the extraction of the target. By contrast, a temporal perturbation to the target produced a significantly greater impairment in performance. It thus appears that the temporal patterning of the display may ultimately prove a defining property for perceiving the point-light walker display as a coherent structure. Indeed, a recently completed study suggests that dynamic symmetry in the movement of the

four limbs is essential for the perception of the point-light target display (Pinto & Bertenthal, 1993).

Acknowledgments—This research was supported by Grant HD16195 from the National Institutes of Health and by a grant from the McDonnell-Pew Program in Cognitive Neuroscience (T89-011245-017). Jeannine Pinto was supported by Predoctoral Fellowship MH18242 from the National Institute of Mental Health.

REFERENCES

- Bertenthal, B.I. (1993). Perception of biomechanical motions by infants: Intrinsic image and knowledge-based constraints. In C. Granrud (Ed.), *Carnegie symposium on cognition: Visual perception and cognition in infancy* (pp. 175–214). Hillsdale, NJ: Erlbaum.
- Bertenthal, B.I., & Kramer, S.J. (1984). The TMS 9918A VDP: A new device for generating moving displays on a microcomputer. *Behavior Research Methods, Instruments, and Computers*, 16, 388–394.
- Bertenthal, B.I., & Pinto, J. (1993). Complementary processes in the perception and production of human movements. In E. Thelen & L. Smith (Eds.), *Dynamical approaches to development: Vol. 2. Approaches* (pp. 209–239). Cambridge, MA: Bradford Books.
- Cutting, J.E., Moore, C., & Morrison, R. (1988). Masking the motions of human gait. *Perception & Psychophysics*, 44, 339–347.
- Duncan, J., & Humphreys, G.W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433–458.
- Farah, M. (1992). Object recognition: You may mistake your wife for a hat, but not for a word. *Current Directions in Psychological Science*, 1, 164–169.
- 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 and Psychophysics*, 14, 201–211.
- Kozlowski, L.T., & Cutting, J.E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception and Psychophysics*, 21, 575–580.
- MacArthur, L.Z., & Baron, R.M. (1983). Toward an ecological theory of social perception. *Psychological Review*, 90, 215–238.
- Marr, D., & Nishihara, H.K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London*, 200B, 269–294.
- Pavlova, M.A. (1989). The role of inversion in perception of biological motion pattern. *Perception*, 18, 510.
- Perrett, D.I., Harries, M.H., Benson, P.J., Chitty, A.J., & Mistlin, A.J. (1990). Retrieval of structure from rigid biological motion: An analysis of the visual responses of neurones in the macaque temporal cortex. In A. Blake & T. Troscianko (Eds.), *AI and the eye* (pp. 181–200). Chichester, England: Wiley.
- Peterson, M.A., & Gibson, B.S. (1991). The initial identification of figure-ground relationships: Contributions from shape recognition processes. *Bulletin of the Psychonomic Society*, 29, 199–202.
- Peterson, M.A., Harvey, E.M., & Weidenbacher, H.J. (1991). Shape recognition contributions to figure-ground reversal: Which route counts? *Journal of Experimental Psychology: Human Perception and Performance*, 17, 1075–1089.
- Pinto, J., & Bertenthal, B.I. (1993). Effects of phase relations on the perception of biomechanical motions. *Investigative Ophthalmology and Visual Science*, 33(Suppl. 1144).
- Pomerantz, J.R. (1981). Perceptual organization in information processing. In M. Kubovy & J.R. Pomerantz (Eds.), *Perceptual organization* (pp. 141–180). Hillsdale, NJ: Erlbaum.
- Sumi, S. (1984). Upside down presentation of the Johansson moving light spot pattern. *Perception*, 13, 283–286.
- Verfaillie, K. (1993). Orientation-dependent priming effects in the perception of biological motion. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 992–1013.
- Webb, J.A., & Aggarwal, J.K. (1982). Structure from motion of rigid and jointed objects. *Artificial Intelligence*, 19, 107–130.
- Wolfe, J.M., Cave, K.R., & Franzel, S.L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 419–433.

(RECEIVED 4/9/93; REVISION ACCEPTED 10/27/93)

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.