

# Person identification from biological motion: Effects of structural and kinematic cues

NIKOLAUS F. TROJE

*Ruhr-Universität, Bochum, Germany  
and Queen's University, Kingston, Ontario, Canada*

and

CORD WESTHOFF and MIKHAIL LAVROV  
*Ruhr-Universität, Bochum, Germany*

Human observers are able to identify a person based on his or her gait. However, little is known about the underlying mechanisms and the kind of information used to accomplish such a task. In this study, participants learned to discriminate seven male walkers shown as point-light displays from frontal, half-profile, or profile view. The displays were gradually normalized with respect to size, shape, and walking frequency, and identification performance was measured. All observers quickly learned to discriminate the walkers, but there was an overall advantage in favor of the frontal view. No effect of size normalization was found, but performance deteriorated when shape or walking frequency was normalized. Presenting the walkers from novel viewpoints resulted in a further decrease in performance. However, even after applying all normalization steps and rotating the walker by 90°, recognition performance was still nearly three times higher than chance level.

Humans, being a highly social species, rely on efficient mechanisms to derive information about individual conspecifics in their immediate environment. We observe what another person is doing, we can attribute biological traits, emotional states, and intentional behavior to an individual, and we can identify other people individually if we have met them before. Socially relevant information is transmitted through many different sensory modalities and perceptual mechanisms, including the way a person moves. In normal situations, we almost never have to rely on animate motion alone, because other sources of information are usually present at the same time. However, if isolated under laboratory conditions the significance of animate motion as an important source of socially relevant information becomes obvious.

The most popular method to achieve this is based on biological motion point-light displays, which were introduced into experimental psychology by Gunnar Johansson more than 30 years ago (Johansson, 1973). Only a small number of dots representing the major joints of a walking person results in a vivid percept of a human body (Johansson, 1976). The use of such point-light walkers showed that besides identifying particular actions (Dittrich, 1993) we can also attribute sex to a walker (Barclay,

Cutting, & Kozlowski, 1978; Cutting, 1978; Kozlowski & Cutting, 1977; Mather & Murdoch, 1994; Troje, 2002), and identify individual persons (Cutting & Kozlowski, 1977; Stevenage, Nixon, & Vince, 1999) solely based on their motion patterns. In some cases, only parts of the body seem to be sufficient for us to recognize specific properties of a performed action and the actor (Pollick, Lestou, Ryu, & Sung-Bae, 2002; Pollick, Paterson, Bruderlin, & Sanford, 2001).

Although the impressive performance of the human visual system to effortlessly organize a small number of moving dots into the coherent percept of a person is well documented, the principles underlying information encoding in biological motion and the mechanisms for its retrieval are poorly understood. A number of studies aimed at understanding the perceptually significant differences between male and female walking patterns have been conducted (Cutting, 1978; Mather & Murdoch, 1994; Troje, 2002), but virtually nothing is known about encoding of person identity. Here, we present the results of an experiment that we designed to investigate the role of a number of different stimulus attributes for person identification from biological motion displays.

The first experiment on person identification from point-light displays was conducted by Cutting and Kozlowski (1977), who showed that point-light displays are sufficient for the recognition of friends. Stimuli were point-light displays of six walkers moving across a computer screen in the sagittal plane. The observers were familiar with the individuals displayed but not with their point-light depictions. The observers performed rather

---

This project was generously funded by the Volkswagen Foundation. We thank Tobias Otto for invaluable technical support and three anonymous reviewers for their comments and suggestions. Correspondence should be addressed to N. F. Troje, Department of Psychology, Queen's University, Kingston, ON, K7M 3N6 Canada (e-mail: troje@post.queensu.ca).

poorly at the beginning, but even though no feedback was given, they improved to a level of correct responses that was well above chance performance. When asked about their strategies, the observers mentioned that characteristics such as speed, rhythm, amount of arm swing, and length of steps helped them to identify the walkers.

Stevenage et al. (1999) investigated the ability to identify walkers under different display conditions (daylight, dusk, point-light displays). Stimuli consisted of video footage of three female and three male walkers whose identifying characteristics, such as face, hair, or body shape, were masked. The display conditions on the screen were varied, with reduced brightness simulating dusk, and reflective tape attached to the major joints used to generate point-light displays. The dependent variable was the number of sessions needed to correctly identify all six walkers. Whereas neither the gender of the observer nor the lighting condition had a significant effect on this criterion, the three female walkers could be identified more easily than the three male walkers. The fact that the display condition had no influence on performance was taken as evidence that the observers relied mostly on dynamic rather than structural information.

In principle, the information carried by biological motion patterns could be contained in a number of different domains. The distinction between motion-mediated structural information and pure kinematic components provides the most basic decomposition (Troje, 2002). A static frame taken from a biological motion sequence appears as an unorganized cloud of individual dots. Only when the dots are set into motion does their articulation become apparent, revealing structure and geometry of the person's body, such as information about the width of the shoulders or leg length. Although mediated by the motion of the dots, this is inherently structural information and must be distinguished from kinematic information, which is contained in the trajectories of single dots. However, just as the retrieval of structural information requires the dots to move, the retrieval of kinematic information seems to be possible only if it is based on a geometry that is consistent with the physics of the human body. If a walker is presented upside-down (Shipley, 2003; Sumi, 1984) or if the trajectories are randomly displaced (Cutting, 1981), recognition of the displays is strongly impaired.

Structure and kinematics of a moving body are highly confounded. Dimensions and masses of a body determine the dynamics of the forces acting on it, which in turn determines the kinematics of its motion. However, it is still possible to generate biological motion point-light displays that constrain diagnostic information to either only the structural domain or only the kinematic domain, but not both. For instance, Mather and Murdoch (1994) worked with artificial point-light walkers and found that the structural feature "shoulder/hip width ratio" is much less informative than the kinematic feature "body sway" for determining the sex of a walker. Hill

and Pollick (2000) examined how temporal properties of a point-light stimulus affect person identification on a more limited domain—namely, the recognition of individuals from their arm movements. It was shown that the sensitivity for identification improved with an increasing level of temporal distinctiveness for a specific sequence. This study is interesting and relevant to our present work, not only in terms of its results but also methodologically. Whereas we used a normalization procedure to separate different aspects of the biological motion stimulus, Hill and Pollick employed exaggeration of spatial and temporal components to measure the perceptual effect of different stimulus components (see also Pollick, Fidopiastis, & Braden, 2001; Pollick, Hill, Calder, & Paterson, 2003).

The distinction between structural and kinematic aspects of biological motion that we employed here adopts a method to decompose our walking data that is described in Troje (2002). We basically defined the structure of a walker in terms of the walker's posture averaged across a whole gait cycle and the kinematics of a walker in terms of what remains if this average posture is subtracted from the motion data. We used this method before to show that Mather and Murdoch's (1994) results could be generalized: Sex classification of point-light walkers that are normalized with respect to their kinematics is much harder than classifying walkers that are normalized with respect to structural features (Troje, 2002). It is important to note that our distinction between structural and kinematic information is not equivalent to Hill and Pollick's (2000) distinction between spatial and temporal features.

An important aspect that has not received much attention in the context of biological motion perception is the viewpoint from which a walker is seen. For the recognition of static objects, this has been an important issue for a long time. The particularities of our ability, as well as our failures, to recognize an object from novel viewpoints have provided ample insight into the mental representations and perceptual mechanisms of object recognition.

Whereas the discussion about whether visual representations of objects are based on viewpoint-dependent two-dimensional projections or on viewpoint-independent structural descriptions was originally very controversial (Biederman & Gerhardstein, 1995; Tarr & Bühlhoff, 1995), researchers now have a more refined view showing that—depending on the particular object class, level of object specificity, or on particular tasks—both image-based representations as well as structural representations can play a role (Foster & Gilson, 2002).

For biological motion, few studies have systematically investigated viewpoint dependence, and most of them were conducted in the context of sex classification. Apparently, the frontal view reveals more information about a walker's sex than does the sagittal view or other views between these extremes (Mather & Murdoch, 1994; Troje, 2002). This seems to be different in face recognition. At least in naming and matching tasks, the most informative

view is neither the frontal nor the sagittal view, but a view somewhere between the two (Bruce, Valentine, & Baddeley, 1987; Troje & Bühlhoff, 1996). Verfaillie (1993) studied the effects of several attributes of point-light displays with a priming paradigm and found that only an identical viewpoint between priming and primed stimulus had a beneficial effect on the reaction time in a recognition task. No effects were found for the direction of the motion, the starting phase inside the walk cycle, or the position of the markers that represented the structure of the walkers. Watson, Johnston, Hill, and Troje (in press) found that nonrigid facial motion provides invariants that can be used to generalize to new viewpoints, whereas information contained in the rigid motion of a head is much more view dependent.

The present study was designed to investigate which parts of the overall information contained in biological motion are responsible for person identification. As stimuli, we used biological motion point-light displays, which were manipulated along two different lines: The point-light walkers were shown from three different viewpoints, and we gradually replaced parts of the individual, diagnostic information, such as size, shape, and step frequency, with averaged values, rendering the respective part of the overall information useless for the identification task. In addition to measuring learning curves, we introduced nonreinforced test sessions at different stages of the learning process. Most of them presented the walkers from the same viewpoint as during training (Type I test sessions). Only at the very end of the experiment did we add a further test session in which we presented the walkers from novel viewpoints (Type II test session).

## METHOD

### Participants

Eighteen observers (12 women, 6 men; 21–40 years of age), all students or employees at the Ruhr University in Bochum, participated in the experiment. All were naïve to the purpose of the study and had no previous experience with point-light displays. The students received course credit for their participation. All observers had normal or corrected-to-normal vision.

### Stimuli

Seven male individuals served as walking-models. All of them were students or staff at Ruhr University and between 25 and 35 years of age. A set of 41 retroreflective markers was attached to their bodies, and they were asked to walk on a treadmill at a comfortable speed. We let them walk for at least 10 min until they felt perfectly comfortable with the setup. We did not tell them when data acquisition was actually ongoing, and we eventually recorded a sequence of 10 sec for each.

For recording, a motion capture system equipped with nine CCD cameras (Oxford Metrics, Vicon 512) was used. The system tracks the positions of the markers in 3-D space with a spatial accuracy on the order of 1 mm and a temporal resolution of 120 Hz. From the 41 markers, we computed the trajectories of 15 “virtual” markers located at the major joints of the body by using commercially available software (Bodybuilder, Oxford Metrics). The data eventually represented 10 sec of stationary treadmill walking in terms of a time

series of 1,200 postures, each consisting of a 45-dimensional vector that contained the 3-D Cartesian positions of the 15 virtual markers.

The average posture, relative size, and walking frequency of each walker were used to create a number of different modifications from the original data. Altogether, five different conditions were used in the experiment:

1. In the veridical (VR) condition, the original data of the walkers as described above were shown to the observers.

2. For the size-normalized (SI) condition, we divided each walker's complete 3-D data by his relative size. To compute relative size, a number of steps were involved. We first computed each walker's average posture by averaging all frames across one full gait-cycle. An overall average posture was then computed from the seven individual average postures. Then, each individual's relative size was determined by the slope of the regression between his average posture and the overall average posture. The seven relative sizes were 1.10, 1.03, 0.98, 0.95, 0.89, 0.98, and 1.07. After we divided each walker's data by his relative size, all now had the same relative size, 1, but the relative positions of the point-lights still differed between the walkers, keeping intact each individual's shape and kinematics.

3. The stimuli for the shape-normalized (SH) condition were derived from the size-normalized data by replacing each walker's individual average posture with the overall average posture. This was achieved by simply subtracting the individual average posture from each frame (computed from the size-normalized data) and then adding the overall average posture to it. The resulting walkers now all had the same average posture. Before shape normalization, the variance of the location of a single dot averaged across the 15 dots was 612 mm<sup>2</sup>.

4. Frequency-normalized (FR) stimuli were obtained directly from the veridical data. We used Fourier analysis to determine each walker's walking frequency. Frequencies for the seven walkers were 1.42, 1.29, 1.04, 1.11, 1.25, 1.19, and 1.13. The data were then resampled, such that they all walked with a frequency of 1.2 Hz, which is the average across the seven walkers. The walkers retained their individual size and shape.

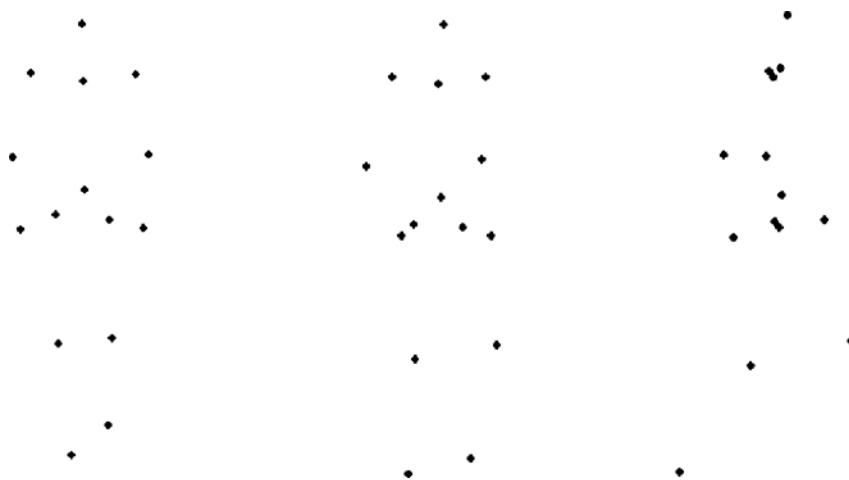
5. For the last (AL) condition, we used the same procedure to normalize for frequency, but this time using the shape-normalized stimuli. The data now lacked diagnostic information about size, shape, and gait frequency, but still contained differences in amplitudes and relative phases of the motion of the single dots.

Note that the VR, SI, SH, and AL conditions are formed by gradually omitting information about size, shape, and gait frequency, respectively.

The point-light stimuli were shown to the observers on a 19-in. CRT monitor (1,240 × 1,024 pixels) as white dots on a black background from one of three possible viewing angles: 0° (frontal view; FV), 30° (half-profile view; HV), or 90° (profile view; PV), as can be seen in Figure 1. The size of the walkers on the screen was about 6 cm, corresponding to a visual angle of 3.4° at a viewing distance of 1 m. We did not model possible occlusions; that is, all 15 markers were visible in every frame and from each viewpoint. Each display lasted for 3 sec, during which about 2.5 full gait-cycles were shown. The starting point within the gait-cycle was randomized on a trial-by-trial basis to rule out the possibility of recognizing a walker by a specific configuration of the points at the start or end of the display.

### Design and Procedure

The participants were randomly assigned to one of three groups according to the viewpoint from which the walkers were shown (view factor with the FV, HV, and PV levels). Each group contained 6 observers. Viewing angles remained constant for each observer, except for the final Type II test session. For the HV and PV groups, half of the participants saw the walkers facing to the left side, and



**Figure 1.** Examples of the point-light displays in frontal view (left), half-profile view (middle), and profile view (right).

half saw them facing to the right side. We did that to counterbalance the possible effects of hemispheric asymmetries, without intending to systematically investigate these effects.

Every experimental session consisted of a specified number of trials. A trial began with the 3-sec stimulus display, followed by the presentation of seven buttons, labeled with fictitious names of the walkers. The participants used the computer mouse to click on one of them in order to name the walker. After they had finished, the next trial followed, separated only by a 1-sec interstimulus interval.

The experiment consisted of a series of nine training sessions, which were interleaved with four Type I test sessions. At the end of the experiment, a single Type II test session was added. The whole experiment was run over 5 days.

In each training session, one of four different types of walking data was used (norm factor with VR, SI, SH, and AL levels), thus gradually decreasing the amount of available diagnostic information. Training sessions consisted of 140 trials, which were subdivided into 10 blocks of 14 trials. In every block, each of the walkers was shown twice in randomized order. If the participants clicked on an incorrect name, the button with the correct name changed its color for 1 sec, thus providing feedback. If the answer was correct, nothing else happened before proceeding to the next trial.

In the Type I test sessions, all five types of stimuli were used (VR, SI, SH, FR, and AL). The sessions consisted of 140 trials (7 walkers  $\times$  5 conditions  $\times$  4 repetitions) in randomized order. The viewpoint in Type I test sessions was the same as that used in the training sessions for each observer. No feedback was given to the participants.

In the Type II test session, only the all-normalized (AL) data of each walker were used, but the viewpoint of the displays was varied: The point-light displays were shown from all three viewpoints. The session consisted of 147 trials (7 walkers  $\times$  3 viewpoints  $\times$  7 repetitions) in randomized order, and no feedback was given to the participants.

The experiment was run over 5 consecutive days. On the first 4 days, there were two training sessions, followed by one Type I test session daily. The training sessions on the 1st day consisted of only veridical (VR) stimuli. Each of the following 3 days started with one session of retraining with the same stimuli as on the previous day, followed by a second training session with gradually normalized stimuli. On the 2nd day, we introduced size-normalization (SI) stimuli; on the 3rd day, shape-normalization (SH) stimuli; and on

the 4th day, frequency-normalization (FR) stimuli. The 5th day started with one session of retraining (AL) followed by a single Type II test session (Figure 2).

## RESULTS

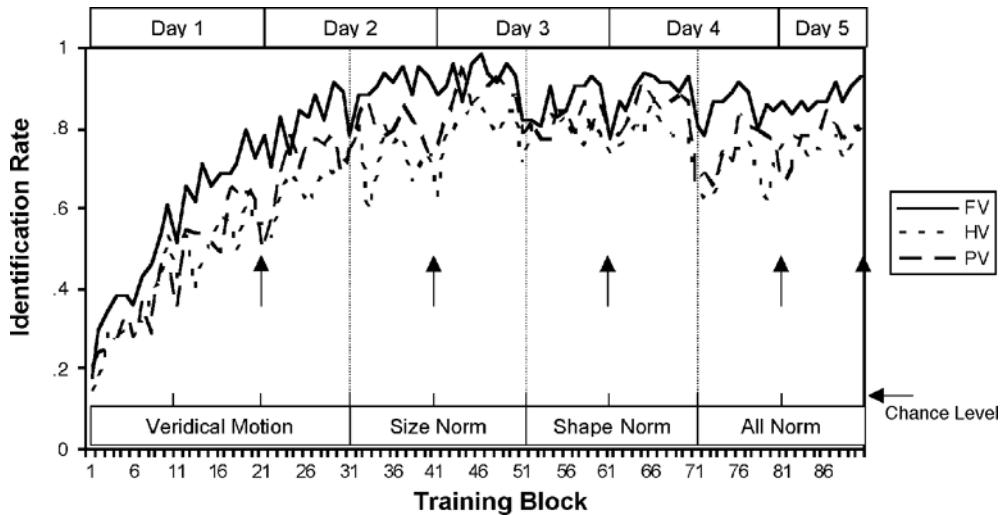
### Training Sessions

Starting from about chance level, all three groups quickly improved their performances in identification of the walkers (Figure 2). At the end of the third training session, the FV, HV, and PV groups reached correct response levels of 88%, 70%, and 75%, respectively (averaged across the last three blocks), and after five sessions performance saturated at a level of about 90%.

Over the complete series of training sessions, there was a significant main effect of view [ $F(2,15) = 7.55$ ,  $p = .005$ ]. The mean values of correct identification were 80% (FV), 67% (HV), and 72% (PV). Bonferroni-corrected post hoc tests ( $p < .01$ ) revealed that the FV group performed significantly better than the HV group. No differences were found between the FV and PV groups or between the HV and PV groups, respectively.

For data analysis, we only used the identification performance of the last three blocks before and the first three blocks after changing the level of the norm condition. For example, the SI level of the norm condition contained the comparison of the results of the last three blocks with veridical motion to the first three blocks with size-normalized stimuli. By doing this, we included all relevant data for comparisons and eliminated data that could distort further analyses—for example, the low performance values at the beginning of the experiment.

A three-way repeated measures analysis of variance (ANOVA) with change (two levels: before and after) and norm (three levels: SI, SH, and AL) as within-subjects factors and view (three levels) as a between-subjects factor was conducted. We found significant effects for all three

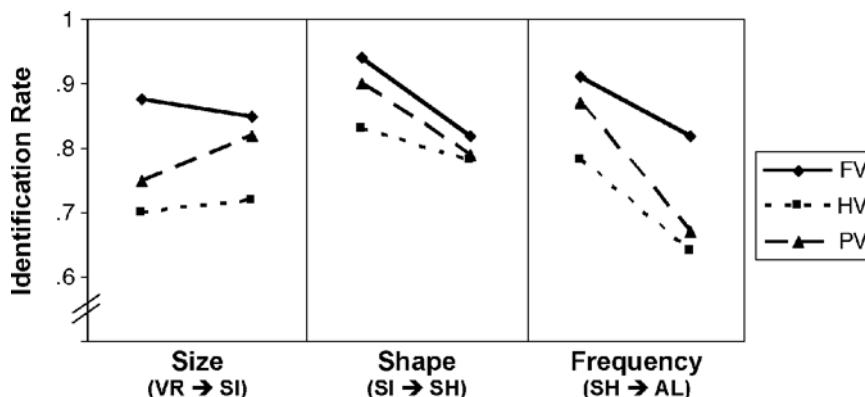


**Figure 2.** Learning curves of the training sessions. The vertical lines represent the introduction of a new kind of stimulus data. The solid vertical arrows represent the insertion of Type I test sessions, whereas the dashed vertical arrow represents the insertion of the Type II test session. The tick marks represent the beginning of a new training session. FV, frontal view; HV, half-profile view; PV, profile view.

factors: change [ $F(1,15) = 15.00, p < .01$ ], norm [ $F(2,30) = 14.00, p < .01$ ], and view [ $F(2,15) = 5.21, p < .02$ ]. There was also a significant interaction between change and norm [ $F(2,30) = 17.12, p < .01$ ], as well as a three-way interaction [ $F(4,30) = 2.75, p < .05$ ]. Post hoc Bonferroni tests ( $p < .05$ ) revealed that, on average, identification performance before the introduction of a new data set (84%) was better than afterward (77%). Furthermore, the FV group (87%) performed significantly better than did the HV group (74%), whereas no differences were found between the FV and PV groups (80%), or between the HV and PV groups. The mean performance level before and after introduction of shape normalization (84%) was significantly better than the mean performance level at the introduction of size nor-

malization, as well as frequency normalization (78% in both cases). The latter effect is simply due to the fact that performance had not reached saturation by the time we introduced the size-normalized versions of the walkers.

More interesting, however, is the interaction between the change and norm factors. To further explore this interaction, we conducted two-way, repeated measures ANOVAs with change as the within-subjects factor and view as the between-subjects factor separately for each condition of norm (Figure 3). For size-normalization (transition from VR to SI), we found no significant effect of change [ $F(1,15) = 0.95, \text{n.s.}$ ]. In contrast, this factor had a highly significant effect when normalizing the stimulus for shape [SI to SH;  $F(1,15) = 29.44, p < .001$ ], as well as when removing frequency information



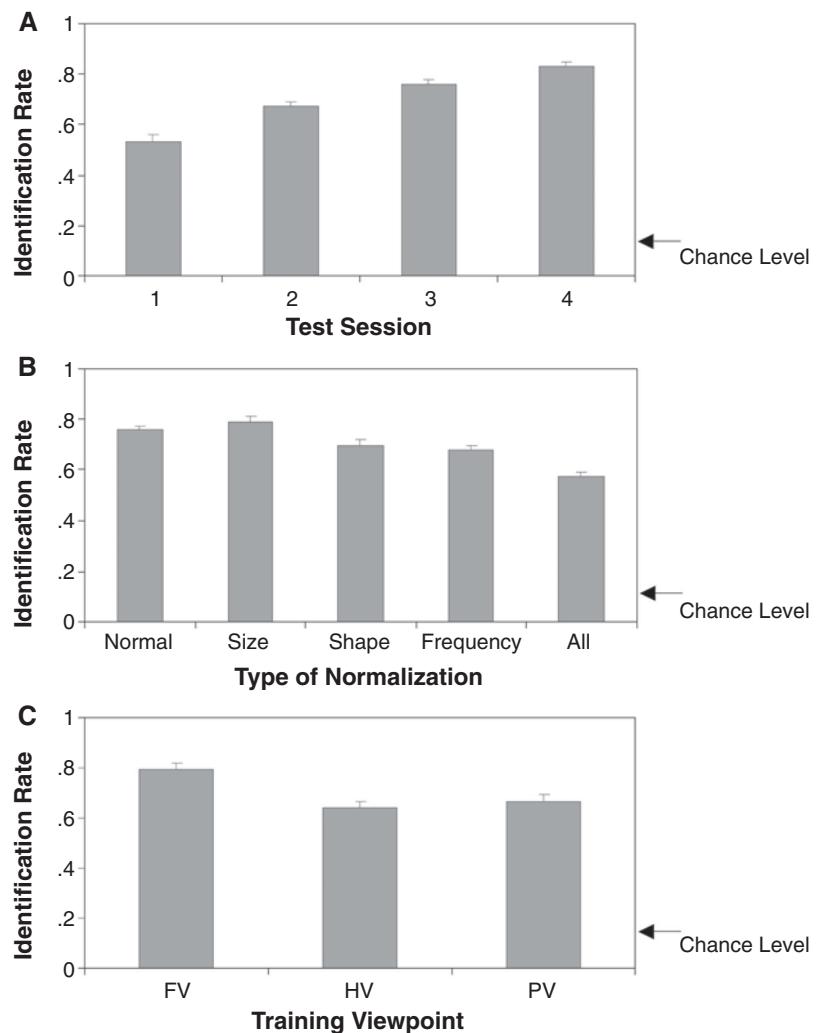
**Figure 3.** Identification performances at the introduction of new normalization conditions. FV, frontal view; HV, half-profile view; PV, profile view; VR, veridical; SI, size normalized; SH, shape normalized; AL, all normalized.

[SH to AL;  $F(1,15) = 28.81, p < .001$ ]. Significance for the view factor was only reached when changing from SH to AL [ $F(1,15) = 6.06, p < .02$ ], even though trends confirming the main effect of view in the three-way ANOVA were observed in all cases. It seems that the frontal view is somewhat advantageous, compared with the other two views. No significant interaction between normalization condition and viewing angle was found in any of the analyses.

#### Type I Test Sessions

A repeated measures, three-way ANOVA with test number (four levels) and type of normalization (norm, five levels) as within-subjects factors and view (three levels) as the between-subjects factor was performed. Identification performances corresponding to the different levels of the three factors are shown in Figure 4.

Significant main effects were found for all three factors: test number [ $F(3,45) = 52.84, p < .001$ ], norm [ $F(4,60) = 47.78, p < .001$ ], and view [ $F(2,15) = 9.32, p = .002$ ]. Bonferroni-corrected post hoc tests ( $p < .02$ ) were performed to test for differences between the factor levels. An increase in performance was found between all four test sessions (53%, 67%, 76%, and 83% identification rate, respectively), as well as between most of the normalization conditions. No difference was found between the VR-SI conditions (76%–79%) and the SH-FR conditions (69%–68%), respectively. Every other comparison between the conditions reached a significant level. In addition, the differences between the three viewpoint groups found in the training sessions were confirmed by the results of the test sessions: The FV group (79%) had a better average identification rate than did the HV group (64%) and even performed better



**Figure 4. Main effects of the three factors used in Type I test sessions on identification performance: (A) number of test session, (B) type of normalization, and (C) viewing angle. Error bars indicate standard errors of the means.**

than did the PV group (67%), but the HV and PV groups showed no significant differences.

In addition to the main effects, we found significant two-way interactions between the test number and norm factors [ $F(12,180) = 3.33, p < .001$ ] and between the norm and view factors [ $F(8,60) = 5.01, p < .001$ ], as well as a three-way interaction of all factors [ $F(24,180) = 2.05, p < .005$ ]. The test number–norm interaction is obviously based on the stepwise introduction of the normalization levels. As the observers were successively confronted with new normalization conditions in the training sessions, they were better able to identify the walkers in the following test sessions, because they had been introduced to these stimuli before.

The interaction of the norm and view factors indicates that different kinds of information are used in the three groups to identify a walking person (Figure 5). In the VR, SI, and SH conditions, we find relatively similar patterns: The frontal view produces the best performance, and the profile view seems to result in better performance than the half-profile view. This is different in the two conditions that lack information about walking frequency (FR and AL). The lack of frequency seems to particularly affect observers from the PV group, whose performance drops to the same or even lower values than those in the HV group.

Taking a closer look at the observed three-way interaction, it appeared that the increase in performance over the four test sessions became higher with increasing steps of normalization, particularly in the PV group, whereas there were different patterns in the other two groups. In other words, the more information we took out of the point-light displays, the more time it took for the PV group to accommodate to the new stimuli.

#### Type II Test Session

The walkers shown in the final Type II test session were normalized according to shape (including size) and walking frequency. In contrast to the Type I tests, walkers were now presented from all three viewing angles. In

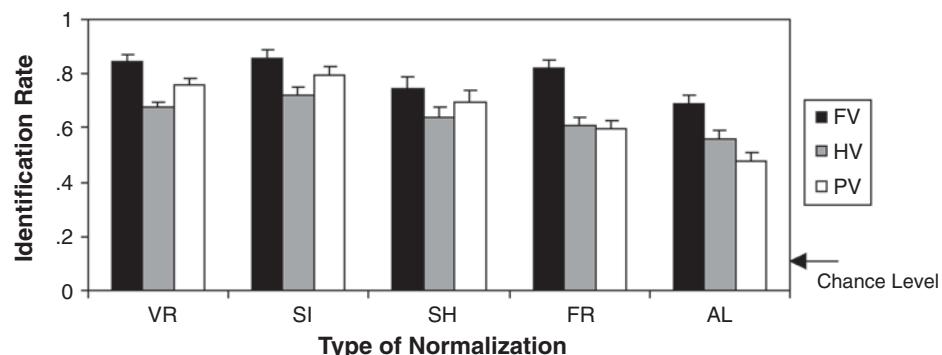
this way, we tested for generalization from one angle to novel viewpoints (Figure 6).

We conducted a repeated measures ANOVA with the training angle (three levels) as between-subjects factor and test angle (three levels) as within-subjects factor. Neither training angle [ $F(2,15) = 2.73, \text{n.s.}$ ] nor test angle [ $F(2,30) = 1.31, \text{n.s.}$ ] revealed a significant main effect. However, there was a highly significant interaction between these two factors [ $F(4,30) = 19.01, p < .001$ ]. As can be seen from Figure 6, the identification performance of the observers was always best when the stimuli were shown from the same viewing angle as in the training sessions. Even though it decreased as the test viewpoint deviated from the training viewpoint, performance was still high above chance level in all conditions.

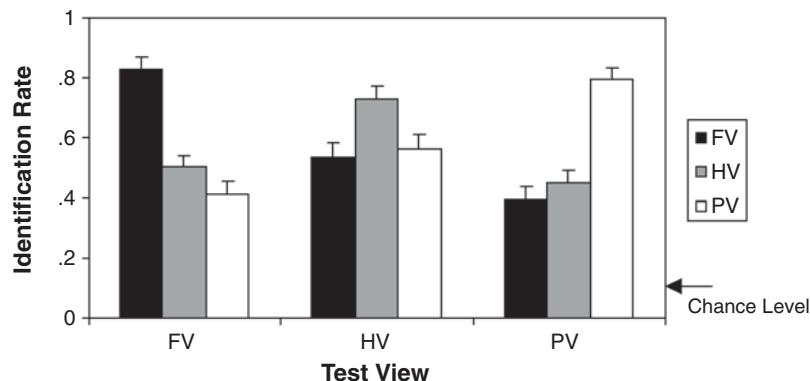
#### DISCUSSION

All participants quickly learned to recognize and name seven individual male walkers from point-light displays. Starting at chance level, performance saturated at a level of almost 90% correct responses after five training sessions. A small but significant frontal-view advantage becomes apparent not only in the learning curve but also from the data obtained from the Type I test sessions. Here, the frontal-view advantage becomes particularly apparent in conditions in which we deprive the observers of information about frequency.

This pattern is consistent with earlier findings on sex classification from biological motion (Mather & Murdoch, 1994; Troje, 2002). Important diagnostic features that help discriminate between males' and females' walk seem to be the rotation of the hip in the fronto-parallel plane, the degree of lateral body sway, the ratio of hip and shoulder width, and the position of the elbows. All these features are expressed in terms of variations in the fronto-parallel plane and are, therefore, best visible from a frontal view. Here, our task was not sex classification but person identification. Even though it is less clear



**Figure 5. Identification performance of the three groups depending on the normalization of the stimuli (norm × view interaction) in Type I test sessions. Error bars indicate standard errors of the means. FV, frontal view; HV, half-profile view; PV, profile view; VR, veridical; SI, size normalized; SH, shape normalized; FR, frequency normalized; AL, all normalized.**



**Figure 6.** Identification performance in the Type II test session as a function of test view and training view. Error bars indicate standard errors of the means. FV, frontal view; HV, half-profile view; PV, profile view.

which kinds of features are most helpful in solving the task, it seems that again motion and static cues in the fronto-parallel plane are more diagnostic than those from other viewing angles.

The pattern we measure here is clearly different from that measured in a variety of face recognition tasks (Bruce et al., 1987; Troje & Bühlhoff, 1996), where we generally find a half-profile view advantage. Because this half-profile view advantage is particularly obvious in experiments investigating the ability to recognize faces from novel viewpoints, it has been interpreted in terms of the fact that faces are approximately bilaterally symmetric. Assuming bilateral symmetry, a nonaccidental view of an object could be used to recover its 3-D affine structure (Troje & Bühlhoff, 1998; Ullman & Basri, 1991; Vetter, Poggio, & Bühlhoff, 1994), resulting in invariants that would facilitate recognition across viewpoints. In principle, the same argument could hold true for walking as well. Even though each single posture is not bilaterally symmetric, the pattern shows glide reflection—that is, the spatiotemporal symmetry obtained from mirror reversal and subsequent phase shifting. It could therefore be used in the same way as proposed in the case of face recognition. In fact, it seems that in the viewpoint generalization experiment (Type II test session), the frontal-view advantage found in the learning curve and in the other test session is canceled. Furthermore, viewpoint generalization from the half-profile view to both the frontal view and the profile view seems to be easier than in the other directions. Even though the half-profile view does not provide advantages for recognition per se, it seems to show an advantage in terms of viewpoint generalization.

The decrease in performance in response to viewpoint changes is very obvious and in general accordance with earlier findings (Verfaillie, 1993). However, the remaining recognition performance is quite substantial. Even under conditions of maximum rotation of the displays between training view and test view (FV to PV and vice versa), recognition performance remains at a level of

about 40%, which is still nearly three times higher than chance level (14%). Obviously, the human visual system is capable of abstracting motion seen from a single viewpoint and then generalizing it to other viewing angles. This means that certain features in human gait patterns can be recognized without reference to particular viewpoints. Because stimuli in the Type II test were normalized with respect to structural features, the invariants mediating this ability must be contained in the kinematics of the motion itself. Frequency is the only aspect of kinematic information that we looked at in the present study, and it seems that even though it does affect recognition, it certainly does not play a major role. One way to further decompose the remaining kinematic information is to look into the Fourier composition of the walking pattern in order to determine which parts of the power spectrum or the phase spectrum are carrying important information for person identification.

The learning curve, as well as the data from the Type I test sessions, provides further insight into which features are used for identification of individual walkers. It is very obvious that the size of the walker is not used at all. This might be an artifact of our particular experiment, in which walkers were presented one at a time and the displays did not provide much reference for comparison. If size had provided information about the identity of the walkers, it might have been difficult to retrieve it from the display. Both the shape of the walker, as well as walking frequency, seemed to provide diagnostic information. Depriving participants from either of them resulted in small but significant recognition drops in the learning curve. The results from the Type I test sessions imply that both cues are rather independent and additive (Figure 4B). However, this comparison is also misleading. Normalizing for shape means to deprive the stimulus of all information about structure, whereas normalizing for frequency alone takes out only one of several aspects of the complex kinematic pattern.

One of the main findings of this study is the fact that even after observers were deprived of information about

shape (including size) and walking frequency, recognition is still about five to six times higher than chance level. The decreases in recognition performance due to normalization for shape and frequency are very small, compared with the overall performance. Because deprivation of structural (shape) information has only a small effect, most of the information used for person identification seems to be extracted from gait-kinematics. The only parameter within this domain that was tested in the present study was walking frequency. Even though this parameter affects performance, it does not seem to play a very major role. Further experiments are required to explore which aspects of the kinematics of human gait are responsible for person identification.

## REFERENCES

- BARCLAY, C. D., CUTTING, J. E., & KOZLOWSKI, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics*, **23**, 145-152.
- BIEDERMAN, I., & GERHARDSTEIN, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition—reply to Tarr and Bültlhoff (1995). *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1506-1514.
- BRUCE, V., VALENTINE, T., & BADDELEY, A. D. (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology*, **1**, 109-120.
- CUTTING, J. E. (1978). A biomechanical invariant of gait perception. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 357-372.
- CUTTING, J. E. (1981). Coding theory adapted to gait perception. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 71-87.
- CUTTING, J. E., & KOZLOWSKI, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, **9**, 353-356.
- DITTRICH, W. H. (1993). Action categories and the perception of biological motion. *Perception*, **22**, 15-22.
- FOSTER, D. H., & GILSON, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings of the Royal Society of London: Series B*, **269**, 1939-1947.
- HILL, H., & POLICK, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Science*, **11**, 223-228.
- 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 motion 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.
- MATHER, G., & MURDOCH, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London: Series B*, **258**, 273-279.
- POLICK, F. E., FIDOPIASIS, C., & BRADEN, V. (2001). Recognising the style of spatially exaggerated tennis serves. *Perception*, **30**, 323-338.
- POLICK, F. E., HILL, H., CALDER, A., & PATERSON, H. (2003). Recognising facial expression from spatially and temporally modified movements. *Perception*, **32**, 813-826.
- POLICK, F. E., LESTOU, V., RYU, J., & SUNG-BAE, C. (2002). Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research*, **42**, 2345-2355.
- POLICK, F. E., PATERSON, H. M., BRUDERLIN, A., & SANFORD, A. J. (2001). Perceiving affect from arm movement. *Cognition*, **82**, B51-B61.
- SHIPLEY, T. F. (2003). The effect of object and event orientation on perception of biological motion. *Psychological Science*, **14**, 377-380.
- STEVENAGE, S. V., NIXON, M. S., & VINCE, K. (1999). Visual analysis of gait as a cue to identity. *Applied Cognitive Psychology*, **13**, 513-526.
- SUMI, S. (1984). Upside-down presentation of the Johansson moving light-spot pattern. *Perception*, **13**, 283-286.
- TARR, M. J., & BÜLTHOFF, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1494-1505.
- TROJE, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, **2**, 371-387.
- TROJE, N. F., & BÜLTHOFF, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research*, **36**, 1761-1771.
- TROJE, N. F., & BÜLTHOFF, H. H. (1998). How is bilateral symmetry of human faces used for recognition of novel views? *Vision Research*, **38**, 79-89.
- ULLMAN, S., & BASRI, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **13**, 992-1005.
- VERFAILLIE, K. (1993). Orientation-dependent priming effects in the perception of biological motion. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 992-1013.
- VETTER, T., POGGIO, T., & BÜLTHOFF, H. H. (1994). The importance of symmetry and virtual views in three-dimensional object recognition. *Current Biology*, **4**, 18-23.
- WATSON, J. S., JOHNSTON, A., HILL, H., & TROJE, N. F. (in press). Viewpoint invariance for facial but not for head movements. *Visual Cognition*.

(Manuscript received March 9, 2004;  
revision accepted for publication August 23, 2004.)