

Critical features for the recognition of biological motion

Antonino Casile

Laboratory for Action Representation and Learning,
Department of Cognitive Neurology, Hertie Institute for
Clinical Brain Research, University Clinic, Tübingen, Germany



Martin A. Giese

Laboratory for Action Representation and Learning,
Department of Cognitive Neurology, Hertie Institute for
Clinical Brain Research, University Clinic, Tübingen, Germany



Humans can perceive the motion of living beings from very impoverished stimuli like point-light displays. How the visual system achieves the robust generalization from normal to point-light stimuli remains an unresolved question. We present evidence on multiple levels demonstrating that this generalization might be accomplished by an extraction of simple mid-level optic flow features within coarse spatial arrangement, potentially exploiting relatively simple neural circuits: (1) A statistical analysis of the most informative mid-level features reveals that normal and point-light walkers share very similar dominant local optic flow features. (2) We devise a novel point-light stimulus (critical features stimulus) that contains these features, and which is perceived as a human walker even though it is inconsistent with the skeleton of the human body. (3) A neural model that extracts only these critical features accounts for substantial recognition rates for strongly degraded stimuli. We conclude that recognition of biological motion might be accomplished by detecting mid-level optic flow features with relatively coarse spatial localization. The computationally challenging reconstruction of precise position information from degraded stimuli might not be required.

Keywords: biological motion, action recognition, motion features, form features, pathways, principal components analysis, neural model

Introduction

Human perception of biological movements (i.e., the movements of living beings) is amazingly robust. This was demonstrated in classical experiments by Johansson (Johansson, 1973, 1976), who showed that subjects can spontaneously recognize complex actions from point-light stimuli, which consist of a small number of illuminated dots moving like the joints of a human actor. If such point-light stimuli are presented dynamically as movies, subjects easily recognize complex actions, whereas the presentation of static frames from these movies does not result in such well-defined percepts.

Recognition of point-light stimuli arises quite early during development (Bertenthal, Proffitt, & Cutting, 1984; Fox & McDaniel, 1982; Pavlova, Krägeloh-Mann, Sokolov, & Birbaumer, 2001). Moreover, several experiments have shown that this visual capability is amazingly robust. Perception is only partially impaired by masking point-light walkers with dynamic noise (Cutting, Moore, & Morrison, 1988; Thornton, Pinto, & Shiffrar, 1998) or by changing the contrast polarity of the dots across frames (Ahlström, Blake, & Ahlström, 1997). Even if only a subset of dots is visible, if the lifetimes of the individual dots are limited, and if the dots are displaced on the skeleton in every frame, substantial recognition performance can be achieved (Beintema & Lappe, 2002; Mather, Radford, & West, 1992; Neri, Morrone, & Burr, 1998; Pinto & Shiffrar, 1999).

The mechanisms underlying the spontaneous robust generalization from normal to point-light stimuli remain

largely unclear, and several hypotheses have been discussed in the literature.

One set of explanations assumes that the brain might dispose of complex computational mechanisms that reconstruct missing information from impoverished stimuli (e.g., by fitting two-dimensional [2D] or 3D models of the human skeleton to the dot positions of point-light stimuli) (Beintema & Lappe, 2002; Marr & Vaina, 1982; Webb & Aggarwal, 1982). Technical implementations show that, in principle, the underlying computational problem can be solved (for recent reviews, see Aggarwal & Cai, 1999; Gavrila, 1999). However, most of the existing algorithms are computationally quite expensive and have no obvious neural implementation.

An alternative explanation assumes that the generalization from normal to point-light stimuli is based on specific features that are shared by both stimulus classes. The precise nature of such features is largely unknown, and it has been discussed whether they are based on form or motion information (Cutting, Proffitt, & Kozlowski, 1978; Mather & Murdoch, 1994; Mather et al., 1992; Troje, 2002).

In this study, we address the problem of finding possible mechanisms for the robust generalization in biological motion recognition by applying methods from image statistics. We extract dominant mid-level motion and form features from normal and point-light stimuli. The dominant mid-level motion features are very similar for both stimulus classes, whereas the dominant form features are quite different. Thus, the extraction of mid-level motion features

provides a computationally simple explanation for the generalization between the two stimulus classes.

For further testing of this hypothesis, we designed a novel point-light stimulus, critical feature stimulus (CFS), that contains the extracted dominant motion features combined with some very coarse positional information. The CFS is spontaneously perceived as a human walker, even though it is inconsistent with the kinematics of the human skeleton. A more detailed psychophysical experiment shows that walking direction can be recognized from the CFS equally well as from similar stimuli that match exactly the kinematics of a human body. Both the spontaneous recognition and the results of the psychophysical results point against a critical role of exact position information for the recognition of point-light walkers.

We finally devised a neural model that accomplishes the recognition of biological motion by extracting the proposed critical motion features. Although based on the extraction of a single type of critical feature, this model reaches substantial performance levels. The proposed model is based on simple neural circuits that can, in principle, be easily implemented by cortical neurons.

Statistical analysis

Normal human movement stimuli and point-light displays are characterized by temporal sequences of specific form or optic flow patterns. We applied principal components analysis (PCA) to extract dominant mid-level form and motion features from movies showing a person performing different actions (e.g., walking). PCA is a common technique for the extraction of informative directions in high-dimensional data spaces.

Method

Two movies were created using joint trajectories that were tracked from real videos (for details, see Giese & Poggio, 2003). One movie showed a full-body 2D walker as a stick figure that matches approximately a human body silhouette (Figure 1a). The second movie showed a point-light walker with 10 dots (Figure 1e). Both movies consisted of 21 frames per walking cycle. From these videos we determined sequences of black-and-white images for the extraction of dominant form features, and sequences of optic flow fields for the extraction of dominant motion features. We assumed that the walker covers an area that corresponds to 9 x 7-deg visual angle. For the extraction of dominant mid-level form features, we sampled each frame (205 x 193 pixels) of the movie with windows of 50 x 50 pixels. This size corresponds to about 3-deg visual angle, and would be in a range that is typical for peri-foveal neurons in area V4 of the macaque (Gattass, Sousa, & Gross, 1988). The sampling window was centered at points of a regular grid (12 pixels between neighboring points), defining 156 overlapping receptive fields. Within each sampling window the pixel values were concatenated into a 2500-

dimensional vector. The vectors collected over all window positions in each frame and over all frames were used to compute a covariance matrix for the PCA. We carried out two separate PCAs, one for the point-light stimulus and one for the full-body walker.

Optic flow fields were generated (1) for the full-body stimulus by computing the local movements from the underlying skeleton model (Figure 1b), and (2) for the point-light walker by finding nearest neighbor correspondences between the dots in subsequent frames (Figure 1f). For both stimulus classes we computed the motion vectors on a grid of 70 x 47 sampling points for each frame of the animation. Similar to the extraction of the form features, we sampled the optic flow fields for each frame with overlapping windows with a size that corresponds to about 3-deg visual angle. This size is within the typical range of peri-foveal receptive fields in the middle temporal visual (MT) area in the macaque (Snowden, 1994). Each sampling window covered an area of 14 x 14 sampling points, and the centers of the local windows were chosen from a regular grid, resulting in 228 overlapping sampling windows. The x and y components of the motion vectors within each sampling window were concatenated into 392-dimensional vectors. These vectors, collected over all window positions and frames, were used to compute a covariance matrix for the PCA. Again, two separate PCAs were computed for the full-body and the point-light walker.

For both, motion and form features, the results did not critically depend on the window size and number of sampling points.

Results

Figure 1c and 1g show the computed dominant form features. The dominant form feature is defined by the eigenvector that corresponds to the largest eigenvalue of the covariance matrix obtained from the luminance values. The dominant eigenvector defines the direction of maximum variance in the high-dimensional feature space that is given by the luminance values in the sampling windows. For visualization purposes, the eigenvectors are plotted as color-coded luminance distributions over the sampling window. The dominant form features for full-body stimuli (Figure 1c) and point-light stimuli (Figure 1g) look very different, as confirmed by a very low correlation coefficient ($r = 0.09$) between the two eigenvectors.

The computed dominant local optic flow features are shown in Figure 1d and 1h. The dominant local optic flow feature is defined as the eigenvector that corresponds to the largest eigenvalue of the covariance matrix that was derived from the optic flow distributions in the sampling windows. This eigenvector defines the direction of maximum variance in the feature space, which is defined by the optic flow fields within the sampling windows. The eigenvector is plotted as optic flow field over the sampling window. The dominant local optic flow features derived from the full-body and the point-light stimulus look amazingly similar, as

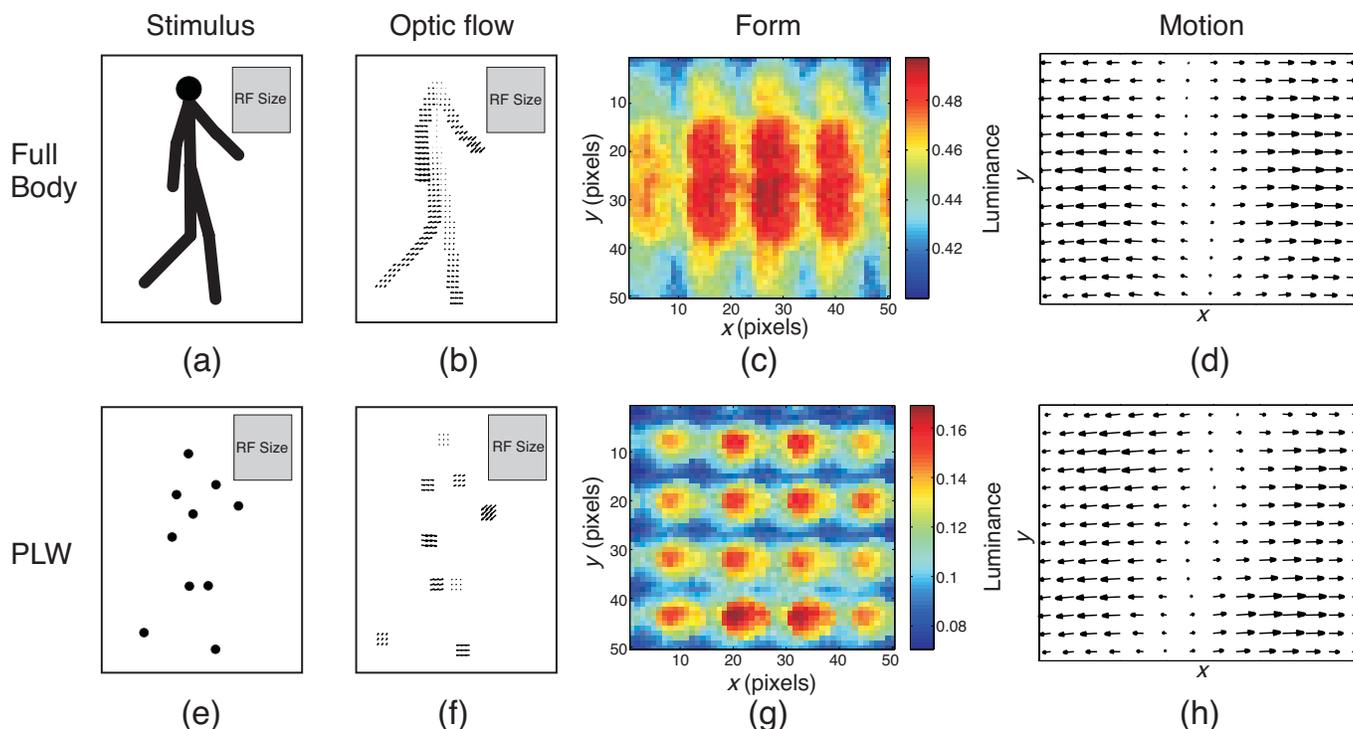


Figure 1. Statistical analysis of mid-level features. (a). Single frame from a full-body biological motion stimulus. Inset shows the size of the receptive field (RF) that was used for the computation of the dominant mid-level form feature. (b). Optic flow field computed from subsequent frames of the movie based on a stick figure model. Inset shows the size of the receptive field for the computation of the dominant local optic flow features. (c). Dominant mid-level form feature for the full-body stimulus extracted by applying principal components analysis (PCA) to the luminance distributions derived from 156 overlapping receptive fields over all frames of the movie. The dominant eigenvector, which corresponds to the feature that explains a maximum amount of variance, is plotted as luminance distribution over the RF (luminance values are color-coded for better visualization). (d). Dominant local optic flow feature extracted by applying PCA to the local optic flow fields derived from 228 overlapping windows and all frames of the movie. The dominant eigenvector is plotted as optic flow field over the RF. (e). Single frame from a point-light biological motion stimulus. (f). Optic flow field computed from a single frame pair of the point-light stimulus using nearest neighbor correspondences. (g). Dominant mid-level form feature for the point-light stimulus extracted by applying PCA and plotted as luminance distribution over the RF. (h). Dominant optic flow feature for the point-light stimulus plotted as optic flow field over the RF.

confirmed by a high correlation coefficient ($r = 0.93$) between the two eigenvectors. An additional important observation is that for both stimuli the dominant mid-level optic flow feature is characterized by strong opponent motion in horizontal direction.

Summarizing, our statistical analysis reveals that full-body and point-light walker stimuli share very similar dominant local optic flow features. Thus, the extraction of these features might be a simple mechanism that can account for the robust generalization from one stimulus type to the other.

A similar statistical analysis can be applied to other human actions leading to similar results. For example, for running, the dominant optic flow features are also very similar for full-body and point-light stimuli, whereas this does not apply to the dominant form features. In this case the extracted dominant mid-level motion features are more complex, such as opponent motion along curved and tilted paths (demo provided with this article).

It is important to stress that this result does not imply that the visual system exclusively uses this dominant feature. In particular for full-body stimuli, which contain a substantial amount of form and contour information, it seems very likely that the recognition of actions also exploits information about the body shape. Contrary to the case of point-light stimuli, in this case, complex perceptual judgments are also possible from individual frames. For example, it has been shown that based on the orientation of the tibia, subjects can distinguish walking and running based on static pictures of stick figures (Todd, 1983).

Psychophysical experiments

Our statistical analysis suggests that opponent motion in horizontal direction might be a critical feature for the recognition of point-light walkers. If this feature carries important information about the presence of a human walker, it should be possible to devise metameric point-light stimuli

that contain this feature, and which are erroneously perceived as human walkers, even though they are not derived from a human body shape. We tested this hypothesis in two sets of psychophysical experiments using a novel stimulus (CFS). This stimulus is illustrated in Figure 2. It contains the extracted dominant motion features combined with some very coarse positional information, whereas other possible cues are minimized by randomization.

The CFS consists of pairs of dots that move in four adjacent rectangular regions. In two of these regions (light gray in Figure 2) the movements of the individual dots are completely random. In the other two regions (dark gray in Figure 2) the vertical components of the motion vectors are completely random, but the horizontal components of the dot movements are sinusoidal and in anti-phase (i.e., specifying opponent motion with a cycle time of about 1 s). The spatial arrangement of the four regions was motivated by the fact that the strongest opponent motion for point-light walkers arises in the regions corresponding to the hands and feet. (Demo provided with this article.)

Experiment 1

Seventeen unpaid subjects took part in this experiment. They were carefully selected so that none of them had any previous experience with point-light stimuli, nor did they know about the Johansson experiment. This is extremely important because subjects that are familiar with the Johansson experiment might be primed to report walking or other human actions if presented with point-light stimuli.

The subjects were presented with the CFS stimulus and had to give a written report about their perceptual impression. They were explicitly told that “nothing” was a valid answer. The stimulus covered an area of about 11 deg by 10 deg of the visual field and it was shown for a total of four walking cycles. Dot size was about 0.3 deg.

The first important experimental result is that, in the presence of a slightly asymmetric CFS stimulus (see upper insets in Figure 2), the majority of the subjects (13 out of 17) perceived the CFS spontaneously as a human walker. The perceived walking directions are indicated by the middle insets of Figure 2. The remaining four subjects reported either seeing nothing or a bunch of dots rotating and jumping. The observed high spontaneous recognition rate indicates that the presence of the extracted critical motion features combined with some very coarse spatial information, defined by the confining rectangular regions, is sufficient to induce the percept of a human walker.

In addition, we conducted two experiments with two different groups of naïve subjects to investigate the role of the motion and form cues in the recognition of the CFS. In the first experiment we presented a CFS without asymmetric displacement of the rectangular regions. The experimental procedure was the same as in the previous experiment. Nine unpaid subjects took part in this experiment. In this

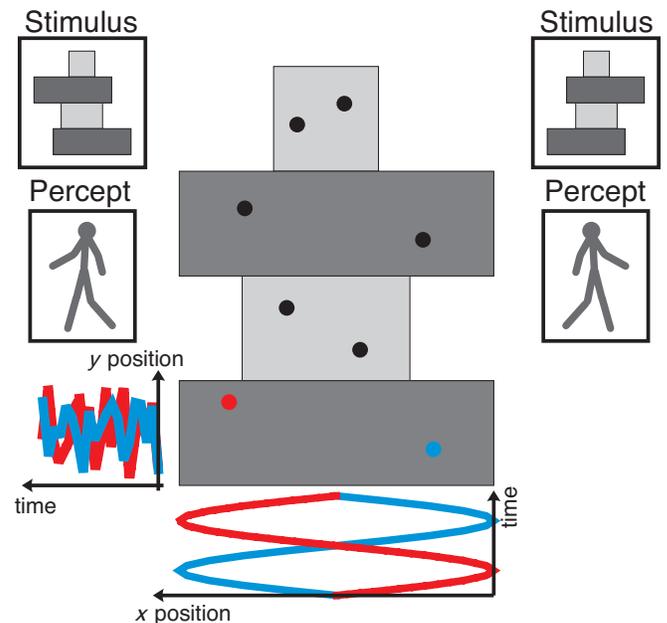


Figure 2. Critical features stimulus used in our psychophysical experiments. Dots in this point-light stimulus are confined to move within the shaded rectangular regions (regions not shown in the actual stimulus). Dot pairs in the dark regions move randomly along the y -axis and have a regular sinusoidal opponent motion along the x -axis. The positions of the dots in the light regions are randomly chosen in every frame. The stick figures (insets, middle row) indicate the percepts that can be elicited, even in completely naïve subjects, by slight displacement of the dark gray regions (insets, upper row). The lower two insets show the trajectories along the x and y axes of the dot pairs contained in the dark gray regions.

case, two subjects spontaneously perceived a human walking and the other four subjects perceived a person performing an action that was compatible with a symmetric opponent motion pattern (spinning or waving with the hands). The remaining three subjects reported seeing nothing, or the algebraic number “eight.” This experiment shows that the symmetry of the CFS is not necessary for inducing the impression of a human action. However, the asymmetry increases the percentage of subjects that perceive a human walker.

In the second experiment we removed the horizontal motion information in the CFS by presenting only dots with completely random motion within the four rectangular regions. Ten naïve subjects took part in this experiment. For this stimulus only two naïve subjects perceived a human person performing actions different from walking. The remaining eight subjects reported seeing either the algebraic number “eight” or nothing. This result shows that opponent motion seems to be critical for generating the impression of a walking human, whereas the mere presence of moving dots within the same four regions is not sufficient.

From the viewpoint of theories that assume that biological motion recognition is accomplished by the reconstruction of body shape from dot positions, the results of [Experiment 1](#) are rather unexpected. The coarse position information defined by the regions of the CFS by itself is not sufficient to fit a human skeleton model. In addition, by the random vertical displacements, the dot positions of the CFS do not comply with the kinematics of a smoothly moving human body. This should make it rather difficult to approximate the point positions of the CFS by a model of the human body. Yet it seems possible that the visual system might use fuzzy templates for the human body shape that fit the CFS in a sub-optimal way. This would predict that recognition performance for the CFS should be lower than for point-light stimuli that exactly match a human skeleton.

Experiment 2

The results of [Experiment 1](#) motivated a second, more quantitative study that compares the CFS with a similar point-light stimulus that matches exactly the shape of a human skeleton. The recognition of the stimulus that complies with the kinematics should be easier than the recognition of the CFS if a reconstruction of the human body shape from point positions is critical for the recognition of point-light walkers.

Stimuli and method

A stimulus that is very similar to the CFS, and that specifies dot positions that are exactly compatible with the human body shape, has been proposed by Beintema and Lappe (2002). Their sequential position stimulus (SPS) is generated by reassigning the dots of a point-light walker to new positions on the walker's skeleton every p -th stimulus frame ($p = 1...4$). The position updates of the dots fulfill the additional constraint that never more than one dot is assigned to the same limb in each frame. The displacement of the dots on the skeleton degrades the local motion information, compared to normal point-light walkers, but does not affect the compatibility of the dot positions with the human body shape. In spite of the degradation of the local motion information, subjects perceive the SPS as human walker.

Seven paid subjects took part in the second experiment. Six of them had previous experience with point-light stimuli and the remaining one was familiarized with the experiment during a short training session. CFS and SPS stimuli were presented in random order. The stimuli consisted of 21 frames, and they contained 1, 2, or 4 dots with a lifetime of 1 frame. Each stimulus condition (direction of walking \times number of dots) was presented 15 times during the experiment. CFS and SPS stimuli were matched with respect to low-level properties (stimulus area, cycle time, and size of the dots) and covered an area of about 9×7.6 -deg visual angle. The dots had a size of 0.2 deg, and

cycle time was about 1.2 s. Subjects were seated at a distance of 75 cm from the monitor (Sony G520 with a refresh rate of 75 Hz). Stimuli were presented using the Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997). Consistent with the experiments by Beintema and Lappe (2002), subjects had to report the perceived direction of walking (right or left) in a 2AFC paradigm.

[Figure 3](#) shows percentages of correct recognition of the direction of walking (upper panel) and response times (lower panel) as a function of the number of stimulus dots for the two stimulus classes (CFS and SPS). Recognition performances for the CFS and the SPS were virtually identical. Applying a two-way repeated-measures ANOVA to the percentages of correct responses, we found no significant effect of the stimulus type, CFS versus SPS, $F(1,6) = 0.6$, $p = .47$, but a significant effect of the number of dots, $F(2,12) = 29.20$, $p < .01$, and no significant interaction, $F(2,12) = 0.1$, $p = .91$. A compatible pattern was found for the response times, with no significant effect of stimulus type, $F(1,6) = 1.77$, $p = .23$, but a significant effect of the number of dots, $F(2,12) = 9.36$, $p < .01$, and no interaction, $F(2,12) = 1.46$, $p = .27$.

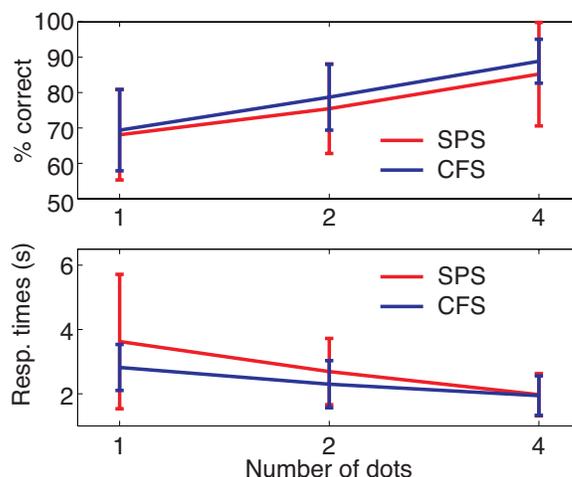


Figure 3. Psychophysical results. Mean percentages of correct recognition of the direction of walking (upper panel) and average response times (lower panel) for the critical features stimulus (CFS) (blue curve) and the sequential position stimulus (SPS) (red curve). Vertical bars indicate standard errors for seven subjects.

Our psychophysical study provides no evidence that supports an advantage for stimuli that match exactly the human body kinematics compared to the CFS. This points against a relevance of precise information about the human body shape for the recognition of walking direction for the two types of degraded point-light stimuli. In addition, the similarity of the experimental results for both types of stimuli suggests that they might be processed by a common mechanism. It seems likely that for both stimulus classes the asymmetry of the stimulus might be important for the

determination of walking direction. The coarse positional information provided by the CFS seems sufficient for accomplishing this task.

Neural modeling

Our psychophysical results demonstrate that stimuli containing the proposed critical optic flow feature with a coarse spatial arrangement tend to be perceived as a person walking. However, these experiments cannot prove that the proposed critical optic flow features are sufficient for the recognition of point-light walkers, because subjects might have used a variety of other cues possibly contained in the two stimuli.

To test how far the proposed critical features are sufficient for the recognition of degraded point-light stimuli, we have devised a neurophysiologically inspired model that exploits only these features. All components of this model can, in principle, be implemented by real neurons. However, for our purposes, it is not critical how far the individual model components really match physiological data.

Sketch of the model

The neural model that we used for our simulations is part of a more elaborated learning-based model for biological motion recognition, which accounts for a variety of experimental results with normal and point-light walkers (Giese & Poggio, 2003). The model is shown schematically in Figure 4. It consists of a hierarchy of neural detectors that extract motion features with different complexity. Feature complexity increases along the neural hierarchy. The tuning properties of the neural detectors are inspired by known properties of cortical neurons. More detailed descriptions of the model can be found in the Appendix and in Casile and Giese, 2003, Giese, 2004, and Giese and Poggio, 2003.

The neural hierarchy consists of four levels. (1) *Local motion energy detectors*. These detectors have small receptive

fields and are selective for different motion directions. For the simulations reported in this study four directions were implemented. (2) *Detectors for horizontal and vertical opponent motion*. These detectors pool the activities of local motion energy detectors with opposite direction preference within two adjacent sub-fields. The sub-field responses are then combined multiplicatively, so that the detector does not respond if only one directional component is present. (3) *Detectors for complex global optic flow patterns*. These are detectors that are modeled by radial basis functions. The selectivity of these detectors is established by training with example movement sequences (right and left walking in our case). The center of each basis function corresponds to the feature vector, extracted at the previous hierarchy level, for one frame of the training movie. Each frame defines a specific instantaneous optic flow pattern that is encoded by the radial basis function. A full walking cycle is encoded by 21 such key frames that are equally spaced in time. This number was not critical for the results. The optic flow patterns that correspond to different key frames of a walking cycle are denoted by the symbols t_1, \dots, t_n in Figure 4. The receptive fields of these detectors are larger than the whole point-light stimulus. (4) *Detectors for complete biological motion patterns*. These detectors sum and temporally smooth the activities of optic flow pattern detectors that belong to the same human action (e.g., walking right or walking left). The activities of these detectors are used to simulate the behavioral response of the model.

To convert the activities of the model neurons into simulated behavioral responses of subjects, we compared the activations of the two neurons at the highest hierarchy level that represent rightward and leftward walking. The simulated percept was assumed to be walking right if the time integral of the activity of the neural detector for rightward walking exceeded the one of the detector for leftward walking. The model response for walking left was simulated in an equivalent way. If none of the neurons was activated, the decision was chosen randomly between right or left.

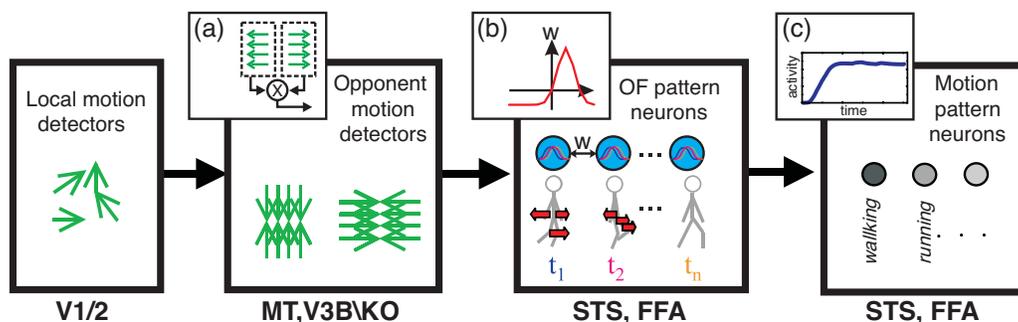


Figure 4. Schematic sketch of the model. The symbols indicate the following brain areas that might fulfill similar computational functions: V1, primary visual cortex; MT, middle temporal area; KO, kinetic occipital area; STS, superior temporal sulcus; FFA, fusiform face area. The symbols t_1, t_2, \dots, t_n indicate presentation times of input frames that are encoded by the radial basis function units that have been trained with optic flow fields that are characteristic for certain input frames. The insets show schematically (a) a detector for opponent motion; (b) the form of the lateral coupling between the detectors for complex optic flow fields as a function of the neuron number; and (c) the response as a function of time of a motion pattern detector at the highest level of the hierarchy.

For our simulations we trained the motion pattern detectors with normal rightward- and leftward-walking point-light stimuli. This choice was motivated by the fact that most subjects in [Experiment 2](#) had substantial previous experience with point-light walkers. Qualitatively similar results were obtained for training with full-body stimuli. The model was tested with rightward- and leftward-walking SPS and CFS stimuli, varying the number of dots (1 to 8) and the lifetimes of the dots (1 to 4 frames). For each combination (number of dots x lifetime of dots x direction of walking), 100 repetitions were simulated, and the dot positions were re-randomized for each trial.

Modeling results

[Figure 5](#) shows the performance of the model (percentage of correct-direction discriminations) as a function of the number and lifetime of dots in the stimulus. The model qualitatively replicates multiple aspects of the psychophysical data: (1) Recognition performances for both types of stimuli (CFS and SPS) are very similar under all considered conditions. (2) Recognition performance increases with the number of dots in the stimulus. (3) Recognition rates for 8 and 4 dots are close to the values obtained in the psychophysical experiment (for CFS and SPS stimuli with 8 dots, psychophysical performance was at ceiling level, and results are thus not reported in [Figure 3](#)). These high recognition rates are astonishing, given that the model exploits only one type of mid-level feature. The recognition rates for 2 dots are lower than human performance. This difference is likely a consequence of the fact that subjects can exploit a variety of features, whereas our model only extracts opponent motion. For the same reason, the model is not able to analyze stimuli with a single dot, because in this case the opponent motion detectors remain silent.

For both the CFS and the SPS, we find no strong increase of performance with the lifetime of dots, in particular for lifetimes above one frame. Such an increase might be expected for a recognition mechanism that is based on

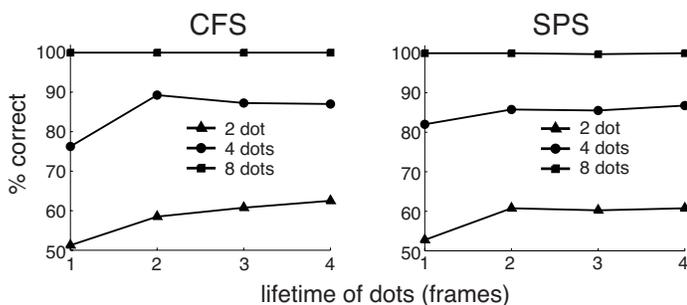


Figure 5. Recognition performances achieved by the model for the CFS (left panel) and the SPS (right panel). Percentages of correct recognition of walking direction are shown as a function of the number and lifetime of the stimulus dots. Standard errors were negligible under all the investigated conditions and are thus not reported in the figure.

local-motion, because long lifetimes should improve the quality of the local motion signals by reducing the number of discontinuities in the motion of the dots.

Our simulation study yields two important results. First, it shows that for both classes of stimuli (CFS and SPS) high recognition rates can be accomplished solely based on the proposed critical motion feature. Although it seems likely that humans exploit a mixture of features for the recognition of biological motion, opponent horizontal motion seems to be a particularly important one. Second, it proves that the SPS contains a considerable amount of horizontal motion information that can be exploited for direction discrimination.

In addition, our model demonstrates that remarkable performance rates for degraded stimuli can be accomplished without complex computational mechanisms, such as closed-loop on-line fitting of articulated models to dot positions. Furthermore, the proposed neural circuits, at least in principle, could be implemented by cortical neurons.

Discussion

In this study we have investigated possible mechanisms for the robust generalization from normal (full-body) articulated motion stimuli to point-light stimuli. We have presented multiple pieces of evidence suggesting that the detection of critical mid-level optic flow features within a specific coarse spatial arrangement might form the basis of this generalization: (I) Normal and point-light stimuli share very similar dominant mid-level optic flow features; (II) the presence of these features with the appropriate spatial arrangement induces the percept of a person walking, even though the stimuli do not comply with the kinematics of the human body; and (III) a neural model that exploits these critical features achieves substantial recognition rates, even for degraded point-light stimuli.

Our results seem to contradict a recent psychophysical study (Beintema & Lappe, 2002) that concludes that the motion information in the SPS is so dramatically degraded that its recognition must be based on the reconstruction of body shape. However, a more detailed statistical analysis seems to disprove this assumption.

The amount of local motion information in the SPS can be quantified using an *index of motion quality* (c.f. Beintema & Lappe, 2002). This quantity was defined as the fraction of dots in the SPS whose motion remains within the 10% range of the veridical motion vectors that would be valid if the dots were not randomly displaced on the skeleton. We computed this index in three different ways: (1) For the full 2D-motion vectors, (2) for the vertical motion components only, and (3) for the horizontal motion components only. In agreement with the study by Beintema and Lappe, we found for the full 2D-motion vectors that less than 2% of the dots remained in the 10% range of the veridical vectors. The same was true if we regarded only the

vertical motion components (< 2%). However, the index of motion quality for the horizontal motion components was much higher (7%), indicating a substantially higher amount of horizontal motion information. Our simulation study confirms that this residual horizontal motion information can be exploited for achieving substantial recognition rates, which are close to psychophysical data if at least 4 dots are present in the stimulus. The asymmetric degradation of horizontal and vertical motion components can be easily understood considering the fact that the limbs of a walker are predominantly vertically oriented. A separate analysis of horizontal and vertical motion components seems physiologically feasible by reading out separately neural ensembles (e.g., in area MT), which are tuned to different preferred directions.

Our model postulates the existence of neural detectors for opponent motion within adjacent receptive subfields. One might ask if this assumption matches experimental data about motion-selective neurons in the brain. Physiological studies, for example in area MT, have revealed a subpopulation of neurons that have receptive fields with antagonistic surrounds. Some of these neurons show enhanced responses if the direction of the movement in the surround is opposite to the direction of the movement in the center (Allman, Miezin, & McGuinness, 1985; Born, 2000). Opponent motion provides an adequate stimulus for such neurons. In addition, neurons that respond selectively to motion discontinuities have also been found in other areas (e.g., V1 and V2; Marcar, Raviguel, Xiao, & Orban, 2000; Reppas, Niyogi, Dale, Sereno, & Tootell, 1997). In monkey area MT it seems that neurons with reinforcing and antagonistic surrounds form separate populations (Born, 2000; Born & Tootell, 1992), suggesting that they might subserve computationally different functions. It is obvious that neurons with non-antagonistic surrounds are suitable for estimating smooth optic flow. The computational role of the neurons with antagonistic surrounds is less clear, and several hypotheses have been discussed (e.g., segmentation of moving objects from the background, the processing of relative motion, or motion parallax). Our study suggests that such neural detectors might also be useful for the processing of biological motion.

Detectors for motion discontinuities, similar to the ones postulated by our model, might also be useful for solving the aperture problem in complex visual scenes. Computational studies show that it is important for the solution of the aperture problem in scenes with multiple moving objects to prevent a combination or smoothing of local motion information across object boundaries (Koch, Marroquin, & Yuille, 1986; Liden & Pack, 1999). Opponent motion detectors may be important for detecting such discontinuities.

Our psychophysical results show that the combination of opponent motion with very coarse positional information is sufficient to induce the percept of a moving person, even in completely naïve subjects. Indeed, the CFS was purposefully designed to minimize other cues. For the de-

tection of a moving human, this limited amount of information seems to be sufficient. For more sophisticated tasks, like identification of gender or emotional content, more detailed information might be required. However, it has also been shown that fine discrimination tasks, like people identification by gait, can be based purely on local motion information (e.g., Giese & Poggio, 2003). In addition, the quantitative comparison between CFS and SPS shows that the detailed form information provided by the SPS does not seem to improve the recognition of walking direction.

The high similarity of the extracted mid-level optic flow features for normal and point-light stimuli was rather unexpected, given that point-light walkers specify a much sparser optic flow field. Even though this study has focused on walker stimuli, the proposed statistical method for the extraction of dominant form and optic flow features applies to any other complex motion stimulus. As an example, we have designed a similar CFS for running ([demo](#) provided with this article).

The importance of motion information for the recognition of biological motion has been pointed out by many previous studies (Mather & Murdoch, 1994; Mather et al., 1992; Troje, 2002). However, the exact nature of the underlying motion features has not so far been clarified, nor have methods been proposed in the psychophysical literature that would allow an identification of such critical features. The detection of mid-level optic flow features with relatively coarse spatial localization provides an elegant explanation for the generalization from normal to point-light stimuli, and also to strongly degraded stimuli like the SPS or the CFS. This explanation seems appealing because it does not require complex computational mechanisms and, in principle, can be implemented with relatively simple neural circuits.

An alternative, although in our view less likely explanation of our results, is a recognition of degraded stimuli based on mechanisms that reconstruct missing information about the body shape (e.g., by fitting articulated models or shape templates to the point positions) (see Giese, [in press](#)). A large body of work in computer vision (Aggarwal & Cai, 1999; Curio & Giese, 2005; Gavrila, 1999) shows that such a reconstruction of missing form information from degraded stimuli is possible in principle. However, most of the existing methods are computationally quite expensive. Algorithms that are based on explicit articulated models typically require the solution of high-dimensional nonlinear optimization and search problems because the position, scaling, and posture of the model are a priori unknown. In addition, the postures specified by monocular visual stimuli are often not unique, requiring methods for multi-hypothesis tracking. A particularly difficult problem is the fitting of articulated shape models in the presence of motion clutter. Psychophysical experiments have shown that biological motion recognition is easily accomplished by human subjects in the presence of moving masking dots (Cutting et al., 1988; Thornton et al., 1998). In technical systems that fit models to feature positions, motion clutter

leads to complex correspondence problems, which have been solved by applying algorithms for search in high-dimensional spaces (Rashid, 1980; Song, Goncalves, & Perona, 2003). Such algorithms typically require many iterative steps. The computational complexity of these methods seems difficult to reconcile with the experimental fact that biological motion recognition in humans and monkeys is very fast, requiring less than 200 ms (Johansson, 1976; Oram & Perrett, 1996). In addition, it remains an open question whether the required algorithms can be implemented with real neurons (cf. Lee & Mumford, 2003).

Our hypothesis of a recognition of point-light stimuli by an analysis of mid-level optic flow features seems compatible with different imaging studies that report activity, which seems compatible for point-light biological motion stimuli, in areas that are typically associated with the dorsal processing stream (e.g., Grossman et al., 2000; Prito, Faubert, Gjedde, & Kupers, 2003; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001). However, other studies also find selective activation by point-light walkers in areas like extrastriate body part area (EBA) and fusiform face area (FFA), which are often assigned to the ventral processing stream (Downing, Jiang, Shuman, & Kanwisher, 2001; Grossman & Blake, 2002). Many studies have failed to find selective activation for point-light walkers in the form-selective area LOC (Grossman et al., 2000; Prito et al., 2003; Vaina et al., 2001). Thus, It remains an open question how exactly form and motion-selective areas interact during the perception of point-light stimuli.

The importance of opponent motion for the recognition of point-light walkers is suggested by fMRI experiments that show an activation of the kinetic occipital area (KO/V3B) for biological motion stimuli (Santi, Servos, Vatikiotis-Bateson, Kuratate, & Munhall, 2003; Vaina et al., 2001). This area has previously been associated with the processing of motion edges and moving objects (Dupont et al., 1997; Orban et al., 1995). A critical role of opponent motion for the detection of point-light walkers seems also consistent with data from the neurological patient AF, who could perceive biological motion in spite of a lesion in the dorsal pathway (Vaina, Lemay, Bienfang, Choi, & Nakayama, 1990). Detailed investigations of the lesion sites suggest that this patient still has area V3B/KO intact (Vaina & Giese, 2002), so that his perception of opponent motion might not be strongly impaired.

Our psychophysical and computational results suggest that relative limb motion might be important for the rec-

ognition of human locomotion. This finding is consistent with psychophysical results in adults (Pinto & Shiffrar, 1999) and infants (Booth, Pinto, & Bertenthal, 2002). In particular, it was shown that infants at the age of about 5 months shift their interest from the absolute and relative motion of individual limbs to the relative motion of contralateral limbs (Booth et al., 2002).

Although in this study we have focused on possible feed-forward mechanisms for achieving a robust recognition of biological motion, we assume that under normal conditions, biological motion recognition is modulated by higher level cognitive representations. Experimental evidence suggests strong influences of top-down processes (Bülthoff, Bülthoff, & Sinha, 1998; Cavanagh, Labianca, & Thornton, 2001; Thornton, Rensink, & Shiffrar, 2002) and potentially representations of biomechanical plausibility (Shiffrar & Freyd, 1990, 1993). In addition, interactions with internal representations of motor programs might play an important role, as suggested by a number of recent psychophysical, neurophysiological, and fMRI studies (Decety & Grezes, 1999; Prinz, 1997; Rizzolatti, Fogassi, & Gallese, 2001; Saygin, Wilson, Hagler, Bates, & Sereno, 2004).

The proposed mechanism (i.e., the detection of critical mid-level motion features) defines a computational hypothesis on how basic visual recognition of normal and impoverished point-light stimuli might be accomplished with high robustness and realistic processing times. However, it seems likely that the human brain integrates a variety of features during biological motion recognition. The proposed critical feature might be particularly important, but more complex tasks like the fine discrimination of actions might require the exploitation of multiple features, or even a modulation of the detection process by high-level cognitive representations.

Appendix

Description of the model

Table 1 shows an overview of the most important properties of the detectors of our hierarchical neural model.

The first hierarchy level models local motion energy detectors. To reduce the computational costs, these detectors were approximated by computing the optic flow from the stimulus sequence. Motion energy signals were computed

Detector	Area	# Detectors	RF Size	Reference
Local motion detectors	V1, MT	1116	≈0.4 deg	Snowden, 1994
Opponent motion detectors	MST, KO/V3B	4x25	4.5 deg	Saito, 1993; Tanaka & Saito, 1989
OF pattern detectors	STS	21	whole stimulus (>8 deg)	Decety & Grezes, 1999; Oram & Perrett, 1994; Perrett et al., 1985; Vaina et al., 2001
Motion pattern neurons	STS	2	whole stimulus (>8 deg)	Decety & Grezes, 1999; Oram & Perrett, 1994; Perrett et al., 1985; Vaina et al., 2001

Table 1. Most important parameters of the neural detectors in our model. RF = receptive field, V1 = primary visual cortex, MT = middle temporal area, MST = medial superior temporal area, KO = kinetic occipital area, OF = optic flow, STS = superior temporal sulcus.

from the optic flow assuming direction selective detectors. Local motion detectors were arranged in a 36×31 grid. In the current implementation we modeled cells with four different preferred directions (0, 90, 180, and 270 deg) and with a speed-tuning that corresponds to band-pass characteristics. Neurons that are selective for local motion energy have been reported in monkey visual cortex in area V1/2 and in area MT (Snowden, 1994).

The output $g_p(x)$ of a local motion detector at position x with preferred direction θ_p to a stimulus with velocity v and direction θ is given by

$$g_p(x) = H(v(x), v_1, v_2) \cdot b(\theta, \theta_p),$$

where H is a rectangular speed-tuning function with $H(v, v_1, v_2) = 1$ for $v_1 < v < v_2$ and $H(v, v_1, v_2) = 0$ otherwise. The function $b(\theta, \theta_p)$ determines the direction-tuning of the motion energy detectors and is given by

$$b(\theta, \theta_p) = \left\{ \frac{1}{2} [1 + \cos(\theta - \theta_p)] \right\}^q.$$

The positive parameter q determines the width of the direction-tuning function. For the simulations presented in this study we chose $q = 2$.

The second level of the model contains neural detectors that are selective for *opponent motion*. The activities of the opponent motion detectors are obtained by combining the responses of the local motion energy units within two adjacent subfields with opposite direction selectivity. The response of each subfield is obtained by pooling the responses of local motion detectors with same-direction preference within the subfield (see Figure 4a). The output $o_l(x)$ of a local opponent motion detector of type l centered at position x is obtained by taking the product of the maxima of the local motion detectors over the two sub-fields, that is,

$$o_l(x) = \sqrt{\max_i(g_p(x_i)) \max_j(g_r(x_j))},$$

where the indices i and j sample the spatial positions of the two subfields with direction preferences p and r .

Partial spatial position invariance of the opponent motion detectors is achieved by pooling the responses of detectors with the same characteristics l at different spatial positions x_k within its receptive field using a maximum operator (Fukushima, 1980; Riesenhuber & Poggio, 1999). Thus, the final output $o_l(x)$ of an opponent motion detector is given by

$$o_l(x) = \max_k(o_l(x_k)).$$

Maximum computation has been found in the visual cortex of monkeys (Gawne & Martin, 2002) and cats (Lampl, Ferster, Poggio, & Riesenhuber, 2004). Results reported in this study were obtained using four types of opponent motion detectors: detectors sensitive for contracting and expanding flows along horizontal and vertical di-

rection. These detectors were arranged in a 5×5 grid covering the whole stimulus area.

In monkeys, opponent motion-sensitive neurons have been reported, for example, in the MT and medial superior temporal (MST) areas (Born, 2000; Tanaka & Saito, 1989). In humans, imaging experiments suggest that opponent motion-sensitive neurons might be located in the kinetic occipital area (KO/V3B) (Orban et al., 1995; Orban et al., 1992).

The next higher level of the motion pathway consists of *optic flow pattern detectors*. The selectivity of these detectors is learned from training sequences. Each of these detectors encodes an instantaneous characteristic optic flow field that is characteristic for one frame of the training stimulus. The optic flow pattern detectors are modeled by Gaussian radial basis functions:

$$G(u) = e^{(u-u_0)^T C(u-u_0)}.$$

The feed-forward input to this layer is given by the instantaneous responses of the opponent motion detectors arranged into the vector u . The centers u_0 of the radial basis functions for each neuron are set during the training. C is a diagonal matrix whose elements are set during the training. Elements corresponding to components of the vector u , whose variance over the training set does not exceed a certain threshold, are set to zero. For the other components the elements C_{ll} are proportional to the inverse of this variance.

The activity of the optic flow pattern detectors provides input signals for the motion pattern neurons that form the highest level of the model hierarchy (see section [Temporal integration](#)).

Sequence selectivity

Biological motion recognition is critically dependent on the temporal order of the presented stimulus frames. This is obvious because presentation of a movie that is scrambled in time does not result in a well-defined percept of biological motion. For this reason the model contains a neural mechanism that makes recognition sequence-selective. Again, for the purpose of this study, it is not important whether the chosen mechanism is really consistent with the circuits in visual cortex. The simulations show that a model with sequence selectivity can extract the relevant information.

One possible neural mechanism of sequence selectivity is based on asymmetric lateral connections between the optic flow pattern detectors (Mineiro & Zipser, 1998): By these lateral connections, the presently active neuron preactivates the neurons encoding future optic flow patterns, and inhibits neurons encoding other patterns. The activity $H_k^l(t)$ of the optic flow pattern neuron encoding the k -th frame belonging to the l -th training sequence obeys the dynamics,

$$\tau \dot{H}_k^l(t) = -H_k^l(t) + \sum_m w(k-m) f(H_k^l(t)) + G_k^l(t),$$

where τ is a time constant ($\tau = 150$ ms). $w(m)$ is an asymmetric interaction kernel, shown in Figure 4b. The function $f(H)$ is a step threshold function, and $G_k^l(t)$ is the feed-forward input as defined in the previous section. It has been shown elsewhere (Mineiro & Zipser, 1998; Xie & Giese, 2002) that for appropriate choice of the interaction kernel, substantial activity arises only if the stimulus frames are presented in the right temporal order. Otherwise, the feed-forward input signals of the network and the recurrent feedback compete in a way that leads to a solution with very small amplitude.

Temporal integration

The highest level of the model consists of *motion pattern neurons* that are selective for complete biological movement patterns like walking right or walking left. These detectors sum the output activities of all optic flow pattern detectors belonging to the same biological movement pattern and integrate over time. The activity $P^l(t)$ of the motion pattern neuron encoding the response to the l -th stored pattern (e.g., walking right) obeys the dynamics:

$$\tau_s \dot{P}^l(t) = -P^l(t) + \sum_k H_k^l(t),$$

where τ_s is a time constant ($\tau_s = 150$ ms) and $H_k^l(t)$ is the activity of the optic flow pattern detector encoding the k -th snapshot of the l -th training sequence. An example of the output of such a detector in presence of a point-light walker is shown in the inset (c) of Figure 4.

Neurons selective for biological motion patterns have been found in the superior temporal polysensory area of monkeys (Oram & Perrett, 1996; Perrett et al., 1985). Imaging studies in humans suggest that such detectors might exist in the superior temporal sulcus (Grossman et al., 2000; Vaina et al., 2001), and potentially also in FFA (Grossman & Blake, 2002).

Acknowledgments

We thank Isabelle Bühlhoff, Ian Thornton, and Lucia Vaina for insightful comments on an earlier version of this manuscript. The authors are supported by the Volkswagenstiftung, Deutsche Forschungsgemeinschaft, and the Human Frontier Science Program. Martin Giese is visiting fellow of the Department of Biomedical Engineering at Boston University.

Commercial relationships: none.

Corresponding author: Martin A. Giese.

Email: martin.giese@uni-tuebingen.de.

Address: Laboratory for Action Representation and Learning, University Clinic, Schaffhausenstr, 113 D-72072, Tübingen, Germany.

References

- Aggarwal, J., & Cai, Q. (1999). Human motion analysis: A review. *Computer Vision and Image Understanding*, 73(3), 428-440.
- Ahlström, V., Blake, R., & Ahlström, U. (1997). Perception of biological motion. *Perception*, 26, 1539-1548. [PubMed]
- Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Annual Review Neuroscience*, 8, 407-430. [PubMed]
- Beintema, J., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences U.S.A.*, 99(8), 5661-5663. [PubMed][Article]
- Bertenthal, B. I., Proffitt, D. E., & Cutting, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motion. *Journal of Experimental Children Psychology*, 37, 213-230. [PubMed]
- Booth, A. E., Pinto, J., & Bertenthal, B. I. (2002). Perception of symmetrical patterning of human gait by infants. *Developmental Psychology*, 38(4), 554-563. [PubMed]
- Born, R. T. (2000). Center-surround interactions in the middle temporal visual area of the owl monkey. *Journal of Neurophysiology*, 84, 2658-2669. [PubMed]
- Born, R. T., & Tootell, R. B. H. (1992). Segregation of global and local motion processing in primate middle temporal visual area. *Nature*, 357, 497-499. [PubMed]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433-436. [PubMed]
- Bülhoff, I., Bülhoff, H., & Sinha, P. (1998). Top-down influences on stereoscopic depth perception. *Nature Neuroscience*, 1(3), 254-257. [PubMed]
- Casile, A., & Giese, M. (2003). Roles of motion and form in biological motion recognition. In O. Kaynak, E. Alpaydin, E. Oja, & L. Xu (Eds.), *Artificial neural networks and neural information processing* (Vol. 2714, pp. 854-862). Berlin: Springer.
- Cavanagh, P., Labianca, A., & Thornton, I. (2001). Attention based visual routines: Sprites. *Cognition*, 80(1-2), 47-60. [PubMed]
- Curio, C., & Giese, M. (2005). Combining view-based and model-based tracking of articulated human movements. Paper presented at the IEEE Computer Society Workshop on Motion and Vision Computing, Beckenridge, Colorado.
- Cutting, J. E., Moore, C., & Morrison, R. (1988). Masking the motion of human gait. *Perception and Psychophysics*, 44(4), 339-347. [PubMed]

- Cutting, J. E., Proffitt, D. E., & Kozlowski, L. T. (1978). A biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception and Performance*, 4(3), 357-372. [PubMed]
- Decety, J., & Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences*, 3(5), 172-178. [PubMed]
- Downing, P., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area for visual processing of the human body. *Science*, 293, 2470-2473. [PubMed]
- Dupont, P., Bruyn, B. D., Vandenberghe, R., Rosier, A., Michiels, J., Marchal, G., et al. (1997). The kinetic occipital region in human visual cortex. *Cerebral Cortex*, 7(3), 283-292. [PubMed]
- Fox, R., & McDaniel, C. (1982). The perception of biological motion by human infant. *Science*, 218(4571), 486-487. [PubMed]
- Fukushima, K. (1980). Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36, 193-202. [PubMed]
- Gattass, R., Sousa, A., & Gross, C. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *Journal of Neuroscience*, 8(6), 1831-1845. [PubMed]
- Gavrila, D. (1999). The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1), 82-98.
- Gawne, T. J., & Martin, J. M. (2002). Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *Journal of Neurophysiology*, 88(3), 1128-1135. [PubMed]
- Giese, M. A. (2004). A neural model for biological movement recognition: A neurophysiologically plausible theory. In L. M. Vaina, S. A. Beardsley, & S. K. Rushon (Eds.), *Optic flow and beyond* (pp. 443-470). Dordrecht: Kluwer.
- Giese, M. A. (in press). Computational principles for the recognition of biological movements. In G. Knoblich, I. M. Thornton, M. Grosjean, M. Shiffrar (Eds.), *Perception of the human body from the inside out*. Oxford: Oxford University Press.
- Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological motion. *Nature Reviews Neuroscience*, 4(3), 179-192. [PubMed]
- Grossman, E., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron*, 35, 1167-1175. [PubMed]
- Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., et al. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, 12(5), 711-720. [PubMed]
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14, 201-211.
- Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. *Psychological Research*, 38, 379-393. [PubMed]
- Koch, C., Marroquin, J., & Yuille, A. (1986). Analog "neuronal" networks in early vision. *Proceedings of the National Academy of Sciences U.S.A.*, 83(12), 4263-4267. [PubMed][Article]
- Lampl, I., Ferster, D., Poggio, T., & Riesenhuber, M. (2004). Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *Journal of Neurophysiology*, 92(5), 2704-2713. [PubMed]
- Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, 20(7), 1434-1448. [PubMed]
- Liden, L., & Pack, C. (1999). The role of terminators and occlusion cues in motion integration and segmentation: A neural network model. *Vision Research*, 39(19), 3301-3320. [PubMed]
- Marcas, V. L., Raiguel, S. E., Xiao, D., & Orban, G. A. (2000). Processing of kinetically defined boundaries in areas V1 and V2 of the macaque monkey. *Journal of Neurophysiology*, 84, 2786-2798. [PubMed]
- Marr, D., & Vaina, L. (1982). Representation and recognition of the movements of shape. *Proceedings of the Royal Society of London B*, 214(1197), 501-524. [PubMed]
- Mather, G., & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London B*, 258(1353), 273-279.
- Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society of London B*, 249(1325), 149-155. [PubMed]
- Mineiro, P., & Zipser, D. (1998). Analysis of direction selectivity arising from recurrent cortical interactions. *Neural Networks*, 10, 353-371. [PubMed]
- Neri, P., Morrone, M., & Burr, D. (1998). Seeing biological motion. *Nature*, 395, 894-896. [PubMed]
- Oram, M. W., & Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to 'biological motion' stimuli. *Journal of Cognitive Neuroscience*, 6, 99-116.
- Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *Journal of Neurophysiology*, 76, 109-129. [PubMed]
- Orban, G. A., Dupont, P., Bruyn, B. D., Vogels, R., Vandenberghe, R., & Mortelmans, L. (1995). A motion area in human visual cortex. *Proceeding of the National Academy of Sciences U.S.A.*, 92, 993-997. [PubMed][Article]

- Orban, G. A., Lagae, L., Verri, A., Raiguel, S., Xiao, D., Maes, H., et al. (1992). First-order analysis of optical flow in monkey brain. *Proceeding of the National Academy of Sciences U.S.A.*, 89, 2595-2599. [PubMed][Article]
- Pavlova, M., Krägeloh-Mann, I., Sokolov, A., & Birbaumer, N. (2001). Recognition of point-light biological motion displays by young children. *Perception*, 30, 925-933. [PubMed]
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437-442. [PubMed]
- Perrett, D. I., Smith, P. A., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., et al. (1985). Visual analysis of body movements by neurones in the temporal cortex of the macaque monkey: A preliminary report. *Behavioral Brain Research*, 16(2-3), 153-170. [PubMed]
- Pinto, J., & Shiffrar, M. (1999). Subconfigurations of the human form in the perception of biological motion displays. *Acta Psychologica*, 102, 293-318. [PubMed]
- Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9(2), 129-154.
- Ptito, M., Faubert, J., Gjedde, A., & Kupers, R. (2003). Separate neural pathways for contour and biological-motion cues in motion-defined animal shapes. *NeuroImage*, 19, 246-252. [PubMed]
- Rashid, R. F. (1980). Towards a system for the interpretation of moving lights display. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(6), 574-581.
- Reppas, J. B., Niyogi, S., Dale, A. M., Sereno, M. I., & Tootell, R. B. H. (1997). Representation of motion boundaries in retinotopic human visual cortical areas. *Nature*, 388, 175-179. [PubMed]
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11), 1019-1025. [PubMed]
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2, 661-670. [PubMed]
- Saito, H. (1993). In T. Ono, L. R. Squire, M. E. Raichle, D. I. Perrett, & M. Fukuda (Eds.), *Brain mechanisms of perception and memory* (pp. 121-140). Cambridge: Oxford University Press.
- Santi, A., Servos, P., Vatikiotis-Bateson, E., Kuratate, T., & Munhall, K. (2003). Perceiving biological motion: Dissociating visible speech from walking. *Journal of Cognitive Neuroscience*, 15(6), 800-809. [PubMed]
- Saygin, A. P., Wilson, S. W., Hagler, D. J., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *Journal of Neuroscience*, 24(27), 6181-6188. [PubMed]
- Shiffrar, M., & Freyd, J. J. (1990). Apparent motion of the human body. *Psychological Science*, 1, 257-264.
- Shiffrar, M., & Freyd, J. J. (1993). Timing and apparent motion path choice with human body photographs. *Psychological Science*, 3(4), 379-384.
- Snowden, R. J. (1994). Motion processing in the primate cerebral cortex. In A. T. Smith & R. J. Snowden (Eds.), *Visual detection of motion* (pp. 51-84). London: Academic Press.
- Song, Y., Goncalves, L., & Perona, P. (2003). Unsupervised learning of human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7), 1-14.
- Tanaka, K., & Saito, H. (1989). Analysis of motion in the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology*, 62(3), 626-641. [PubMed]
- Thornton, I. M., Pinto, J., & Shiffrar, M. (1998). The visual perception of human locomotion. *Cognitive Neuropsychology*, 15(6/7/8), 535-552.
- Thornton, I. M., Rensink, R. A., & Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception*, 31, 837-853. [PubMed]
- Todd, J. T. (1983). Perception of gait. *Journal of Experimental Psychology: Human Perception and Performance*, 9(1), 31-42. [PubMed]
- Troje, N. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2(5), 371-387, <http://journalofvision.org/2/5/2>, doi:10.1167/2.5.2. [PubMed][Article]
- Vaina, L. M., & Giese, M. (2002). Biological motion: Why some motion impaired stroke patients "can" while others "can't" recognize it? A computational explanation [Abstract]. *Journal of Vision*, 2(7), 332, <http://journalofvision.org/2/7/332>, doi:10.1167/2.7.332.
- Vaina, L. M., Lemay, M., Bienfang, D., Choi, A., & Nakayama, K. (1990). Intact "biological motion" and "structure from motion" perception in a patient with impaired motion mechanisms: A case study. *Visual Neuroscience*, 5, 353-369. [PubMed]
- Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences U.S.A.*, 98(20), 11656-11661. [PubMed][Article]
- Webb, J., & Aggarwal, J. (1982). Structure from motion of rigid and jointed objects. *Artificial Intelligence*, 19, 107-130.
- Xie, X., & Giese, M. (2002). Nonlinear dynamics of direction-selective recurrent neural media. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 65(5 Pt 1), 1539-3755. [PubMed]