

# Roles of motion and form in biological motion recognition

Antonino Casile and Martin Giese

Laboratory for Action Representation and Learning  
University Clinic, Tübingen

{casile,martin.giese@tuebingen.mpg.de}

**Abstract.** Animals and humans recognize biological movements and actions with high robustness and accuracy. It still remains to be clarified how different neural mechanisms processing form and motion information contribute to this recognition process. We investigate this question using simple learning-based neurophysiologically inspired mechanisms for biological motion recognition. In quantitative simulations we show the following results: (1) Point light stimuli with strongly degraded local motion information can be recognized with a neural model for the (dorsal) motion pathway. (2) The recognition of degraded biological motion stimuli is dependent on previous experience with point light stimuli. (3) Opponent motion features seem to be critical for the recognition of these stimuli.

## 1 Introduction

Biological movements, like locomotion of humans and animals, are perceived visually with amazing sensitivity and robustness. This was demonstrated in a seminal experiment by Johansson [5]. He demonstrated that complex movements can be recognized from strongly impoverished stimuli that consisted of a small number of moving dots (cf. section 2). Though the form information conveyed by these "point light stimuli" is highly impoverished, subjects perceived the moving dots as human beings and could correctly recognize the executed actions. This has been interpreted as evidence that biological motion recognition is primarily based on the analysis of motion information [6]. Physiological evidence suggests that form and motion information are processed in two separate visual pathways that depart from the primary visual cortex. Form information is processed predominantly in a ventral pathway that includes areas V2, V4 and IT in monkeys. Motion information is processed predominantly in a dorsal pathway that includes areas as MT and MST. This implies that, according to the classical view, biological motion stimuli are mainly processed in the dorsal pathway.

This view has been recently challenged on the basis of results from patients who, in spite of lesions in the dorsal pathway, still could recognize biological motion stimuli [9]. Additional evidence that seems to contradict an involvement of the dorsal pathway has been obtained in experiments using point light displays with strongly degraded local motion information [1].

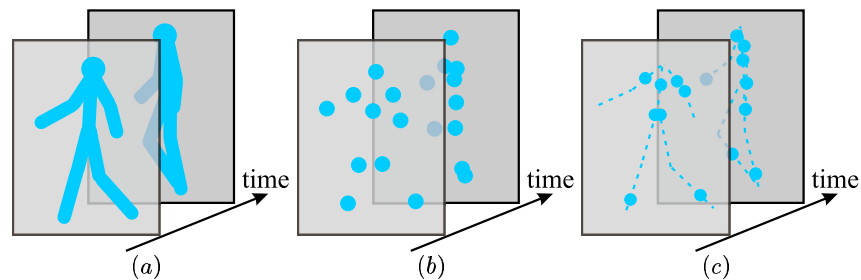
In this paper we use a learning-based neural network model for the recognition of biological movements. The model includes two separate processing streams for the analysis of form and motion information and allows to test the computational relevance of the two pathways. The model is consistent with neurophysiological data and reproduces a variety of experimental results on biological motion recognition [2]. We show that recognition of previously learned

patterns can be accomplished with either pathway alone and depends on the previous learning history. However, generalization between regular (full-body) biological motion stimuli and point light stimuli could be accomplished only in the motion pathway of the model. For this generalization opponent motion features seem to be highly important.

## 2 Stimuli

In psychophysical experiments different types of stimuli have been used to study biological movement recognition (cf. figure 1). Subjects can easily recognize the direction and type of locomotion (e.g. walking vs. running) from movies of regular (full-body) walkers, and even from static pictures of such walkers (figure 1 (a)).

When movies of point light walkers (cf. figure 1 (b)) are shown to **naive** subjects, who have never seen such stimuli before they can recognize easily the type and direction of locomotion. If, however, individual frames from such movies are presented as static pictures naive subjects can not recognize the locomotion pattern, and often even do not detect the presence of a human body. This fact was interpreted as evidence for the hypothesis that point light stimuli are recognized using predominantly the dorsal pathway. Of course, after some training, observers can learn to categorize correctly individual static dot configurations. Subjects that have received such previous training will be called **expert** subjects in the following.



**Fig. 1.** Stimuli used in our simulations. (a) Regular (full-body) walker; (b) Point light walker; (c) Sequential position stimulus (SPS).

Recently a novel point light stimulus with strongly degraded local motion information has been proposed by Beintema and Lappe [1]. The dots<sup>1</sup> of their **sequential position stimulus (SPS)** are positioned randomly on the limbs of the walker, i.e. not necessarily on the major joints. They were randomly reassigned to a new limb every  $p$  frames ( $p = 1...4$ ). Figure 1 (c) shows an SPS that consists of 8 dots. (The bones of the skeleton are sketched only for illustrative purposes; they are not shown in the real stimulus.) Despite the fact that the local motion information of the SPS is strongly perturbed, subjects could judge the direction of walking with high accuracy. From this it has been concluded that perception of the SPS must be based on the reconstruction of the form of the walker exploiting predominantly mechanisms in the ventral pathway [1], possibly by fitting a kinematic model of the skeleton.

<sup>1</sup> Depending on the experiment, 1 to 8 points were shown.

### 3 Model Description

In this section we give a brief overview of our model. A complete description of the details exceeds the scope of this paper and can be found in [2].

The form and the motion pathway of the model consist of a hierarchy of neural units that are selective for motion and form features with different levels of complexity. The complexity of the extracted features and the position and scale invariance of the detectors increase along the hierarchy. The parameters of the neural detectors are chosen so that they match available neurophysiological results. The highest levels of the hierarchy contain neurons that represent biological motion patterns in terms of learned snapshots (in the form pathway) and characteristic optic flow patterns (in the motion pathway). These neurons are trained with examples of biological motion sequences.

#### 3.1 Form pathway

The form pathway analyzes biological motion by extracting the form information contained in individual "snapshots" from movement sequences. This is accomplished by recognizing body shapes using a neural model for stationary object recognition (cf. [8]). The first level of the form pathway, corresponding to brain areas V1 and V2, models simple cells in primary visual cortex using Gabor filters. This stage extracts **oriented contours** with two different spatial scales and 8 different orientations. The next level of the form pathway contains neurons that detect oriented **bars** irrespective of their exact positions within the receptive field and of their spatial scale. Such behavior has been observed for complex cells in primary visual cortex (e.g. area V2) and for neurons in area V4. The model achieves position and scale invariance by pooling the responses of the Gabor filters with same preferred orientation, but different spatial positions and scale using a maximum operation [8]. The next-higher level of the form pathway contains neurons that are selective for snapshots from movement sequences. The feed-forward inputs of these **snapshot neurons** are modeled by radial basis function units:  $G(u) = \exp(-[u - u_0]C[u - u_0])$ , where  $u$  is the response vector of the complex cells in the previous layer. The centers  $u_0$  and covariance matrix  $C$  of the basis functions are learned by training with example movement sequences<sup>2</sup> (see [2]).

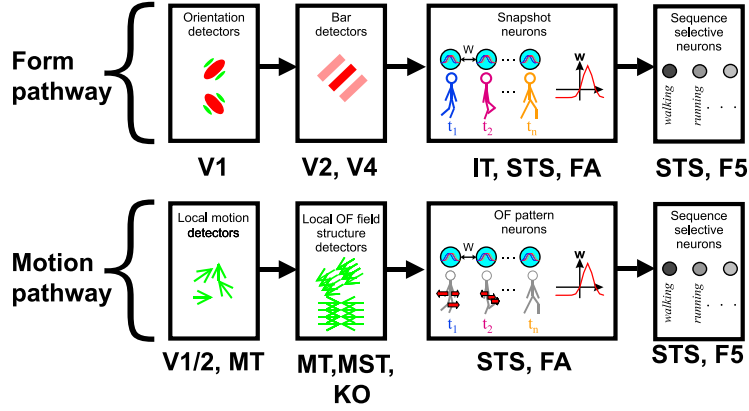
#### 3.2 Motion pathway

The motion pathway recognizes biological movements by analyzing optic flow information. The first level of this pathway consists of **local motion detectors** that compute a motion energy signal that is derived from local optic flow vectors. These neurons model speed- and direction-selective cells in areas V1/2 and MT. The model contains detectors for 4 different motion directions and two speed regimes.

The second level of the motion pathway models neurons which are selective for opponent motion. Such neurons may be located in areas MT, MST and KO. The activities of the opponent motion detectors are obtained by combining the responses of two adjacent subfields with opposite direction selectivity in a

---

<sup>2</sup> Each training movement sequence was represented by 21 snapshot neurons.



**Fig. 2.** Overview of the model. See section 3 for details.

multiplicative manner. The response of each subfield is obtained by summation of the responses of local motion detectors with same direction preference. Position invariance is obtained by pooling the responses of opponent motion detectors with different positions using a maximum operation. An inclusion of additional detectors for translational motion at this hierarchy level did not improve the recognition performance for the tested stimuli.

The next-higher level of the motion pathway consists of **optic flow pattern detectors** which are analogous to the snapshot neurons in the form pathway. These neurons are selective for complex global optic flow patterns that arise during individual instances of biological movements. The existence of this type of neurons is a prediction of the model. Their feed-forward input is modeled by radial basis function units that receive their inputs from the opponent motion detectors. They are trained like the snapshot neurons in the form pathway.

### 3.3 Sequence selectivity

The analysis of individual snapshots is not sufficient to account for biological motion recognition. Only if the snapshots are presented in the correct temporal order a percept of biological motion arises, implying that biological motion recognition is sequence-selective. One possible neural mechanism of sequence selectivity is based on asymmetric lateral connections between the snapshot neurons in the form pathway and the optic flow pattern detectors in the motion pathway [7]: By these lateral connections the presently active neuron preactivates the neurons encoding future body configurations, and inhibits neurons encoding other body configurations. The activity  $H_k^l(t)$  of the snapshot (or OF 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  $\tau$  is a time constant,  $w(m)$  is an asymmetric interaction kernel,  $f(\cdot)$  is a monotonic nonlinear threshold function, and  $G_k^l(t)$  is the feed-forward input of the neuron according to section 3.1. It is shown elsewhere [7,3] that for appropriate choice of the interaction kernel (see insets in fig 2) substantial activity arises only if the stimulus frames are presented in the right temporal order.

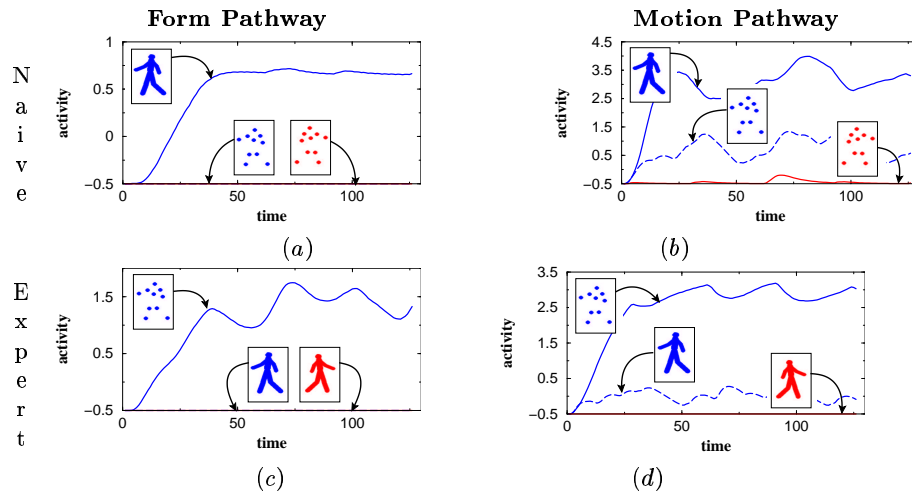
The highest level of both pathways consists of **sequence selective neurons** which sum the output activities of all the snapshot neurons (or OF pattern detectors) and smooth them over time. The activity  $P^l(t)$  of the sequence-selective neuron encoding the response to the  $l$ -th stored pattern obeys the dynamics:  $\tau_s \dot{P}^l(t) = -P^l(t) + \sum_k H_k^l(t)$ , where  $\tau_s$  is a time constant and  $H_k^l(t)$  is the activity of the snapshot neuron or the OF pattern detector encoding the  $k$ -th snapshot of the  $l$ -th training sequence. Sequence selective neurons might be found in the Superior Temporal Sulcus (STS) and maybe area F5 in monkeys.

## 4 Simulation results

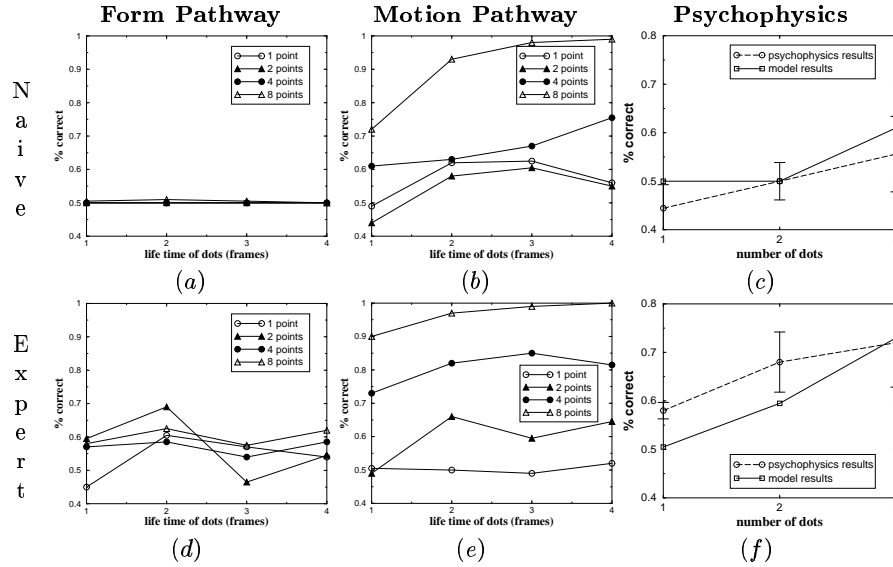
In this section we present data showing the generalization properties of the model (figure 3) and the recognition performances for strongly impoverished stimuli (figure 4). All simulations reported here were obtained with a model containing only opponent motion detectors on the second level of the motion pathway. We also tested a model that contained additional translation detectors modeling translation-selective neurons, e.g. in area MT. These detectors did not improve the recognition performance. This implies that opponent motion seems to be the critical feature for the recognition of the tested point light stimuli.

### 4.1 Generalization: regular vs. point light stimuli

In a first set of simulations we tried to reproduce Johansson's main result [5]: Naïve subjects who have seen only regular walkers can easily recognize point light walkers. Both pathways of the model were trained with regular walker stimuli, and were tested either with regular walkers or point light stimuli walking in the same or in opposite direction.



**Fig. 3.** Generalization to different types of stimuli then previously learned. For each figure the inset in the upper left corner shows the pattern presented during the training. The other insets show patterns presented during testing. Solid lines indicate the response of the sequence selective neurons for the training patterns. The dashed curves show the responses for the other stimulus type, walking in two different directions. Upper (lower) panels indicate the responses for a simulated naïve (expert) subject.



**Fig. 4.** Percentages of correct direction discrimination of an using either the form pathway (left column) or the motion pathway (central column). The rightmost column shows psychophysics results obtained with 3 naive (upper row) or 5 expert (lower row) subjects and recognition percentages predicted by the model. Error bars are  $\pm 1SE$ .

Figures 3 (a) and (b) show that regular stimuli elicit activity in both pathways. On the contrary, only the motion pathway is activated when point light stimuli are presented. Moreover, this activity is highly selective for the direction of walking. This result implies that generalization from regular to point light stimuli is possible in the motion pathway of the model, but not in the form pathway.

It is an interesting to test whether a similar generalization occurs when the model is trained with point light walkers and tested with regular walker stimuli. Both pathways of the model were trained with a point light walker and tested with point light walkers or regular stimuli walking in opposite directions. As shown in figures 3 (c) and (d) generalization occurs in the motion pathway, but not in the form pathway. Again the activity in the motion pathway is selective for the walking direction.

These results indicate that generalization to novel types of stimuli can be achieved in the motion pathway much better than in the form pathway – at least for our model.

## 4.2 Recognition of the Sequential Position Stimulus (SPS)

In a second series of experiments we tested the capability of the model to recognize stimuli with strongly impoverished motion information (cf. Figure 1 (c)). In particular we tested the hypothesis that only form information can be used for recognizing the sequential position stimulus (SPS) [1]. We used SPS containing either 1, 2, 4 and 8 points with lifetimes  $t_q$  ranging from 1 to 4 frames. The parameters of the stimulus were closely matched with the experiments in [1].

Two different sequence selective neurons were trained with point light stimuli walking to the right or the left side. The simulated response of the subject was defined by the sequence selective neuron that was, on average, more active. The solid lines in Figure 4 show the average percentages of correct discriminations between stimuli walking leftwards and rightwards over 100 presentations obtained with the model.

Significant differences were found between the simulated naive and expert subjects. For naive subjects recognition rate was always at chance level (50%) when only form information was exploited (panel (a)). A recognition rate above chance level (up to almost 100%) was obtained with the motion pathway (panel (b)). Only expert subjects achieved a recognition rate above chance level (about 60%) using form information (panel(d)). Also for these simulated subjects the best performance was obtained with the motion pathway (panel(e)). Even for stimuli containing just 2 dots recognition was better than chance for lifetimes larger than one frame.

A first important implication of these simulations is that high recognition performances for these impoverished stimuli can be achieved with the motion pathway alone. This disproves that the form pathway is necessary for the recognition of SPS.

A major prediction from these simulations is that a considerable difference for the recognition of SPS should exist between naive and expert observers. We tested this prediction psychophysically. We exactly replicated the experiment from [1] using completely naive subjects, and subjects that had previous experience with point light walkers, but not with the SPS. Panels (c) and (f) show the comparison between our psychophysical results and the predictions from the model for SPS with one frame life time. Consistent with the prediction expert subjects show a higher performance than naive subjects<sup>3</sup>. The deviations between model and data for stimuli with 1 and 2 dots in panel(f) might be caused by special recognition strategies of the expert observers<sup>4</sup>.

## 5 Discussion

In this paper we have used a neurophysiologically inspired model to test how form and motion (optic flow) information contribute to the recognition of biological motion stimuli. From our theoretical analysis we conclude that for the recognition of regular walker stimuli either pathway alone is computationally sufficient. Generalization to novel types of biological motion stimuli (regular vs. point light) was possible only in motion pathway of our model. The recognition of point light stimuli (even with strongly degraded motion information) did not

---

<sup>3</sup> We were not able to replicate the high recognition rates reported in [1]. This difference might be due to the fact that our subjects did not have previous experience with SPS.

<sup>4</sup> For example, we observed that the tilt of the spatial region that is covered by the dots over many frames allows a reliable prediction of walking direction. This strategy is not related to biological motion perception, and therefore is not predictable by our model.

require the form pathway. The recognition of point light stimuli required detectors for opponent motion features, whereas detectors for translational motion were not necessary. This is consistent with fMRI results showing selective activity in the opponent motion sensitive area KO during observation of biological motion (e.g. [10]). It is also consistent with data on patients with lesions in the dorsal pathway that could recognize biological motion when the lesion spared area KO [3].

The proposed model is constrained by neurophysiological results, and by the fact that it reproduces many other experimental results on biological motion recognition [2]. However, we cannot exclude that the form pathway could be changed in a way that improves its performance.

We conclude that biological movement recognition under normal conditions is likely based on a fusion of cues, rather than on motion or form information alone. This seems to be particularly the case for stimuli that have been learned previously. Our theoretical study predicts a dominance of motion information for more demanding tasks, like recognition of point light and sequential position stimuli. This conclusion is consistent with results from fMRI experiments and data from neurological patients [9,4].

## References

1. J. A. Beintema and M. Lappe. Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences*, 99(8):5661–5663, April 2002.
2. Martin Giese and Tomaso Poggio. Neural mechanisms for the recognition of biological motion. *Nature Reviews Neuroscience*, 4(3):179–192, March 2003.
3. Martin A. Giese and Lucia M. Vaina. Pathways in the analysis of biological motion: computational model and fMRI results. In *Proceedings of the 24th European Conference on Visual Perception (ECVP)*, August 2001.
4. Emily D. Grossman and Randolph Blake. Brain areas active during visual perception of biological motion. *Neuron*, 35:1167–1175, September 2002.
5. Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14:201–211, 1973.
6. George Mather, Kirstyn Radford, and Sophie West. Low-level visual processing of biological motion. *Proc. R. Soc. Lon. B: Biological Sciences*, 249(1325):149–155, August 1992.
7. Paul Mineiro and David Zipser. Analysis of direction selectivity arising from recurrent cortical interactions. *Neural Networks*, 10:353–371, 1998.
8. Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–1025, November 1999.
9. Lucia M. Vaina, Marjorie Lemay, Don C. Bienfang, Albert Y. Choi, and Ken Nakayama. Intact "biological motion" and "structure from motion" perception in a patient with impaired motion mechanisms: a case study. *Visual Neuroscience*, 5(4):353–369, October 1990.
10. Lucia M. Vaina, Jeffrey Solomon, Sanjida Chowdhury, Pawan Sinha, and John W. Belliveau. Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences*, 98(20):11656–11661, September 2001.