

Measurement of generalization fields for the recognition of biological motion

M.A. Giese ^{a,*},¹, M. Lappe ^b,²

^a*Center for Biological and Computational Learning, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and Dept. of Cognitive Neurology, University Clinic, D-72076 Tübingen, Germany*

^b*Laboratory for Computational and Cognitive Neuroscience Dept. for General Zoologie und Neurobiology, Ruhr-University, D-44780 Bochum, Germany*

Abstract

The human visual system processes complex biological motion stimuli with high sensitivity and selectivity. The characterization of spatio-temporal generalization in the perception of biological motion is still a largely unresolved problem. We present an experiment that investigates how the visual system responds to motion stimuli that interpolate spatio-temporally between natural biological motion patterns. Inspired by analogous studies in stationary object recognition, we generated stimuli that interpolate between natural perceptual categories by morphing. Spatio-temporal morphs between natural movement patterns were obtained with a technique that allows to calculate linear combinations of spatio-temporal patterns. The weights of such linear combinations define a linear metric space over the set of generated movement patterns, so that the spatio-temporal similarity of the motion patterns

can be quantified. In our experiments, we found smooth and continuous variation of the categorization probabilities with the weights of the prototypes in the morphs. For bipedal locomotion patterns we could accurately predict the perceived properties of the morphs by linear combinations of the perceived properties of the prototypes. Such predictions were not possible for morphs between locomotion and very dissimilar movements. We conclude that the visual system shows generalization within classes of motion patterns with similar basic structure, such as bipedal locomotion.

Key words: biological motion, recognition, morphing, linear combination, prototype

* Corresponding author. *Email address:* martin.giese@tuebingen.mpg.de

¹ Martin Giese was supported by the Deutsche Forschungsgemeinschaft Gi 305/1-1, the Deutsche Volkswagenstiftung, and Honda R&D Americas, Inc. Research at CBCL is sponsored by a grant from Office of Naval Research under contract No. N00014-93-1-3085, Office of Naval Research under contract No. N00014-95-1-0600, National Science Foundation under contract No. IIS-9800032, and National Science Foundation under contract No. DMS-9872936. Additional support is provided by: AT&T, Central Research Institute of Electric Power Industry, Eastman Kodak Company, Daimler-Benz AG, Digital Equipment Corporation, Honda R&D Co., Ltd., NEC Fund, Nippon Telegraph & Telephone, and Siemens Corporate Research, Inc.

² M. Lappe and the Lab for Computational and Cognitive Neuroscience is supported by the BMBF program "Biofuture".

1 Introduction

Classical psychophysical experiments demonstrate that the human visual system is highly selective for biological motion patterns. We can recognize complex movements, such as locomotion or dancing, from strongly impoverished stimuli that consist only of a small number of moving illuminated dots ("point light walkers") (e.g. Johansson, 1973; Dittrich, 1993). The recognition of point light walkers is very selective. Subjects can not only differentiate between different actions. They can also to extract subtle details, such as gender or the familiarity of the walker from point light stimuli (e.g. Kozlowski and Cutting, 1977; Mather and Murdoch, 1994). However, the recognition of point light walkers is very robust and does not break down even in presence of strong background noise and motion clutter (e.g. Cutting et al., 1988; Thornton et al., 1998). The principles that underlie the efficient encoding of such complex spatio-temporal patterns still remain to be uncovered.

During the last two decades much research has been dedicated to the analysis of stationary object recognition. Theoretical, psychophysical and neurophysiological results have provided support for the hypothesis that complex three-dimensional objects are represented in the basis of learned prototypical example views (e.g. Poggio and Edelman, 1990; Bülthoff and Edelman, 1992; Logothetis et al., 1995; Riesenhuber and Poggio, 1999; Rolls and Milward, 2000), (see e.g. Riesenhuber and Poggio (2000); Tarr and Bülthoff (1998) for reviews). It seems an interesting hypothesis that complex movement patterns might be encoded on the basis of

similar principles, on the basis of prototypical example movements. In fact, several results like view dependence in the recognition of biological motion (Verfaillie, 1992, 1993; Verfaillie et al., 1994; Bühlhoff et al., 1998; Daems and Verfaillie, 1999) seem to support this hypothesis.

To evaluate the feasibility of the hypothesis of prototype-based representations of complex patterns it is important to study how such representations generalize. Generalization characterizes how well new patterns are encoded that are similar, but not identical with the prototypes. Good generalization implies that a representation "interpolates" well between the stored examples. Such interpolation is important in order to encode whole classes of similar complex patterns with a small number of stored examples.

The generalization properties of stationary object recognition have been quantified by Bühlhoff and Edelman by measuring *generalization fields* for learned artificial objects (paper clips) (Bühlhoff and Edelman, 1992). In these experiments subjects were trained with a particular two-dimensional view of a synthetic three-dimensional object. Afterwards, they were tested with new views in which the object was rotated against the prototypical view. The regime of orientations for which the subjects still were able to recognize the object defines the "generalization field" of the prototypical view. The size of the generalization field provides a measure for the degree of generalization with respect to orientation changes of the object. The classification probabilities of the individual paper clips varied gradually with the orientation indicating smooth interpolation properties of the underlying representa-

tion. The same technique has been used in monkey experiments to characterize the view-variance of the responses of neurons in area IT of the macaque (Logothetis et al., 1995). It seems interesting to apply the concept of generalization fields also to biological motion perception in order to characterize how the visual system responds to motion stimuli that deviate from natural biological movements in terms of their spatial and temporal properties.

Many studies on stationary form recognition and face perception have used *morphing techniques* to generate stimuli that vary continuously between natural or learned prototypical perceptual categories (e.g. Benson and Perrett, 1991; Busey, 1998; Perrett et al., 1998; Leopold et al., 2001). One advantage of using morphing over heuristic variations of form features is that complex combinations of such features can be changed together in a consistent way that makes the morph successively more similar to another perceptually meaningful pattern. This change occurs along the "shortest path" in the sense of a metric that is dependent on the morphing algorithm. Another advantage of stimulus generation by morphing between prototypes is that such techniques allow the definition of abstract perceptually meaningful dimensions or "axes" in pattern spaces. For example, for the morphs between a male and a female face the weight of the male prototype defines a male-female dimension in the space of face images. Such spaces created by morphing are closely related to the concept of perceptually defined pattern spaces, such as "face spaces" (Valentine, 1991), that characterize the topology of the perceptual similarity between complex shapes. Previous work on the exaggeration of movement patterns suggest that the

concept of perceptual spaces might also be applicable to biological motion patterns (Pollick et al., 2001a,b).

In this paper we apply morphing techniques for the study of biological motion perception. By morphing between prototypical movements classes of biological movements can be generated that vary along different perceptually interpretable dimensions. This makes it possible to study generalization fields by adding different amounts of other prototypes to a natural biological motion patterns. Our technique provides a new approach to study the perceptual effects of spatio-temporal distortions of natural movements. It has been shown before that purely temporal or purely spatial modifications of biological movement patterns can improve the discrimination between different categories of complex movements (Hill and Pollick, 2000; Pollick et al., 2001a).

Our experiments are based on a special technique for spatio-temporal morphing. This technique, called "spatio-temporal morphable models" (STMMs), has been applied before in the context of computer vision (Giese and Poggio, 1999, 2000a,b) and provides a possibility to generate new artificial biological movement patterns by *linear combination* of prototypical example movements. By changing the weights of the linear combination STMMs allow to morph continuously between remarkably different biological movements, like for instance "walking" and "running". At the same time, the method defines a metric space over a set of similar complex movements. As example, the linear combination of the prototypical patterns

”walking” and ”running” can be written formally:

$$\text{new motion pattern} = a \cdot \text{walking} + b \cdot \text{running}$$

The linear weights (a, b) define a point in a two-dimensional space of complex motion patterns that is defined by the two prototypes. The dissimilarity of the spatio-temporal structure of two movement patterns 1 and 2 can be quantified by the Euclidean distance between the associated linear weights vectors (a_1, b_1) and (a_2, b_2) . This makes it possible to quantify how much the spatio-temporal structure of a pattern must be changed in order to induce differences in the perceived motion category. The same considerations remain valid if more than two prototypical motion patterns are linearly combined. STMMs can thus be used to define distances between biological motion patterns that are embedded in higher-dimensional pattern spaces.

In this paper, we present an application of this technique for the measurement of the generalization fields of natural biological motion patterns. As generalization field of a motion pattern we define all points in pattern space that lead to the perception of the same type of motion. For our experiments, we generated two classes of motion patterns by morphing between classes of prototypes for which we expected good and bad generalization. Subjects had to classify the stimuli and had to rate several perceived properties of these motion patterns. Our results indicate relatively good interpolation between bipedal locomotion patterns, whereas we got no such interpolation if the prototypes included locomotion and completely different movements,

like physical exercises. This results seems compatible with the interpretation that the visual system can generalize within a limited regime of spatio-temporal deformations of natural biological motion patterns.

2 Methods

2.1 Stimulus generation

Stimuli were generated by tracking biological motion from video sequences. The first class of prototypical motions (SIM) contained only locomotion patterns (walking, running, limping, and marching). The second class of movements (DIF) contained one locomotion pattern (walking) and different types of physical exercises that are not similar to locomotion (aerobics, boxing and knee-bends). From these two sets of prototypical patterns the stimulus trajectories were generated by motion morphing.

2.1.1 Tracking of the prototypical movement patterns

The patterns were filmed using a Kodak VX 1000 camera from the side, the actor facing orthogonal to the view direction of the camera. The actor was moving on a line orthogonal to the camera axis. The closest distance between this line and the camera was 6 m. All movements were executed periodically, but only a single cycle of the movements was used for motion morphing. Dependent on the specific

movement, one cycle lasted for about 30 recorded frames. The frame rate of the camera was 30 frames per second.

To track the trajectories of the movements, first the translation of the whole body was subtracted by hand-marking the hip position in a number of frames and fitting the translation of the hip by a linear function of time. The fitted translation was then subtracted resulting in movement that looks like a person performing the movements on a tread mill. 12 feature points illustrated in Figure 1 were tracked manually. A single cycle of the movement was defined by the segment of the image sequence between the frames that were characterized by a maximum extension of the extremities.

The tracked trajectories were time-normalized and smoothed by fitting them with a second order Fourier series. Prior testing revealed that an inclusion of higher order Fourier components did not increase the quality of the appearance of the biological motion pattern, but leads to more noisy trajectories. The spatial scaling of the patterns was normalized by rescaling them so that the distance between head and hip was always one. It has been verified that this distance is approximately constant in the original trajectories. The obtained spatially and temporally normalized trajectories were used as prototypes for the motion morphing.

Please insert Figure 1 about here.

2.1.2 Motion morphing

In computer graphics a variety of techniques have been developed for editing, deforming, blending and morphing natural motion trajectories that are typically acquired using motion capture systems (e.g. Bruderlin and Williams, 1995; Perlin, 1995; Unuma et al., 1995; Witkin and Popović, 1995; Amaya et al., 1996; Gleicher, 1997; Wiley and Hahn, 1997; Cohen et al., 1998; Lee and Shin, 1999; Brand, 2000). Such techniques are used for producing naturally looking movements for the animation of synthetic figures.

The technique of *spatio-temporal morphable models* (Giese and Poggio, 1999, 2000a) that we used for the generation of our stimuli is closely related to such techniques, and allows the generation of morphs between biological movements by linearly combining the movement trajectories of prototypical motion patterns in space-time. Linear combinations of movement patterns are defined on the basis of spatio-temporal correspondences. The concept of spatio-temporal correspondence is illustrated in Figure 2 (a). Complex movement patterns can be characterized by trajectories of feature points, in our case the 2D coordinates of the joints of the moving figure. The trajectories of the prototypical movement pattern n can be characterized by the time-dependent vector $\mathbf{x}_n(t)$. The correspondence field between two trajectories \mathbf{x}_1 and \mathbf{x}_2 is defined by the spatial shifts $\boldsymbol{\xi}(t)$ and the temporal shifts $\tau(t)$ that transform the first trajectory into the second. The transformation is speci-

fied mathematically by the equation:

$$\mathbf{x}_2(t) = \mathbf{x}_1(t + \tau(t)) + \boldsymbol{\xi}(t) \quad (1)$$

By introduction of spatial and temporal shifts the spatio-temporal morphable model allows to interpolate smoothly between motion patterns with significantly different spatial structure, but also between patterns that differ with respect to their timing.

Please insert Figure 2 about here.

The correspondence shifts $\boldsymbol{\xi}(t)$ and $\tau(t)$ are calculated by solving an optimization problem that minimizes the spatial and temporal shifts under the constraint that the temporal shifts define a new time variable that is always monotonically increasing. For further details about the underlying algorithm we refer to Giese and Poggio (1999, 2000a).

Figure 2 (b) shows schematically the proceeding for generating linear combinations of spatio-temporal patterns. First the correspondence between each individual prototypical patterns and a reference pattern is established. The reference pattern is also a biological motion pattern and can be identical with one of the prototypes. Signifying the spatial and temporal shifts between prototype n and the reference pattern by $\xi_n(t)$ and $\tau_n(t)$, linearly combined spatial and temporal shifts can be defined by the two equations:

$$\begin{aligned}\boldsymbol{\xi}(t) &= \sum_{n=1}^N w_n \boldsymbol{\xi}_n(t) \\ \tau(t) &= \sum_{n=1}^N w_n \tau_n(t)\end{aligned}\tag{2}$$

The weights w_p define the contributions of the individual prototypes to the linear combination. We always assume convex combinations with $0 \leq w_n \leq 1$ and $\sum_p w_p = 1$. After linearly combining the spatial and temporal shifts the trajectories of the morphed pattern can be recovered by morphing the reference pattern in space time using the spatial and temporal shifts $\boldsymbol{\xi}(t)$ and $\tau(t)$. The space-time morph is defined by equation (1) where \mathbf{x}_1 is the reference pattern and \mathbf{x}_2 has to be identified with trajectory of the linearly combined pattern.

During the development of the algorithm we observed that linear combinations of locomotion patterns, like walking, running, limping and marching look very natural, at least if all prototypes are recorded with the same view angle and if the stimuli are presented as stick figures Giese and Poggio (2000a). Morphing between fundamentally different movements, like for instance knee bends and walking, leads to morphs that have no meaningful perceptual interpretation and which look very unnatural. These informal observations motivated us to study morphs between two different set of prototypes. One set of stimuli in our experiment, indicated by "SIM", was generated from the four locomotion patterns for which we expected the morphs to look relatively natural. The second class, "DIF", consisted of four very dissimilar patterns, some of them being fundamentally different from locomotion. We expected morphs between these prototypes to look much less natural. This

second pattern class served as control condition that was suitable to study morphing patterns that look highly unnatural.

Separately for the two classes SIM and DIF, we generated morphs with from either two or three prototypical patterns in the class. The weights c_p were always chosen from the set $[0, 0.2, 0.4, 0.6, 0.8]$. Appropriate combinations of these values were chosen in order to ensure that $\sum_p w_p = 1$. For each of the two data sets, SIM and DIF, we generated 52 stimuli: the 4 prototypes, 24 morphs between 2 prototypes, and 24 morphs that contained three prototypes. All trajectories were pre-calculated and stored before the experiment.

2.1.3 Stimulus presentation

The stimuli were presented as point light walkers with 12 dots on a Silicon Graphics workstation with 72 frames per second. The stimulus dots were white and had a diameter of 0.5 degrees of visual angle. The size of the whole figure was about 5×12 deg. The whole figure subtended 8×20 deg. The center position of the figure was randomized uniformly within an interval of ± 2 deg horizontally and vertically.

2.1.4 Subjects

We tested 7 subjects for the data set SIM and 6 subjects for the data set DIF. All subjects were from the Institute for General Zoology and Neurobiology in Bochum

and the center for Computational and Biological Learning at M.I.T. Their ages ranged from 25 to 38 years, and two of them were female. All subjects had normal or corrected-to-normal vision. Two subjects were familiar with the purpose of the experiment, the others were naive.

2.1.5 Procedure

Subjects watched the computer screen from a distance of 40 cm. They were briefed about the experimental procedure and had an opportunity to practice the experiment for about 10 trials. Each stimulus was presented once, for as long as the subject wished to see it. During the presentation the subject was allowed to adjust the speed of the display such that the pattern appeared natural. The task of the subject was to give three judgments for each stimulus. The first was to classify the pattern by pressing one out of four keys that identified the four prototypes of the respective pattern class SIM and DIF. The keys were the initial letters of the patterns (w)alking, (r)unning, (l)imping, (m)arching, (a)erobics, (k)nee bend, and (b)oxing. After classification, the subjects had to rate the "naturalness" of the perceived movement pattern on a scale from 1 to 9. Finally, the subject was asked to adjust the minimum and the maximum speed of the pattern for which is still appeared "natural". The midpoint between the minimum and maximum speed was used as measure of the "optimal natural speed" of the pattern. We asked subjects to adjust a whole range of speeds because it turned out to be difficult for them to select a single speed that appeared maximally natural. The different prototypes, and

the two and three pattern combinations were presented in randomized order within each block. The experiment took about 20 minutes per subject.

3 Results

All subjects were able to classify the prototypical motion patterns easily. Some of the morphs, in particular in the class "DIF", were perceived as unnatural (e.g. "jerky motion", "unnatural postures", etc.). In the following, we first introduce a graphical representation for the pattern space that is defined by the morphing technique. This graphical representation will be used for the illustration of the results throughout the rest of the paper. We then present the results on pattern classification, the adjusted optimum speed, and the perceived naturalness of the patterns.

3.1 Representation format

For the presentation of the results we introduce a special 2D graphical illustration of the 4D pattern space of the motion morphs. This representation tries to illustrate the topology that is imposed by the linear weight vectors onto the space of generated motion patterns. Morphs which correspond to linear weight vectors that are close in the 4D Euclidean pattern space are (in most cases) mapped onto points that are close in this 2D graphical representation. Figure 3 illustrates this representation format for the four prototypes "walking" (W), "running" (R), "limping" (L), and "marching" (M). Each stimulus is represented by a hexagon. The four capitals W, R, L, and M indicate the hexagons of the four prototypes. The prototypes "running" (R), "limping" (L), and "marching" form an equal-sided triangle, "walking" (W) is placed in the center of this triangle. Morphs between any two prototypes are rep-

resented by the hexagons that lie on the lines between these letters. The weight of the prototype in the morph decreases with the distance from the respective letter. For example, the stimulus labelled (a) in Figure 3 is a morph between walking and running with the weight combination $w_w = 0.6$ and $w_r = 0.4$. All other hexagons correspond to three-pattern morphs. The three-pattern morphs that contain the prototype in the center of the figure are presented by hexagons within the triangle formed by the three adjacent prototypes. For example, the hexagon labelled (b) in Figure 3 is a morph between walking, running, and limping with the weight combination $w_w = 0.6$, $w_r = 0.2$, $w_l = 0.2$. Three-pattern morphs that do not contain the central prototype are mapped onto points outside the outer triangle. Hexagon (c) represents a morph between marching, running, and limping with the weight combination $w_m = 0.6$, $w_r = 0.2$, $w_l = 0.2$. Hexagon (d) is a morph between marching, running, and limping with the weight combination $w_m = 0.4$, $w_r = 0.4$, $w_l = 0.2$.

Please insert Figure 3 about here.

The relative distances between the points in the diagram and the letters that symbolize the prototypes are monotonically related to the weights of the prototypes in the linear combination. However, the two distances are related in a non-trivial non-linear way. This monotonic relationship is very useful to illustrate the distribution of the different experimental measures over the space of morphing patterns.

3.2 *Classification*

The upper panel of Figure 4 shows the results for the classification task for the pattern class SIM. Each dot represents a trial for which a subject classified the morphing pattern as the prototype given by the label in the left upper corner of the panels. The classification probability is thus proportional to the number of dots at a certain location of the figure. Consistent with our expectations, the prototypes were categorized correctly in almost all cases. Patterns that are closer to a particular prototype in the pattern space, in the metric defined by the linear weights, are classified with higher probability as this prototype. The classification probabilities vary smoothly and gradually with the contribution of the prototypes to the morph. This indicates that the perceptual similarity of the morphs with the prototypes seems to vary smoothly and gradually with the linear weights. This can be interpreted as evidence for the smooth structure of the underlying pattern space. A possible alternative outcome of this experiment could have been that the intermediate patterns lead to very unclear percepts that are randomly assigned to different prototypes, which seems not to be the case.

Since we used a forced choice task the classification data alone does not allow to decide whether the morphs were seen as instantiations of the categories defined by the prototypes, or if they were just assigned to the category that was judged as perceptually most similar. The fact that the classification probabilities of morphs with contributions of more than two prototypes vary also gradually with the weights

of the prototypes in the morph can be seen as indication of the multidimensionality of the underlying perceptual space of biological motion patterns.

The lower panel of the figure shows the classification results for the class DIF, that was generated using walking and different physical exercises as prototypes. Qualitatively, the results are very similar. In this case, most of the morphs were perceived as unnatural by the subjects (see below). However, we find a gradual variation of the classification probability with the linear weights. We conclude from this that the subjects at least were able to assign the morphs to categories with spatio-temporally similar trajectories, even though in this case the morphs were not seen as typical representants of the category assigned in the forced choice task.

Please insert Figure 4 about here.

3.3 *Generalization fields*

The fact that the classification probabilities of biological motion patterns vary gradually with the "coordinates" of motion patterns in the pattern space defined by motion morphing makes it possible to define *generalization fields for biological motion*. The generalization field of a pattern is the area in the pattern space for which patterns are classified as the same biological motion percept. We can get a clearer impression of the sizes of these generalization fields for the different prototypical movements in our experiment by replotting the classification results by placing the prototype always in the center of the triangle in the diagrams. This makes the areas

around the prototypes for different patterns more comparable.

Please insert Figure 5 about here.

Figure 5 shows the measured generalization fields for the pattern classes SIM and DIF. Interestingly, the generalization fields for the pattern "walking" is significantly larger than the generalization fields for the other prototypes. This result might be related to the familiarity of this pattern. In everyday situations subjects might see boxing much more rarely than people walking. Another explanation is that this result reflects the topology of the pattern space. Walking might be, in the metric defined by the features extracted by the visual system, more similar to most points of the generated pattern space than the other prototypes. The latter explanation is supported by recent theoretical work (Giese and Poggio, 2002). A biologically plausible neural model was trained and tested with the same stimuli. Even though all prototypes were trained in exactly the same way the model reproduces the over-proportional size of the generalization field for walking that we found in our experiment.

3.4 *Natural speeds*

The subjects reported also the speed regimes for which the patterns seemed natural. Since normal biological motion patterns are associated typically with a set of characteristic speeds it seems possible that the typical speed of motion patterns is encoded together with their spatio-temporal structure.

Please insert Figure 6 about here.

Figure 6 shows the measured optimal speeds for the patterns calculated from the average of the adjusted minimum and maximum speeds. The figure shows color-coded plots of the optimum speeds of the patterns using the graphical representation discussed above. The color of the pixels codes for the ratio of the adjusted optimum speed and the initial presentation speed, which was about one movement cycle per second. The positions of the pixels indicates the location of the morph in the pattern space.

The pattern "running" is associated with particularly high adjusted speeds, and the pattern "knee bends" with very small speeds. Patterns in the neighborhood of the individual prototypes are typically associated with similar optimum speeds. This result indicates that properties that are associated with the prototypes seem to be transferred to patterns close to the prototypes in pattern space. This indicates generalization. It seems reasonable to assume that the influence of the perceptual properties of the prototypes onto the perceived properties of the morph increases with the weight of the prototype in the linear combination. As test of this hypothesis we plotted the optimum speed of the patterns of the class SIM as function of the weight of the prototype "running". If our hypothesis is true the perceived speed of the pattern should gradually increase with the contribution of the running prototype to the morph. Figure 7 (a) shows that a highly significant correlation exists between the contribution of the running prototype and the adjusted optimal speed of the patterns (correlation coefficient $r^2 = 0.6$, $t = 8.7$, $N = 52$, $p < 0.001$). Correlations

between the optimal speed of the patterns and the weights of the individual prototypes were also found when the prototypes were not associated with the minimum or maximum optimal speed. This is shown in Figure 7 (b) where the adjusted optimal speed is plotted as a function of the weight of the prototype "boxing" for the pattern class DIF. In this case also a significant correlation exists ($r^2 = 0.3$, $t = 4.54$, $N = 52$, $p < 0.001$).

Please insert Figure 7 about here.

We wanted to provide a more rigorous test of the hypothesis that the prototypes determine perceived properties of similar patterns by generalization. For this purpose, we tried to predict the perceived properties of the morphed patterns by linearly combining the properties of the prototypes using the same linear weights with which the prototypes contribute to the linear combination in the motion morph. With v_n signifying the perceived optimum speed of prototype n the prediction was given by $\hat{v} = \sum_n w_n v_n$, where the w_n are the weights of the individual prototypes in the morphable model. Figure 8 shows the results from this prediction as a function of the adjusted optimum speeds for both pattern classes SIM and DIF. In both cases the linear combination of the optimal speeds of the prototypes predicts accurately the perceived optimal speed of the pattern (correlation coefficients $r^2 = 0.54$, $t = 7.7$, $N = 52$, $p < 0.01$ for the class SIM, and $r^2 = 0.8$, $t = 14.0$, $N = 52$, $p < 0.01$ for the class DIF). This result suggests that the perceived properties of a larger class of motion patterns can in principle be predicted relatively accurately from the perceived properties of few prototypical example patterns by a simple voting scheme

for which the prototypes contribute to the perceived property according to their similarity with the stimulus.

Please insert Figure 8 about here.

Please insert Figure 9 about here.

3.5 *Ratings of naturalness*

Naturalness ratings can be used as a measure for the generalization from natural to spatio-temporally deformed patterns. Naively, one might assume that artificial biological motion patterns that are generated by spatio-temporal morphing should appear less natural than natural biological motion patterns that are directly tracked from video sequences. To test this hypothesis we analyzed the naturalness ratings in a similar way as the perceived optimal speeds. Fig. 9 shows the naturalness ratings for the presented biological motion patterns from both pattern classes. The naturalness ratings vary gradually with the weights of the prototypes. For the pattern class SIM the perceived naturalness of the morphs "interpolates" between the perceived naturalness of the prototypes. This can be seen as a sign of generalization, i.e. as an indication of the gradual transfer of the properties of the prototypes to other motion patterns that are close in the pattern space. Interestingly, the prototypes are associated with different naturalness ratings. In particular marching is perceived as much less natural than the other patterns, potentially because it is much less familiar. On the other hand, running and walking are associated with very high naturalness rat-

ings. Consistent with the hypothesis of a transfer of perceived properties to similar patterns in the morphing space, the perceived naturalness increases with the contribution of morphing to the morph. Fig. 10 (a) shows that for the pattern class SIM the naturalness ratings, on average, as function of the linear weight of the pattern "walking" in the morph. The correlation between the linear weight and the naturalness ratings is significant ($r^2 = 0.33$, $t = 5.2$, $N = 52$, $p < 0.01$).

Please insert Figure 10 about here.

We tested the quality of the interpolation between the prototypes more rigorously by trying to predict the naturalness ratings by weighted linear combinations of the naturalness ratings of the prototypes. Like for the prediction of the optimal speeds, the predicted naturalness values were given by the relationship $\hat{S} = \sum_n w_n S_n$, where S_n is the naturalness rating of prototype n , and w_n the weight of the prototype in the linear combination. Fig. 11 (a) shows that, for the pattern class SIM, the naturalness of the pattern can be efficiently predicted from the naturalness values of the prototypes ($r^2 = 0.74$, $t = 11.8$, $N = 52$, $p < 0.01$). This provides additional evidence for generalization with respect to the perceived properties within this pattern class. In particular, we did not find a significant reduction in perceived naturalness for the morphs compared to the prototypes. This was unexpected because the prototypes, as natural patterns, could be expected to look more natural than the morphs. Since the morphs were generated by spatio-temporal interpolation in most cases they likely do not correspond to naturally occurring complex movements.

Please insert Figure 11 about here.

The same analysis for the pattern class DIF, which contained not only locomotion patterns, leads to different results. Fig. 9 (b) shows the naturalness ratings for this pattern class. Only the prototypes and the patterns that are very close in pattern space look natural. Morphs with significant contributions of multiple prototypes tend to be perceived as unnatural. This becomes particularly evident in Fig. 10 (b) where the naturalness ratings are plotted as a function of the linear weight of "walking" in the linear combination. A clear dip exists for intermediate weights of walking. Inspection of Fig. 9 (b) shows that this dip is caused by the low naturalness of patterns that are morphs with significant contributions of walking and other prototypes. Correspondingly, the correlation between the weight of walking and the naturalness rating in this pattern class is non significant ($r^2 = 0.002$, $t = 0.32$, $N = 52$, $p > 0.1$).

The attempt to predict the naturalness ratings of the morphs from the naturalness values of the prototypes is not successful for this pattern class. The correlation between the predictions shown in Figure 11 (b) and the real naturalness ratings is almost zero ($r^2 = 0.001$, $t = 0.26$, $N = 52$, $p > 0.1$). This result is consistent with the low perceived naturalness of morphed patterns compared to the prototypes. For this pattern class no efficient interpolation between the prototypes seems to occur.

Please insert Figure 12 about here.

4 Discussion

We have presented an experiment that investigates generalization properties of biological motion perception using a new class of stimuli that were generated by spatio-temporal morphing between natural movements patterns. For this purpose, we applied the technique of spatio-temporal morphable models that allows to calculate linear combinations of movement patterns. The weights of these linear combinations define a metric linear space over the class of generated biological motion patterns. This space, on one hand, provides a metric that is suitable for a quantification of the spatio-temporal similarity of the biological motion patterns. On the other hand, we can introduce variations of the patterns along different dimensions in this metric space. This made it possible to study the perceptual effects of spatio-temporal changes of the patterns that make the patterns more similar to other perceptually meaningful interpretations. In our experiments we used such parameterized classes of motion patterns to study how well the perceptual system generalizes if natural biological motion patterns are distorted in space-time.

The idea of abstract metric spaces of complex patterns spanned up by morphing has been used extensively before in the study of stationary object recognition (e.g. Valentine, 1991; Benson and Perrett, 1991; Busey, 1998; Perrett et al., 1998; Leopold et al., 2001). Our work extends this approach to the recognition of complex movements. By introduction spatio-temporal deformations of natural motion patterns that interpolate between different perceptual categories, such as walking

and running, our work is different from earlier experiments that have studied how the distinction between different perceptual categories is improved by purely temporal (e.g. Hill and Pollick, 2000) or purely spatial exaggerations of motion patterns (Pollick et al., 2001a). Such exaggeration studies are also based on the idea of an underlying continuous space of motion patterns.

In experiment we found different degrees of generalization in the two pattern classes. Within the class of bipedal locomotion patterns we found relatively good generalization. This was indicated by two results: (1) by fact that we could predict the perceived properties of the morphs in many cases relatively accurately from the perceived properties of the prototypes, and (2) by the relative high naturalness ratings for the morphs. The same results were not obtained for the second pattern class including locomotion and non-locomotion patterns. In this case, the morphs had lower naturalness ratings than the prototypes, and a prediction of perceived properties from the prototypes was not possible. It was astonishing that no specific signs of a discontinuity in pattern space were observed for morphs between walking and running, since biomechanical analyses indicate a phase transition between these two gaits (Diedrich and Warren, 1995). Interestingly, we found smooth interpolation between relatively dissimilar locomotion patterns such as walking and running, or running and marching. This indicates that the visual system can interpolate well between more dissimilar patterns than just between style differences of the same type of locomotion, like the walk of a male and a female.

In the metric defined by our morphing technique "walking" was characterized by

relative large generalization field compared with the other prototypes in the classification task. This might be a property of the geometrical structure of the pattern space within our parameterization since we found the same effect with a theoretical neural model that treated all prototypes in exactly the same way, so that familiarity effects could be excluded (Giese and Poggio, 2002).

The result that for locomotion patterns we could predict the properties of the morphs from the perceived properties of the prototypes is compatible with a prototype-based representation of biological motion. A simple voting scheme might be sufficient to derive the perceived properties of motion patterns from the stored perceptual properties of the prototypes. Weighting ensures that the prototypes that are most similar to the stimulus have the strongest influence on the perceived property. Such voting schemes could be implemented easily by reading out population codes from neurons that encode prototypical spatio-temporal patterns together with characteristic properties of the represented motion.

However, the stimulus set used in our experiment does not allow to decide finally about the hypothesis of a prototype-based representation of biological motion. We can not exclude that some of the morphs resemble natural motion patterns that are not identical with the prototypes, and which the subjects also have stored. A rigorous evaluation of the hypothesis of an example-based encoding would require to use stimuli that are completely novel, i.e. unnatural, and which result in morphing patterns that do not resemble naturally occurring movements. The methods that we have presented in this paper would be applicable in the same way to such novel

biological motion stimuli.

Acknowledgements

We thank B. Knappmeyer for helpful comments and A. Benali for help with the data acquisition. Thanks to the Max Planck Institute for Biological Cybernetics for support during publication of this work.

References

- Amaya, K., Bruderlin, A., and Calvert, T. (1996). Emotion form motion. In Davis, W. A. and Bartels, R., editors, *Graphics Interface '96*, pages 222–229. Canadian Human-Computer Communications Society.
- Benson, P. and Perrett, D. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *Journal of Cognitive Psychology*, 3(1):105–135.
- Brand, M. (2000). Style machines. In *Proceedings of SIGGRAPH 2000, New Orleans, Louisiana, USA*, pages 23–28.
- Bruderlin, A. and Williams, L. (1995). Motion signal processing. In *Proceedings of SIGGRAPH 95, Los Angeles*, pages 97–104.
- Bülthoff, H. H. and Edelman, S. (1992). Psychophysical support for a 2D-view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences (USA)*, 89:60–64.
- Bülthoff, I., Bülthoff, H. H., and Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nature Neuroscience*, 1:254–257.
- Busey, P. (1998). Physical and psychological representations of faces: Evidence from morphing. *Psychological Science*, 9:476–483.
- Cohen, M., Rose, M., and Bodenheimer, B. (1998). Verbs and adverbs: Multidimensional motion interpolation. *IEEE Transactions on Computer Graphics and Applications*, 18:32–40.
- Cutting, J. E., Moore, C., and Morrison, R. (1988). Masking the motions of human

- gait. *Perception and Psychophysics*, 44:339–347.
- Daems, A. and Verfaillie, K. (1999). View-dependent priming effects in the perception of human actions and body postures. *Visual Cognition*, 6:665–693.
- Diedrich, F. J. and Warren, W. H. (1995). Why change gaits? dynamics of the walk-run transition. *Journal of Experimental Psychology: Human Perception and Performance*, 21:183–202.
- Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception*, 22:15–22.
- Giese, M. A. and Poggio, T. (1999). Synthesis and recognition of biological motion patterns based on linear superposition of prototypical motion sequences. In *Proceedings of IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes (MVIEW '99)*, Fort Collins, CO, pages 73–80. IEEE Computer Society, Los Alamitos.
- Giese, M. A. and Poggio, T. (2000a). Morphable models for the analysis and synthesis of complex motion pattern,. *International Journal of Computer Vision*, 38:59–73.
- Giese, M. A. and Poggio, T. (2000b). Quantification and classification of locomotion patterns by spatio-temporal morphable model. In *Third IEEE Workshop on Visual Surveillance, Dublin, Ireland, July 1, 2000*, pages 73–80. IEEE Computer Society, Los Alamitos.
- Giese, M. A. and Poggio, T. (2002). Neural mechanisms for the recognition of biological movements and actions. *Nature*, (submitted).
- Gleicher, M. (1997). Motion editing with spacetime constraints. In *Symposium on*

- Interactive 3D Graphics*, pages 139–148, 195.
- Hill, H. and Pollick, 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 and Psychophysics*, 14:201–211.
- Kozlowski, L. T. and Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception and Psychophysics*, 21:575–580.
- Lee, J. and Shin, S. Y. (1999). A hierarchical approach to interactive motion editing for human-like figures. In *Proceedings of SIGGRAPH 99, Los Angeles*, pages 39–48.
- Leopold, D. A., OToole, A. J., Vetter, T., and Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, 4:89–94.
- Logothetis, N. K., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5:552–563.
- Mather, G. and Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London B*, 258:273–279.
- Perlin, K. (1995). Real time responsive animation with personality. *IEEE Transactions on Visualization and Computer Graphics*, 1:5–15.
- Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D. A., Yoshikawa, S., Burt, D. M., Henzi, S. P., Castles, D. L., and Akamatsu, S. (1998). Effects of sexual

- dimorphism on facial attractiveness. *Nature*, 294:884–887.
- Poggio, T. and Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343:263–266.
- Pollick, F. E., Fodopastis, C., and Braden, V. (2001a). Recognizing the style of spatially exaggerated tennis serves. *Perception*, 30:323–338.
- Pollick, F. E., Patterswon, H. M., and Bruderlin, A. Sanford, A. J. (2001b). Perceiving affect from arm movement. *Cognition*, 82:B51–B61.
- Riesenhuber, M. K. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2:1019–1025.
- Riesenhuber, M. K. and Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, 3, Supp.:1199–1204.
- Rolls, E. T. and Milward, T. (2000). A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Computation*, 12:2547–2572.
- Tarr, M. J. and Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, 67(1-2):1–20.
- Thornton, I. M., Pinto, J., and Shiffrar, M. (1998). Why change gaits? dynamics of the walk-run transition. *Cognitive Neuropsychology*, 15:535–552.
- Unuma, M., Anjyo, K., and Takeuchi, R. (1995). Fourier principles for emotion-based human figure animation. *Proceedings of SIGGRAPH '95, Los Angeles*, 29:91–96.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology*,

43A:161–204.

Verfaillie, K. (1992). Variant points of view on viewpoint invariance. *Canadian Journal of Psychology*, 46:215–235.

Verfaillie, K. (1993). Orientation-dependent priming effects in the perception of biological motion. *Journal of Experimental Psychology: Human Perception and Performance*, 19:992–1023.

Verfaillie, K., De Troy, A., and Van Rensbergen, J. (1994). Transsaccadic integration of biological motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20:649–670.

Wiley, D. and Hahn, J. (1997). Interpolation synthesis of articulated figure motion. *IEEE Transactions on Computer Graphics and Applications*, 11:39–45.

Witkin, A. and Popović, Z. (1995). Motion warping. In *Proceedings of SIGGRAPH 95, Los Angeles*, pages 105–108.

Figure Legends

- (1) Example images from the recorded sequence for "walking" and tracked feature points (white dots).
- (2) (a) Spatio-temporal correspondence between two trajectories $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ is defined by the spatial shifts $\boldsymbol{\xi}(t)$ and the temporal shifts $\tau(t)$ that map the two trajectories onto each other. These shifts are indicated by the thick black lines. The solid curve shows the approximation of the trajectory \mathbf{x}_1 that results from rewarping $\mathbf{x}_2(t)$ with the correspondence shifts. (b) Method for generating linear combinations of complex movement from spatio-temporal correspondence fields.
- (3) Graphical illustration of the pattern space of morphing patterns: The four letters W, R, L, and M symbolize the four prototypes "walking" (in the center of the triangle), "running", "limping", and "marching". The positions of the points within the triangle symbolize the relative contributions of the four prototypes to the morph. The distances of the points from the for prototypes W, R, L and M are monotonically related to the weights of the prototypes in the linear combination. Closer distance from the prototype codes for a higher weight of the prototype in the morph. Two-pattern morphs are indicated by the points on the lines that connect the four letters W, R, L and M. Morphs between three patterns that include the central prototype (W) are indicated by points within the large triangle. The points outside this triangle indicate morphs between three patterns that do not include the central prototype. The hexagons

with small Latin letters represent examples for linear combinations that are explained in the text.

- (4) Classification results for the pattern set SIM (upper panel) and the pattern set DIF (lower panel). Each dot presents a trial for which a subject has classified the motion pattern as the prototype that is given by the label in the left upper corner of the panel. The figures are based on the results of 7, respectively 6 subjects.
- (5) Generalization fields for the pattern classes SIM (upper panel) and DIF (lower panel) obtained by replotting the classification results in a way that makes sure that the classified prototype is always in the center of the triangle. The figure shows results of 7, respectively 6 subjects.
- (6) (a) Means of the adjusted optimal speeds for the patterns of the class SIM. The value 1.0 corresponds to the initial presentation speed of all patterns at the beginning of the trials. The figure shows average results over 7 subjects.
(b) Means of the adjusted optimal speeds for the patterns of the class DIF. The value 1.0 corresponds to the initial presentation speed of all patterns at the beginning of the trials. The figure shows average results over 6 subjects.
- (7) Relationship between the weight of the prototype "running" in the linear combination and the adjusted optimum speed for the pattern class SIM (left panel) and the pattern class DIF (right panel). The figure shows average results over 7, respectively 6 subjects.
- (8) Predicted and measured optimal speed of the biological motion pattern for the pattern class SIM (left panel) and the pattern class DIF (right panel). The

- figure shows results of 7, respectively 6 subjects.
- (9) Naturalness ratings for the two pattern classes SIM (upper panel) and DIF (lower panel) averaged over 7, respectively 6 subjects.
 - (10) Relationship between the weight of the prototype "walking" in the linear combination and the perceived naturalness of the motion pattern for the pattern classes SIM (left panel) and DIF (right panel). The figure shows average results over 7, respectively 6 subjects.
 - (11) Predicted and measured naturalness ratings for the biological motion patterns from the pattern classes SIM (left panel), and the pattern class DIF (right panel). The figure shows results for 7, respectively 6 subjects.

Figures

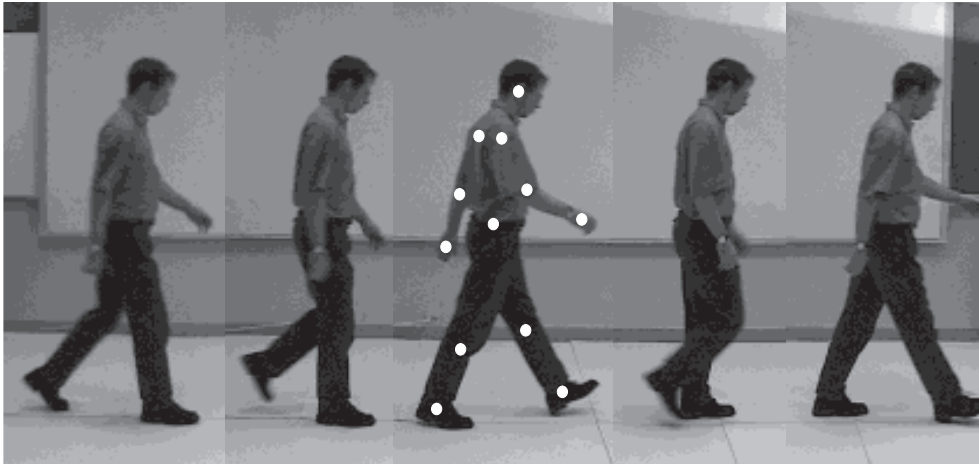


Fig. 1.

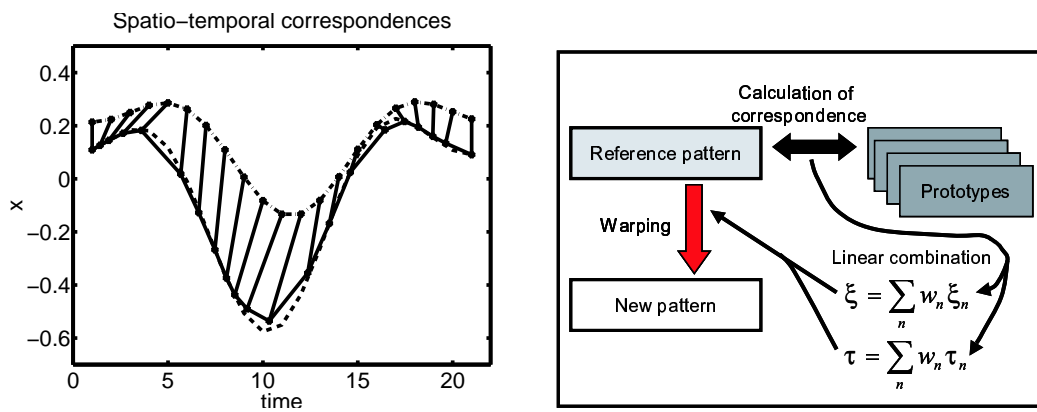


Fig. 2.

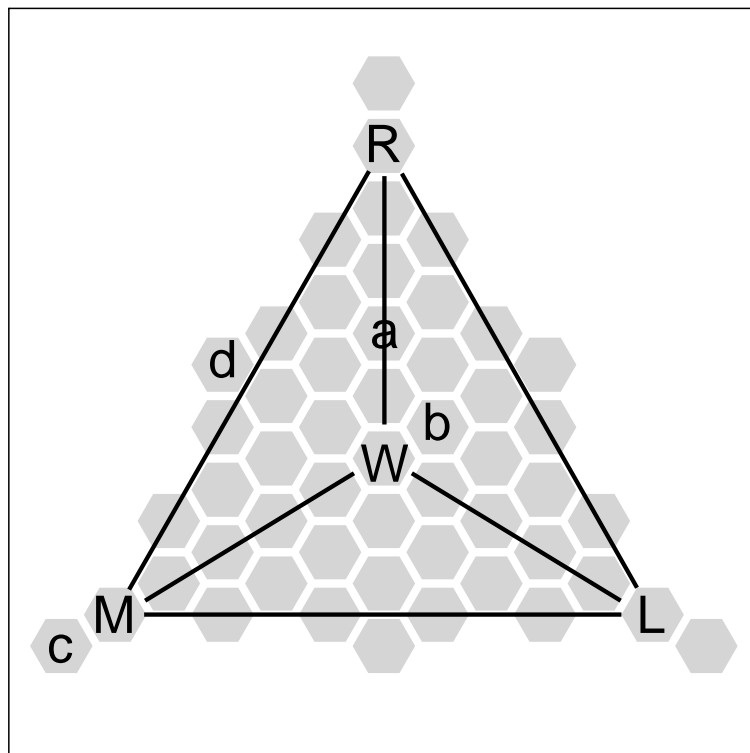


Fig. 3.

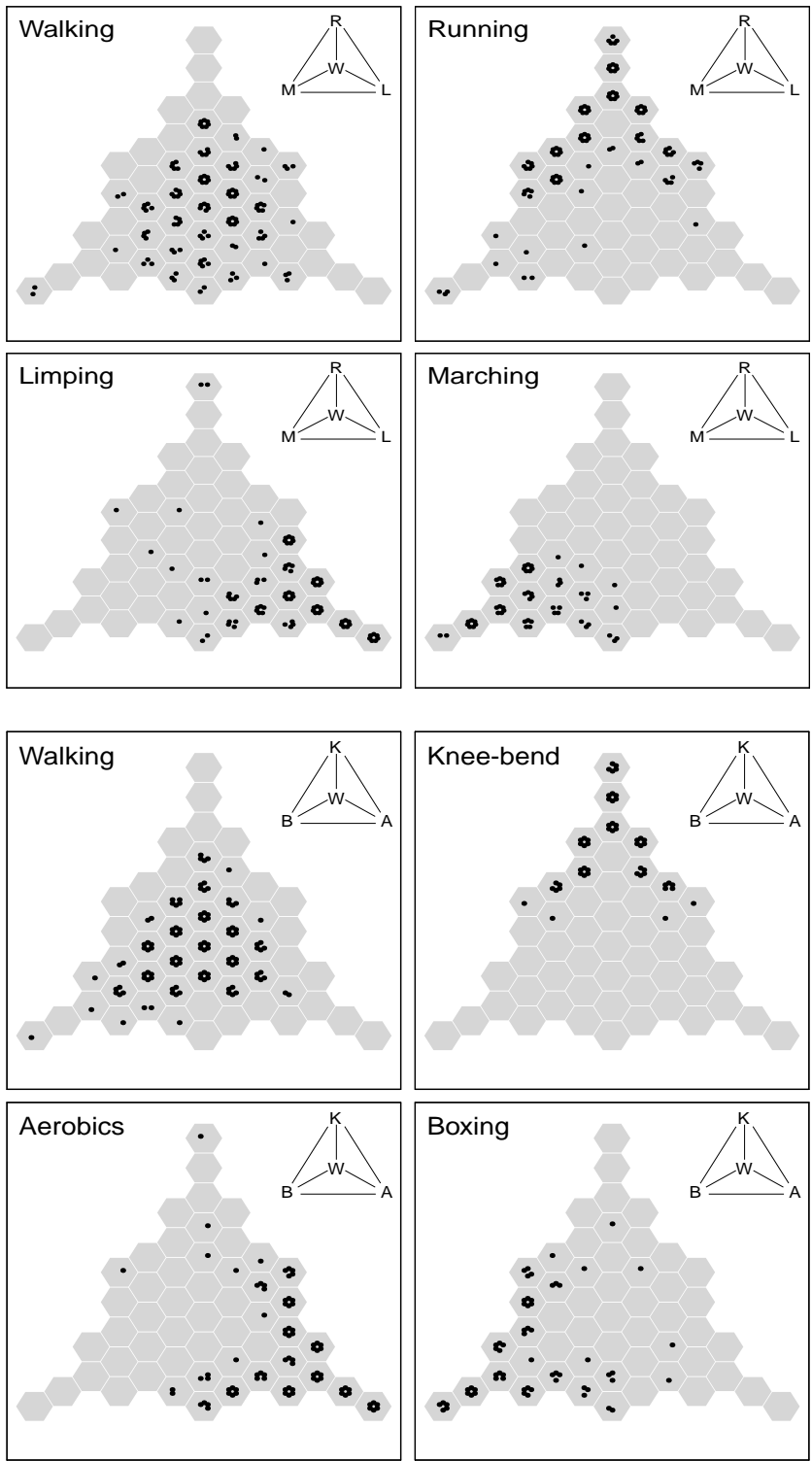


Fig. 4.

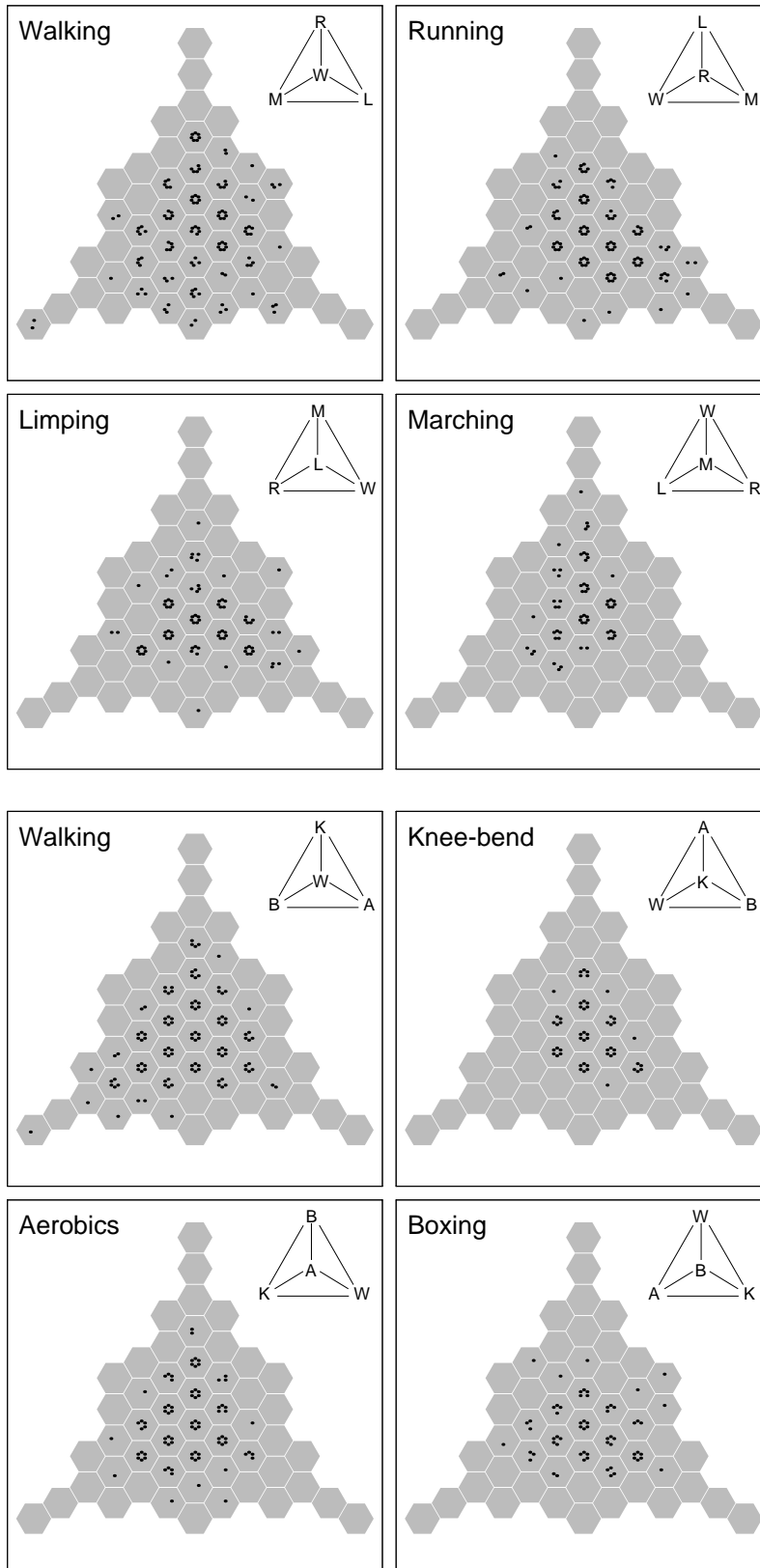


Fig. 5.

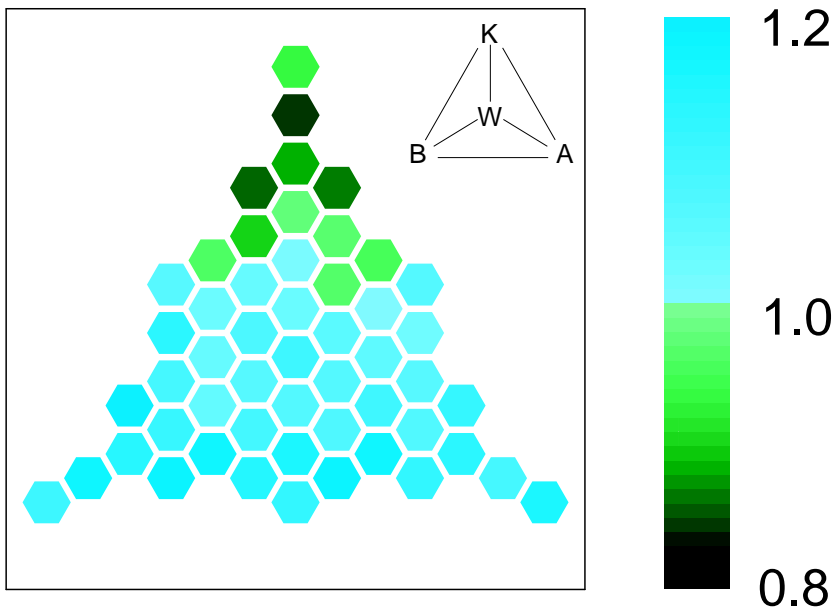
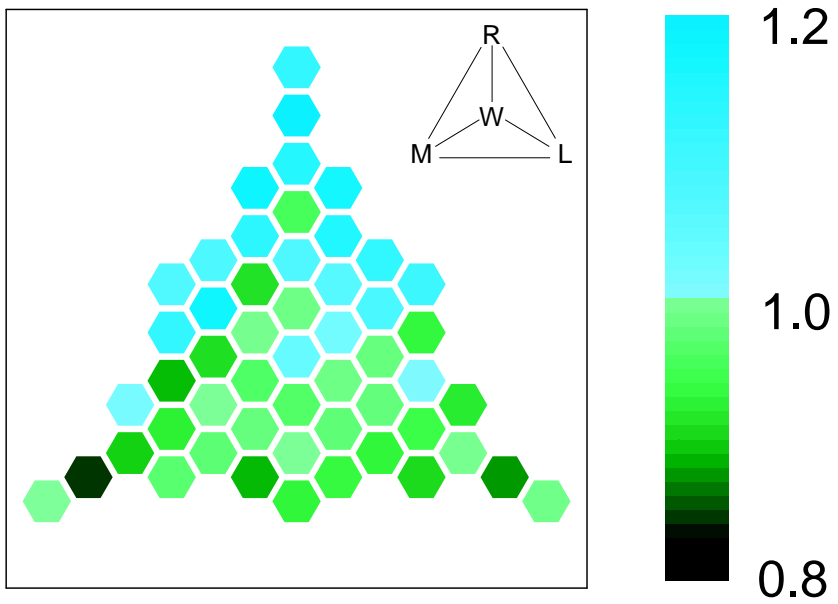


Fig. 6.

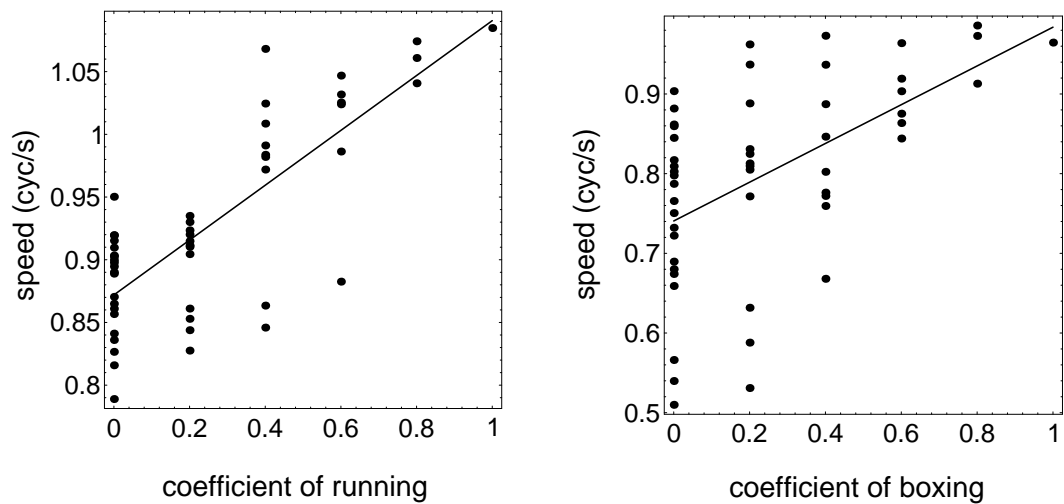


Fig. 7.

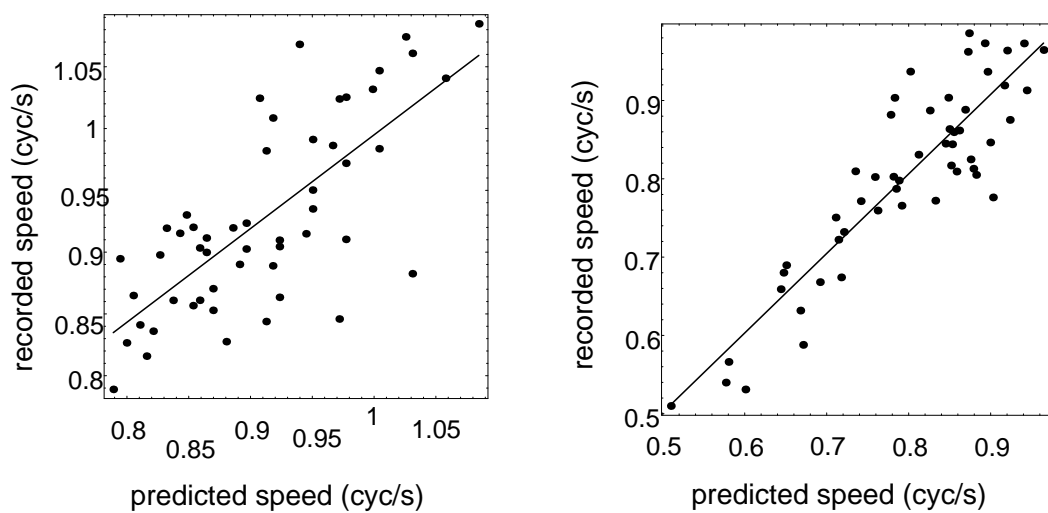


Fig. 8.

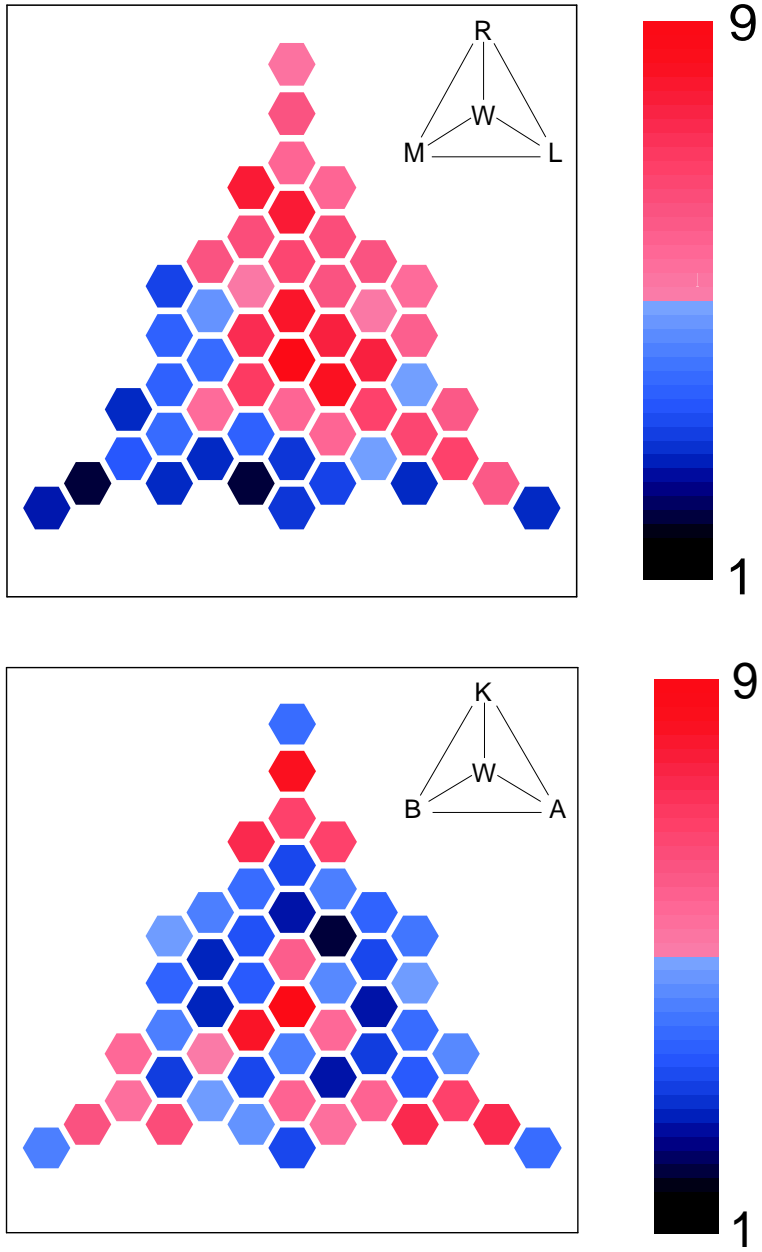


Fig. 9.

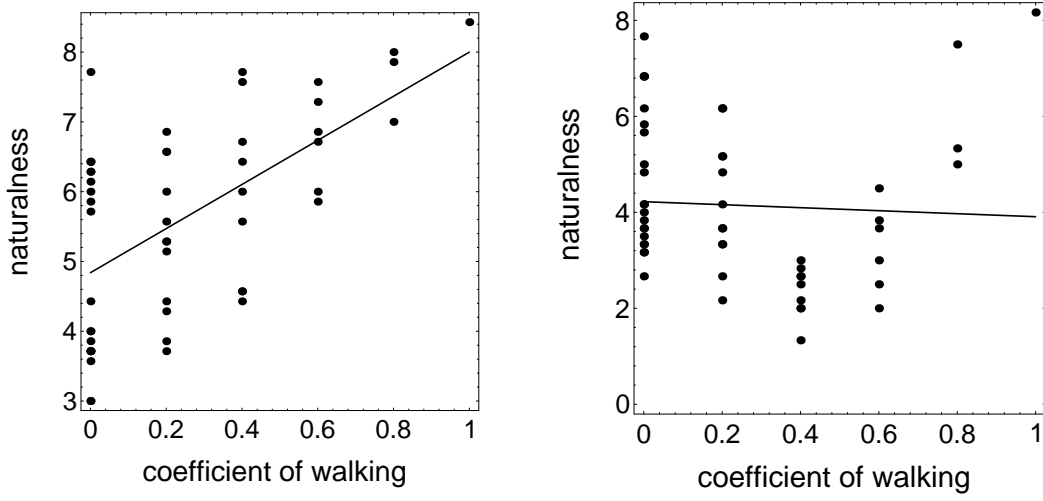


Fig. 10.

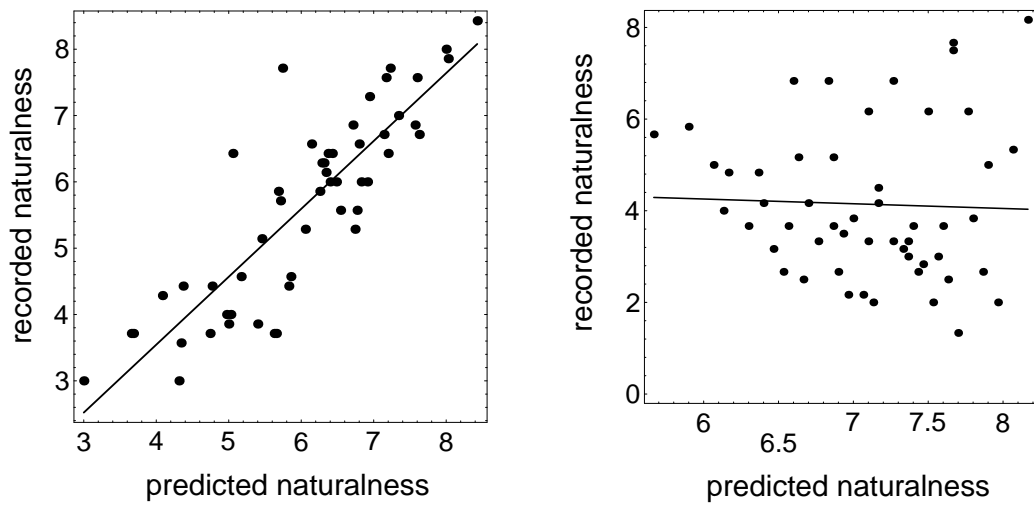


Fig. 11.