Metrics of the perception of body movement

Martin A. Giese

Laboratory for Action Representation and Learning,
School of Psychology, University of Bangor, UK, &
Hertie Institute for Clinical Brain Research,
Tübingen, Germany

Ian Thornton

Department of Psychology, University of Swansea, UK, &
MPI for Biological Cybernetics,
Tübingen, Germany

Shimon Edelman

Department of Psychology, Cornell University,
Ithaca, NY, USA

Body movements are recognized with speed and precision, even from strongly impoverished stimuli. While cortical structures involved in biological motion recognition have been identified, the nature of the underlying perceptual representation remains largely unknown. We show that visual representations of complex body movements are characterized by perceptual spaces with well-defined metric properties. By multidimensional scaling, we reconstructed from similarity judgments the perceptual space configurations of stimulus sets generated by motion morphing. These configurations resemble the true stimulus configurations in the space of morphing weights. In addition, we found an even higher similarity between the perceptual metrics and the metrics of a physical space that was defined by distance measures between joint trajectories, which compute spatial trajectory differences after time alignment using a robust error norm. These outcomes were independent of the experimental paradigm for the assessment of perceived similarity (pairs-comparison vs. delayed match-to-sample) and of the method of stimulus presentation (point-light stimuli vs. stick figures). Our findings suggest that the visual perception of body motion is veridical and closely reflects physical similarities between joint trajectories. This implies that representations of form and motion share fundamental properties and places constraints on the computational mechanisms that support the recognition of biological motion patterns.

Keywords: biological motion, MDS, metric, motion morphing, similarity, distance

Introduction

Biological movements can be recognized, both by humans and animals, with minimal effort and high precision, even from point-light stimuli (Blake & Shiffrar, 2007; Johansson, 1973). Such performance requires perceptual representations that reflect the subtle differences between actions, for example, to support the recognition of emotional state or gender from movement (Dittrich, Troscianko, Lea, & Morgan, 1996; Kozlowski & Cutting, 1977; Pollick, Paterson, Bruderlin, & Sanford, 2001). At the same time, these representations must be robust against irrelevant variations, such as changes of viewpoint. While electrophysiological and imaging studies have identified neural structures involved in motion recognition (e.g., Decety & Grèzes, 1999; Grossman & Blake, 2002; Puce & Perrett, 2003; Rizzolatti & Craighero, 2004; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001), the computational nature of visual representations of body movements is still largely unknown. Specifically, it is unclear which spatio-temporal

properties determine the perceived similarity of complex body movements and actions.

Much more is known about the visual recognition of static objects, a process that combines sensitivity to subtle shape details with tolerance to irrelevant changes, e.g., of viewpoint and illumination. Many studies support the hypothesis that perceptual representations of shape can be characterized by continuous perceptual spaces with well-defined metric properties reflecting physical similarities of shapes. This idea underlies classical theories of shape categorization (Ashby & Perrin, 1988; Edelman, 1999; Nosofsky, 1992) and has motivated the concept of a "face space" that is central to many theories of face recognition. Studies with face morphs yield converging evidence suggesting that both perceptual performance and the responses of face-selective neurons vary gradually and systematically with the location of faces in a face space (Leopold, Bondar, & Giese, 2006; Rhodes, Brennan, & Carey, 1987; Valentine, 1991). Further evidence for the relevance of a face space is provided by studies of high-level after-effects showing that adaptation to an "anti-face," generated by extrapolation in face space, results in

an after-effect that facilitates the perception of the original face (Leopold, O'Toole, Vetter, & Blanz, 2001; Webster, Kaping, Mizokami, & Duhamel, 2004).

Direct evidence for metric perceptual shape spaces was obtained in studies that used multidimensional scaling (MDS) (Borg & Groenen, 1997; Shepard, 1987) to map perceptual similarity judgments for three-dimensional shapes onto geometrical configurations in low-dimensional embedding spaces. Stimuli generated by morphing between prototypical shapes resulted in recovered configurations in the constructed perceptual space that closely matched the configurations in the morphing space, as defined by the weights assigned to the individual prototypes in the morph (Cutzu & Edelman, 1996, 1998; Sugihara, Edelman, & Tanaka, 1998). Thus, the metric of the perceptual space reflects the physical similarities between shapes. When applied to neural activity data from monkey inferotemporal cortex, the same technique yielded metric "neural spaces" in which the stimulus configurations strongly resembled the original layout in morph space (Op de Beeck, Wagemans, & Vogels, 2001).

In the present paper, we address the question whether complex spatio-temporal patterns, such as movement trajectories, are also represented by a perceptual space with a metric that is determined by physical similarity. The underlying hypothesis is illustrated schematically in Figure 1: Three motion patterns (for example walking, running and limping) are characterized by the corresponding joint trajectories (Figure 1A). The similarities between these trajectories define a *physical space,* and each locomotion pattern corresponds to a single point in this space. At the same time, these motion patterns elicit perceptual impressions, which might be characterized as points in a low-dimensional *perceptual space* that can be inferred by analysis of the viewer's perceptual judgments (Figure 1B). This perceptual representation of motion patterns is only indirectly linked to the parameters of the viewed movement (e.g., joint trajectories), because it is derived from fundamentally different signals (e.g., neural activity patterns in the higher visual cortex). This implies that the mapping from physical space into the perceptual space might be given by a rather complex transformation that would not be easy to model. The approach we adopt here circumvents this difficulty by analyzing the metric structure of the two spaces separately, and quantifying their similarity (see Edelman, 1999; Shepard & Chipman, 1970, for details).

The present study addresses the following questions:

1. Does a readily interpretable, low-dimensional metric perceptual space for complex motion patterns exist, similar to shape spaces for static patterns?
2. Is this space related in a systematic way to the physical movement space that is defined by the joint trajectories?
3. Is the perceptual representation *veridical* in the sense that the perceptual space shares metric properties with the physical space?

A particularly useful relationship between perceptual and physical space is obtained when distance ranks between movement patterns are preserved in the perceptual space
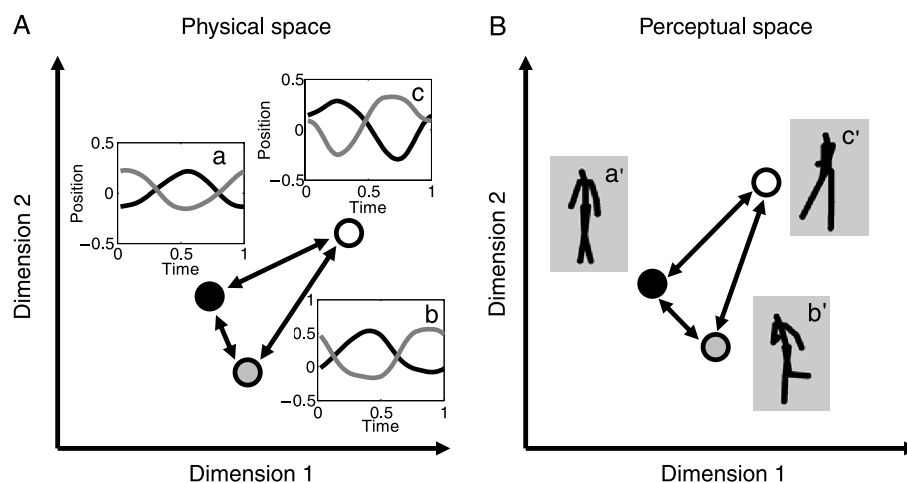

Figure 1. Schematic illustration of a veridical relationship between a physical movement space and a visual perceptual space. (A) Three motion patterns (e.g., walking, running and marching) are defined by the corresponding joint trajectories. Each pattern is mapped onto a single point (indicated by the disks with different colors) in the physical space. Distances in this space are determined by the physical distances between joint trajectories. (B) The same patterns result in the perception of locomotion patterns. We assume that each pattern can be represented as point in a low-dimensional metric perceptual space. The distances in this space are determined by the perceived similarities of the motion patterns. For the shown example, the mapping between perceptual and physical space is a *second order isomorphism*: Pattern pairs with larger distance in physical space are mapped onto point pairs with larger distance in the perceptual space (i.e., $d(a, b) < d(a, c) < d(b, c)$ implies $d(a', b') < d(a', c') < d(b', c')$).

(cf. Figure 1). This would imply, in particular, that the perceptual system maps patterns that are farther apart in physical movement space onto points that are likewise farther from each other in perceptual space. A mapping between physical and perceptual spaces that preserves distance ranks is called a *second-order isomorphism* (Edelman, 1999; Shepard & Chipman, 1970). This class of mappings provides a representation that is particularly useful for the classification and categorization of motion patterns based on their physical similarities.

In contrast to static shapes, it is not obvious which physical similarity measures might best capture behaviorally relevant differences between dynamic movement patterns. A further question that we need to address is thus which physical distance measures are most appropriate for characterizing the similarities between physical and perceptual spaces?

We approach these questions by constructing perceptual spaces through the application of MDS to perceptual similarity judgments for movement patterns. The movement patterns were generated by motion morphing between three natural locomotion patterns, presented as point-light stimuli or as stick figures. The recovered low-dimensional configurations in perceptual space were compared to configurations in physical spaces, which were constructed, also by MDS, from physical distance measures for the joint trajectories. We considered a variety of different distance measures, in an attempt to determine which physical distances are reflected most closely in the perceptual metrics. Our analysis reveals important computational constraints on visual motion perception and suggests structural similarities between perceptual representations of motion and shape.

A part of this work has been published previously in abstract form (e.g., Giese, Thornton, & Edelman, 2003).

Methods

Our study includes two main experiments and one control experiment. Parameterized classes of motion patterns were created by *motion morphing* (Giese & Poggio, 2000), applying a method that generates new movements by linear combination of the trajectories of three prototypical gait patterns (walking, running, and marching). The stimulus sets consisted of 7 movements that formed simple, low-dimensional configurations in the space defined by the morphing weights. The perceived similarities of these patterns were assessed using two different experimental paradigms. Based on the perceived similarities of pairs of the motion patterns, the metrics of the *perceptual space* was reconstructed by multidimensional scaling (MDS). The recovered configurations in perceptual space were first compared to the original configuration in morphing weight space. The result of this comparison depends on details of the morphing

algorithm, which determines the meaning of the morphing weights. In order to obtain a characterization of the actual *physical movement space* that is independent of the morphing method, we computed the distances between the joint trajectories of the presented stimuli and constructed their metric configurations in physical space by applying MDS. An entire family of physical distance measures was tested, with the goal of finding the physical metric that most closely approximates the configurations in the perceptual space.

Experiment 1 used motion morphs that were generated from prototypical locomotion patterns, which had been tracked from normal video. Morphs were generated between the two-dimensional joint trajectories in the image plane. In this experiment, the test patterns formed a triangular configuration in morphing space (Figure 2A). Stimuli were point-light walkers (Johansson, 1973) presenting the figure from a side view (Figure 2B). The assessment of similarity in this experiment was based on a compare pairs-of-pairs paradigm (CPP).

In order to verify that the basic results were not critically dependent on the presented configuration in morphing space, we performed a control experiment using exactly the same procedure as in Experiment 1, but an L-shaped configuration instead of a triangular one in morphing space.

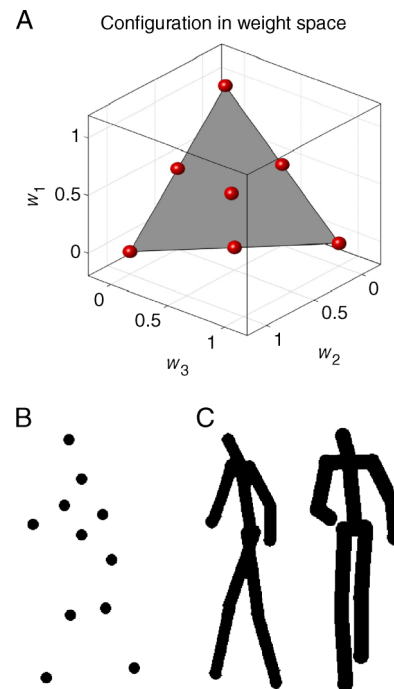Experiment 2 was based on prototype movements that were recorded by motion capture. Morphs, forming again

Figure 2. Stimuli. (A) Displayed motion patterns in Experiments 1 and 2 formed a triangular configuration in the space of morphing weights w_i. (B) Stimuli in Experiment 1 were point-light walkers with 11 dots presented in a side view. (C) Stimuli in Experiment 2 were stick figures shown with different views.

a triangular configuration, were created from three-dimensional joint trajectories. Stimuli were stick figures that were rendered from different view angles to test the view dependency of the underlying perceptual representation (Figure 2C). Assessment of perceived similarity was based on a delayed match-to-sample paradigm (DMTS).

Prototype movements

The same three locomotion patterns (walking, running and marching) were recorded from the same actor for Experiments 1 and 2. Trajectories for Experiment 1 were tracked from normal video showing the actor locomoting orthogonally to the view axis of the camera. Joints were tracked by hand-marking of the positions of 11 body points in individual frames. The prototypical locomotion patterns were filmed using a Kodak VX 1000 camera with a frame rate of 30 frames per second. The closest distance between the locomotion line and the camera was 6 m. Only a single gait cycle was used for the generation of the stimuli. The translation of the body center was subtracted by fitting the two-dimensional hip trajectory by a linear function of time. Subtraction of this translation results in movement that looks like a person on a treadmill. Start and end of the gait cycles were determined by the frames with maximum extension of the legs. Tracked trajectories were time-normalized and smoothed by fitting them with a second order Fourier series. The size of the patterns was normalized by rescaling to keep the distance between hip and head constant (see Giese & Lappe, 2002, for details).

Three-dimensional joint trajectories for Experiment 2 were recorded with a VICON 612 motion capture system with 6 cameras. The 3D positions of 41 reflecting markers were recorded with a sampling frequency of 120 Hz and a spatial accuracy below 1 mm. The resulting trajectories were processed using commercial software by VICON. Multiple steps were recorded for each prototypical gait, and one representative step was selected as the basis for the motion morphing. The average motion of all hip markers was subtracted from all marker positions to generate a movement looking like a walker on the treadmill. The distance between head and hip was scaled to a constant to normalize the size of the moving figure. Trajectories were smoothed by fitting 4th order Fourier series to the data. Each gait cycle was resampled to time-normalize the data to a cycle time of about 1.6 s. From the marker positions 13 joint positions were computed, which were connected by straight lines to draw the stick figures (Figure 2C).

Motion morphing

The prototypical locomotion patterns were interpolated by applying a motion morphing algorithm that generates new trajectories by a linear combination of the prototype trajectories in space–time (Giese & Poggio, 2000). Formally, the new trajectory was given by the equation:

$$
\begin{aligned}
\text{New motion pattern} = {} & w_1 \cdot (\text{Prototype 1}) \\
& + w_2 \cdot (\text{Prototype 2}) \quad (1) \\
& + w_3 \cdot (\text{Prototype 3}).
\end{aligned}
$$

The morphing weights were non-negative and satisfied the condition $w_1 + w_2 + w_3 = 1$, thus defining a two-dimensional plane in weight space (Figure 2A).

In computer graphics, a variety of motion morphing techniques have been proposed (Bruderlin & Williams, 1995; Gleicher, 2001; Lee & Shin, 1999; Rose, Cohen, & Bodenheimer, 1998; Unuma, Anjyo, & Takeuchi, 1995; Wiley & Hahn, 1997; Witkin & Popovic, 1995), some of which have been used in psychophysical experiments (Giese & Lappe, 2002; Hill & Pollick, 2000; Troje, 2002). The algorithm that we applied (Giese & Poggio, 2000) has been demonstrated to produce naturally looking morphs even for quite dissimilar prototypical movements and for highly complex body movements such as martial arts techniques (Giese & Poggio, 2000; Ilg, Bakir, Mezger, & Giese, 2004; Mezger, Ilg, & Giese, 2005). Psychophysical experiments using similar stimuli showed that this algorithm produces interpolated movements whose naturalness ratings lie between those of the prototypes (Giese & Lappe, 2002). This rules out the potential concern that the present results may hold only for artificial-looking interpolated movements.

Stimuli

Point-light walkers consisted of 11 black dots presented on a gray background with a diameter of 0.4 degrees visual angle (Figure 2B). Each figure, presented as side view, subtended an area of approximately 4 by 10 degrees of visual angle. The stimuli were presented without time limitation, until the observer responded.

Stick figures consisted of 12 black line segments, approximately 0.2 degrees wide (Figure 2C). These figures were projected onto the image plane in parallel projection with four different view angles (90, 112.5, 135, and 157.5 deg relative to the locomotion direction). These figures subtended approximately an area of 2 by 5 degrees of visual angle. These stimuli were presented sequentially. The first figure was always located at the centre of the screen for two step cycles. The second stimulus was randomly positioned on the circumference of a virtual circle (diameter 8 deg) centered on the screen and was presented until the observer responded.

Stimuli were presented at 25 frames per second on a Macintosh G4 with a 21-in. color monitor with screen resolution of 1152 × 780 pixels. Stimuli were viewed binocularly, but non-stereoscopically, from a distance of 70 cm. Stimuli were presented in random order, and the

initial phases of each individual gait pattern were randomized in each trial. Stimuli were generated in MATLAB 6.1 and were stored as frame sequences. The frames were played using a stimulus presentation software that was custom written in C using routines based on published work (Pelli & Zhang, 1991; Rensink, 1990; Steinman & Nawrot, 1992).

Procedure

The experimental paradigms were adopted from Cutzu and Edelman (1996, 1998). For the compare pairs-of-pairs paradigm (CPP), two pairs of stimuli were presented simultaneously, one in the upper and one in the lower half of the screen. Observers had to indicate by a key press the pair in which the stimuli appeared more similar. In total, the 7 stimuli of the triangle configurations resulted in 210 unique pairs of pairs, each of which was presented once.

For the delayed match-to-sample paradigm (DMTS), two stimuli were presented in sequence, showing the same or two different gaits. The first stimulus was always one out of four tested views, randomly chosen. The second pattern always showed one of the other views. Observers had to indicate whether the two observed actions were same or different, *irrespective of viewpoint*. With 4 different view angles, 378 unique stimulus pairs were generated. To balance the fraction of trials in which the pair showed the same action with different views against the trials with different actions, the same-action pairs were repeated five times, resulting in a total of 546 trials.

Participants

Fourteen subjects took part in Experiment 1 and eleven in Experiment 2. In addition, 19 participants took part in the control experiment. Participants were recruited from the Tübingen community by the Max Planck Institute for Biological Cybernetics. All observers had normal or corrected-to-normal vision, and none of them participated in more than one experiment. Participants were naive as to the purpose of the experiment until data collection had been completed. They were tested individually and gave written informed consent to participate in the study and were paid for participation. Approval for the applied procedure had been obtained from the ethics board of the Max Planck Institute for Biological Cybernetics in Tübingen.

Physical trajectory distances

A physical space for the presented motion stimuli can be constructed in a straightforward manner by using the morphing weights $w_1, \ldots, w_3$ as axes of a low-dimensional pattern space. Distances between trajectories are then characterized by the Euclidean distances between the corresponding weight vectors. However, this characteriza- tion of the physical space is unsatisfactory because the interpretation of the morphing weights depends in a complex manner on the set of morphed patterns and on details of the applied morphing algorithm. Morphing weights thus do not provide a characterization of physical similarity between joint trajectories that is independent from the morphing method. Potential differences between the metric structure of the perceptual space and that of the space of morphing weights could thus be explained either by the perceptual metrics being incompatible with the physical similarities of the stimuli, or by the inability of the morphing weights to capture the perceptually relevant physical characteristics of the movement. This problem arises in conjunction with all known technical algorithms for motion morphing because all of them apply heuristic methods for interpolating between trajectories in space–time rather than taking into account specific constraints derived from perception.

In order to obtain a characterization of the physical space that is independent of the morphing method, we tested an entire family of trajectory distance measures that do not rely on motion morphing. Based on the computed distances, we constructed the corresponding physical spaces by MDS. In principle, there are infinitely many ways of defining distances between trajectories. This is schematically demonstrated for a simplified example in Figure 3A. It illustrates two trajectories with different timing, so that that the trajectory $x_2(t)$ can be derived from the trajectory $x_1(t)$ by *time warping,* i.e., by deformation of the time axis using a smooth monotonic warping function $\tau(t)$. Formally this deformation can be written as $x_2(t) = x_1(\tau(t))$. It is assumed that $\tau(0) = 0$ and $\tau(T) = T$, T being the total duration of the movements, which is assumed to be equal.

One way of defining the distance between the two trajectories is to use the amount of temporal deformation as a measure. The temporal deformation can be quantified by the deviation of the time warping function from the identity, for example, using the expression

$$D_{12} = \sqrt{\frac{1}{T} \int_0^T |\tau(t) - t|^2 \mathrm{d}t}. \tag{2}$$

Alternatively, one can use the Euclidean distance between the trajectories for each fixed point in time to construct a distance measure without trying to determine the timing shifts between the trajectories. This results in the distance measure

$$D_{12} = \sqrt{\frac{1}{T} \int_0^T |w_1(t) - w_2(t)|^2 \mathrm{d}t}, \tag{3}$$

which has been quite frequently used in the literature. It is important to note that this distance measure does not quantify timing differences explicitly. Instead, they
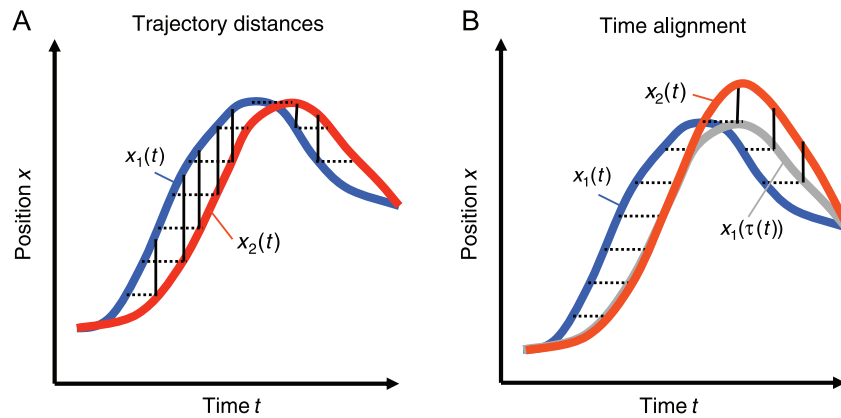
Figure 3. Time alignment and spatio-temporal distance. (A) Two trajectories that differ only in their timing, implying the relationship $x_2(t) = x_1(\tau(t))$ with a time warping function $\tau(t)$. A distance measure can be derived, for example, based on the time shifts between the trajectories (dashed lines), or as function of the spatial shifts for each point in time (solid lines). (B) Usually, only parts of the differences between trajectories can be accounted for by temporal alignment. The gray curve shows the best possible temporal alignment of the trajectory $\tau_1(t)$ with the trajectory $\tau_2(t)$, which is obtained by an adequate deformation of the time axis. The remaining deviation between the trajectories (solid lines) needs to be accounted for by spatial differences. The physical distance measure that fits best the perceptual metric depends only on those spatial shifts, but not on the temporal shifts (indicated by the dashed lines).

influence the distance indirectly through the induced spatial differences (solid lines in Figure 3). In fact, by choosing the time warping function, the two trajectories are brought into spatio-temporal correspondence (Giese & Poggio, 2000), defining for each point on trajectory $x_1(t)$ one point on the second trajectory $x_2(t)$ that corresponds to it. There are many other ways to define such correspondence. For example, one might associate points of the two trajectories that are related to each other by shifts in space *and* time. Likewise, there are multiple ways in which timing difference and spatial differences can be combined into a single value of the distance measure.

The family of distance measures in our study was defined as a weighted combination of temporal and spatial differences, as suggested above. Assuming that $x_1(t)$ and $x_2(t)$ signify the joint (position) trajectories, which are multidimensional because they contain all major joints of the viewed moving figure, the tested distance measures were given by the expression

$$d(x_1, x_2) = \sqrt[q]{(D_{12} + D_{21})/2} \quad \text{with}$$

$$D_{ij} = \frac{1}{T} \int_0^T |x_i(\tau(t)) - x_j(t)|^q dt + \lambda \frac{1}{T} \int_0^T |\tau(t) - t|^q dt,$$

$$(4)$$

with $i, j = 1, 2$. The first equation ensures that the distance measure d is symmetric with respect to x_1 and x_2. T is the duration of one gait cycle. The time warping function $\tau(t)$ was determined by dynamic time warping (Giese & Poggio, 2000; Rabiner & Juang, 1993), an algorithm that determines an optimal time alignment between the two trajectories. For the trajectories depicted in Figure 3A, this

algorithm would compute the time shifts indicated by the dashed lines, resulting in a vanishing contribution of the first term in the expression for D_{ij} in Equation 4. Choosing instead $\tau(t) = t$ defines distance measures without time alignment. By choosing positive values for the parameter λ, the distance measure takes explicitly into account overall timing differences between the two trajectories. The choice $\lambda = 0$ corresponds to physical distances that, except for time warping, depend only on spatial differences between the trajectories. Usually, only a part of the differences between trajectories can be removed by time alignment (cf. Figure 3B). In this case, one obtains nonzero contributions of both terms of the expression D_{ij}.

Distance norms

The parameter q in Equation 4 makes it possible to specify different distance norms (Kreyszig, 1989). Setting $q = 2$ results in the common Euclidean norm. Large values of q correspond to distances that emphasize the influence of outliers. This is illustrated in Figure 4 that shows two hypothetical trajectories. The trajectory $x_1(t)$ is just a linear trend. The other trajectory follows this trend with a positive spatial offset of 10. Around $t = 40$, however, this trajectory has some outliers, the biggest deviation appearing for $t = 40$. Without these outliers, the distance measure $D_{12}(q) = \sqrt[q]{\frac{1}{T} \int_0^T |x_1(t) - x_2(t)|^q dt}$ provides the same value $D_{12}(q) = 10$ independent of the value of the parameter q (assuming $q \geq 1$). In the presence of the outliers, for the Euclidean norm that corresponds to the choice $q = 2$, the computed distance $D_{12}(2) = 12.16$ is larger than the offset between the trajectories. For the choice $q = 1$, one obtains the value $D_{12}(1) = 10.99$ that is closer to the true offset. The reason is that the square term
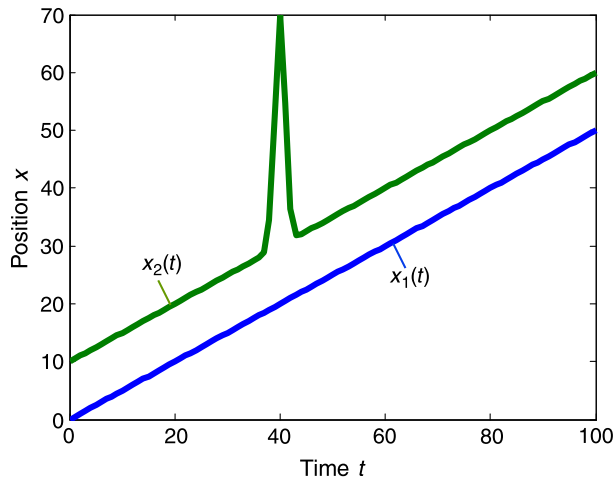
Figure 4. Data set with outliers. Two trajectories showing the same linear trend. The trajectory $x_2(t)$ has a spatial offset and is displaced by 10 spatial units upwards. Around $t = 40$ this trajectory has outliers and takes the maximum value 70. The maximum vertical distance between the two trajectories is 50, arising for $t = 40$.

in the Euclidean norm amplifies the influence of outliers compared to the linear term for $q = 1$. Choice of very large values of q results in distance values that are mainly determined by the outliers, resulting in a distance that is close to the largest distance between the trajectories determined time-point by time-point. For the given example, this maximum distance is given by $|x_2(40) - x_1(40)| = 50$. For example, for $q = 100$ one obtains a distance value $D_{12}(100) = 47.74$ that is close to this maximum.

The value of q in Equation 4 thus controls how robust the constructed metric in physical space is against outliers. For the limit $q \rightarrow \infty$ (infinity norm), the distance is determined by the time point(s) with the largest spatial difference between the trajectories over the entire time interval $[0, T]$. Small values $q < 2$ correspond to distance measures that average over the entire time course of the trajectory and which are more robust against outliers than the Euclidean norm.

Applying MDS, we constructed physical spaces for different combinations of the parameters λ and q. By comparing the stimulus configurations in these physical spaces with the ones in perceptual space, we sought a physical distance that closely approximates the perceptual metric.

Statistical analysis

The perceptual judgments were used to compute similarity matrices. For the CPP task, we counted how often a stimulus pair was perceived as more similar than all other pairs. For the DMTS task, we counted how often a stimulus pair was classified as identical action, combining the data of all subjects. (Consistent but somewhat noisier results were obtained with a similarity matrix constructed from reaction times.) Each similarity matrix was normal-

ized by subtracting its minimum over all entries and rescaling the matrix to maximum one. The distance matrix for the MDS analysis was obtained by subtracting the entries of this matrix from one. Results from different subjects for the CPP task were combined using INDSCAL analysis (Carroll & Chang, 1970) implemented in SAS (SAS Inst. Inc.). The results of the DTMS task were analyzed by computing a common similarity matrix over all participants. The perceptual space for this task and the physical spaces were reconstructed by regular metric MDS (Borg & Groenen, 1997).

Because MDS yields low-dimensional metric configurations that are determined up to an arbitrary scaling, rotation, and reflection, the resulting configurations needed to be aligned prior to the quantification of their similarities. The recovered metric configurations were aligned by Procrustes transformation (Borg & Groenen, 1997; Cutzu & Edelman, 1996). For two given sets of data points, this transformation determines an optimal combination of scaling, rotation and reflection that aligns the second data set with the first (see Appendix A for further details).

In Experiment 2, configurations were aligned with a three-dimensional space, defined by the morphing weights (2 dimensions) and the view angle (third dimension). View angle and morphing weights have different units, implying that their relative scaling with respect to each other in the perceptual metrics is unknown. The unknown scaling factor was estimated using an iterative procedure that combines regular Procrustes alignment with a line search for this additional parameter (see Appendix A).

The statistical significance of the similarities between recovered and original configurations was tested using the bootstrap method (Efron & Tibshirani, 1993), which compares the estimated similarity to the mean and standard deviation of a distribution of values obtained in a Monte Carlo simulation that randomizes the data by reshuffling the rows of the similarity matrices (Cutzu & Edelman, 1998). Random data sets consisted of 100 samples, from which means, $\bar{d}_{P,random}$, and standard deviations, $SD(d_{P,random})$, of the similarity measures were computed. Values of a d' equivalent were computed according to the relationship:

$$d' = (d_{P,data} - \bar{d}_{P,random})/SD(d_{P,random}). \quad (5)$$

The statistics for testing the significance of the similarities was based on a t test for this d' equivalent.

Results

Experiment 1: Morphs derived from two-dimensional trajectories in the image plane

In Experiment 1, stimuli were presented as point-light walkers. The dot trajectories were obtained by motion
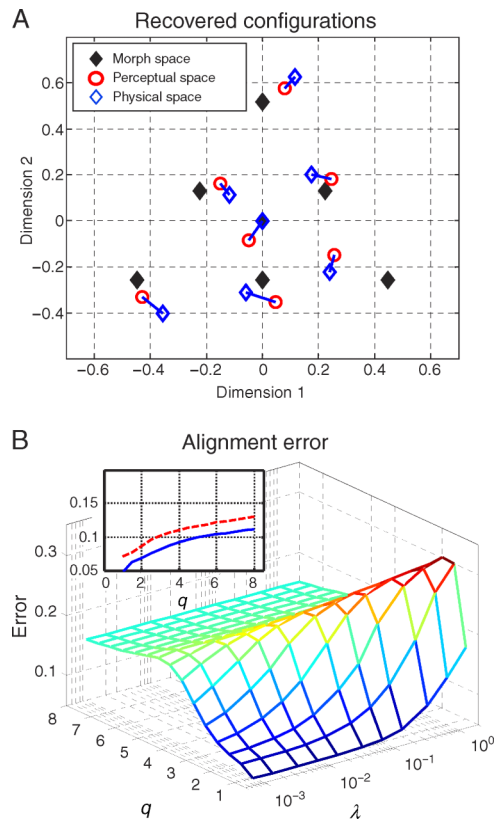
Figure 5. Results from Experiment 1 (pair comparison paradigm) and comparison with configurations in physical space. (A) Stimulus configuration defined in space of morphing weights (filled diamonds) compared with the recovered configuration in the perceptual space (circles). The corresponding configuration in physical space computed from trajectory distances with optimized parameters ($\lambda = 0$, $q = 1$) is indicated by the open diamonds. (B) Alignment errors (Procrustes distance) between configuration in perceptual space and configurations in physical space constructed from trajectory distances varying the parameters λ and q in Equation 4. The inset shows alignment errors for space–time distances ($\lambda = 0$) with (solid line) and without time warping ($\tau(t) \equiv t$) (dashed line).

morphing of the two-dimensional joint trajectories in the image plane, defining a triangular pattern in the space of morphing weights. Perceived similarities were assessed by a compare-pairs-of-pairs paradigm (CPP). The perceptual metric was reconstructed jointly form the data of all subjects by individually weighed MDS (INDSCAL). The optimal embedding dimension determined by a scree test was two, thus matching the dimensionality of the stimulus configuration in morphing space.

Reconstructed configuration in perceptual space

Figure 5A shows the recovered configuration in perceptual space after re-alignment with the triangular configuration in morphing weight space via a Procrustes

transformation (Borg & Groenen, 1997). The recovered configuration (circles) is quite similar to the configuration in morphing-weight space (solid diamonds). However, some systematic deformations, in particular in the lower right corner, indicate that the perceptual space is not linearly related to the space of morphing weights. This nonlinearity might be due to the choice of prototypes or to specific properties of the motion morphing method. Despite this deformation, the Procrustes distance (Borg & Groenen, 1997), i.e., the Euclidean distance after Procrustes alignment and normalization, between the two configurations is relatively small ($d_P = 0.41$).

To determine whether the similarity of the recovered configuration in perceptual space to the original configuration in morphing space is significant, we conducted a bootstrap analysis (see Methods). This analysis indicated that the resulting value for the Procrustes distance is highly significant (d' equivalent 8.15; $p < 0.001$), cf. Table 1. All data were also analyzed using the coefficient of congruence (Borg & Groenen, 1997) as another measure for the similarities of the recovered configurations. The results of this analysis (not shown) are highly consistent with the results obtained for the Procrustes distance.

Configurations in physical space

The Procrustes distances between the reconstructed configuration in perceptual space and configurations in physical space were computed for different combinations of the parameters λ and q in Equation 4. Figure 5B shows that the best agreement between the configurations in physical and perceptual space, measured by the Procrustes distance, was obtained for $q = 1$. This implies that the perceptual similarity of movements is characterized by an integration over time that is robust against outliers rather than by an error measure that emphasizes the contribution of individual time points where the spatial distances

Perceptual vs. morph space

d_P data	d_P random	d' equivalent	t_{99}	p
Experiment 1				
0.41	1.1 ± 0.09	8.15	81.5	<0.001
Control experiment (L configuration)				
0.47	0.63 ± 0.01	1.62	16.2	<0.001
Experiment 2				
2D embedding space				
0.15	0.95 ± 0.31	2.61	26.1	<0.001
3D embedding space				
0.51	0.91 ± 0.04	9.74	97.3	<0.001

Table 1. Results from the bootstrap analysis for the similarity of configurations in perceptual and morphing space. Similarity was assessed by computation of the Procrustes distance d_P. For details about the bootstrap analysis, see Methods.

Physical vs. perceptual space

d_P data	d_P random	d' equivalent	t_{99}	p
Experiment 1				
0.05	0.97 ± 0.28	3.3	32.9	<0.001
Control experiment (L configuration)				
0.35	1.03 ± 0.31	2.15	21.5	<0.001
Experiment 2				
Distance of 2D trajectories (2D embedding space)				
0.61	1.02 ± 0.31	1.3	13.0	<0.001
Distance of 3D trajectories (2D embedding space)				
0.23	1.05 ± 0.27	3.0	30.0	<0.001
Distance of 3D trajectories (3D embedding space)				
0.26	0.91 ± 0.06	11.7	117	<0.001

Table 2. Results from the bootstrap analysis for the similarity of configurations in physical and perceptual space. For Experiment 2, physical distances between 2D joint trajectories in the image plane and the motion-captured 3D trajectories were compared. All configurations in physical space were reconstructed with the optimized parameters ($\lambda = 0$ and $q = 1$) for the distance measure defined by Equation 1.

between trajectories are large. For small values of λ, towards large values of q the alignment error levels off. This behavior might be explained by the fact that above a certain critical value of q, the distance measure is dominated by the maximum differences over the whole time course further increases of q resulting only in minor additional changes.

In addition, the Procrustes distance between the configurations increases with the parameter λ and attains a minimum for $\lambda = 0$. This implies that the physical distances that contain an extra term that measures overall timing differences between the two trajectories resulted in worse fits of the perceptual metrics. The best approximation was achieved with distances that, apart from prior time alignment, depend only on spatial displacements between the trajectories time-point by time-point. This does not imply that timing information, in general, is irrelevant for the perceptual metric. It just shows that *overall timing changes* that affect all joints synchronously have only a moderate influence on the perceptual metric. Temporal variations that affect the relative timing between different joints cannot be modeled by simple time alignment (as defined by the function $\tau(t)$) and are thus captured by spatial differences in the first term of the distance measure (2).

A more detailed analysis suggests, in addition, that it is crucial to include a nontrivial time-warping function in the distance computation. The inset in Figure 5B shows that physical distances with time alignment (solid curve) result in consistently better approximations of the perceptual metrics than distances without time warping ($\tau(t) \equiv t$; dashed curve). The perceptual system thus seems to compensate efficiently for time warping between the trajectories. The open diamonds in Figure 5A indicate the reconstructed configuration in physical space for the optimal parameter values ($q = 1$, $\lambda = 0$) after Procrustes alignment with the perceptual configuration. The two configurations are extremely similar, as confirmed by the very small and highly significant Procrustes distance between them ($d_P = 0.05$; d' equivalent 3.3; $p < 0.001$) (cf. Table 2 for further details.) The structure of the perceptual metrics can therefore be approximated very accurately by a physical metric that is based on the distances between joint trajectories.

Control experiment: L configuration in morph space

In an additional control experiment with 11 subjects using the same paradigm and analysis, we verified that these results remain valid for another configuration in morphing space. The specified pattern was an "L" in
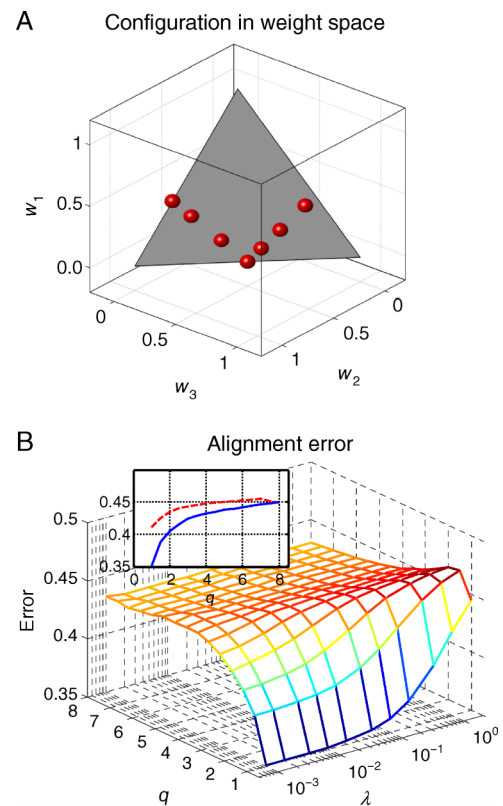

Figure 6. Control Experiment with L configuration in morphing space. (A) Configuration in space of morphing weights. (B) Alignment errors (Procrustes distance) between configuration in perceptual space and configurations in physical space constructed from trajectory distances varying the parameters λ and q. Inset shows the alignment errors for space–time distances ($\lambda = 0$) with (solid line) and without time warping ($\tau(t) \equiv t$) (dashed line). Conventions as in Figure 5B.

weights space (Figure 6A). Paradigm and analysis were as in Experiment 1.

While perceptual data were more noisy than in Experiment 1, the Procrustes distance between configurations in perceptual and physical space was comparable to the triangular configuration (see Table 1). Most importantly, the alignment error between configurations in perceptual space and physical space was again minimal for the parameter choices $\lambda = 0$ and $q = 1$ (Figure 6B; Table 2). The inset in Figure 6B shows the comparison between the alignment errors for trajectory distances for $\lambda = 0$ with (solid line) and without time warping, i.e., for $\tau(t) \equiv t$ (dashed line). Best alignment was obtained for a physical distance with time warping and $q = 1$. This confirms the results of Experiment 1 for another configuration in morphing space.

Experiment 2: Morphs derived from motion captured three-dimensional trajectories

The movements for Experiment 2 were generated by morphing between three-dimensional trajectories recorded by motion capture (see Methods). Again, stimuli formed a triangle configuration in the space of morphing weights (Figure 2A). Trajectories were used to animate a stick figure (Figure 2C) that was rendered with four different view angles to study the influence of view dependency.

Configuration in perceptual space

The recovered configuration in a *two-dimensional* perceptual space (assuming two embedding dimensions for the MDS procedure) is shown in Figure 7A. Crosses with identical color indicate different views of the same motion patterns (combination of morph weights). Different views of the same motion are clustered in perceptual space. This indicates that subjects effectively ignored the view angle in their similarity judgments. Open circles with matching colors indicate the centroids of the view clusters. The centroids define a configuration that is very similar to the triangular configuration in morphing space (open diamonds), as confirmed by a small and highly significant Procrustes distance between the two configurations ($d_P = 0.15$; d' equivalent 2.6; $p < 0.001$), cf. Table 1. Thus, results very similar to those of Experiment 1 were obtained for stick figures and motion morphs that had been generated from three-dimensional trajectories. As in Experiment 1, the recovered configuration in perceptual space shows gradual deformations compared to the configuration in weight space.

The reconstructed *three-dimensional* configuration in perceptual space is shown in Figure 7B (red spheres), aligned with the configuration of the stimuli in a three-dimensional space (blue spheres) that is defined by the morphing weights (two independent dimensions) and the view angle (third dimension). Along the first two
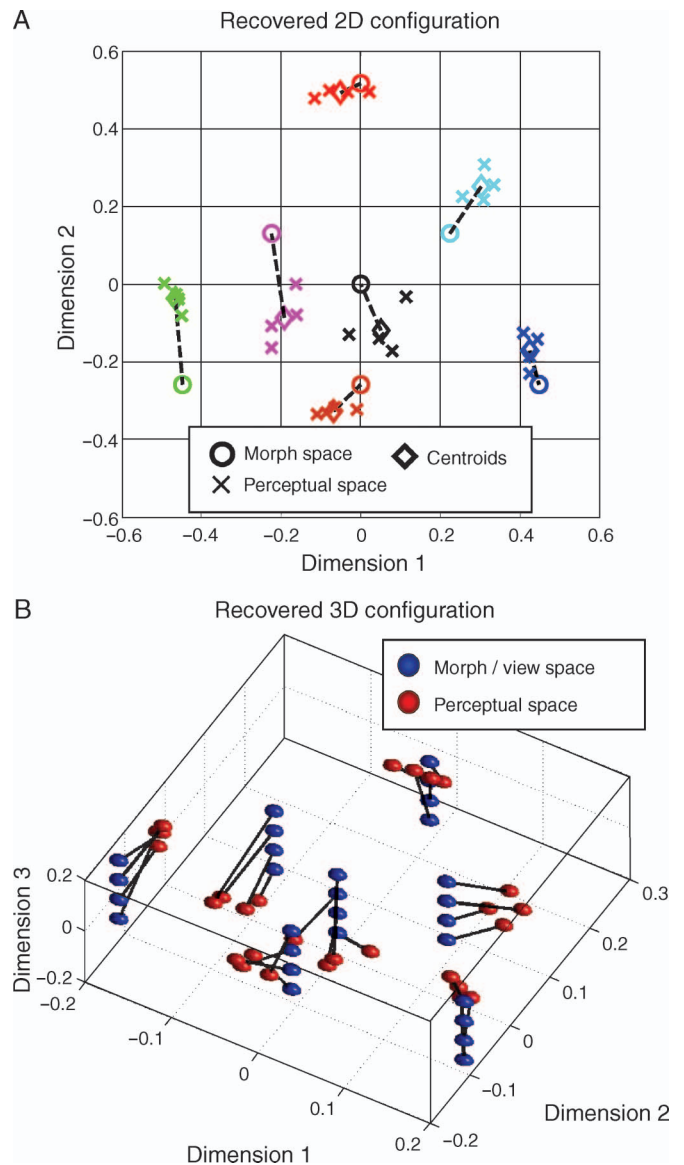

Figure 7. Results from Experiment 2 (delayed match-to-sample paradigm with viewpoint variation). (A) Two-dimensional configuration recovered from the perceptual data, aligned with the configuration in weight space (circles). Crosses with same color indicate the same locomotion pattern with different viewpoints. Open diamonds with same color indicate the centers of these clusters. (B) Three-dimensional configuration recovered from the perceptual data (red spheres) aligned with the stimulus configuration in the three-dimensional space (blue spheres) that is defined by the morphing weights (dimensions 1 and 2) and the view angle (dimension 3).

dimensions, the recovered configuration is very similar to the configuration in the space of morph weights. This supports the relevance of these two dimensions for the underlying perceptual representation. Along the third recovered dimension, which is aligned with the view angle dimension of the stimuli, the data do not show any systematic ordering. This explains why the Procrustes distance between the original and the recovered configuration

($d_P = 0.51$; d' equivalent 9.74; $p < 0.05$) is larger than for the two-dimensional configurations. This result also confirms that subjects effectively ignored the viewpoint of the moving figure, as it was required by the task.
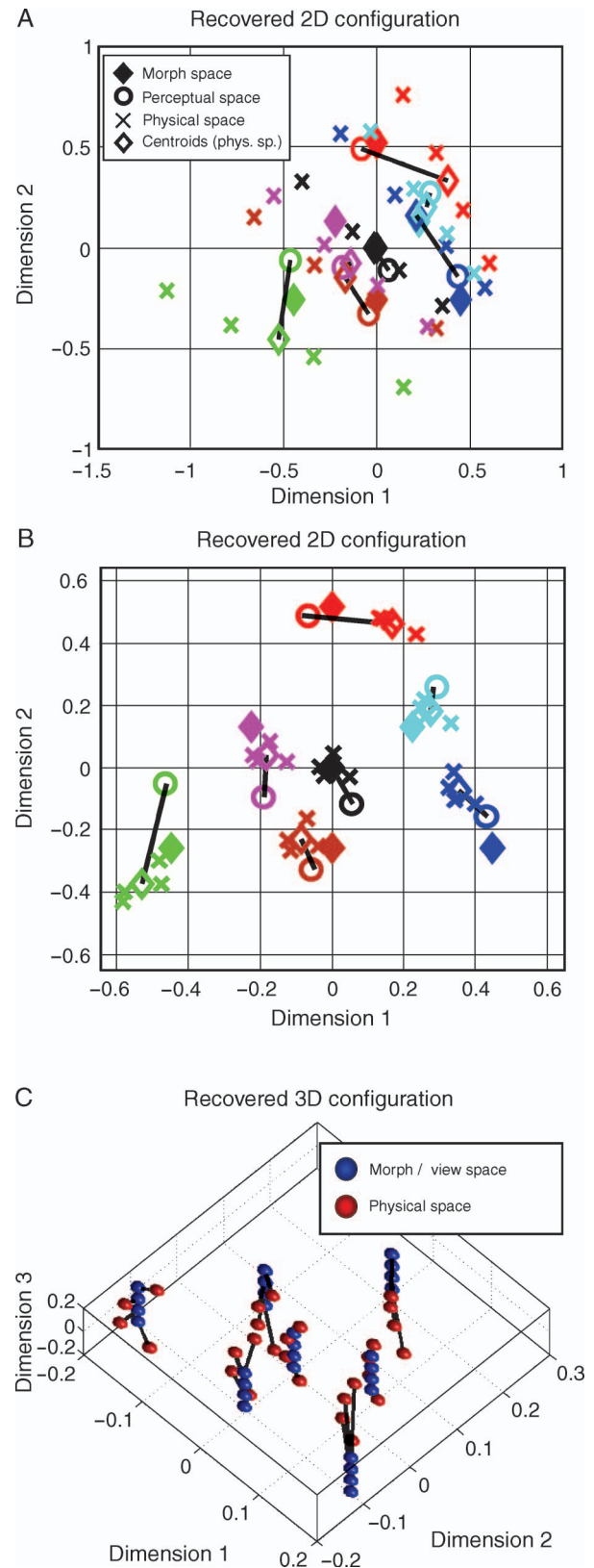
Configurations in physical space

As the stimuli in Experiment 2 were based on three-dimensional trajectories, this data set allowed us to compare physical distance measures derived from three-dimensional trajectories and two-dimensional trajectories in the image plane.

The configuration in a physical space derived from two-dimensional trajectories that had been obtained by projecting the three-dimensional trajectories in the image plane is shown in Figure 8A. As for the distances derived from the trajectories in the image plane in Experiment 1, optimal alignment between the configurations in perceptual and physical space was obtained for a distance measure described by Equation 4 with $q = 1$ and $\lambda = 0$. Figure 8A shows the recovered two-dimensional configuration in physical space using the same conventions as in Figure 7A. In contrast to the psychophysical data, points representing different view angles of the same motion are widely scattered for the recovered two-dimensional configuration. The centroids of the points belonging to the same actions are indicated by circles. The Procrustes distance between these centroids and the corresponding points of the configuration in perceptual space are large, although still significant (Table 2). This result indicates that distances between the two-dimensional joint trajectories are not suitable for reproducing the perceptual metrics of motion recognition in the presence of viewpoint changes.

In contrast, the two-dimensional configuration in the physical space recovered from distances between the three-dimensional trajectories (Figure 8B) does closely resemble the configuration in the perceptual space (Figure 7A). Motion patterns presented from different viewpoints are clustered, and the centroids of those clusters closely match the cluster centers in the perceptual space, resulting in a small Procrustes distance between the two configurations ($d_P = 0.23$; cf. Table 2). This indicates that the perceptual metric may be based on the three-dimensional distances between joint trajectories.

However, this interpretation contradicts the result obtained by reconstruction of the three-dimensional configuration in the physical space from distances between three-dimensional joint trajectories: The recovered configuration (for $q = 1$ and $\lambda = 0$), aligned with a

Figure 8. Configurations in physical space for viewpoint Experiment 2. (A) Two-dimensional configuration constructed from distances (with optimized parameters: $\lambda = 0$, $q = 1$) between the two-dimensional joint trajectories in the image plane. (Conventions are as in Figure 7A.) (B) Two-dimensional configuration reconstructed from distances between three-dimensional joint trajectories. (C) Three-dimensional configuration in physical space reconstructed from the distances between the three-dimensional joint trajectories (red spheres), aligned with the stimulus configuration in the three-dimensional space (blue spheres) that is defined by the morphing weights and view angle.

three-dimensional space, which is defined by the morphing weights and the view angle, is shown in Figure 8C. Similar to the corresponding configuration in perceptual space (Figure 7B), the configuration in physical space shows clustering of different views of the same movement along the dimensions that are aligned with the morphing weights. However, when compared to Figure 8B the recovered configuration shows a clear ordering along the third dimension that is aligned with the view angle. This results in a very good alignment between the configurations in morphing space and physical space ($d_P = 0.23$; see Table 2). This good alignment along the view-angle dimension contrasts with the perceptual data (Figure 7B) where no such ordering was observed, causing significantly worse alignment of the two configurations (with $d_P = 0.51$; see above).

We also computed the alignment between the three-dimensional configurations in physical and perceptual space. Due to the differences between the two configurations along the view-angle dimension, the alignment error is somewhat larger than for the alignment of the configurations in physical and morphing space ($d_P = 0.26$; Table 2).

In summary, these results argue against the notion that the perceptual metrics can be adequately described by a simple measurement of distances between three-dimensional trajectories. Instead, a pooling over multiple view angles seems to be necessary in order to account for the experimental results. Whether such pooling is based on a view-based representation, or on the reconstruction of the three-dimensional movement trajectories and a subsequent elimination of one dimension in accordance with the task cannot be decided with our experiments.

Discussion

This study investigated the metric properties of perceptual representations of complex full-body movements. Stimuli were generated by motion morphing. Applying multidimensional scaling (MDS) to perceptual similarity judgments, we found that perceptual representations of movement patterns reside in perceptual spaces with well-defined metric properties. The stimulus configurations in such perceptual spaces closely resembled the true movement configurations in morphing weight space. Even better approximations were obtained by constructing physical spaces from distance measures based on joint trajectories. This implies that perceptual representations of complex body movements are *veridical* in that they closely reflect the metric of movements in the physical world. This finding was insensitive to the details of movement generation (e.g., based on morphing two- or three-dimensional trajectories) and to the manner of visual stimulus presentation (stick figures vs. point-light stimuli).

In addition, comparable results were obtained for two completely different experimental paradigms: comparison of pairs and delayed match-to-sample.

The veridicality of visual representations of complex body movements parallels similar results for static shape stimuli: In the past, it has been shown that shape spaces generated by morphing get mapped by subjects into perceptual shape spaces that reflect the original physical similarities between stimuli, as parameterized by the morphing weights (Cutzu & Edelman, 1996; Op de Beeck et al., 2001). Our results for motion patterns showed somewhat higher variability than experiments with static patterns. This might be explained by the increased difficulty of the judgments that required an integration of information over time.

In computer graphics, it has been known for a long time that visual representations of body movements can be generated by interpolation between motion-captured trajectories, parameterizing movement style by continuous morphing parameters (e.g., Unuma et al., 1995; Wiley & Hahn, 1997). Psychophysical studies employed such methods to investigate motion categorization (Giese & Lappe, 2002; Li, Ostwald, Giese, & Kourtzi, 2007), gender classification (Troje, 2002), and caricature and high-level after-effects (Giese, Knappmeyer, & Bülthoff, 2002; Hill & Pollick, 2000; Jordan, Fallah, & Stoner, 2006; Troje, Sadr, Geyer, & Nakayama, 2006). While those studies exploited the monotonic relationship between morphing parameters and perceptual dimensions, they did not reconstruct the metrics of the perceptual space.

One previous study applied MDS to perceptual judgments of arm movements, aiming at the reconstruction of a two-dimensional emotion space (Pollick, Paterson, et al., 2001). Interestingly, one of the recovered dimensions correlated with kinematic measures for the speed and duration of the movements. This result seems compatible with the veridicality of the visual representation of body movements. However, because movement time was normalized in our experiments, the observed structure similarity between perceptual and physical metric must have been based on more subtle features of body movements than total movement time or average speed.

Another set of related studies applied classifiers to joint trajectory data for gender and affect classification, comparing performance with psychophysical data from humans (Pollick, Lestou, Ryu, & Cho, 2002; Troje, 2002). These studies show that three-dimensional joint trajectories provide sufficient information for solving such classification tasks. In addition, human subjects typically performed less efficiently than such classifiers, indicating that they did not exploit all information that is available in the three-dimensional trajectories. This seems compatible with our result from Experiment 2, where the information about the view-angle dimension was not reflected in the reconstructed perceptual metric, even though it could be easily retrieved from the three-dimensional trajectory data, as shown by the construction of the three-dimensional

physical space (Figure 8C). Even if the agreement between human and classifier performance were more perfect, deriving specific constraints for multidimensional feature spaces from the classification errors of binary classifiers is extremely difficult due to the ill-posed nature of the underlying inverse problem. The present application of methods that directly address the metric properties of perceptual representations of body movements seems thus to be an important step towards a deeper understanding of the underlying computational mechanisms.

It seems likely that veridical perception of motion in the present study depended on the chosen prototypical patterns resulting in morphs that were all perceived as natural (Giese & Lappe, 2002). Choosing more dissimilar prototype movements might have resulted in unnatural-looking morphs, disrupting the continuity of the perceptual space. However, recent experiments suggest that even for very unnaturally looking patterns continuous perceptual representations can be learned after sufficient training (Jastorff, Kourtzi, & Giese, 2006).

Because the parameterization of the stimuli in terms of morphing weights depends on the morphing algorithm and on the choice of the prototype trajectories, we also compared the configurations in perceptual space with configurations in physical spaces that were constructed from various physical distance measures for movement trajectories. Testing a family of such measures, we found that good approximations of the configurations in perceptual space were obtained with trajectory metrics that (a) included time alignment of the compared trajectories; (b) contained no terms that measure directly the overall timing differences between trajectories (parameter $\lambda = 0$), implying that the distance (except for time alignment) depends only on spatial differences; and Equation 5 involved a robust average of the differences over the entire time course (parameter $q = 1$). The high similarity between configurations in perceptual and physical space (for optimized parameter values) shows that perceptual representations of complex body movements faithfully reflect the physical similarities of joint movement trajectories, if measured by the right distance measure.

The detailed comparison between physical distances derived from two- and three-dimensional trajectories suggests that measures derived from the trajectories in the image plane account well for the perceptual metrics in the absence of viewpoint changes. In the presence of viewpoint changes, such measures did not reproduce the perceptual metrics for the task that required observers to ignore the view angle. Likewise, distance measures derived from the three-dimensional trajectories did not capture the properties of the perceptual metrics because the recovered configurations faithfully represented the view angle, contrary to the psychophysical data. However, such distances captured the perceptual metrics well with appropriate pooling over multiple view directions. In the analysis of Experiment 2, such pooling was modeled by embedding the three-dimensional stimulus configuration

in the morph weight/view space into a two-dimensional physical space, resulting in an excellent reproduction of the perceptual metrics.

Integration of information over multiple views is a common element of many theories for the recognition of three-dimensional objects (Edelman, 1999; Perrett & Oram, 1998; Riesenhuber & Poggio, 1999) and may be relevant also for motion recognition (Giese & Poggio, 2003). Whether the recognition of body movements is based on the pooling of view-based two-dimensional representations or on the reconstruction of the full three-dimensional geometry (Aggarwal & Cai, 1999; Marr & Vaina, 1982; O'Rourke & Badler, 1980; Webb & Aggarwal, 1982) is unclear. Electrophysiological recordings have revealed biological motion-selective neurons in the STS that show view-dependence (Jellema & Perrett, 2006; Puce & Perrett, 2003). In addition, several psychophysical studies show view dependence of biological motion recognition (e.g., Bülthoff, Bülthoff, & Sinha, 1998; Jacobs & Shiffrar, 2005; Jokisch, Daum, & Troje, 2006; Troje, 2002), and it has been shown that the two-dimensional information in known views of actions can even override stereoscopically provided veridical depth information (Bülthoff et al., 1998). Yet, the available data seem insufficient to determine conclusively how two- vs. three-dimensional cues are integrated in biological motion and action recognition.

In the current studies, the best approximations of the perceptual metrics were obtained with physical distances that aligned the trajectories by time warping. This suggests that the perceptual system might include an efficient mechanism that compensates for timing differences between action patterns. Such mechanisms can be realized with relatively simple neural circuits (Giese & Poggio, 2003; Hopfield & Brody, 2001), and would contribute substantially to the robustness of action recognition since they result in generalization between actions with slightly different timings. Sequence alignment by time warping and related techniques are central in many engineered systems for the recognition of speech (Rabiner & Juang, 1993) and in computer vision algorithms for action recognition (e.g., Bobick, 1997; Veeraraghavan, Roy-Chowdhury, & Chellappa, 2005; Yacoob & Black, 1999). An interesting interpretation of the robustness against time warping is motivated by recent results in motor control that suggest a separate control of the spatial path and the timing of arm movements (Biess, Liebermann, & Flash, 2007). The visual system might thus have learned to categorize movements with the same path but different timings.

Apart from time alignment, the spatial differences in the physical distance measure (2) were critical for the reproduction of the perceptual metrics. A substantial influence of the spatial structure on the recognition of properties from biological motion is consistent with other studies, e.g., on person identification, that varied the amount of available shape and dynamic information (e.g.,

Troje, Westhoff, & Lavrov, 2005; Westhoff & Troje, 2007). In these studies, an influence of dynamic cues, like gait speed or the timing (parameterized by the phases of a Fourier representation) was also found. Effects of timing and spatial information were also found in studies varying the expressiveness of facial and body movements by spatial and temporal exaggeration (Hill & Pollick, 2000; Pollick, Fidopiastis, & Braden, 2001). In addition, movement speed has been shown to influence the accuracy of the perception of structure and identity (Barclay, Cutting, & Kozlowski, 1978; Beintema, Oleksiak, & van Wezel, 2006; Jacobs, Pinto, & Shiffrar, 2004), and is an important determinant for the expression of emotion by movements (Pollick, Paterson, et al., 2001; Sawada, Suda, & Ishii, 2003).

The minimal impact of overall timing differences in our experiments might be explained by the normalization of the total duration of the movements in our stimuli. This manipulation largely eliminates the information provided by the average movement speed. The remaining timing differences between the three prototypical locomotion patterns might be too small for the perceptual system to extract reliable information. Another study shows that subjects can learn to categorize motion patterns based just on their timing, keeping the spatial information constant by time warping (Li et al., 2007). In these experiments a substantial amount of time warping had to be applied in order to make the stimuli distinguishable for the participants. This indicates that the visual system might be relatively insensitive to overall timing changes that affect all parts of a moving figure synchronously.

In contrast to such low sensitivity to overall timing changes, the visual system does appear to be highly tuned to detect relative timing changes between different parts of a moving figure. For example, scrambling the phases of the individual dots of a point-light walker abolishes recognition of biological motion patterns (Bertenthal & Pinto, 1994). Such relative timing changes are not adequately modeled by time warping, and thus mainly affect the spatial difference term in Equation 4. For future studies, it seems an interesting question to study systematically the dependence of the perceptual metric on changes of the global vs. relative timing.

Finally, one might ask if our results also have implications for the ongoing debate as to whether biological motion recognition is based on form or motion features. Some psychophysical experiments suggest that local motion information might be central in the recognition of biological motion (Casile & Giese, 2005; Mather, Radford, & West, 1992; Thurman & Grossman, 2008), and several computational models show the feasibility of biological motion recognition from motion features (Giese & Poggio, 2003; Hoffman & Flinchbaugh, 1982; Little & Boyd, 1998; Webb & Aggarwal, 1982). At the same time, it is obvious that normal action stimuli, and even stick figures, contain substantial amounts of form information that can be exploited for recognizing action

from body shapes (Todd, 1983). Many models have been proposed that accomplish action recognition by recognizing temporal sequences of body shapes, either using three-dimensional shape models or two-dimensional form templates (e.g., Chen & Lee, 1992; Giese & Poggio, 2003; Hogg, 1983; Lange & Lappe, 2006; Marr & Vaina, 1982; O'Rourke & Badler, 1980; Rohr, 1994). It has been proposed that biological motion recognition integrates both motion and form information potentially at the level of the *superior temporal sulcus* patterns (Giese & Poggio, 2003; Peuskens, Vanrie, Verfaillie, & Orban, 2005). This view has been challenged by the alternative hypothesis that biological motion recognition exploits exclusively form information, local motion information being essentially irrelevant except for segmentation (Lange & Lappe, 2006). This alternative view seems difficult to reconcile with recent experiments demonstrating biological motion recognition from stimuli that prevent the extraction of form information from individual frames (Singer & Sheinberg, 2008), and the fact that the most informative features for the detection of point-light walkers seem to coincide with dominant motion features, rather than with the most informative body shapes (Casile & Giese, 2005; Thurman & Grossman, 2008). However, it reiterates the importance of the question how form and motion features influence the perceptual metric.

Can we say anything about this question based on the results discussed in this paper? The joint trajectories, on which our physical metric is based, specify both, form and local motion information, and both could thus have contributed to perceptual similarity judgments. As we did not apply specific methods for degrading either type of information, it seems difficult to draw strong conclusions their relative influence from our data. We should note, however, that we did find very similar results for stick-figure and point-light stimuli, even though these might be expected to weight form and motion features quite differently. In addition, neural modeling suggests that invariance against time warping and speed variations can be accounted for by mechanisms based on form, or based on local motion features (Giese & Poggio, 2003). While the methodology discussed in this paper might not be optimal for exploring this broader issue, it might be an interesting idea for future research to combine methods for the investigation of the perceptual metrics with specific techniques for varying the information conveyed by form and motion features (e.g., Beintema & Lappe, 2002; Beintema et al., 2006; Casile & Giese, 2005; Singer & Sheinberg, 2008).

More generally, we note that precise neural mechanisms that implement distance computation and time alignment and the neural level remain to be uncovered. Neurophysiological studies (e.g., Vangeneugden, Pollick, & Vogels, 2006) and functional imaging studies might benefit from the present methods for the analysis of perceptual and physical metrics for spatio-temporal patterns.

Appendix A

The recovered configurations in perceptual and physical space were aligned with the configuration in morphing space by a Procrustes transformation. Let $\mathbf{X}$ be the matrix of positions of the recovered configuration and $\mathbf{Y}$ be corresponding points in morphing space, where the coordinates of each point are given by the corresponding row of the matrix. The *orthonormal* Procrustes transformation determines a linear transformation defined by the scaling factor s and the orthonormal matrix $\mathbf{T}$ with $\mathbf{T}'\mathbf{T} = \mathbf{I}$ that minimizes the error function:

$$L(s, \mathbf{T}) = \min_{\mathbf{T},s} \| \mathbf{Y} - s\mathbf{X}\mathbf{T} \|_F, \tag{A1}$$

where $\| \ \|_F$ is the Frobenius norm. It can be shown (Borg & Groenen, 1997) that this optimum is given by the matrix $\mathbf{T} = \mathbf{U}\,\mathbf{V}'$. The last two matrices are defined by the singular value decomposition $\mathbf{Y}'\,\mathbf{X} = \mathbf{V}\,\Sigma\,\mathbf{U}'$ with the diagonal matrix Σ and two orthonormal matrices fulfilling $\mathbf{U}'\,\mathbf{U} = \mathbf{I}$ and $\mathbf{V}'\,\mathbf{V} = \mathbf{I}$.

For Experiment 2, the recovered three-dimensional configurations had to be aligned with a three-dimensional configuration in the stimulus space that was defined by the two-dimensional space of morphing weights and the view angle. Morphing weights and view angles are different physical dimensions that are related to each other by a scaling factor that is unknown a priori. An alignment using the classical Procrustes transformation that assumes equal weights of all dimensions would not be appropriate in this case. Instead we estimated the relative scaling of morphing weights and view-angle dimension together with the other parameters of the Procrustes transform. For this purpose, we minimized an error function of the form

$$L(s_1, s_2, \mathbf{T}) = \min_{\mathbf{T}} \left\| \mathbf{Y} - \mathbf{X} \begin{bmatrix} s_1\mathbf{I} & \mathbf{0} \\ \mathbf{0}' & s_2 \end{bmatrix} \cdot \mathbf{T} \right\|_F, \tag{A2}$$

with two separate scaling factors. This problem has no closed-form solution. We used an iterative procedure that combines a line search for the ratio $\gamma = (s_1/s_2)$ with the solution of an orthonormal Procrustes problem for γ fixed:

$$\begin{aligned} L(s_1, \mathbf{T}) &= \min_{\mathbf{T},s_1} \left\| \mathbf{Y} - s_1\mathbf{X} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}' & \gamma \end{bmatrix} \cdot \mathbf{T} \right\|_F \\ &= \min_{\mathbf{T}} \| \mathbf{Y} - s_1\widetilde{\mathbf{X}} \cdot \mathbf{T} \|_F. \end{aligned} \tag{A3}$$

This problem has the same form as Equation A1 and can be solved by singular value decomposition. The solution of this combined optimization problem defines a Procrustes transformation matrix is *orthogonal* but *not othonormal*.

References

Aggarwal, J. K., & Cai, Q. (1999). Human motion analysis: A review. *Computer Vision and Image Understanding, 73,* 428–440.

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review, 95,* 124–150.

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics, 23,* 145–152. [PubMed]

Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 5661–5663. [PubMed] [Article]

Beintema, J. A., Oleksiak, A., & van Wezel, R. J. (2006). The influence of biological motion perception on structure-from-motion interpretations at different speeds. *Journal of Vision, 6*(7):4, 712–726, http://journalofvision.org/6/7/4/, doi:10.1167/6.7.4. [PubMed] [Article]

Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. *Psychological Science, 5,* 221–225.

Biess, A., Liebermann, D. G., & Flash, T. (2007). A computational model for redundant human three-dimensional pointing movements: Integration of independent spatial and temporal motor plans sim-

plifies movement dynamics. *Journal of Neuroscience, 27,* 13045–13064. [PubMed] [Article]

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology, 58,* 47–73. [PubMed]

Bobick, A. F. (1997). Movement, activity and action: The role of knowledge in the perception of motion. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 352,* 1257–1265. [PubMed] [Article]

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling.* New York: Springer-Verlag.

Bruderlin, A., & Williams, L. (1995). Motion signal processing. *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques* (pp. 97–104). New York: ACM Press.

Bülthoff, I., Bülthoff, H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nature Neuroscience, 1,* 254–257. [PubMed]

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via and N-way generalization of the Eckart–Young decomposition. *Psychometrika, 35,* 283–319.

Casile, A., & Giese, M. A. (2005). Critical features for the recognition of biological motion. *Journal of Vision, 5*(4):6, 348–360, http://journalofvision.org/5/4/6/, doi:10.1167/5.4.6. [PubMed] [Article]

Chen, Z., & Lee, H.-J. (1992). Knowledge-guided visual perception of 3-D human gait from a single image sequence. *IEEE Transactions on Systems, Man and Cybernetics, 22,* 336–342.

Cutzu, F., & Edelman, S. (1996). Faithful representation of similarities among three-dimensional shapes in human vision. *Proceedings of the National Academy of Sciences of the United States of America, 93,* 12046–12050. [PubMed] [Article]

Cutzu, F., & Edelman, S. (1998). Representation of object similarity in human vision: Psychophysics and a computational model. *Vision Research, 38,* 2229–2257. [PubMed]

Decety, J., & Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences, 3,* 172–178. [PubMed]

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception, 25,* 727–738. [PubMed]

Edelman, S. (1999). *Representation and recognition in vision.* Cambridge, MA: MIT Press.

Efron, B., & Tibshirani, R. (1993). *An introduction to the Bootstrap.* New York, London: Chapman and Hall.

Giese, M. A., Knappmeyer, B., & Bülthoff, H. H. (2002). Automatic synthesis of sequences of human movements by linear combination of learned example patterns. In H. H. Bülthoff, S. W. Lee, T. Poggio, & C. Wallraven (Eds.), *Biologically motivated computer vision* (pp. 538–547). Berlin: Springer.

Giese, M. A., & Lappe, M. (2002). Measurement of generalization fields for the recognition of biological motion. *Vision Research, 42,* 1847–1858. [PubMed]

Giese, M. A., & Poggio, T. (2000). Morphable models for the analysis and synthesis of complex motion patterns. *International Journal of Computer Vision, 38,* 59–73.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements [Abstract]. *Nature Reviews Neuroscience, 4,* 179–192. [PubMed]

Giese, M. A., Thornton, I. M., & Edelman, S. E. (2003). Metric category spaces of biological motion. *Journal of Vision, 3*(9):83, 83a, http://journalofvision.org/3/9/83/, doi:10.1167/3.9.83.

Gleicher, M. (2001). Comparing constraint-based motion editing methods. *Graphical Models, 63,* 107–134.

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35,* 1167–1175. [PubMed] [Article]

Hill, H., & Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Science, 11,* 223–228. [PubMed]

Hoffman, D. D., & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics, 42,* 195–204. [PubMed]

Hogg, D. (1983). Model-based vision: A program to see a walking person. *Image and Vision Computing, 1,* 5–20.

Hopfield, J. J., & Brody, C. D. (2001). What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 1282–1287. [PubMed] [Article]

Ilg, W., Bakir, G. H., Mezger, J., & Giese, M. A. (2004). On the representation, learning and transfer of spatio-temporal movement characteristics. *International Journal of Humanoid Robotics, 1,* 613–636.

Jacobs, A., Pinto, J., & Shiffrar, M. (2004). Experience, context, and the visual perception of human movement. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 822–835. [PubMed]

Jacobs, A., & Shiffrar, M. (2005). Walking perception by walking observers. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 157–169. [PubMed]

Jastorff, J., Kourtzi, Z., & Giese, M. A. (2006). Learning to discriminate complex movements: Biological versus artificial trajectories. *Journal of Vision, 6* (8):3, 791–804, http://journalofvision.org/6/8/3/, doi:10.1167/ 6.8.3. [PubMed] [Article]

Jellema, T., & Perrett, D. I. (2006). Neural representations of perceived bodily actions using a categorical frame of reference. *Neuropsychologia, 44,* 1535–1546. [PubMed]

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14,* 201–211.

Jokisch, D., Daum, I., & Troje, N. F. (2006). Self recognition versus recognition of others by biological motion: Viewpoint-dependent effects. *Perception, 35,* 911–920. [PubMed]

Jordan, H., Fallah, M., & Stoner, G. R. (2006). Adaptation of gender derived from biological motion. *Nature Neuroscience, 9,* 738–739. [PubMed]

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics, 21,* 575–580.

Kreyszig, E. (1989). *Introductory functional analysis with applications.* New York: Wiley.

Lange, J., & Lappe, M. (2006). A model of biological motion perception from configural form cues. *Journal of Neuroscience, 26,* 2894–2906. [PubMed] [Article]

Lee, J., & Shin, S. Y. (1999). A hierarchical approach to interactive motion editing for human-like figures. *Proceedings of the 26th annual conference on computer graphics and interactive techniques* (pp. 39–48). New York: ACM Press.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature, 442,* 572–575. [PubMed]

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience, 4,* 89–94. [PubMed] [Article]

Li, S., Ostwald, D., Giese, M., & Kourtzi, Z. (2007). Flexible coding for categorical decisions in the human brain. *Journal of Neuroscience, 27,* 12321–12330. [PubMed] [Article]

Little, J. J., & Boyd, J. E. (1998). Recognizing people by their gait: The shape of motion. *Videre, 1,* 2–32.

Marr, D., & Vaina, L. (1982). Representation and recognition of the movements of shapes. *Proceedings of the Royal Society B: Biological Sciences, 214,* 501–524. [PubMed]

Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society B: Biological Sciences, 249,* 149–155. [PubMed]

Mezger, J., Ilg, W., & Giese, M. A. (2005). Trajectory synthesis by hierarchical spatio-temporal correspondence: Comparison of different methods. *Proceedings of the 2nd symposium on Applied Perception in Graphics and Visualization* (pp. 25–32). New York: ACM Press.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 1992,* 25–53.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience, 4,* 1244–1252. [PubMed]

O'Rourke, J., & Badler, N. (1980). Model-based image analysis of human motion using constraint propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2,* 522–536.

Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research, 31,* 1337–1350. [PubMed]

Perrett, D. I., & Oram, M. W. (1998). Visual recognition based on temporal cortex cells: Viewer-centred processing of pattern configuration. *Zeitschrift für Naturforschung C: Journal of Biosciences, 53,* 518–541. [PubMed]

Peuskens, H., Vanrie, J., Verfaillie, K., & Orban, G. A. (2005). Specificity of regions processing biological motion. *European Journal of Neuroscience, 21,* 2864–2875. [PubMed]

Pollick, F. E., Fidopiastis, C., & Braden, V. (2001). Recognising the style of spatially exaggerated tennis serves. *Perception, 30,* 323–338. [PubMed]

Pollick, F. E., Lestou, V., Ryu, J., & Cho, S. B. (2002). Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research, 42,* 2345–2355. [PubMed]

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition, 82,* B51–B61. [PubMed]

Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 358,* 435–445. [PubMed] [Article]

Rabiner, L., & Juang, B. H. (1993). *Fundamentals of speech recognition.* Englewood Cliffs, NJ: PTR Prentice Hall.

Rensink, R. A. (1990). Toolbox-based routines for Macintosh timing and display. *Behavior Research Methods, Instruments, & Computers, 22,* 105–117.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19,* 473–497. [PubMed]

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025. [PubMed]

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27,* 169–192. [PubMed]

Rohr, K. (1994). Towards model-based recognition of human movements in image sequences. *Computer Vision, Graphics, and Image Processing. Image Understanding, 59,* 95–115.

Rose, C., Cohen, M. F., & Bodenheimer, B. (1998). Verbs and adverbs: Multidimensional motion interpolation. *IEEE Computer Graphics and Applications, 18,* 32–40.

Sawada, M., Suda, K., & Ishii, M. (2003). Expression of emotions in dance: Relation between arm movement characteristics and emotion. *Perceptual and Motor Skills, 97,* 697–708. [PubMed]

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323. [PubMed]

Shepard, R. N., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology, 1,* 1–17.

Singer, J., & Sheinberg, D. L. (2008). A method for the real-time rendering of formless dot field structure-from-motion stimuli. *Journal of Vision, 8*(5):8, 1–8, http://journalofvision.org/8/5/8/, doi:10.1167/8.5.8.

Steinman, S. B., & Nawrot, M. (1992). Real-time color-frame animation for visual psychophysics on the Macintosh computer. *Behavior Research Methods, Instruments, & Computers, 24,* 439–452.

Sugihara, T., Edelman, S., & Tanaka, K. (1998). Representation of objective similarity among three-dimensional shapes in the monkey. *Biological Cybernetics, 78,* 1–7. [PubMed]

Thurman, S. M., & Grossman, E. D. (2008). Temporal "Bubbles" reveal key features in point-light biological motion perception. *Journal of Vision, 8(3):28,* 1–11, [PubMed] [Article]

Todd, J. T. (1983). Perception of gait. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 31–42. [PubMed]

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision, 2*(5):2, 371–387, http://journalofvision.org/2/5/2/, doi:10.1167/2.5.2. [PubMed] [Article]

Troje, N. F., Sadr, J., Geyer, H., & Nakayama, K. (2006). Adaptation aftereffects in the perception of gender from biological motion. *Journal of Vision, 6*(8):7, 850–857, http://journalofvision.org/6/8/7/, doi:10.1167/6.8.7. [PubMed] [Article]

Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Perception & Psychophysics, 67,* 667–675. [PubMed]

Unuma, M., Anjyo, K., & Takeuchi, R. (1995). Fourier principles for emotion-based human figure animation. Motion signal processing. *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques* (pp. 91–96). New York: ACM Press.

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 11656–11661. [PubMed] [Article]

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology A, 43,* 161–204. [PubMed]

Vangeneugden, J., Pollick, F., & Vogels, R. (2006). *Responses of macaque superior temporal sulcus neurons to a parameterized set of dynamic images of visual actions.* Paper presented at the Society for Neuroscience 36th Annual meeting, Atlanta, GA.

Veeraraghavan, A., Roy-Chowdhury, A. K., & Chellappa, R. (2005). Matching shape sequences in video with applications in human movement analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*, 1896–1909. [PubMed]

Webb, J. A., & Aggarwal, J. K. (1982). Structure from motion of rigid and jointed objects. *Artificial Intelligence, 19,* 107–130.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature, 428,* 557–561. [PubMed]

Westhoff, C., & Troje, N. F. (2007). Kinematic cues for person identification from biological motion. *Perception & Psychophysics, 69,* 241–253. [PubMed]

Wiley, D. J., & Hahn, J. K. (1997). Interpolation synthesis of articulated figure motion. *IEEE Computer Graphics and Applications, 17,* 39–45.

Witkin, A., & Popovic, Z. (1995). Motion warping. *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques* (pp. 105–108). New York: ACM Press.

Yacoob, Y., & Black, M. J. (1999). Parameterized modeling and recognition of activities. *Journal of Computer Vision and Image Understanding, 73,* 232–247.