Decomposing biological motion: A framework for analysis and synthesis of human gait patterns

Nikolaus F. Troje

Ruhr-Universität, Bochum, Germany

Biological motion contains information about the identity of an agent as well as about his or her actions, intentions, and emotions. The human visual system is highly sensitive to biological motion and capable of extracting socially relevant information from it. Here we investigate the question of how such information is encoded in biological motion patterns and how such information can be retrieved. A framework is developed that transforms biological motion into a representation allowing for analysis using linear methods from statistics and pattern recognition. Using gender classification as an example, simple classifiers are constructed and compared to psychophysical data from human observers. The analysis reveals that the dynamic part of the motion contains more information about gender than motion-mediated structural cues. The proposed framework can be used not only for analysis of biological motion but also to synthesize new motion patterns. A simple motion modeler is presented that can be used to visualize and exaggerate the differences in male and female walking patterns.

Keywords: gender classification, recognition, social recognition, animate motion

Introduction

The human visual system is extremely sensitive to animate motion patterns. We quickly and efficiently detect another living being in a visual scene, and we can recognize many aspects of biological, psychological, and social significance. Human motion, for instance, contains a wealth of information about the actions, intentions, emotions, and personality traits of a person. What our visual system seems to solve so effortlessly is still a riddle in vision research and an unsolved problem in computer vision. Little is known about exactly how biologically and psychologically relevant information is encoded in visual motion patterns. This study aims to provide a general framework that can be used to address this question. The approach is based on transforming biological motion data into a representation that subsequently allows for analysis using linear statistics and pattern recognition. To demonstrate the potential of this framework, we construct a sex classifier and compare its performance with the performance of human observers that classify the same stimuli.

Some 30 years ago, Gunnar Johansson (1973, 1976) introduced to experimental psychology a visual stimulus display designed to separate biological motion information from other sources of information that are normally intermingled with motion information. Johansson attached small point lights to the main joints of a person's body and filmed the scene so that only the lights were visible in front of an otherwise homogeneously dark background. Using these displays, he demonstrated the compelling power of perceptual organization from biological motion of just a few light points.

A large number of studies have since used Johansson's point-light displays. It has been demonstrated that biological motion perception goes far beyond the ability to recognize a set of moving dots as a human walker. Point-light displays contain enough information to recognize other actions as well (Dittrich, 1993), to determine the gender of a person (Barclay, Cutting, & Kozlowski, 1978; Hill & Johnston, 2001; Kozlowski & Cutting, 1977; Mather & Murdoch, 1994, Runeson, 1994), to recognize emotions (Dittrich, Troscianko, Lea, & Morgan, 1996; Pollick, Paterson, Bruderlin, & Sanford, 2001), to identify individual persons (Cutting & Kozlowski, 1977; Hill & Pollick, 2000), and even one's own walking pattern (Beardsworth & Buckner, 1981). However, whereas many studies exist that demonstrate the capability of the human visual system to detect, recognize, and interpret biological motion, there have been virtually no attempts to solve the question of how information about the moving person is encoded in the motion patterns. Only for gender recognition are there a few investigations addressing the nature of the informational content mediating this ability. In this study, we will also use gender classification of walking patterns as an example. However, the proposed framework can be generalized to solve other pattern classification problems based on biological motion.

One way to approach the question of where diagnostic information is hidden in a sensory stimulus is through psychophysical experiments. In such studies, the stimulus is manipulated along different dimensions in order to measure the effect of such manipulations on recognition performance. The first study on gender recognition from biological motion was conducted by Kozlowski and Cutting (1977). They demonstrated that

DOI 10:1167/2.5.2 Received January 24, 2002; published September 10, 2002 ISSN 1534-7362 © 2002 ARVO

observers are able to classify point-light walkers shown in saggital view with a performance of 63% correct recognition. Additionally, they introduced a number of manipulations: increased or reduced arm swing amplitudes, unnaturally fast or slow walking speeds, and occlusion of either the lower or the upper part of the body. All manipulations considerably reduced recognition performance. With unnatural arm swings, performance dropped almost to chance level. Showing only the lower body impaired recognition to a larger extent than showing only the upper body. None of the manipulations caused a shift in perception into a defined direction, making the percept either more male or more female. Only for the speed manipulation did there seem to be a trend to perceive fast walkers more female, which, however, did not reach a statistically significant level.

Barclay et al. (1978) conducted a similar study investigating the influence of four different parameters. The initial experiment focused on the influence of exposure duration. The results show that two complete gait cycles are required to determine gender from biological motion. Shorter exposure times result in reduced performance. In a second experiment, speed was altered, but rather than recording different walking speeds from the model walkers as in Kozlowski and Cutting's (1977) study, they used just one recording showing a walker at his most comfortable walking speed and presented this stimulus with different play-back speeds to the observers. This manipulation had a strong effect and gender recognition was almost at chance level. The third manipulation consisted of blurring the discrete dots of the point-light walker to such an extent that the walker appeared as a single blob that changed shape during walking. This caused gender recognition performance to decrease also to chance level. Finally, the authors tested gender recognition with walkers that were presented upside-down. Interestingly, in this case, recognition performance dropped significantly below chance. If a female walker was turned upside down, the display tended to be perceived as a man and an inverted man tended to be perceived as a woman. Whereas all other manipulations only resulted in a general decrease in recognition performance, inversion of a point-light walker clearly induced defined shifts in perceived gender.

Barclay et al. (1978) proposed that their finding was due to the fact that the ratio of shoulder width and pelvis width differ between men and women. Men tend to have wider shoulders than hips, whereas this ratio is reversed in women. If, upon inversion, the walker's shoulders are seen as if they were hips and the hips are seen as if they were shoulders, then observers' responses would reverse with respect to the true gender of the walker. Given this scenario, the question remains how shoulder and hip width could be measured. Because the walker was presented in a side view, neither shoulder nor hip width could be determined directly from the stimulus. However, due to a torsional twist of the upper body, both shoulder and hip perform elliptical motions in the saggital plain. The amplitude of those ellipses depends on the widths of shoulder and pelvis, and, therefore, may have provided a diagnostic cue.

If the extent of movement at the shoulder and the hip is an important cue for gender recognition, artificial walkers that differ only in those attributes should be classified accordingly. Cutting (1978a, 1978b, 1978c) developed a generative model of human gait and showed that this is indeed the case. The isolated cue apparently provided diagnostic information about a walker's gender.

However, biological motion contains more information that can serve for gender classification. In principle, biological motion can provide two sources of information. One is motion-mediated structural information, and the second is truly dynamic information. In contrast to a static frame of a point-light walker, motion reveals the articulation of the body. Setting a point-light walker into motion immediately uncovers information about which segments are rigid, where the joints are located, and, therefore, about the lengths of the connecting segments. The resulting information is structural, static information about the geometry of the body. Motion is only needed as a medium to obtain this information and could be replaced by other cues. A static view of a point-light walker in which the connections are explicitly drawn (stick figure) combined with information to disambiguate the 2D projection (e.g., using stereo displays) would, in principle, provide the same information.

In addition to motion-mediated structural information, biological motion also contains truly dynamic information. The amplitude and velocity of the arm swing or the torsion of the trunk are simple examples for information that is clearly different from structural information. It should be noted, however, that although representing two different sources of information, structural and dynamic information might not be independent. The amplitude of the elliptical motion of shoulders and hips as a function of the respective widths, as discussed above, provides an illustrative example of this fact.

The role of motion-mediated structural information and dynamic cues for gender recognition from biological motion was explicitly addressed in a series of experiments conducted by Mather and Murdoch (1994). The static cue they concentrated on was the ratio of the width of the hip and the width of the shoulder. The dynamic cue that was manipulated differed from the one used by Cutting (1978b). Whereas Cutting emphasized differences in motion of hips and shoulders in the saggital plane, Mather and Murdoch focused on differences in lateral body sway. Men show a larger extent of lateral sway of the upper body than women do (Murray, Kory, & Sepic, 1970). Mather and Murdoch (1994) generated stimuli that showed artificial point-light walkers with well-defined structural measures (shoulder and hip width) and well-defined dynamic cues (lateral sway of shoulder and hip). The walkers were shown from different viewing angles and subjects had to indicate the perceived gender. Setting structural and dynamic cues into conflict, the authors could show that the dynamic cue clearly dominated the structural cue.

In summary, the different studies on gender recognition from biological motion show that information about a walker's gender is not a matter of a single feature. Barcley et al. (1978), as well as Cutting (1978a), have identified the elliptical motion of shoulder and hip in the saggital plain to be an important cue to gender. Mather and Murdoch (1994) focused on the extent of lateral body sway. Kozlowski and Cutting (1977) showed that seeing only parts of the body could provide enough information about the gender of its owner to yield classification performances above chance. Gender recognition appears to be a complex process with a holistic character that takes into consideration hints and cues that are distributed over the whole display and that are carried both by motion-mediated structural information and by pure dynamics. Other studies employing different tasks confirm the holistic nature of biological motion perception (Bertenthal & Pinto, 1994; Lappe & Beintema, 2002).

Most of the studies summarized above aimed to investigate particular properties of the stimulus that were suspected to be promising candidates to carry information for gender discrimination. The role of such stimulus properties for gender recognition was, in turn, scrutinized by means of psychophysical experiments. In this study, we chose a different approach to the question of how information is encoded in biological motion patterns. Here we want to treat the problem as a pattern-recognition problem. With no a priori assumptions about possible candidate cues, we attempted to construct a linear classifier that can discriminate male from female walking patterns. We can then, in turn, scrutinize the classifier to determine which cues have been used. The cues may be simple features or complex holistic cues that are described in terms of correlation patterns between different parts and motions of the body. Any attribute or combination of attributes that changes when moving along an axis perpendicular to the separation plane defining the classifier is diagnostic for gender classification. Attributes that change while moving within the separation plane do not contribute any information to the gender classification problem.

A prerequisite to generate a linear classifier for gender discrimination or other stimulus features from human motion is a data structure within which linear operations are effectively applicable. The problem is similar to attempts to construct linear models of classes of images. In the domain of object recognition and human face recognition, such representations have been termed "linear object classes" (Vetter, 1998; Vetter & Poggio, 1997) or "morphable models" (Giese & Poggio, 2000; Jones & Poggio, 1999; Shelton, 2000). The latter term expresses the fact that the linear transition from one item to another represents a well-defined smooth metamorphosis between the items. Another term that has been used for the same class of models in the context of human face recognition is "correspondence based representations" (Troje & Vetter, 1998; Vetter & Troje, 1997). This term focuses on morphable models' reliance on establishing correspondence between features across the data set, resulting in a separation of the overall information into range-specific information on the one hand and domain specific information on the other hand (Ramsay & Silverman, 1997).

The use of linear techniques to describe human motion data has been employed in a number of studies, both in computer vision and in animation. Some of these techniques focus on recognition of actions and blending between actions. Others concentrate on the recognition and generation of emotion and other stylistic features within a set of instances of an action. In the context of this study, we want to define an action as a set of motion instances that are structurally similar. Extrapolating Alexander's (1989) definition of a gait, we define an *action* as a pattern of motion characteristics described by quantities of which one or more change discontinuously at transitions to other actions. Instances of the same action can be smoothly transformed into each other, with all transitions being valuable representations of the particular action. The definition implies structural similarity between instances of the same action, and, therefore, a means to define correspondence in space and time between two or more instances in a canonical and unambiguous way. Systematic differences between motion instances of an action are referred to as *styles*. Styles can correspond to emotions, personality or biological features, such as age or gender. According to the above definitions, the stylistic state space of an action is expected to be continuous and therefore defines smooth transitions between all instances of an action. Warping between actions, in contrast, requires the definition of additional constraints in order to achieve unambiguous correspondence.

Most of the existing systems for recognition, classification, synthesis, and editing of biological motion are based on data representations with a continuous smooth behaviour. A number of different techniques have been used to achieve this behaviour. Brand and Hertzmann's (2000) "style machines" are based on a hidden Markov model, that is, a probabilistic finite-state machine consisting of a set of discrete states, state-to-state transition probabilities, and state-to-signal emission probabilities (see also Wilson & Bobick, 1995). Rose, Bodenheimer, and Cohen (1998) presented a model using radial basis functions and low-order polynomials that both provide blending between actions and interpolation within stylistic state spaces. A number of models are based on frequency domain manipulations. Fourier techniques (Davis, Bobick, & Richards, 2000; Davis, 2001; Unuma, Anjyo, & Takeuchi, 1995; Unuma & Takeuchi, 1993) are suitable for periodic motions, such as locomotion patterns. Multiresolution filtering (Bruderlin & Williams, 1995) applies to a wider spectrum of movements but is restricted to modify and edit existing motion, rather than creating new motions through interpolation between existing motions. If the latter is required, multiresolution filtering has to be combined with time-warping techniques (Witkin & Popovic, 1995). Time warps are required to align corresponding signal features in time. Depending on the complexity of the action, time warps are parameterized in terms of simple uniform scaling and translation (e.g., Wiley & Hahn, 1997; Yacoob & Black, 1997) by using nonlinear models, such as B-splines (Ramsay & Li, 1989; Ramsay & Silverman, 1997), or by fitting nonparametric models by means of dynamic programming (Bruderlin & Williams, 1995; Giese & Poggio, 1999; 2000).

The dimensionality of the resulting linear spaces are not necessarily reflecting the number of degrees of freedom within the set of represented data. Some of the above cited techniques therefore use principal components analysis (PCA) to reduce the dimensionality to a degree that stands in a reasonable relation to the size of the available data set. PCA can be used on different levels. For instance, Yacoob and Black (1997) apply PCA to a set of "atomic activities," which are registered in time and then represented by concatenating all measurements (joint angles) of all frames of the sequence. Ormoneit, Sidenbladh, Black, and Hastie (2000) use a similar approach (see also Bobick, 1997; Ju, Black, & Yacoob, 1996; Li, Dettmer, & Shah, 1997). Rosales and Scarloff (2000) apply PCA to a set of postures, each posture being represented only by measurements of a single frame.

Linear motion models have been applied to a number of different problems, such as motion editing (Brand & Hertzmann, 2000; Bruderlin & Williams, 1995; Gleicher, 1998; Guo & Roberge, 1996; Wiley & Hahn, 1997), retargeting motion from one character to another (Gleicher, 1998), tracking a human figure from video data (Ju et al., 1996; Ormoneit et al., 2000; Rosales & Scarloff, 2000), recognizing activities (Yacoob & Black, 1997), speech (Li et al., 1997) or gait patterns (Giese & Poggio, 2000). Giese and Poggio's model, which is in many respects similar to ours, is able to discriminate between different gaits (running and walking), but also to discriminate limping from walking. Whereas running and walking have to be considered two different actions according to the above definition, limping and walking are two styles of the same action. Other than this work, Davis' (2001) work on visual categorization of children and adult walking styles is the only one that we are aware of that applies linear motion modelling to the recognition of stylistic aspects within an action.

Although linear motion models have become common within the animation and computer vision community, there exist only few studies that use such models for psychological studies on motion perception. An exception is the work by Hill, Pollick, and colleagues (Hill & Pollick, 2000; Pollick, Fidopiastis, & Braden, 2001). Both studies show that extrapolations in linear motion spaces are perceived as caricatured instances that are recognized even better than the original sequences. The results imply that the topology of perceptual spaces used for biological motion recognition is similar to the one implicit in artificial linear motion spaces that are based on a distinction between range-specific information on the one hand and domain-specific information on the other hand.

Our approach to linearize human walking data employs many of the techniques summarized above. Starting with motion capture data from a number of human subjects, we first reduce the dimensionality of each subject's set of postures using PCA in a way similar to that described by Rosales and Scarloff (2000). This results in a low-dimensional space spanned by the first few eigenpostures. As postures change during walking, the corresponding coefficients change sinusoidally. The temporal behaviour of the sequence is well described by simple sine functions, and the decomposition becomes very similar to previous work on Fourier decomposition of walking data (Unuma et al., 1995). The eigenposture approach, however, is more general because it is not based in the frequency domain and thus can be used for nonperiodic motions as well. The main difference is that time warping, which reduces to simple uniform scaling in the case of our walking data, has to be parameterized using a more complex model.

Based on the outlined linearization of biological motion data, we are primarily interested to recognize and characterize stylistic features within an action. The action we are using is human walking. The stylistic variations we are investigating are the differences relating to the walker's gender. The aim of this study is twofold. First, we want to quantitatively characterize the differences in walking style between men and women. We test the success of our approach in terms of a linear classifier operating on the proposed linear representation of a set of human walking data. Second, we compare the performance of the linear classifier to the performance of human observers in a gender classification task. By depriving both the linear classifier as well as our human observers from parts of the information contained in the walking patterns, we want to find out which aspects of the stimulus are diagnostic and relevant for solving the gender classification task.

Walking Data

Twenty men and 20 women, most of them students and staff of the psychology department at the Ruhr-University, served as models to acquire motion data. Their ages ranged from 20 to 38 years (average age, 26 years). A set of 38 retroreflective markers was attached to their body. Participants wore swimming suits and most of the markers were attached directly onto the skin. Others, such as the ones for the head, the ankles, and the wrists, were attached to elastic bands, and the ones on the feet were taped onto the subjects' shoes.

Participants were then requested to walk on a treadmill. They could adjust the speed of the belt so that they felt most comfortable. To ensure that they did not feel too much under observation and that they did not "perform" in an unnatural manner, we let them walk for at least 5 min before we started to record 20 steps (i.e., 10 full-gait cycles) from each of them. Participants were not notified when recording started.

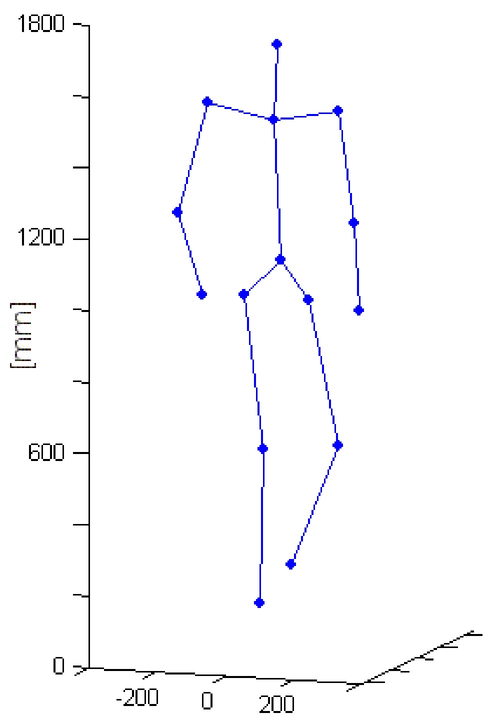


Figure 1. The movie illustrates the 15 marker positions used in the computations. The markers are located at the major joints of the body (shoulders, elbows, wrists, hips, knees, ankles), the sternum, the center of the pelvis, and the center of head.

Data were recorded using a motion capture system (Vicon; Oxford Metrics, Oxford, UK) equipped with 9 CCD high-speed cameras. The system tracks the three-dimensional trajectories of the markers with spatial accuracy in the range of 1 mm and a temporal resolution of 120 Hz.

From the trajectories of the 38 original markers, we computed the location of "virtual" markers positioned at major joints of the body. The 15 virtual markers used for all the subsequent computations were located at the joints of the ankles, the knees, the hips, the wrists, the elbows, the shoulders, at the center of the pelvis, on the sternum, and in the center of the head (Figure 1). Commercially available software (BODYBUILDER, Oxford Metrics) for biomechanical modeling was used to achieve the respective computations.

The Algorithm

The walk of an individual subject can be regarded as a time series of postures. Each posture can be specified in terms of the positions of the 15 markers. Because three coordinates are needed for each marker's position, the representation of a posture is a 45-dimensional vector $\mathbf{p}=(m1_x, m1_y, m1_z, m2_x \ldots m15_z)^T$ (we take the transpose because we regard $\mathbf{p}$ to be a column vector).

A walker needs about 12 s to perform 20 steps, thus providing about 1,400 single postures. Of course, this set of postures is highly redundant. For instance, if the left wrist is in front of the torso, it is very likely that the right foot is also in front of the torso, whereas the right wrist and the left foot are both behind it.

One way to capture redundancy within a data set is principal components analysis. PCA is a linear basis transformation that basically decomposes the original data so that any number of components accounts for as much as possible of the data's variance. Mathematically, the principal components are the eigenvectors of the covariance matrix of the original data set. The corresponding eigenvalues express the variance covered by the individual components. Redundancy in a data set means that the data occupy only a part of the space. PCA can capture the redundancy only in cases in which the data lie within a low-dimensional linear subspace of the original space. If they are occupying a low-dimensional but still nonlinear manifold, PCA will not be able to recover all of the redundancy within the data set.

For the moment, let us consider only a single walker. The data of a particular walker consist of about 1,400 postures sampled while the walker performed 10 gait cycles. We applied PCA separately to the postures of each walker. On average, across all 40 walkers, the first principal component already covers 84% of the overall variance. The first four principal components taken together account for more than 98% of the overall variance (Figure 2). Apparently, PCA is very successful in capturing the redundancy in the data. Each posture $\mathbf{p}$ can be described as a linear combination of the average posture $\mathbf{p}_0$ plus a weighted sum of the first four PCs

$$\mathbf{p} = \mathbf{p}_0 + \sum_i c_i \mathbf{p}_i \tag{1}$$

with $\mathbf{p}_i$ denoting the ith principal component and c_i denoting the respective score.

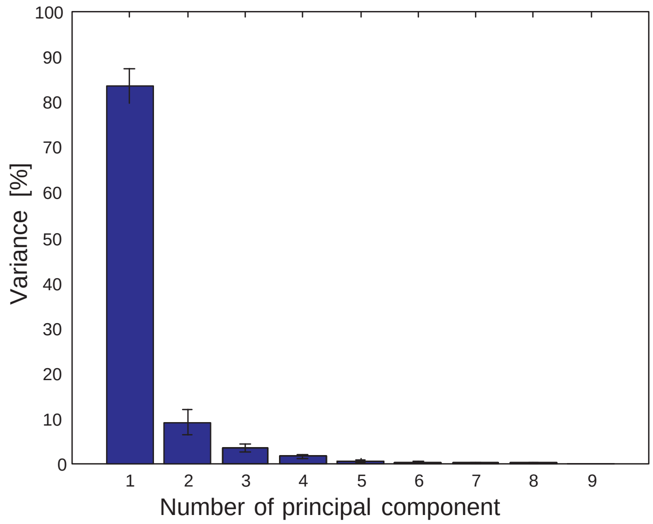


Figure 2. The variance covered by the first few eigenpostures. The bars represent the mean (with standard deviations shown) across all 40 walkers.

In order to distinguish the outcome of this analysis from a second PCA that is introduced later in this study, we call the principal components as derived from an analysis across postures the “eigenpostures” of a particular walker. Given the mean posture and the first four eigenpostures, each posture can now be described simply by the four weights. Note that the eigenpostures are specific for each walker.

Walking is a time series of postures. If we can model the temporal behavior of the first four components, we have modeled the walk. In fact, the temporal behavior of the components is very simple and can be nicely modeled with pure sine functions (Figure 3). On average, across all walkers, the quality of a simple sinusoidal fit as given by the coefficients of determination is 0.99, 0.95, 0.94, and 0.90 for the first four eigenpostures, respectively. Each sine function is characterized by its frequency, its amplitude, and its phase. The frequency of the first two PCs always equates the fundamental frequency of the walking and the frequency of the third and fourth PC is the second harmonic. The amplitudes are just scaling factors that can be multiplied with the PCs. What remain are the phases. Because we are interested only in the relative phases of the PCs, we set the phase of the first PC to be zero and change the phases of the other components accordingly.

To fully describe the walk of a single walker, we now need the average posture $\mathbf{p}_0$, the first four eigenpostures $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, $\mathbf{p}_4$, the fundamental frequency ω, and the phases of the second, third, and fourth PC with respect to the first component, φ_2, φ_3, and φ_4:

$$\begin{aligned}\mathbf{p}(t) &= \mathbf{p}_0 + \mathbf{p}_1 \sin(\omega t) + \mathbf{p}_2 \sin(\omega t + \varphi_2) \\ &\quad + \mathbf{p}_3 \sin(2\omega t + \varphi_3) + \mathbf{p}_4 \sin(2\omega t + \varphi_4).\end{aligned} \tag{2}$$

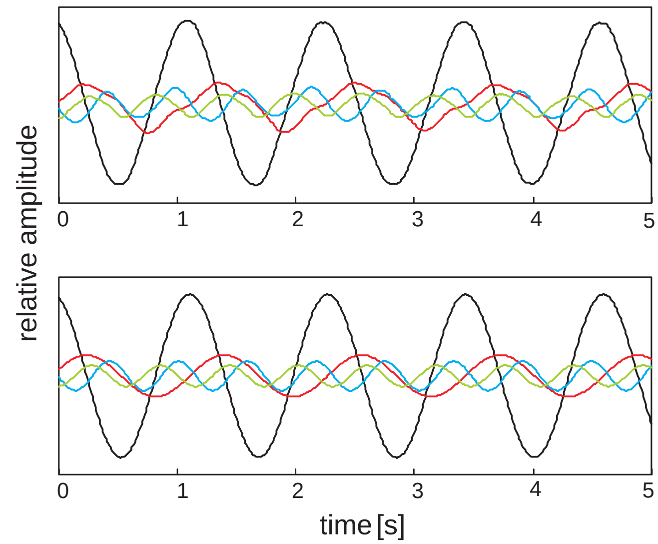


Figure 3. The upper panel shows the coefficients of the first four eigenpostures changing over time for 600 frames of a single walker. The lower panel is the corresponding fit using sine functions. The coefficients of determination of the fits for this particular walker are 0.99, 0.95, 0.94, and 0.94 for the first four eigenpostures, respectively.

This description is specific for each walker, and, therefore, should also contain an index for the particular walker j:

$$\begin{aligned}\mathbf{p}_j(t) &= \mathbf{p}_{j,0} + \mathbf{p}_{j,1} \sin(\omega_j t) + \mathbf{p}_{j,2} \sin(\omega_j t + \varphi_{j,2}) \\ &\quad + \mathbf{p}_{j,3} \sin(2\omega_j t + \varphi_{j,3}) + \mathbf{p}_{j,4} \sin(2\omega_j t + \varphi_{j,4}).\end{aligned} \tag{3}$$

Because the average posture and all the eigenpostures are 45-dimensional vectors, the overall number of parameters is 5*45 + 4 = 229. Therefore, a 229-dimensional vector $\mathbf{w}_j$ encoding all the parameters provides a full representation of an individual’s walking pattern $\mathbf{p}_j(t)$.

The nice property of this representation is that it is morphable. If compared across different walkers, both the average posture and the eigenpostures are very similar. They show walker-specific variations but they also contain similar structure. This becomes evident when looking at the covariance matrices. The average correlation across all possible 40*39/2 pairs of average postures $\mathbf{p}_{i,0}$ and $\mathbf{p}_{j,0}$ is 0.998. The corresponding numbers for the first four principal components are 0.95, 0.88, 0.85, and 0.73, respectively. This high correlation shows that the components principally encode similar aspects of the walk while still representing the individual differences between walkers.

This result justifies treating the 229-dimensional vector describing the walk $\mathbf{w}_j$ of a walker j as a point in a linear space of the same dimension and, thus, the application of linear methods. Even though the dimensionality of this description is tremendously reduced compared to the original motion capture data, 229 is still a large number of variables for a concise and compact model. In particular for the purpose of constructing linear classifiers with the ability to reasonably generalize to new walking samples, we have to reduce the dimensionality to a degree that is considerably smaller than the number of items in the data set. In an attempt to reduce redundancy within the set of 40 walkers that make up our database, we computed a PCA across the walkers. In contrast to the similar computation on the level of the postures of a single walker, the problem arises that the entries of a walk vector $\mathbf{w}_j$ are not homogenous. Whereas most of the entries encode positions (e.g., in millimeters), there is one entry that encodes the fundamental frequency (e.g., in Hz) and three more that account for the phases of the PCs (e.g., in degrees). PCA is very sensitive to relative scaling. For instance, its outcome would be very different depending on whether the phase would be given in radiants or in degrees or whether the positional measures would be in millimeters or centimeters. We therefore whitened the data by dividing each entry by the standard deviation based on the 40 corresponding entries before subjecting the data to a PCA:

$$\mathbf{W}' = \mathrm{diag}(1/\mathbf{u})\,\mathbf{W}\,. \tag{4}$$

$\mathbf{W}$ is a 229 x 40 matrix containing all the walker data with one walker per column: $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_{40})$. $\mathbf{u}$ is a vector containing the 229 standard deviations computed from the rows of $\mathbf{W}$. $\mathbf{W}'$ is the resulting whitened data matrix.

Computing a PCA on the whitened data $\mathbf{W}'$ results in a decomposition of each walker $\mathbf{w}_j$ into an average walker $\mathbf{w}_0$ and 39 weighted components that we call the eigenwalkers:

$$\mathbf{w}_j = \mathbf{w}_0 + \sum_i k_{i,j}\mathbf{v}_i \tag{5}$$

or in Matrix notation:

$$\mathbf{W} = \mathbf{W}_0 + \mathbf{VK}\,. \tag{6}$$

$\mathbf{W}_0$ denotes a matrix with the average walker $\mathbf{w}_0$ in each of its 40 columns. The matrix $\mathbf{V}$ containing the eigenwalkers as column vectors $\mathbf{v}_i$ is obtained by pre-multiplying the matrix $\mathbf{V}'$ containing the eigenvectors of the covariance matrix of $\mathbf{W}'$ with $\mathrm{diag}(\mathbf{u})$, therefore multiplying each entry with the corresponding standard deviation of this element:

$$\mathbf{V} = \mathrm{diag}(\mathbf{u})\,\mathbf{V}'\,. \tag{7}$$

The matrix $\mathbf{K}$ containing the weights (or the scores) $k_{i,j}$ is obtained by solving the linear equation system:

$$\mathbf{VK} = \mathbf{W} - \mathbf{W}_0\,. \tag{8}$$

Each walker j can now be represented in a space spanned by the first n eigenwalkers $\mathbf{V}_n = (\mathbf{v}_1, \mathbf{v}_2, ... \mathbf{v}_n)$ in terms of the respective score vector $\mathbf{k}_j = (k_{1,j}, k_{2,j}, ... k_{n,j})^T$. The dimensionality of this representation (i.e., the number of eigenwalkers used) can be treated flexibly depending on the particular requirements of the application. With increasing dimensionality, the representation becomes more accurate in terms of its reconstruction quality. On the other hand, a large ratio between the dimensionality and the number of items available for learning invariants becomes unfavourable for classification purposes.

Linear Gender Classification

The representation derived above provides a linear framework for the analysis of the informational content of gait patterns and the extraction of diagnostic parameters. Our database is still comparatively small and many interesting psychological or biological attributes may not yet be fully represented. However, it contains exactly 20 men and 20 women. If the linearization is successful, we can hope to find the attributes that differ between walking men and women to be spread along a straight axis in the space spanned by the eigenwalkers. Using the redundancy inherent in the set of walkers, we can hope to derive a low-dimensional classifier that would correctly classify new walkers. Besides training a linear classifier on the full representation, classifiers can be constructed that use only different parts of the overall information. Their performances can be used to evaluate the role of those parts for gender classification. For instance, it is easy to separate structural information from dynamic information. The average posture p_0 can be regarded to encode structural information comprising both information about the lengths of the body's segments and their average positions. The eigenpostures, in contrast, encode dynamic information. Using different sorts of input information we tested (1) how the two classes separate and (2) how a linear classifier based on a linear discriminant function would generalize to new instances that have not been used for training.

In the following $\mathbf{x}_j$ denotes a column vector with the data of a particular walker used as input for classification. Accordingly, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m)$ is the matrix containing the data set of m=40 different walkers. $\mathbf{x}_j$ can stand for the whole walker representation ($\mathbf{x}_j$=$\mathbf{w}_j$), or only for parts of it, for instance, only for the structural or only for the dynamic part of the representation. The row vector $\mathbf{r}$ contains the expected output of the classifier. It has m entries with $r_j = 1$ if walker j is a man and $r_j = -1$ if the walker is a woman.

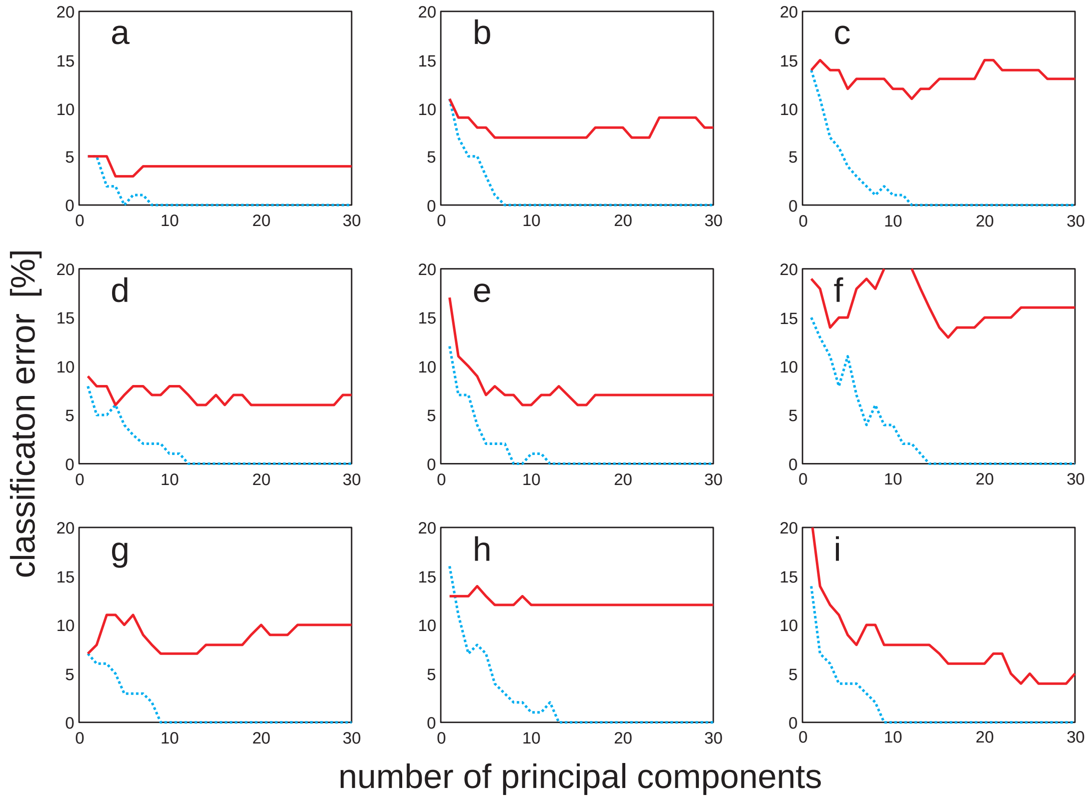


Figure 4. Results of applying the classifier to different versions and parts of the walking data. The dashed blue curve depicts separation performance in terms of the number of misclassifications as a function of the number of components used. The solid red curve shows misclassifications in the generalization test. The following input data have been used:

a: full 229 dimensional description of the walkers with their original size
b: 229 dimensional description, size-normalized
c: only the 45 entries of $\mathbf{p}_{j,0}$, size-normalized
d: four eigenpostures, their phases and the fundamental frequency, size-normalized
e: only first eigenposture, size-normalized
f: only second eigenposture, size-normalized
g: only third eigenposture, size-normalized
h: only fourth eigenposture, size-normalized
i: first, second, and third eigenposture, size-normalized

To test for the ability to separate men's and women's walks in the space, we first ran a PCA by computing the eigenvectors of the covariance matrix of **X**. As described above in more detail, this results in a decomposition of **X** into:

$$\mathbf{X} = \mathbf{X}_0 + \mathbf{VK}\,. \tag{9}$$

$\mathbf{X}_0$ denotes a matrix with the average input data $\mathbf{x}_0$ in each column. The matrix **V** contains the principal components as column vectors $\mathbf{v}_i$ and **K** denotes a matrix containing the scores similarly to the notation used above.

A linear discriminant function **c** is now computed by solving the equation

$$\mathbf{cK} = \mathbf{r} \tag{10}$$

Next we reordered the PCs spanning the walking space by the weights with which they contribute to the discriminant function. For the following computations, component number i no longer is the ith principal component but is the component with the ith highest weight in the discriminant function **c**. We then evaluated the ability to separate male and female walks of discriminant functions of increasing dimension n. A walker j was considered to be classified correctly if

$$\operatorname{sgn}\left(\sum_{i=1}^{n} \mathbf{c}_i k_{i,j}\right) = \operatorname{sgn}\left(r_j\right). \tag{11}$$

Otherwise, the walker j was considered to be misclassified. The dotted lines in Figure 4 depict the percentage of misclassifications as a function of n. If n is large, separation is perfect due to the mismatch between the number of items to be classified and the dimensionality of the space. Depending on the information provided for the classification, perfect separation is reached at dimensions between $n = 4$ and $n = 14$.

More interesting than the ability to find a separating plane is the degree to which the corresponding classifier can generalize to new instances of walking patterns. Lacking a whole new set of data that we could use to test the linear classifier, we ran a single-elimination jack-knife procedure: One of the 40 walking patterns was taken out and a linear classifier was computed on the remaining 39 walkers as described above. After having done so, the remaining walker was projected first onto the principal components derived from the other 39 walkers. The resulting score vector was then projected onto the discriminant function in the subspace spanned by the first n components. Classification was considered to be correct if the projection had the expected sign. The same procedure was repeated with all 40 walkers. The results are plotted as a function of n in Figure 4 (solid lines). Typically, in the generalization test, misclassification reaches a minimum if n has about the size needed to achieve perfect separation in the previous step. If the dimension of the classifier gets much higher, the error increases slightly due to overlearning.

The procedure was applied to different sets of input data. First, we applied it to the full 229 dimensional description of the walker described in the previous section. The results are plotted in Figure 4a. Full separation is reached using only 5 components. Classification performance in terms of generalization to new walkers is very effective. The best classifier needs only 4 components and produces only 3 misclassifications (out of 40 items), corresponding to an error rate of 7.5%.

Visualizing the changes between male and female walkers on which the classifier picked up (see next section for details), we suspected that differences in overall size between men and women are strongly contributing to the good classification performance. To further investigate the role of size, we defined the relative size s_j of each walker j by finding a least-square solution to the equation

$$\mathbf{p}_{j,0} = s_j \frac{1}{n} \sum_k \mathbf{p}_{k,0} \qquad (12)$$

with $\mathbf{p}_{j,0}$ being the average posture of walker j. Using just the s_j as an input for linear classification, only 5 walkers are misclassified corresponding to an error rate of 12.5%.

Although size might be a diagnostic feature for gender classification, we are more interested in other parameters and in particularly in motion-based cues. For further calculations, we normalized the walker data by their size. To achieve this, for each walker j the average posture $\mathbf{p}_{j,0}$ as well as the four eigenpostures $\mathbf{p}_{j,1}$, $\mathbf{p}_{j,2}$, $\mathbf{p}_{j,3}$, and $\mathbf{p}_{j,4}$ were divided by s_j.

Figure 4b illustrates the results of training and testing a classifier that uses the size-normalized version of the full 229-dimensional representation. Complete separation is obtained with 7 components (dotted curve). Generalization is optimal with 6 components resulting in 7 misclassifications (17.5%).

The size-normalized full representation still contains both structural information in terms of the average posture $\mathbf{p}_{j,0}$ and dynamic information in terms of the principal components $\mathbf{p}_{j,1}$, $\mathbf{p}_{j,2}$, $\mathbf{p}_{j,3}$, and $\mathbf{p}_{j,4}$, their respective phases, and the fundamental frequency. In order to evaluate the roles of structural and dynamic information, we submitted only the respective parts of the full representation to the classifier. Figure 4c shows the results obtained from training and testing the classifier with data that contain, for each walker, only the 45 entries of $\mathbf{p}_{j,0}$. Performance in this case is not very good. Twelve components are needed for complete separation and the best generalization performance with 11 misclassifications (27.5%) requires 12 components.

Better performance is obtained if only the dynamic information is used for classification. Figure 4d presents the results of a calculation with the four principal components, their phases and the fundamental frequency being used as input parameters. As before, full separation requires 12 components. Optimal generalization is obtained with only 4 components and reduces the error rate with 6 misclassification to only 15%.

Except for size, the structural information encoded in the average posture does not appear to contribute much information to gender classification. Which parts of the dynamic information are the most relevant ones? Kozlowski and Cutting (1977) mentioned a trend in their data hinting to a possible role of walking frequency. We cannot confirm this. In our data, walking frequencies are virtually identical in men and women. On average, men walked with 0.836 Hz (standard deviation: 0.07 Hz), whereas women walked with 0.845 Hz (standard deviation: 0.09 Hz). Recall that the walkers were allowed to freely adjust the speed of the treadmill to a setting that would feel most comfortable.

The relative phases of the eigenpostures do not make a significant contribution to gender classification either. If the values φ_2, φ_3 and φ_4 are used as input for classification, best separation still produces 14 misclassifications (35%) and best generalization is obtained with 2 components and 15 misclassifications (37.5%).

The role of the four eigenpostures can also be examined separately. Figures 4e-4h show classification performance based on single eigenpostures. Using the first eigenposture alone results in a classification performance that is almost as good as the one obtained with all four eigenpostures (15% misclassifications, 9 components). Using only the third eigenposture also yields good classification performance. The good performance of single eigenpostures implies that the advantage of dynamic information is not simply a matter of the larger number of variables (4 x 45 for dynamic information, 45 for structural information) accounting for it. The best performance that we could obtain was achieved with a classifier based on the first three eigenpostures (Figure 4i). Using 24 components the classification error could be reduced to 4 misclassifications (10%). Three of the four walkers that were misclassified in this case are the same that were also misclassified by the classifier trained with all information, including size information. Those three walkers were also among the misclassifications of all other classifiers.

Synthesizing Walking Patterns

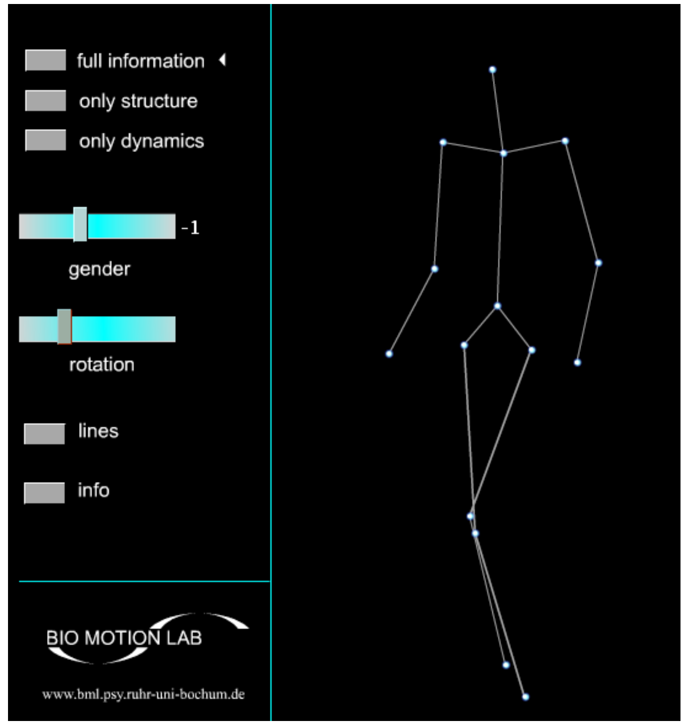
We proposed a representation of human walking data that is suitable for linear analysis of the data with straightforward methods from linear statistics and pattern recognition. The proposed representation is, at least approximately, a complete representation. Virtually no information is lost when transforming the raw motion data into this representation. This has the consequence that the mapping of the raw data into our linearized representation is bijective and therefore invertible. Any point in the 229-dimensional walking space or any low-dimensional eigenwalker-based derivation from it can be mapped back into an explicit description of a walking pattern. Our framework can therefore not only be used for data analysis but also for the synthesis of motion patterns.

The rule to achieve this was actually already given above. A particular vector $\mathbf{w}_j$ in the walking space has to be decomposed into its constituting components $\mathbf{p}_{j,0}$, $\mathbf{p}_{j,1}$, $\mathbf{p}_{j,2}$, $\mathbf{p}_{j,3}$, $\mathbf{p}_{j,4}$, ω_j, $\varphi_{j,2}$, $\varphi_{j,3}$, and $\varphi_{j,4}$. The walk, explicitly described in terms of a time series of postures is then given by Equation 3.

The invertibility of the representation can be used to visualize what is happening along the different classifiers that have been developed in the previous section. For a given classifier $\mathbf{c}$, the differences that a walker undergoes along the discriminant function can be illustrated by displaying walkers $\mathbf{w}_{c,\alpha}$ corresponding to different points along this axis as point-light displays or stick figure animations. Demonstration 1 allows you to visualize and to interactively manipulate a walker display by changing the value of α:

$$\mathbf{w}_{\mathbf{c},\alpha} = \mathbf{w}_0 + \alpha \mathbf{V}\mathbf{c}\,. \tag{13}$$

As above, $\mathbf{w}_0$ denotes the average walker. The matrix $\mathbf{V}$ contains the first few eigenwalkers, one in each column. As α changes from negative to positive values, the appearance of the walker changes its gender. The dimensionality of the eigenwalker space used to compute the respective linear classifiers is n=10. The value of α is scaled in terms of standard deviations (z-scores). A walker resulting from setting α = 6 or α = -6 is therefore an extrapolation into a region of the walker space, which is far away from any real walker. Changing the value of α from negative to positive values evokes a clear percept of a change in the gender of the walker.



Demonstration 1. An interactive demonstration that allows the user to synthesize walkers for different classifiers and gender weightings. Click anywhere in the image to activate the demonstration.

It is interesting that by exaggerating the differences between male and female walks in these animations one discovers the existence of a behavioral pattern that is well established in many animal species. Male animals often try to make themselves bigger than they really are. Mechanisms to achieve this include ruffling fur or feathers, or adopting postures and movement patterns that would make them appear more voluminous. The same purpose seems to rule the differences between male and female walking patterns in humans. Men tend to hold their elbows further away from the body resulting in a posture that requires much more room than the average posture taken on by women. In the dynamic domain, men show a pronounced lateral sway of the upper body that also has the effect of occupying more room than women need.

Gender Classification in Human Observers

In order to compare the performance of the artificial classifier with human gender classification performance, we visualized the motion data of the 40 walkers in terms of point-light displays. A number of observers were presented with these stimuli and were asked to indicate the gender of the walkers.

Participants

Twenty-four students of the Department of Psychology of Ruhr-University participated in the experiment. All had normal or corrected-to-normal vision. They received credit for the participation in the experiment.

Stimuli

For each of the 40 walkers, several versions of point-light displays were generated. All of them were normalized with respect to their size (Equation 12). The duration of each walking sequence was 7 s. The 15 markers were depicted as small white dots on a black background displayed on a computer screen. The full display subtended 5 deg (vertically) of visual angle. The renderings differed in the viewpoint from which the walker was seen and in the type of information provided. Three different viewpoints were used: frontal view (0 deg), 3/4 view (30 deg or -30 deg), and profile view (90 deg or -90 deg). For each walker and each viewpoint, three different sequences were generated. The first one ("full info") showed the original walking data. The second set of stimuli ("structure-only") was generated by combining the individual average postures $\mathbf{p}_{j,0}$ with averaged motion data. This was obtained by computing averaged eigenpostures $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, and $\mathbf{p}_4$ as well as average values for the phases φ_2, φ_3, and φ_4 and for the fundamental frequency ω. The components were then combined with the individual average postures according to Equation 3. The stimuli are therefore normalized with respect to dynamic information and contain only structural information to be used for gender classification. Finally, a third set ("dynamic-only") was generated by replacing each individual's average posture $\mathbf{p}_{j,0}$ with the average across all walker's postures, therefore normalizing for the structural information and providing only dynamic information:

$$\mathbf{p}_0 = \sum_j \mathbf{p}_{j,0} \,.$$

Design and Procedure

Twenty-four participants were divided into three groups of equal size. One group was presented with only 0-deg walkers; the second group saw only 30-deg walkers, and the last group only 90-deg walkers. The experiment was run in two blocks, each consisting of 80 trials. The first block showed two instances of each walker's veridical motion. The order was randomized for each observer. In the second block 40 structure-only and 40 dynamic-only trials were presented in randomized order. Observers had to indicate whether a walker appeared to be male or female by pressing one of two keys on the computer's keyboard. Subjects were required to respond during the 7 s while the stimulus was presented. The display was repeated if no response was made. An inter-trial interval of 3 s, during which the screen remained black, separated the trials. We measured error rates and evaluated them in terms of an ANOVA with the factors VIEW (0, 30, and 90 deg) and INFO (full, dynamic-only, and structure-only).

Results

Figure 5 shows the results. Both factors were highly significant (VIEW: $F(2,21)=26.4$, $p<.001$; INFO: $F(2,42)=29.3$, $p<.001$). Performance is best with error rates around 25% when a walker is seen in frontal view and declines gradually with increasing deviation from that viewing angle. The effects with respect to the information provided are such that depriving observers from diagnostic structural information hardly impairs performance whereas depriving observers from dynamic information results in a severe drop in performance. A Scheffé post-hoc test confirms that the difference between performance in the structure-only condition and the other two conditions is statistically reliable ($p<.01$).

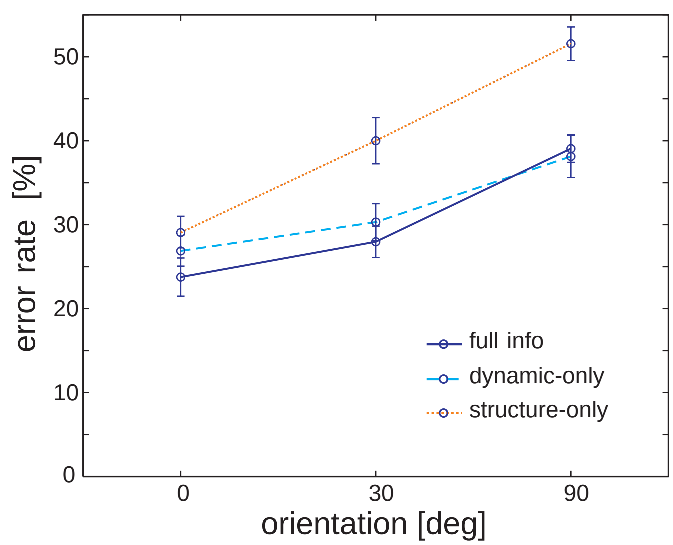


Figure 5. Results of psychophysical classification of the 40 walkers shown from three different viewpoints. The three lines depict results using stimuli showing the veridical walker (solid line), the dynamic-only (dashed line), or the structure-only versions of the walkers.

The ANOVA also shows a significant interaction between the factors VIEW and INFO ($F(4,42)=3.1$, $p<.05$) indicating that deprivation of dynamic information has a much stronger effect if the walker is shown in profile view as compared to frontal view presentation. In the profile view condition, performance drops from an error rate of 39% in the full-info condition all the way down to chance level (52% error rate) in the structure-only condition. In the frontal view condition, error rate increases from 24% in the full-info condition to 29% in the structure-only condition. The relatively small difference between the performances obtained in full-info and structure-only conditions with frontal view stimuli is still statistically significant (paired t test: n=8, p<.05).

Discussion

The psychophysical results show a pattern similar to the results from the simulations presented in the previous section. Performance of both human observers and the artificial classifier is mainly carried by dynamic information. If this part of the overall information is not provided, performance declines significantly. Depriving the stimulus of diagnostic structural information, on the other hand, has only a comparatively weak effect on both human and artificial gender classification.

An item analysis reveals additional parallels between the psychophysically derived results and the artificial classifier. We ordered the 40 walkers according to the number of misclassifications that they received in the psychophysical experiment. The rankings were computed separately for data resulting from full-info, structure-only, and dynamic-only presentations, collapsing data from all three VIEW groups. The three walkers that were consistently misclassified by all the artificial classifiers were at positions 1, 3, and 12 for the full-info data, at positions 1, 24, and 36 for the structure-only data, and at positions 1, 4, and 6 for the dynamic-only data.

To further compare the outcome of the psychophysical results with the various artificial classifiers, we sorted the 40 walkers by means of the value of the projection of this walker on the respective discriminant function multiplied with a value of 1 if the walker was a man and a value of -1 if the walker was a woman:

$$z_j = r_j \sum_{i=1}^{n} \mathbf{c}_i k_{i,j} \, . \tag{14}$$

z_j is a measure for how well a walker j (represented in terms of the scores $k_{i,j}$) with gender r_j was classified by the linear classifier $\mathbf{c}$. Table 1 lists the correlation coefficients of a rank correlation between the three ranks obtained from the psychophysical data, and the ranks obtained for classifiers corresponding to the data presented in Figures 4a-4d and 4i.

The psychophysically obtained rankings were computed separately for full-info trials, dynamic-only trials, and structure-only trials. The classifiers used correspond to the ones illustrated in Figures 4a-4d and Figure 4i, respectively. n indicates the number of eigenwalkers used to construct the classifier. n was always chosen such as to yield optimal generalization performance (see "Linear Gender Classification" for details). Correlation coefficients larger than 0.373 are significant ($\alpha = 0.01$).

Table 1. Correlation Coefficients Obtained From a Spearman Rank Correlation Between the Number of Misclassifications Received by the Individual Walkers in the Psychophysical Experiment and a Measure for the Confidence of the Classification by Five Different Linear Classifiers.

	Psychophyics		
Linear Classifier	Full-info	Structure-only	Dynamic-only
Full-info plus size n=4	0.2820	-0.0379	0.4473
Full-info n=6	0.5158	0.1970	0.6525
Structure-only n=12	0.5114	0.4595	0.3760
Dynamic-only n=4	0.3484	0.0250	0.5602
First, second, and third eigenposture n=24	0.3773	0.1008	0.5353

The number of misclassifications obtained in the psychophysical experiment and the confidence measure for the artificial classifications correlate to a high degree if the information provided to the human observers and to the artificial classifier is similar. The rankings obtained from providing full information and from the trials with dynamic-only information also show large correlations. The pattern of misclassifications obtained when providing human observers with only structural information correlates to the one obtained from a linear classifier provided with the same information but is very different from the one obtained by training the classifier with full-info or dynamic-only information.

Whereas we provided the artificial classifier with the full three-dimensional information, the human observers were presented with two-dimensional projections of the walker. This might be one reason that the artificial classifiers performed considerably better than the human observers. The best performance that was reached by our observers was 76% correct responses in the case of frontal views of the veridical walkers. The artificial, linear classifier, in contrast, reached a performance of 90% correct classifications.

The results provided by the psychophysical data compare well with data from previous studies. Kozlowski and Cutting (1977), as well as Barclay et al. (1978), showed only saggital views of point light walkers to their observers and yielded correct gender classification rates between 63% and 65%. The performance that we measured in saggital view was 62% correct classification. We can also confirm parts of the results obtained by Mather and Murdoch (1994), who used artificial walker stimuli in a gender classification task and found that, first, frontal views result in much better performance and, second, that dynamic stimulus attributes are more important than structural stimulus attributes. In contrast to the findings of Mather and Murdoch, however, the different role of structural and dynamic information becomes much more evident in the saggital view and almost disappears in the frontal view. Mather and

Murdoch had concentrated on lateral body sway as an example for a dynamic cue and the hip/shoulder ratio as an example for a structural cue. Due to the artificial nature of the stimuli, both cues were not detectable from a saggital view. It is therefore not surprising that the dominance of the dynamic cue over the structural cue was only apparent in frontal view but not in sagittal view.

General Discussion

Human locomotive motion is a complex spatio-temporal pattern that is ruled by biomechanical as well as functional constraints. Many of these constraints are modified by individual characteristics and personality traits of the actor. The human visual system is capable of decoding information about the characteristics of a walker by visually analyzing the motion pattern. Here we provided a framework for transforming human walking data into a representation that allows us to treat the analysis of biological motion as a linear pattern recognition problem. To demonstrate its ability to extract perceptually relevant information, we constructed a linear classifier capable of discriminating between male and female walkers. Using different modifications or only parts of the overall information as input data for classification, we examined the respective roles of different aspects of the data for gender classification.

Simply measuring the size of a walker results in a relatively reliable gender estimation. Measuring absolute size requires an absolute scale and although available in our motion capture data, this cue is generally not readily available for human vision or in computer vision. We therefore ignored this source of information and normalized the size of all walking data. Providing the classifier with either only structural information or only dynamic information showed that the dynamics contain more reliable diagnostic cues than the structure. Walking speed (stride frequency) did not provide a diagnostic cue. We found this result surprising. Considering animate locomotion to be articulated pendular motion, an inverse quadratic correlation between size and stride frequency is expected (Alexander, 1989; Troje & Jokisch, submitted). Because size is a diagnostic cue to gender classification in our data set, we would have expected that stride frequency would also be diagnostic. Although subjects were allowed to adjust the speed of the belt in order to walk as comfortable as possible, the lack of gender dependent frequency differences may have to do with the particular situation of walking on a treadmill rather than on solid ground.

Scrutinizing on the role of the different eigenpostures shows that the first and the third component are providing more information than the second and the fourth. The advantage of the first component over the second is probably simply a consequence of the larger variance covered by the first component. The same reason may account for higher contribution of the third component as compared to the fourth. Whereas the first two components account for the fundamental frequency, the third and fourth components represent the second harmonic. It seems that although having less power, the second harmonic carries as much information as the fundamental frequency.

Comparing the classification behavior of the model with the performance of human observers yields several similarities. Human observers also seem to rely more on dynamic information than on structural information. However, this difference is much more pronounced when the walkers are shown in saggital view and almost disappears in frontal view. Whereas the predominance of dynamic information over structural cues is in accordance with earlier work by Mather and Murdoch (1994), the dependence on viewpoint seems to contradict their results. In Mather and Murdoch's study, however, structural and dynamic cues were represented only by single features that were chosen so that they were not distinguishable in saggital view. Here, in contrast, we manipulated the walking pattern of real walkers by normalizing either for structural or for dynamic information, preserving not only a single feature in the complementary domain but the whole array of available information. On average, men and women do show clear differences in body structure. This has been shown by Barclay et al. (1978), and it is also clearly visible in the animations of the structure-only walkers (Demonstration 1). However, the variance within the two classes is so large that they overlap to an extent that renders body structure a cue, which is less reliable than the dynamics of the walking pattern.

Gender classification was used as an example to test how suitable the proposed linearization of motion data is for classification purposes. Other attributes of a walker such as age, weight, emotional state, or personality traits could be treated similarly. However, the database that we used would have to be extended to better represent such attributes. At this point, the sample of walkers is still quite homogenous and does not span a statistically representative range of age, weight, and other attributes. Given an extended database, it is straightforward and absolutely analogous to the gender classification problem to extract the diagnostic features conveying information about other attributes from walking patterns.

In principle, the model can also be extended to other actions. Each action, however, requires its own formulation. For example, a model for running could be obtained in a similar manner as the walking model. However, at least within the framework outlined here, it would not make sense to try to describe both walking and running patterns within the same model. Dynamic models of gait production (Golubitsky, Stewart, Buono, & Collins, 1999; Golubitsky, Stewart, Buono, & Collins, 1998) show that the transition between walking and running is characterized by a singularity, and, therefore, represent two distinctively different actions. Empirical data supporting this view can be found in Alexander and Jayes (1980). The sensitivity of our model to small but meaningful variations in the style of an action depends to a large degree on the structural similarity of the items spanning the space, which, in turn, defines the correspondence between items. Each item in the space must match any other item in a canonical, unambiguous way. Of course, it is possible to smoothly blend between different actions, but the definition of the correspondence on which such a blend is based, remains somewhat arbitrary. In contrast, the correspondence between items that belong to the same action can be defined in a canonical and unambiguous way by means of the naturally occurring transitions between structurally similar items. A system that could be used both for action recognition as well as for the classification of stylistic features would ideally separate those two steps. A model describing different actions within the same motion space could be used for action recognition on a basic level (Rosch, 1988; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Knowing which particular action the system is confronted with would then elicit a recognition module for stylistic features within an action-specific linear motion model on a subordinate level.

Correspondence-based representations result in a separation of the overall information into range-specific information on the one hand and domain-specific information on the other hand. Applied to the current model, the range-specific information is the positional information contained in the average posture as well as in the eigenpostures. The domain-specific information is the information about when things are happening. This information is contained in the phases and frequencies corresponding to the eigenpostures.

The domain-specific part of the walking data has a comparatively simple description that is possible only because the amplitudes of the eigenpostures change sinusoidally in time. The frequency of the first two eigenpostures is the fundamental walking frequency and the frequency of the third and fourth component equals the second harmonic of the fundamental frequency. If φ_2 (i.e., the phase difference between the sine functions describing the temporal behavior of the first and the second eigenposture) would be exactly 90 deg and if the same would be true for the difference between φ_3 and φ_4, then the four-dimensional PCA decomposition would be similar to a second-order Fourier decomposition. Both decompositions are based on the same model:

$$\begin{aligned} p(t) = \mathbf{p}_0 + \mathbf{p}_1 \sin(\omega t) + \mathbf{p}_2 \sin(\omega t + \varphi_2) \\ + \mathbf{p}_3 \sin(2\omega t + \varphi_3) + \mathbf{p}_4 \sin(2\omega t + \varphi_4). \end{aligned} \tag{15}$$

However, whereas PCA considers the $\mathbf{p}_i$ to be the basis and constrains them to be orthogonal, Fourier analysis considers the sine functions to be an orthogonal basis and therefore requires φ_2 and also φ_4-φ_3 to equal 90 deg. Both can, in general, not be achieved at the same time. It is therefore interesting that the temporal behavior of the orthogonal basis constituted by the first four eigenpostures approximates a Fourier decomposition to a very high degree. In fact, both φ_2 and φ_4-φ_3 assume values very close to 90 deg (φ_2: mean 91, STD 5.3; φ_4-φ_3: mean 91, STD 3.8).

Hence, a decomposition of the walking data using Fourier analysis instead of PCA would yield similar results. The Unuma et al. (1995) "rescaled Fourier functional model" could have been used in a similar way to design linear classifiers for gender recognition and other similar tasks. However, applying PCA as a first step to reduce dimensionality in the description of postures is much more general and can be applied to nonperiodic motions in a similar way. The main addition that would be needed to derive a linear model for other motions lies in the parameterization of the temporal behavior of the scores. The only parameters needed to describe the domain-specific information in our case are the frequencies and phases of the components. Hence, time warping reduces to simple uniform scaling and translation in the case of our walking data. For a general parameterization that would also apply to nonperiodic actions, more complex models have to applied. A very flexible solution is, for instance, the use of B-spline functions (Ramsay, 1998; Ramsay & Silverman, 1997). Nonparametric solutions have been demonstrated by Giese and Poggio (2000).

Another important point has yet to be discussed. Biological motion is articulated motion and has several commonalities with pendular motion (Aggarwal, Cai, & Sabata, 1998; Cutting, 1978a, 1981). The distal part of a limb's bone moves on a spherical trajectory around the proximal end that is fixed at the joint's position. For this reason, it seems reasonable to describe the movements of a body in terms of joint angles: The position of a given point on the body is not represented in terms of its allocentric Cartesian coordinates in 3-D space but rather in polar coordinates with respect to a coordinate system, which is fixed to the "parent" part, that is, the part which provides the more proximal articulation. Transforming positional data into joint angle data thus seems a reasonable step toward linearizing such data.

However, this requires knowledge about the hierarchy of the articulation. In the context of many applications, this information will be available anyway. In other cases, this might be a problem. For video-based tracking purposes, for instance, it might be relatively easy to segment a walking figure from the steady background; however, it might not be as straightforward to identify particular parts of the body and recover its full hierarchy beforehand.

Cartesian representations have many advantages as opposed to joint angle representations because they do not need information about the articulation of a body.

Joint angles can be relatively easily derived from motion capture data with markers placed close to joint positions. Nonetheless, even in this case, many constraining assumptions have to be made in order to define a biomechanical model which, when applied to the raw motion capture data, yields the exact joint locations. In cases, however, in which the motion information comes from feature points that cannot be precisely positioned in the course of a well-controlled motion capture session, joint angles might not be accessible directly. Cartesian representations, on the other hand, correspond to the raw data format that a motion capture system outputs anyway. Data from markers positioned on any part of the body can be used as well as data from markers positioned at or near joints. In particular for markerless, video-based motion tracking, a simple model that does not rely on information about the articulation of the body has many advantages. The same is true for any model of the human visual system. Off-joint point-light displays can easily be interpreted, and it seems unlikely that the human visual system relies on joint angle representations.

We proposed a framework for transforming human gait data into a representation that allows such data to be applied to simple linear methods from statistics and pattern recognition. We tested this approach by designing classifiers that discriminate between male and female walking patterns with a performance that is even better than the performance achieved by human observers. We do not know how the human visual system solves the problem of extracting information from biological motion patterns, but it is interesting that the behavior of the artificial gender classifiers reflects aspects of human visual performance, such as the dominance of dynamic information above structural information. Using a generative model rather than some kind of feature space for visual motion recognition fits the idea of using the same brain systems for both the analysis and synthesis of motion patterns (Prinz, 1997). This idea has recently received strong support. Both the discovery of mirror neurons in the prefrontal cortex of monkeys (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996) and the finding that imagery and observation of movements can activate brain areas that have previously been considered to accomplish mainly motoric functions strongly suggest that a common neuronal basis exists for the visual analysis of biological motion and for planning and execution of motor commands. Detailed psychophysical experiments could provide more insight into the principles according to which the human visual system processes biological motion patterns.

Acknowledgments

Commercial relationships: none.

References

Aggarwal, J. K., Cai, Q., & Sabata, B. (1998). Nonrigid motion analysis: Articulated and elastic motion. *Computer Vision and Image Understanding, 70*, 142-156.

Alexander, R. M. (1989). Optimization and gaits in the locomotion of vertebrates. *Physiological Reviews, 69*, 1199-1227. [PubMed]

Alexander, R. M., & Jayes, A. S. (1980). Fourier analysis of forces exerted in walking and running. *Journal of Biomechanics, 13*, 383-390. [PubMed]

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]

Beardsworth, T., & Buckner, T. (1981). The ability to recognize oneself from a video recording of one's movements without seeing one's body. *Bulletin of the Psychonomic Society, 18*, 19-22.

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

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. Series B: Biological Sciences, 352*, 1257-1265. [PubMed]

Brand, M., & Hertzmann, A. (2000). Style machines. *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques,* New Orleans, 183-192.

Bruderlin, A., & Williams, L. (1995). Motion signal processing. *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, Los Angeles, 97-10.

Cutting, J. E. (1978a). A biomechanical invariant of gait perception. *Journal of Experimental Psychology: Human Perception & Performance, 4*, 357-372. [PubMed]

Cutting, J. E. (1978b). Generation of synthetic male and female walkers through manipulation of a biomechanical invariant. *Perception, 7*, 393-405. [PubMed]

Cutting, J. E. (1978c). A program to generate synthetic walkers as dynamic point-light displays. *Behavior Research Methods & Instrumentation, 10*, 91-94.

Cutting, J. E. (1981). Coding theory adapted to gait perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 71-87.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9*, 353-356.

Davis, J., Bobick, A., & Richards, W. (2000). *Categorical representation and recognition of oscillatory motion patterns.* Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC, June 13-15, 2000, *628-635*.

Davis, J. W. (2001). *Visual categorization of children and adult walking styles.* Paper presented at the International Conference on Audio- and Video-based Biometric Person Authentication, Halmstad.

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception, 22*, 15-22. [PubMed]

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

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593-609. [PubMed]

Giese, M. A., & Poggio, T. (1999). *Synthesis and recognition of biological motion patterns based on linear superposition of prototypical motion sequences.* Paper presented at the IEEE Workshop on Multi-View Modeling and Analysis of Visual Scene, Fort Collins, CO.

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.

Gleicher, M. (1998). Retargetting motion to new characters. *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques,* Orlando, 32-42.

Golubitsky, M., Stewart, I., Buono, P. L., & Collins, J. J. (1999). Symmetry in locomotor central pattern generators and animal gaits. *Nature, 401*, 693-695. [PubMed]

Golubitsky, M., Stewart, I., Buono, P. -L., & Collins, J. J. (1998). A modular network for legged locomotion. *Physica D, 115*, 56-72.

Guo, S., & Roberge, J. (1996). A high-level control mechanism for human locomotion based on parametric frame space interpolation. In R. Boulic & G. Hegron (Eds.), *Computer animation and simulation '96* (pp. 95-107). New York: Springer.

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology, 11*, 880-885. [PubMed]

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

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

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. *Psychological Research, 38*, 379-393. [PubMed]

Jones, M. J., & Poggio, T. (1999). Multidimensional morphable models: A framework for representing and matching object classes. *International Journal of Computer Vision, 29*, 107-131.

Ju, S. X., Black, M. J., & Yacoob, Y. (1996). *Cardboard people: A parameterized model of articulated image motion.* Paper presented at the Proceedings of the International Conference on Face and Gesture, Vermont.

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

Lappe, M., & Beintema, J. (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]

Li, N., Dettmer, S., & Shah, M. (1997). Visually recognizing speech using eigensequences. In M. Shah & R. Jain (Eds.), *Motion-based recognition* (pp. 345-371). Amsterdam: Kluwer Academic Publishing.

Mather, G., & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 258*, 273-279.

Murray, M. P., Kory, R. C., & Sepic, S. P. (1970). Walking patterns of normal women. *Archives of Physical Medicine and Rehabilitation, 51*, 637-650. [PubMed]

Ormoneit, D., Sidenbladh, H., Black, M. J., & Hastie, T. (2000). Learning and tracking of cyclic human motion. *Proceedings of NIPS 2000 (Neural Information Processing Systems),* Denver, CO, *894-900*.

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

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

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology, 9*, 129-154.

Ramsay, J. O. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society, 60*, 365-375.

Ramsay, J. O., & Li, X. (1989). Curve registration. *Journal of the Royal Statistical Society, 60*, 351-363.

Ramsay, J. O., & Silverman, B. W. (1997). *Functional data analysis*. New York: Springer.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131-141. [PubMed]

Rosales, R., & Scarloff, S. (2000). Specialized mappings and the estimation of human body pose from a single image. *IEEE Computer Society Workshop on Human Motion,* Austin, TX, 19-24.

Rosch, E. (1988). Principles of categorization. In A. Collins & E. E. Smith (Eds.), *Readings in cognitive science* (pp. 312-322). San Mateo: Morgenkaufmann.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

Rose, C., Bodenheimer, B., & Cohen, M. F. (1998). Verbs and adverbs: Multidimensional motion interpolation using radial basis functions. *Computer Graphics and Applications, 18*, 32-40.

Runeson, S. (1994). Perception of biological motion: The KSD-principle and the implications of a distal versus proximal approach. In G. Jansson, W. Epstein & S. S. Bergström (Eds.), *Perceiving events and objects* (pp.383-405). Hillsdale, NJ: Erlbaum.

Shelton, C. R. (2000). Morphable surface models. *International Journal of Computer Vision, 38*, 75-91.

Troje, N. F., & Jokisch, D. *Biological motion as a cue for size.* Manuscript submitted for publication.

Troje, N. F., & Vetter, T. (1998). Representations of human faces. In C. Taddei-Ferretti & C. Musio (Eds.), *Downward processing in the perception representation mechanism* (pp. 189-205). Singapore: World Scientific.

Unuma, M., Anjyo, K., & Takeuchi, R. (1995). Fourier principles for emotion-based human figure animation. *Proceedings of the 1995 Conference on Computer Graphics and Interactive Techniques,* Los Angeles, CA, August 6-11, 1995, *91-96*.

Unuma, M., & Takeuchi, R. (1993). Generation of human motion with emotion. *Computer Animation '93*, 77-88.

Vetter, T. (1998). Synthesis of novel views from a single face image. *International Journal of Computer Vision, 28*, 103-116.

Vetter, T., & Poggio, T. (1997). Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*, 733-742.

Vetter, T., & Troje, N. F. (1997). Separation of texture and shape in images of faces for image coding and synthesis. *Journal of the Optical Society of America A, 14*, 2152-2161.

Wiley, D. J., & Hahn, J. K. (1997). Interpolation synthesis for articulated figure motion. *Proceedings of the Virtual Reality Annual International Symposium*, 39-45.

Wilson, A. D., & Bobick, A. F. (1995). *Learning visual behaviour for gesture analysis.* Paper presented at the Proceedings of IEEE International Symposium on Computer Vision, Coral Gables, FL.

Witkin, A., & Popovic, Z. (1995). Motion warping. *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques,* Los Angeles, 105-108.

Yacoob, Y., & Black, M. J. (1997). Parameterized modeling and recognition of activities. *Computer Vision and Image Understanding, 73*, 232-247.