WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

The influence of form and motion information on the perception of biological motion

Der Einfluss von Form- und Bewegungsinformation auf die Wahrnehmung von biologischer Bewegung

Inaugural-Dissertation zur Erlangung des Grades eines Doktors der
Geisteswissenschaften der Philosophischen Fakultät der Westfälischen
Wilhelms-Universität zu Münster (Westf.)

vorgelegt von
Simone Kuhlmann
aus Bochum
am 23.Juni 2009

Tag der mündlichen Prüfung: 18.12.2009

Dekan: Herr Prof. Dr. Christian Pietsch

Referent: Herr Prof. Dr. Markus Lappe

Korreferent: Herr Prof. Dr. Norbert Sachser

Contents

List of Figures

List of Tables

Part I

General Introduction

Chapter 1

Biological Motion

Contents

To humans vision is one of the, if not the most important sensory input. Our visual system constantly perceives and analyses a wide range of visual input. Within the visual system the perception and interpretation of movement plays an important role. Our visual system is very sensitive for movement, and we tend to direct our attention toward moving objects. Motionless objects in a scene are not easily detectable, and several camouflage strategies in the animal kingdom are based on this fact. Movement helps us to recognize different objects in a scene and to obtain object properties such as their relative depth. For human beings perceiving and interpreting the movement of other living creatures, particularly humans, is of great importance. We are fast and efficient in recognizing movements of other living beings and we are able to achieve many aspect of biological, psychological or social meaning.

We can imagine that for the early human the correct interpretation of the movement of prey and other humans was necessary for surviving. In order to not become prey themselves and to react appropriately it was also important to detect an approaching enemy. Even though nowadays we hunt our food in supermarkets and do not need to fear approaching predators at the next corner, interpreting biological movements still plays a significant role in our daily life. We interact with other humans by reacting to their actions. In the interaction with other humans we use facial expressions, gesturing or posture changes as a mean for nonverbal communication. The sign language, a fast consequence of gestures, has enabled deaf people to use a communication form, which in principle equals speaking.

Movements of living creatures are often very complex and can consist of a complicated sequence of single movements and slightly different movements and gestures can have a different meaning and impact on others. Therefore, the perception and interpretation of biological movements is a highly complex task for the visual system. Even though many species, including our own, are very sensitive for the movement patterns of other living creatures. A very impressive example of the ability of the visual system to recognize human motion even under strongly reduced conditions showed the Swedish psychologist Gunnar Johansson at the beginning of the 70's. Johansson's assumption was that humans would have a special capability for the perception of biological movement. He constructed representations of humans, on which only the joints were to be seen as point-lights (Johansson, 1973). Johansson showed that naive human observers could readily recognize the animated point-light displays as moving human figures, whereas a set of static points, as it is to be seen in a static image, remains insignificant. Not only were observers able to recognize portrayed actions, they could easily disentangle sets of point-lights if more than one person was shown. Johansson termed this phenomenon the perception of biological motion.

1.1 Research on the Perception of Biological Motion

Over the last years, numerous psychophysical, physiological, as well as neuropsychological studies have been conducted to investigate the perception of biological motion. Since Johanssons first study on biological motion (Johansson, 1973) it has been demonstrated that point-light displays provide sufficient information to derive very subtle meanings about the person filmed. Biological motion can be perceived even within masks of dots (Cutting et al., 1988; Bertenthal & Pinto, 1994; Thornton et al., 1998) and also complicated characteristics of the movement can be recognized by the view of point-light representations. Despite the lack of usable information, recognizing the gender of a person (Kozlowski & Cutting, 1977; Mather & Murdoch, 1994), the current mood (Dittrich et al., 1996) and even the identification of familiar persons (Cutting & Kozlowski, 1977; Troje,

2002) is quite possible for example.

Until today the exact processing mechanisms for biological motion are not completely enlightened. Many investigations are based on the assumption that the perception of biological motion is obtained by movement signals (Johansson, 1973; Cutting, 1981; Mather et al., 1992). In addition, there are approaches, which brings out the role of the form information. Studies, which argue against this great importance of movement information, are for example clinical case studies (Vaina et al., 1990; McLeod et al., 1996). Motion perception in their patients was strongly impaired by damages in important movement-processing areas. They were however still able to perceive biological motion. A more exact enlightening of these two aspects form and movement information takes place in section 1.3.

1.2 Stimuli Used in the Research of Biological Motion

The discovery of photographic techniques during the previous century provided new opportunities for the investigation of the biological movements. Now it was possible to capture instants or complete sequences of the movements of humans or animals in still-frame pictures and by that to look into details of biological movements. The American photographer Eadward Muybridge for example, showed with his image-series of a galloping horse for the first time that there is a moment in the horses gait, in which all four hooves are off the ground at once (Scientific American, Oct. 19, 1978). Muybridge used a series of 12 stereoscopic cameras, 21 inches apart to cover the 20 feet taken by one horse stride, taking pictures at one thousandth of a second (see Figure 1.1).

A number of different techniques have been employed to investigate the perception of biological movements. For example, Shiffrar and co-workers (Shiffrar

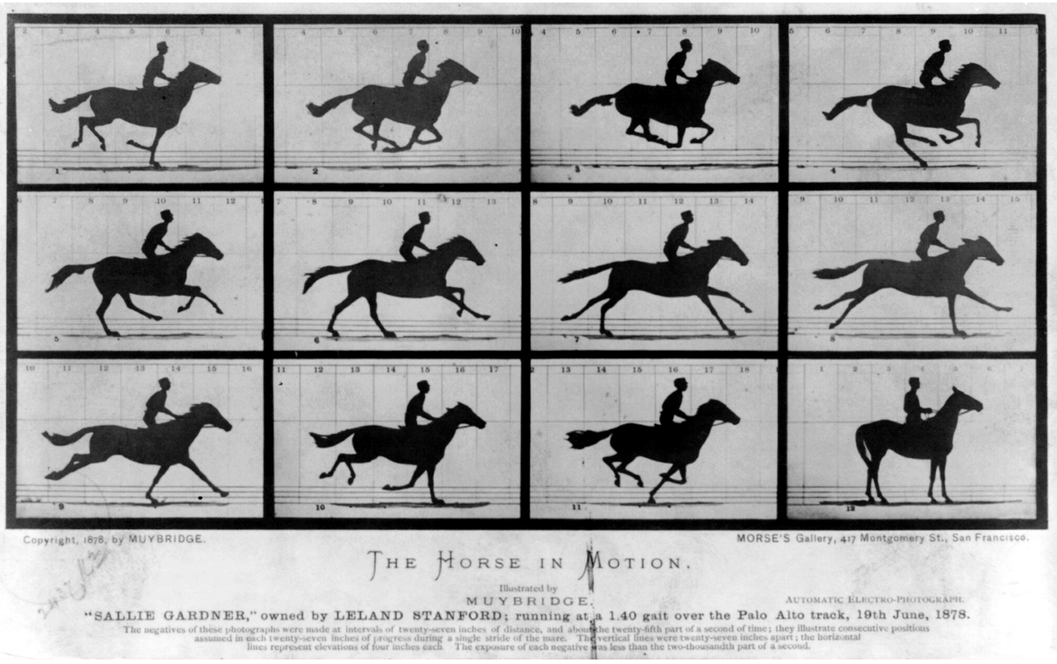

Figure 1.1: Eadward Muybridge's - The horse in motion. (published in Scientific American, vol. 39, no. 6, 19 October 1878; 'A Horse's Movement Scientifically Considered'

& Freyd, 1990, 1993; Chatterjee et al., 1996) presented pairs of alternating photographs of people in different postures and asked their subjects to describe the motion between the two images. They showed that when realistic photographs of a human body are sequentially presented at slow temporal rates, observers perceive biomechanically plausible paths of apparent motion even when those paths are not the shortest possible. Apparent motion is the illusion of real motion induced by a rapid succession of still frames. Important are here the distance of the objects to each other as well as the temporal interval, in which the frames are presented. Both apparent motion and real motion are thought to be processed by the same neural substrate.

Hodgins et al. (1998) investigated the sensitivity to biological motion for different geometric models, including stick figures, polygonal models, and models with muscles, flexible skin, or clothing. They found the viewer's perception of motion characteristics affected by the geometric model used. In the study of Knoblich & Prinz (2001) subjects had to predict the outcome of an action watching whole body displays either of themselves or of somebody else. The predictions were more accurate when subjects watched themselves acting. Kuhlmann & Lappe (2006) used movies of actions in natural scenes in different degrees of blurring. They found that reductions of local form and local motion information can be compensated by global form change and global motion. A detailed definition of local and global aspects of form respectively motion information can be found in section 1.3.

However, the most common technique for studying biological motion perception is the point-light-technique and the stimulus Johansson (1973) developed is still in use in more or less modified form. In these classic point-light displays only the major joints of the human body (ankles, knees, hips, shoulders, elbows and wrists) are represented by a set of light points (see 1.2).

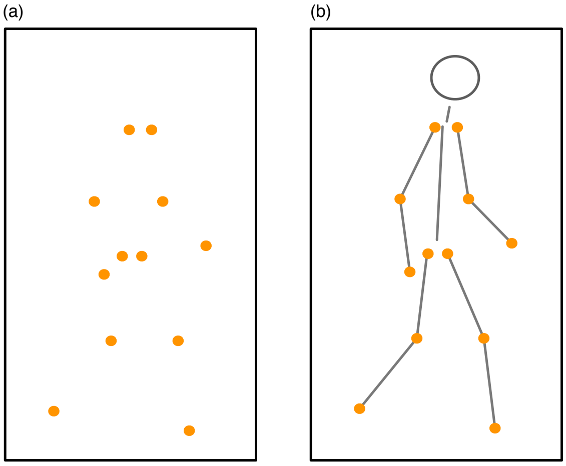

Figure 1.2: Classic point-light representation of a walking human; the light-points represent the major joints of the body (ankles, knees, hips, shoulders, elbows and wrists). In the actual animation, only the dots are visible as shown in (a) - in (b) the lines are added to make the fixed point positions on the body obvious.

Representing biological motion by point-lights has obvious advantages, because it allows a large degree of control over the stimulus and it restricts the visual information. Such an accumulation of points carries only two-dimensional information and hardly offers structural references, since the points are not connected. The classic Johansson stimulus offers only little form information about the actual outlines of the figure or its structure. Presented in motion, the individual joints and therewith the individual limbs are well defined by movement vectors. The figure moves in front of a uniform background; consequently no information about the environment is mediated. As information about the shape of the body was largely removed Johansson concluded that the information in a point-light display is carried mainly by the motion of the points over time.

While Johansson and others recorded the movements of real actors, Cutting (1981) developed a computer program to generate an artificial walker based on a mathematical algorithm, closely similar to a real walker. This stimulus was only available for walking, but had the advantage of being easy to modify and replicate.

Modified version of the classic point-light display like the inter-joint displays still give adequate representations of human motion (Cutting, 1981; Dittrich, 1993). Even though Cutting (1981) found with computer generated stimuli the normal Johansson stimuli more general and better fitting for gender recognition, the study of (Dittrich, 1993) found no differences between normal and inter-joint displays and agued against such generalization at least for real-life displays.

Beintema & Lappe (2002) used a limited lifetime technique to create novel point-light stimuli in which the presence of local motion information is strongly reduced (see Figure 1.3). These sequential position stimuli directly pit motion and form information against each other. A small number of light points is placed on random positions on the limbs of the body rather than on the joints. Each light point remains at its position on the body for a limited time. After that time, the

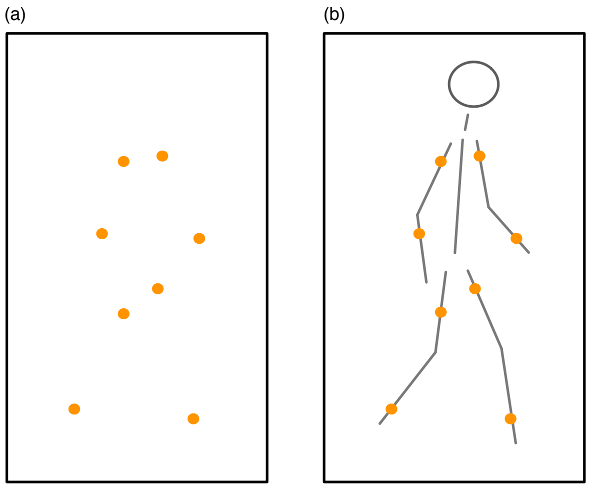

Figure 1.3: Sequential position point-light representation of a walking human; the light-points are located at random position on the limbs. In the actual animation, only the dots are visible as shown in (a) - in (b) the lines are added to make the point positions on the body obvious.

point extinguishes and a new one appears at a different random position. In these stimuli, the form of the body is sampled over time more completely than in classic point light stimuli. Each individual image, however, gives only very limited form information. The amount of form information can be adjusted by varying the number of dots displayed simultaneously and by varying the presentation time of the stimulus and by that the total number of dots in the stimulus. The amount of local motion information, on the other hand, can be adjusted by varying the lifetime of each dot. If the lifetime is restricted to only a single frame, the probability for an individual point to create local motion in the direction of the limb movement is reduced to about 2 %.

1.3 Features Underlying Biological Motion

Biological motion patterns are highly complex and containing many degrees of freedom. The human body consists of several, in itself rigid elements, the extremities, which are connected by joints. The distance between the individual joints remains constant within a movement. The type of the joints limits the directions of motion of the individual extremities, and also how far the movement can go into a certain direction.

The human figure is defined by both local and global features, which can be used for the perception process. Although local and global are difficult to define as absolute terms, most studies of the visual perception of human movement have defined local analysis as the computations conducted on individual points. Global analyses are conducted over larger areas. In the temporal domain, local motion processes are thought to be restricted to a window of 50 ms or less (Baker & Braddick, 1985), while global motion processes may operate over much longer intervals.

A human figure provides local form information such as orientation of individual edges (Figure 1.4a) and global form information about the shape of the body (Figure 1.4b). In the visual system, local form cues are detected in early visual areas whereas global shape generally involves higher areas. For instance, if subjects expect a human figure they may recognize a static set of point-lights as the global form of a human.

If such a figure starts to move, additional information about the movement is obtainable. At the local level, this represents the movement of individual points with certain direction and velocity (Figure 1.4c). These local motion signals are likely to activate local motion detectors. At the global level, the change of the posture over time and thereby the change of the human form gives the global motion information (Figure 1.4d).

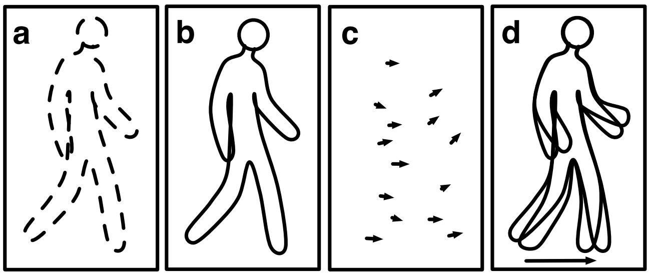

Figure 1.4: Schematic demonstration of local and global features of a human figure - (a) local form (b) global form (c) local motion (d) global motion.

Many studies dealing with biological motion used point-light walkers similar to Johansson's. For a point-light walker the individual points provide no local form information. The fact that we are able to perceive biological motion in these point-lights demonstrates that local form signals are not necessary for biological motion perception.

Global form, on the other hand, does seem to be important. Bülthoff et al. (1998) showed that low-level stereo-depth perception is overridden by top-down influences from global form. Bertenthal & Pinto (1994) provided evidence of a global form analysis by using masks of dots with trajectories identical to walker dots but with different, randomly selected positions. Walker and noise differed only in the global spatial configuration. Subjects were still able to recognize the walking figure, despite the identical motion signals. Casile & Giese (2005) compared orientation discrimination of a different version of a point-light stimulus to a regular point light walker, in which individual points followed sinusoidal trajectories not strictly consistent with the human skeleton. The information on walker orientation in this stimulus is contained in the offsets between the upper and lower body parts, which clearly constitutes a global form cue.

The relevance of local motion is debated. Studies in which the local motion information has been directly manipulated have so far yielded unclear, or at least debatable, results. Mather, Radford, & West (1992) examined the perception of biological motion in random noise. They manipulated the efficiency of low-level motion detectors by inserting temporal delays between subsequent frames (inter-stimulus-intervals) or reversing dot contrast. They found impaired discriminability in coherence or direction tasks when the walker was presented in noise and suggested that local motion is the basis for perception of biological motion. Later, Thornton et al. (1998) showed that observers were able to recognize biological motion using the same set-up as Mather, Radford, & West (1992) but longer presentation times. They found an effect of inter-stimulus-intervals only on short stimulus durations. Ahlström et al. (1997) investigated several properties of biological motion perception in their study and presented evidence, which suggested that perception of biological motion does not rely on first-order motion. They used biological motion defined by luminance, texture or reversal of dot contrast and found no difference in performance. Beintema & Lappe (2002) used a stimulus with near absent local motion. The points of their stimuli had random locations anywhere on the four limbs (arms and legs) and were reallocated every frame to a new random position. With such jumps there is nearly no local motion information (motion vector and trajectory information) carried by each point, without altering the temporal sampling of the sequence. They found an advantage of available local motion only for the recognition of point-light walker in noise, which argues for a role of local motion as an aid for segregation.

The importance of global motion and of the global dynamics of the stimulus motion has recently been demonstrated by Shipley (2003). He found that the unfamiliarity of the dynamic relations by upside-down presentation reduced recognition. Troje (2002) separated movement dynamics and body structure on point-light walkers. His study on gender recognition showed a greater importance for movement dynamics than for body structures.

The influence of local and global information was investigated by Kuhlmann & Lappe (2006) in an action recognition task. They used movies of different actions in natural scenes that were blurred (low-pass filtered) to different degrees. By the mean of blurring the local features are strongly reduced, whereas global form or motion information is mostly unaffected. Results revealed that reductions of local form and local motion information by blurring can be compensated by global form change and global motion.

Also other features may play a role, like size or depth information or even higher level influence as for example the semantic meaning. Motor or sensory representations may influence the perception of actions, as well.

How is the perception of biological motion achieved? What information is relevant and what not? Until now, there are no satisfactory answers to these questions. But there are promising theories. We believe that our brain uses a template-matching process for the recognition of biological motion as suggested by Lange (2006). Our attempt in this study was to investigate the properties of such templates.

1.4 Cortical Representation of Biological Motion

The visual information processing takes place over two parallel and hierarchically organized subcortical and cortical pathways, in which different aspects of the scenery are processed. Visual information crosses the retina and is led over the lateral geniculate nuclei (LGN) to the primary visual cortex (V1) and further on to the secondary visual cortex (V2). The further processing takes place over two anatomically and functionally different processing ways (Ungerleider & Mishkin, 1982) , the dorsal and the ventral path (see illustration 1.5). Proofs for such a division were derived from comparative investigations from patients with lesions in the ventral or the dorsal system (Ungerleider & Mishkin, 1982).

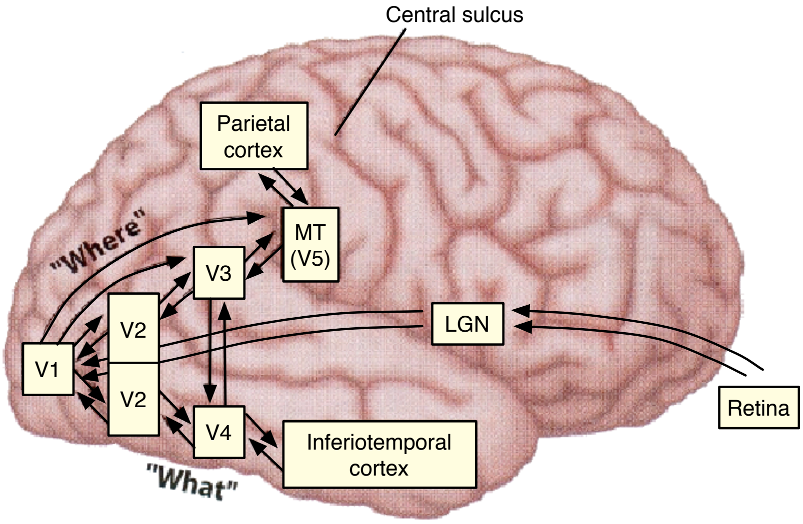

Figure 1.5: The visual information is processed in two anatomic and functional different pathways, termed 'where' (dorsal) and 'what' (ventral) pathway. Both pathways consist of several visual areas. Abbreviations: LGN - lateral geniculate nuclei; V1 - primary visual cortex; V2, V3, V4 and V5 - higher visual areas; MT - middle temporal area. Adapted and modified from Gazzaniga et al., 1998.

The ventral (temporal) path starts with the p-cells of the retina and leads over the parvocellular layers of the LGN to V1. This path runs from V2 to the temporal cortex. It is responsible among others for the processing of form and color information as well as recognizing objects (Livingstone & Hubel, 1988) and faces (Perrett & Oram, 1998). It deals with the ”what” and is often named as ”form” - path. The dorsal (parietal) path has its origin in the m-cells of the Retina and leads over the magnocellular layers of the CGL to different layers in V1. This path leads dorsal into the parietal Cortex. It runs from V2 to the dorsal part of V3 and then to area V5, known as the middle temporal area (MT) in monkeys, and to area V5A, known as the middle superior temporal area (MST). The dorsal path is above all responsibly for recognizing and analyzing spatial relations of objects to each other and for their movement in the environment. It processes the ”where” and is thus regarded as ”motion” - path. Within these paths a gradual more and more specific information processing takes place. At present however, such a strict separation of the two paths as Ungerleider & Mishkin (1982) suggested is no longer assumed. Connections between the two paths and the existence of areas, which receive entrance from both paths, is known.

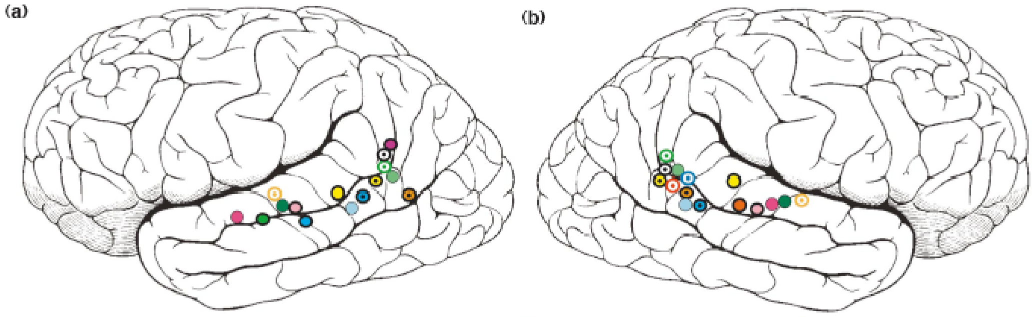

Figure 1.6: Activation of the STS for different kinds of biological motion stimuli. a): left hemisphere, b): right hemisphere. Adapted and modified from Allison et al. (2000). (The individual points represent different studies).

Due to the assumption that the perception of biological motion is mediated by movement signals (Johansson, 1973; Cutting, 1981), also the cortical representation of biological movement was suspected in areas of the movement path. In order to identify these areas, different physiological and neuropsychological studies were performed. Representations of biological movement were presented to the subjects (for example Bonda et al., 1996, Grossman & Blake, 2001, and Vaina et al., 2001). Apart from activation in clear motion areas, also activity in other areas appeared, particularly in the superior temporal Sulcus (STS, see illustration 1.6). The STS is thought to be an intersection between the dorsal and the ventral pathway. Grossman and colleagues found on humans a selective activation in the posterior part of the superior temporal Sulcus (STSp), when viewing biological movement (Grossman et al., 2000). Upside down presentation of point-light-displays produced a clearly lower activation (Grossman & Blake, 2001). Neurophysiological investigations at Macaque (Oram & Perret, 1994, 1996) showed active neurons in the anterior part of the superior temporal Sulcus (STSa), which did not only react selectively to biological movement, but also coded for a certain combination of form and direction. These neurons showed no reaction with other kinds of movement, however.

As mentioned before beside the STS also different other areas are involved in the processing of biological motion (see Figure 1.7). Grossman & Blake (2002) used point-light animations of biological motion, to examine the extent to which form and motion pathways are involved in the perception of biological motion. They showed that in the ventral path point-light displays of human bodies produce weaker answers than faces or complete bodies. The lateral occipital complex (LOC) and the extrastriate body area (EBA) were found active but did not contain neural signals selective for biological motion. Whereas, parts of the fusiform and occipital face area (FFA and/or OFA) were capable of differentiating biological motion from scrambled biological motion. In a study

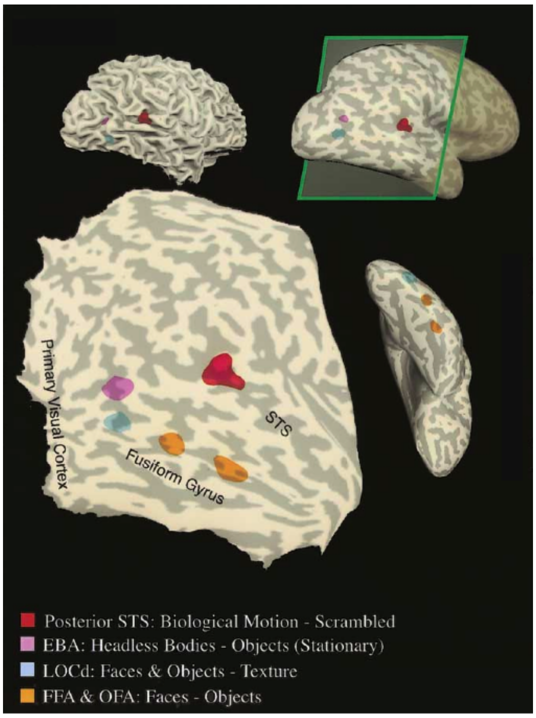

Figure 1.7: Brain areas active in response to biological motion stimuli. Adapted from Grossman and Blake, 2003

from Michels et al. (2005) activation in the FFA and OFA as well as in the EBA was found by presenting classic point-light stimuli and sequential position stimuli. These activations were increased for stimuli containing primarily form information. These results indicate that FFA and OFA as well as EBA in the ventral stream are used for processing biological motion stimuli. The activation of these areas is dependent on the amount of form information.

Motion sensitive areas react nonspecific to biological and not-biological movement (Grossman & Blake, 2002; Michels et al., 2005). These include among others the mediotemporal or the mediosuperior temporal area (MT and/or MST), the lateral occipital complex (LOC) and the kinetic occipital region (KO).

1.5 Computational Models

A better understanding of biological motion recognition and the relevance of the different cues can be gained from computational models (Aggarwal & Cai, 1999; Gavrila, 1999).

One of the first models was proposed by Cutting (1981). His vector coding theory assumes a hierarchical processing of the movement information from the center of the body outgoing to the distal parts of the body. This hierarchical processing resulted in a dependency of the distal joints to the more proximal joints, and thereby implied a dependency of the processing of the movement vectors from individual joints from other joints. However, studies with inter-joint displays (Dittrich, 1993) or studies where individual joint-points were omitted during the presentation (Mather et al., 1992; Pinto & Shiffrar, 1999) could not support this model. The necessary rigidity for a movement-based model is reached in the models of Webb and Hoffman (Hoffman & Flinchbaugh, 1982; Webb & Aggarwal, 1982) by the assumption that all points, which belong to one extremity always move in one plane. Thus, the possible combinations of the joint points could be limited, and led to recognizing rigid connections between the points. These models come to their limits, if the rigidity perception from a stimulus is made impossible. That is for example the case, if one sees a figure only through a very narrow aperture Shiffrar et al. (1997). Therefore, Aggarwal et al. (1998) suggested a so-called A-priori model for human action recognition that performs a comparison with an internally existing template. Here a Top-down process is proposed. Bobick & Davis (2001) first recovered the momentary shape of an actor through a simple local motion segmentation process. Then a space-time pattern of the movement is produced which consists of the sequence of the actor's postures over time. This spatiotemporal shape serves as template and is used for identification. In similar spirit but more related to human perceptual processes, the template-matching model suggested by Lange & Lappe (2006) used a global form template and its temporal evolution for the analysis of biological motion from point-light stimuli. They found similarities between psychophysical data and their model. Giese & Poggio (2003) proposed a template-matching model with two separate hierarchical bottom-up processes; one based on local motion the other on local form analysis. They showed that their model could also explain perception of biological motion in noise with results similar to psychophysical data (Neri et al., 1998). Lee & Wong (2004) recently presented another template-matching model with a form template similar to the one Lange, Georg and Lappe used. In contrast to Lange & Lappe (2006) they used point-light templates instead of stick-figure templates.

1.6 Objective of this Thesis

Biological motion stimuli are complex stimuli containing motion and form cues and involving rigid and non-rigid elements. As described above, there is still disagreement about the fundamental perceptual processes underlying perception of biological motion. Taken together current research results, it is still unclear if form or motion information builds the basis of the perception process. A convincing approach is the use of template-matching for the recognition of biological motion. The template matching model of Lange & Lappe (2006) assumes static template cells at the first stage. Possible neural correlates may be EBA or FFA, which are sensitive to static postures of human bodies (Downing et al., 2001; Peelen & Downing, 2005). At the next stage the temporal order is analyzed. For this stage STS would be a possible neural correlate as it is sensitive to the global motion of a point-light walker (Grossman et al., 2000; Vaina et al., 2001).

The attempt of this study is to investigate the properties of such templates particularly to clarify the role of form and motion information for the perception of a biological motion. As tasks forward/backward discrimination and action recognition are used, both tasks, which require global motion integration and cannot be solved solely on the basis of spatial integration as for example coherence or direction discrimination. Chapter 2 investigates the possible aid of local motion in perspective projections by the mean of different views and different walker types. Performance on forward/backward walking discrimination will be compared of walkers in orthographic and perspective projection when view orientations and point lifetime are varied. Chapter 3 describes an experiment in which action recognition is compared for different stimuli-types. The performance on classic and sequential position stimuli with either a point-lifetime of one frame (no valid local motion) or with a point-lifetime of four frames (valid local motion information) is examined. In chapter 4 a novel stimulus is used. In this stimulus counter-changing luminance is used to avoid local motion. The results of each chapter are discussed in the context of other psychophysical studies and finally, I will present a general conclusion.

Part II

Experimental Section

Chapter 2

Perception of Limited Lifetime Biological Motion from Different Viewpoints

Contents

2.1 Abstract

Studies with time-limited point-lights suggested that biological motion does not require local motion detection. These studies used walkers seen from the side, but biological motion perception excels also when walkers are oriented towards the observer, or in intermediate, half-profile views. In perspective projection, the local motion of points on the body provides a cue to the 3D structure of the walker. Thus, local point motion that was irrelevant for walkers in profile view may become important for biological motion perception in perspective projection. Performance on forward/backward walking discrimination was compared of walkers in orthographic and perspective projection when view orientations and point lifetime was varied. There was no difference between orthographic and perspective projections. Walkers with point lifetime 1 allowed forward-backward discrimination reliably in non-profile views, suggesting that local image motion is not required. Discrimination performance became extremely difficult in the frontal view, however. Follow-up experiments that tested lifetime, view orientation, and specific information from the feet indicated that this dependence on viewing angle can be explained by the reliance of the forward/backward discrimination on information about the movement of the lower legs, which is difficult to ascertain in the frontal view.

2.2 Introduction

Our visual system is highly sensitive to the movement patterns of other living creatures. This ability is so well developed that we obtain an immediate, vivid percept of a walking human, already from seeing just a few points attached to the joints of a moving body (Johansson, 1973). Point-light displays contain both form and motion information. Each point at each time provides position information about a single spot on the body. Integration of the positions of multiple points, either per frame or over time, yields form information about the configuration of the body. At the same time, the temporal evolution of the positions of each single point provides local motion, acceleration, and trajectory information for that point.

The limited-lifetime technique can be used to investigate the contributions of motion, acceleration, and trajectory of individual points, while leaving global form intact. In limited-lifetime stimuli each single point is shown only for a limited number of successive image frames, after which it is extinguished. The number of frames that a point lives determines whether this point offers motion, acceleration, or trajectory information to the viewer. If the lifetime is limited to only a single frame the point cannot offer motion information because it is not moving with the limb between frames. The minimum lifetime for motion is two frames because then apparent motion sensors can be activated. A higher lifetime may improve the local motion sensing by spatio-temporal integration. If the point moves in a straight line the motion measurement will become more robust. If the point moves along a curved trajectory, on the other hand, simple spatio-temporal integration would introduce errors since the motion direction is changing between each pair of frames. Lifetimes longer than two frames also offer acceleration information, i.e. how the local motion changes over time. Lastly, the longer the lifetime the more information about the trajectory of the point is available. The trajectory is the curve in space that the point traverses over time and is independent of direction or speed of the motion of the point. The trajectory cannot be calculated at any moment in time but is a shape that must be estimated from observing the positions of a point over time.

The limited-lifetime technique was first applied to biological motion by Neri et al. (1998) who used a lifetime of two frames in a walker with only six points placed randomly on the main joints of the body. Beintema & Lappe (2002) examined the role of local motion and global form with limited-lifetime walkers in which the individual points appear at random locations on the limbs of the body. Local motion of these points was manipulated by limiting the lifetime of the points, i.e., the number of frames that a point moves along with a single spot on the body. When lifetime was reduced to one frame only, nearly no local motion information was present because the points did not follow the movement of the body. Nevertheless, naive observers spontaneously recognized these animations as human walkers (Beintema & Lappe, 2002) and could reliably judge the facing direction and the coherency of a walker, as well as discriminate between forward and backward walking (Beintema et al., 2006). Thus, local motion was not necessary for these tasks.

Lange et al. (Lange & Lappe, 2006; Lange et al., 2006) have suggested that a template matching analysis of the body configuration may underlie biological motion recognition. In this model, the positions of points in each stimulus frame are matched to templates of the human body in different postures. Local image motion from individual points is not used. The motion of the body is derived from analyzing the evolution of the best-matching body postures over time.

Thus, from experimental observations and computational considerations local image motion does not appear necessary for biological motion analysis. However, experiments that used the limited-lifetime technique have so far only used profile views of walking in orthographic projection (Figure 2.1 A). It is important to test the usage of local image motion in other view orientations and in perspective projection because the combination of profile view and orthographic projection is a special case for two reasons.

The first reason is the difference between orthographic and perspective projection. In orthographic projection, a point $P = (X, Y, Z)^T$ on the body is projected onto a point $p_{orth} = (x, y)^T$ in the image so that

$$p_{orth} = \begin{pmatrix} X \\ Y \end{pmatrix}. \tag{2.1}$$

Here, the projection is without loss of generality assumed to be along the Z-axis. Image coordinates (x, y) directly correspond to world coordinates (X, Y), and the depth coordinate Z is lost in the projection. The image motion v_{orth} of image point p_{orth} is

$$v_{orth} = \frac{d}{dt} p_{orth} = \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix}. \tag{2.2}$$

Therefore, the image motion is independent of the motion-in depth, $\dot{Z}$, of point P along the Z-axis. Any information about the motion-in depth component of the point on the body therefore has to be gleaned from the motion along the X- and Y-axes. This requires knowledge of the structure of the human body, as, for instance, provided by a template of the body. The visual information in image point positions and image point motions is mathematically insufficient to estimate body posture and movement (Ullman, 1984) and perceptual recognition can only be achieved when additional assumptions about the structure or movement of the body are introduced. This can be done either by assuming explicit body models (Rashid, 1980; Marr, 1982; Chen & Lee, 1992; Aggarwal & Cai, 1999) or biomechanical constraints on the body motions (Webb & Aggarwal, 1982; Hoffman & Flinchbaugh, 1982).

The mathematical insufficiency of the visual position and local image motion signals for biological motion recognition also holds for perspective projection. However, unlike in orthographic projection, the position and image motion signals in perspective projection contain information about the Z (depth) component of the walker. In perspective projection, point P is projected onto p_{persp} so that

$$p_{persp} = f \frac{1}{Z} \begin{pmatrix} X \\ Y \end{pmatrix}, \tag{2.3}$$

where f is the focal length of the projection. The image motion v_{persp} of image point p_{persp} is

$$v_{persp} = \frac{d}{dt} p_{persp} = f \frac{1}{Z^2} \begin{pmatrix} \dot{X}Z - X\dot{Z} \\ \dot{Y}Z - Y\dot{Z} \end{pmatrix} = f \frac{1}{Z} \begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} + f \frac{\dot{Z}}{Z^2} \begin{pmatrix} X \\ Y \end{pmatrix}. \quad (2.4)$$

Therefore, the image motion in perspective projection consist of a part that is specified by the motion $(\dot{X}, \dot{Y})^T$ of P in X and Y directions and a part that is specified by the motion-in-depth, $\dot{Z}$.

The comparison of the two projections shows that in orthographic projection all information in both the positions p_{orth} of image points and the motions v_{orth} of image points is related only to the X and Y coordinates of the body. Information about the Z component of the body structure and its motion is missing from the stimulus and can only be reconstructed by using external knowledge of the body structure. In perspective projection, on the other hand, both the positions p_{persp} of image points and the motions v_{orth} of image points carry information about the depth Z. Most importantly, p_{persp} and v_{persp} carry independent information about depth, because p_{persp} depends on Z, i.e., the position in depth of point P, and v_{persp} depends on Z and also on $\dot{Z}$, i.e., the motion of P in depth. Therefore, in perspective projection the local image motion of a point may convey information over and above the information conveyed by the point positions. Hence one must ask, whether local image motion, which has previously been shown to not contribute to perception in the orthographic projection, will contribute in the case of perspective projection.

The second reason why the combination of profile view and orthographic projection is a special case has to do with the shape and limb movement of the walker. In profile view, the movement of the limbs is almost exclusively in parallel to the image plane. Since there is little motion along the depth axis, the lack of information about Z-axis motion in the orthographic projection is of no influence.

In fact, in orthographic projection in the profile view, the depth distribution of the light points of the stimulus is entirely ambiguous and the stimulus is mathematically indistinguishable from a flat arrangement of light points in a single depth plane. For a template matching recognition procedure it would be sufficient to match the stimulus frames to two-dimensional templates. The true three-dimensional structure of the body becomes visually more apparent when the walker is shown in other view orientations and in perspective projection. For instance in the half-profile view (Figure 2.1 B), the movement of the limbs is directed in depth, and, because of the perspective projection, the visual speed of the limb movement gets smaller when the limb is further away then when it is closer to the observer. Thus, in these stimuli, visual speed is an independent cue to distance and hence to the three-dimensional structure of the stimulus.

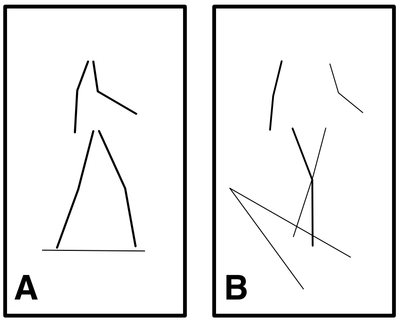

Figure 2.1: A: 2D orthographic projection, profile view and B: 3D perspective projection, half-profile view

In perspective projection, visual speed is also informative about the depth structure of the walker in profile view. Consider, for example, the movement of the shoulders. The shoulder nearer to the observer will move faster than the shoulder further from the observer. Thus, in perspective projection the visual motion of points on the body provides a cue to the 3D structure of the walker. In orthographic projection, the speed of point movement is independent of the distance to the observer.

Limited-lifetime experiments with walkers in profile view in orthographic projection showed no influence of local point motion on biological motion perception. However, in perspective projection, and in view orientations other than the profile view, local point motion carries information about the 3D structure of the walker. Thus, local point motion that is irrelevant in orthographic projection may become important for biological motion perception in perspective projection. We wanted to test whether this is the case.

From Johanssons demonstrations and a number of further studies (Verfaillie, 1993; Mather & Murdoch, 1994; Troje et al., 2005) it is known that observers not only readily recognize profile views of point-light walkers, but also point-light walkers seen in other view orientations. In this case, point-light actions convey a strong impression of depth even if static low-level depth cues are missing (Vanrie et al., 2004). The depth percept conveyed by a point-light walker even dominates over conflicting disparity depth cues (Bülthoff et al., 1998). It is possible that local motion information, which is not necessary in the profile view, aids the depth perception process in other views by exploiting the relationship between speed and depth in the light point motion (Ullman, 1984). On the other hand, depth perception of 3D walkers could also be achieved by template-matching without exploiting local motion signals. Such template matching could either use 2D templates for particular viewpoints or full 3D representations of the walker.

In the present study, 3D limited-lifetime walkers were used to investigate the role of local motion in the perception of biological motion for the case of differently oriented 3D walkers. We asked observers to discriminate between a display of a forward walking figure and the same display in reversed order (similar to backward walking). In profile view this task is easy even with lifetime 1, so that it does not require local image motion (Beintema et al., 2006). We were interested whether this also holds true for other viewing angles. Specifically, as described above, image motion signals might convey information about the motion-in-depth of a point. If this is indeed the case, one would expect a difference in performance for non-profile views between orthographic and perspective projection. Moreover, if local point motion is important for biological motion perception in non-orthographic views, we would expect to find an advantage for lifetime 2 over lifetime 1 in perspective projection.

2.3 General Methods

2.3.1 Subjects

Seven subjects (24-35 years, 3 females) participated in the experiments. All of them were experienced with psychophysical experiments involving biological motion stimuli. Apart from authors SK and MdL, the participants were naive to the objective of the experiments.

2.3.2 Stimuli

Stimuli displayed walking human figures, which consisted of white points (0.15 x 0.15 deg) on a black background. Width and height of the stimulus subtended approximately 5 x 9 degree visual angle. The stimuli were based on the 3D joint positions of nine walking humans (5 male and 4 female) recorded using MotionStar Wireless (Ascension Technology Corp., Burlington, USA). The forward translation was subtracted giving the impression of walking on a treadmill. Walking speed was normalized so that a complete walking cycle, consisting of two steps, took about 1.4 seconds. The stimulus sequence was either presented in normal (forward walking) or reversed (backward walking) frame order. The walker started from a random phase in the step-cycle and was shown for one complete walking cycle of 1.4 s. All walkers were presented in perspective and orthographic projection.

For the limited-lifetime walkers the points on the walker were assigned a random position on one of 8 limb segments (upper and lower parts of the arms and legs). The possible positions were distributed uniformly across the segments, each segment defined by the line connecting joints. The lifetime of a point, defined as the number of frames before the point was relocated to another location on the body, could be varied. Relocating the points to a new random location on the limbs after a limited number of frames disturbs the continuous motion and removes the local motion information (motion vector and trajectory information) carried by each point, without altering the temporal sampling of the sequence. The points were relocated in an asynchronous fashion.

The total number of points per trial is an important parameter for the performance (Beintema et al., 2006). It is calculated by multiplying the number of points per frame with the number of frames seen in the trial. This calculation is independent of whether the points stay on the same limb position over successive trials or not, since in both cases each frame provides a certain number of points that signal the current posture. For example, a stimulus with 4 points per frame provides over 8 frames a total of 32 points no matter if the lifetime is 1 or 8. In the former case, new point positions on the limbs are chosen in each frame. In the latter case, the same point locations on the limbs are used in each frame but because the body posture changes over those 8 frames each point provides new body posture information over the last frame. Thus, the total amount of body posture signals is the same in both cases but the latter condition, in addition, provides local motion and trajectory signals of each point. Conditions with either 128, 512, or 384 points per trial were used .

Depending on the experiment, different combinations of the following stimulus conditions were used. Limited-lifetime walkers had two, four or twelve points per frame. For the two points per frame condition the total number of points per trial amounts to either 128 points with a frame duration of 22.2 ms or 512 points with a frame duration of 5.56 ms (i.e. 4 or 1 multiples of the 180 Hz at which the monitor displayed). For the four points per frame condition the total number of points per trial amounts to either 128 points with a frame duration of 44.4 ms or 512 points with a frame duration of 11.1 ms (i.e. 8 or 2 multiples of the 180 Hz at which the monitor displayed). For the twelve points per frame condition the total number of points per trial amounts to 384 points with a frame duration of 44.4 ms (i.e. 8 multiples of the 180 Hz at which the monitor displayed).

It is important to note that because of visible persistence the apparent number of simultaneously present points on the screen was higher than the number of points presented in each frame. Visible persistence describes the apparent duration of a point that is briefly flashed. It has been shown that brief flashes of light, such as the points that were presented for durations between 5.56 ms and 44.4 ms, remain visible for longer temporal intervals, up to 100 or 200 ms (Brown et al., 1974; Coltheart, 1980). Therefore, the stimuli appeared to consist of more than the 2 or 4 points, which they physically contained. It is not known at what level of the visual pathway visible persistence is created, or whether it contributes to form recognition. For our analysis we focus on the number of points that are physically provided in each frame since this is the source of information present in the stimulus.

As a further stimulus condition, classic Johansson walkers were used which consisted of 12 light-points, displaying the joints of shoulders, elbows, wrists, hips, knees and ankles. In the last experiment a modification of the classic walker was used, where the foot-point could be positioned at different locations on the lower limb (a more detailed description can be found in the method section of experiment 4).

2.3.3 Procedure and Experimental Set Up

Stimuli were displayed on an Iiyama Vision Master CRT Monitor (40 x 30 cm, 800 x 600 pixel) at a vertical refresh-rate of 180 Hz. The subjects were seated in a darkened room with their eyes about 70 cm in front of the monitor. They were asked to fixate a red fixation point in the middle of the screen. The walkers were presented in the center of the screen. Walkers measured 5 x 9 degrees of visual angles. When the walker disappeared the subjects had to press a response key. Thereafter a new trial started and a new walker appeared after 200 ms. The subjects task was to detect the walking direction (forward/backward) of the walker; pressing the ‘up’ (forward) and ‘down’ (backward) arrow keys of the keyboard.

2.3.4 Data Analysis

The proportion of correct responses was assessed. T-tests or repeated measures analysis of variance (significance level = 0.05) on the d’ values were conducted for statistical testing. The Scheffé-test was used as a posteriori procedure. For all post hoc tests an alpha significance level of 0.05 was used. Error bars in the figures give the standard error of the mean.

2.4 Experiment 1

In the first experiment the question was whether local motion information could improve performance on forward/backward discrimination in perspective projection, when the walkers are presented in different view orientations.

2.4.1 Methods

Stimuli were limited-lifetime walkers with points lifetime one (no local motion signals) or two (with local motion signals) frames. The number of points per frame and the total number of points per trial was also varied, since these are parameters that are known to influence the performance in the profile view (Beintema et al., 2006). The limited-lifetime walkers had either two or four points per frame. A further stimulus was the classic Johansson walker, which consisted of 12 light-points, displaying the joints of shoulders, elbows, wrists, hips, knees and ankles.

All walkers, the limited-lifetime walkers as well as the classic walkers were presented in perspective projection as well as in orthographic projection. Both walker types were also presented in three orientations, the profile view (0°), the half profile view (45°), and the frontal view (90°).

Each experiment session had 96 different conditions for the limited-lifetime walkers (2 lifetime x 2 points per frame x 2 points per trial x 2 play-directions x 2 projections x 3 orientations) and 12 for the classic walker (2 play-directions x 2 projections x 3 orientations) with 9 repetitions for each condition. One experiment session consisted therefore of 972 trials and took about 20 minutes. In each session all limited-lifetime walker and classic walker conditions were presented in randomized order. Each subject conducted three experiment sessions. The task was to detect whether the walkers walk forwards or backwards. The answers were giving by the ‘up’ and ‘down’ arrow keys of the keyboard.

2.4.2 Results

Whereas orthographic projections do not contain direct information about the structure and motion in depth, the perspective projection does. The performances for the orthographic and perspective projections were compared for the different

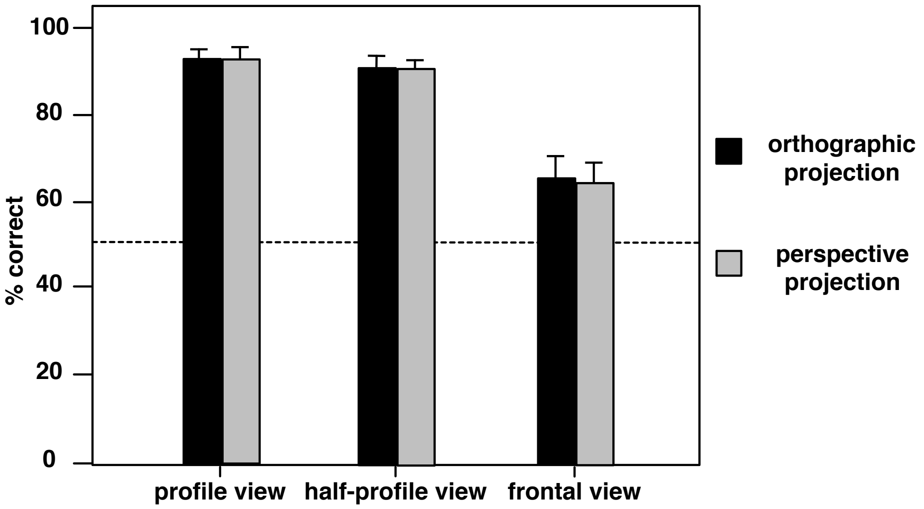

Figure 2.2: Experiment 1: Comparison of discrimination performance in orthographic and perspective projection.

views (Figure 2.2). For both cases equally good performance was found for profile and half-profile view and poor performance for the frontal view. To test for statistical differences a 2-way ANOVA on the factors projection and viewing angle was conducted (2 x 2 design with repetition on subjects). The main effect of viewing angle was statistically significant (F(1,6) $=$ 10.3, p$<$ 0.01). Importantly, there was no significant difference between the two projection types (F(1,6) $=$ 663.8, p$=$ 0.27). This indicates that the participants had no advantage of the perspective projection. Theoretically, the perspective projection also yields information about the motion-in-depth of the individual points, as explained in the Introduction. To test this possibility directly, a more detailed analysis of the lifetime conditions was performed.

The results for perspective projection in profile view are illustrated in Figure 2.3 A. In all conditions the performance with a lifetime of one frame was as good as with two frames. Performance was generally better with 512 points per trial than with 128 points per trial. With 512 points per trial the performance for the

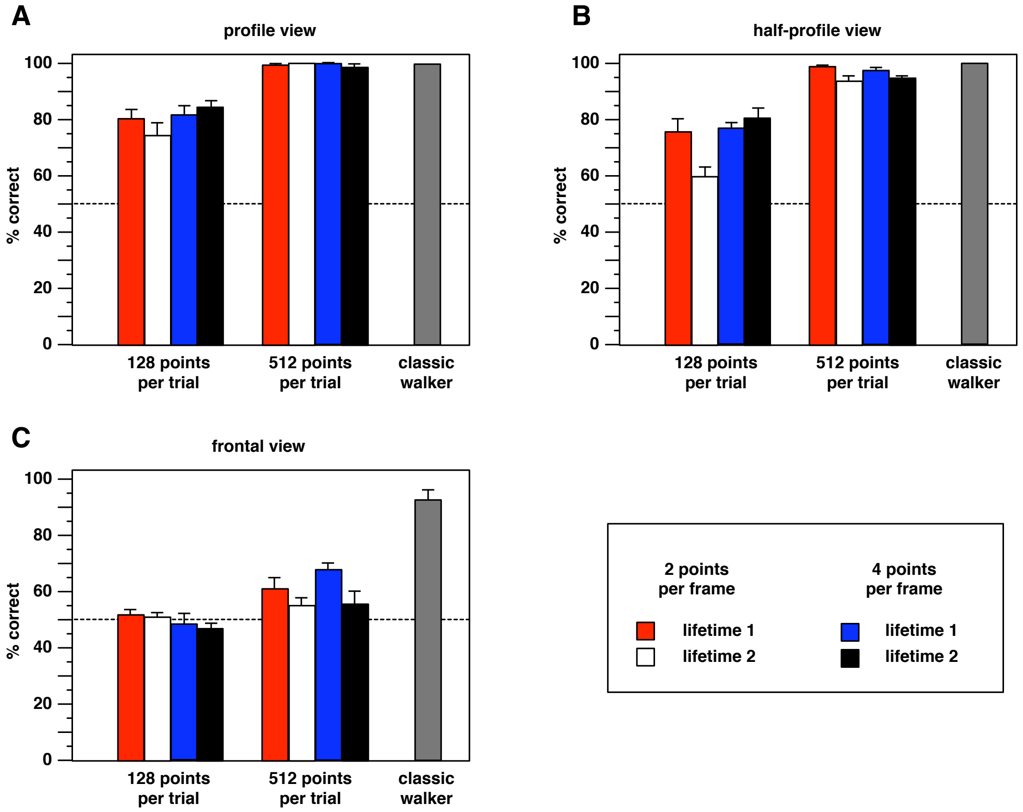

Figure 2.3: Experiment 1: Averaged percentage of correct responses for the profile, half-profile and frontal view in perspective projection. Data of the limited-lifetime walkers are split by the factors points per trial, points per frame and lifetime. Results of the classic walker are added for comparison. Error bars represent the standard error over subjects.

limited-lifetime walkers was as good as the performance for the classic walker independent of the number of points per frame. The results for the half-profile view were similar to the results for the profile view (Figure 2.3 B). A higher lifetime had no positive effect on the performance, and performance for 512 points per trial approached that of the classic walker.

To test for statistical differences between profile and half-profile view within the limited-lifetime conditions a 3-way repeated measures ANOVA with the factors lifetime, points per frame and viewing angle was conducted (2 x 2 x 2 design with repetition on subjects). To exclude ceiling effects from the testing, only the 128 points per trial conditions for the profile and half-profile view were included. There were significant main effects of viewing angle (F(1,6) = 10.3, p= 0.02), and points per frame (F(1,6) = 22.8, p= 0.003), but not for lifetime (F(1,6) = 1.9, p= 0.2). There was a significant interaction between points per frame and lifetime (F(1,6) = 16.5, p= 0.007). No other interactions were significant.

For the frontal view (see Figure 2.3 C) the performance for limited-lifetime walkers with 128 points per trial was not different from chance level (t-test; $p > 0.05$). In the two conditions with 512 points per trial and lifetime 1 the performance was significantly above chance level (t-test; for 2 points per frame $p = 0.01$ and for 4 points per frame $p = 0.004$) but still worse than the performance for the classic walker.

For comparison, the results for orthographic projection are shown in Figure 2.4. Results were very similar to those of the perspective projection, consistent with the overall analysis provided in Figure 2.2. For the profile and half-profile views, the performance with a lifetime of two frames was in no condition better that that with one frame. Performance was generally better with 512 than with 128 points per trial and approached that of the classic walker, independent of the number of points per frame. A 3-way repeated measures ANOVA with the factors lifetime, points per frame and viewing angle (2 x 2 x 2 design with repetition on subjects) on the 128 points per trial condition showed significant main effects of viewing angle (F(1,6) = 15.5, p= 0.008), points per frame (F(1,6) = 19.7, p= 0.004) and lifetime (lower performance with lifetime two, F(1,6) = 12.5, p= 0.01) and no significant interactions. For the frontal view (Figure 2.4 C) the performance walkers with 128 points per trial was not different from chance level (t-test; $p > 0.05$). The performance for the 512 points per trial conditions was

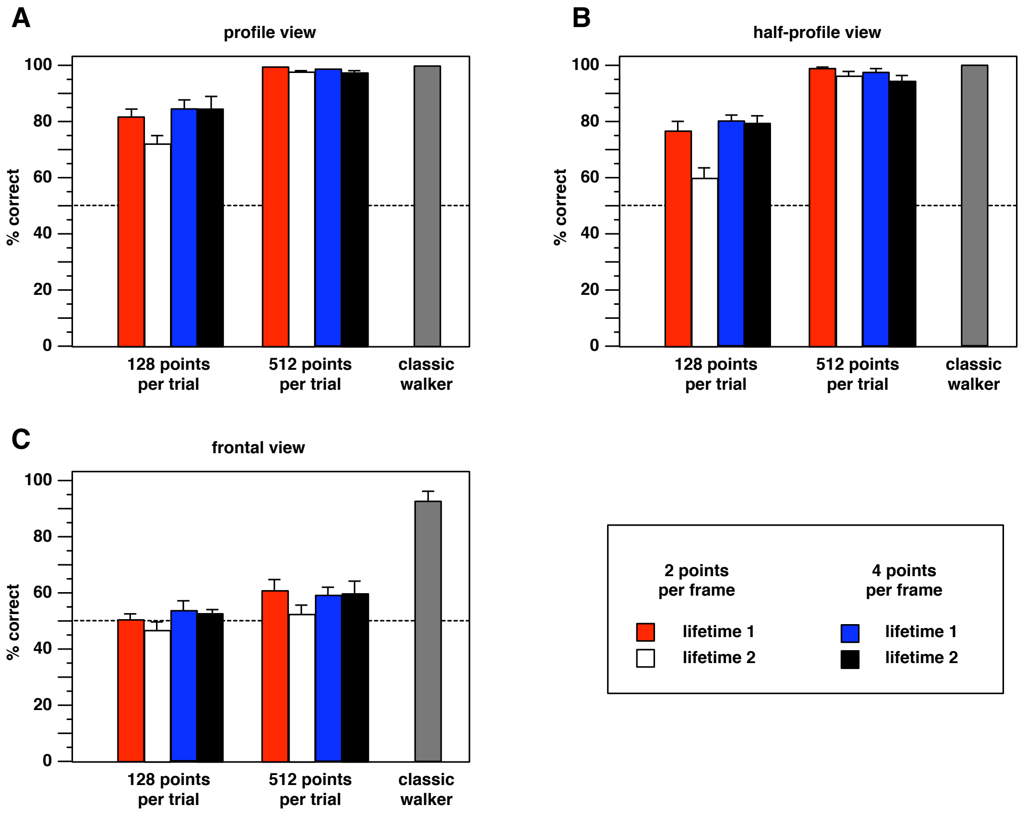

Figure 2.4: Experiment 1: Averaged percentage of correct responses for the profile, half-profile and frontal view in orthographic projection. Data of the limited-lifetime walkers are split by the factors points per trial, points per frame and lifetime. Results of the classic walker are added for comparison. Error bars represent the standard error over subjects.

lower than the performance for the classic walker, but significantly above chance level (t-test; $p < 0.01$), except for the condition with 2 points per frame and lifetime 2.

2.4.3 Experiment 1B

The results indicate that performance did not benefit from local motion information. However, in the previous experiment local motion was present only for 2

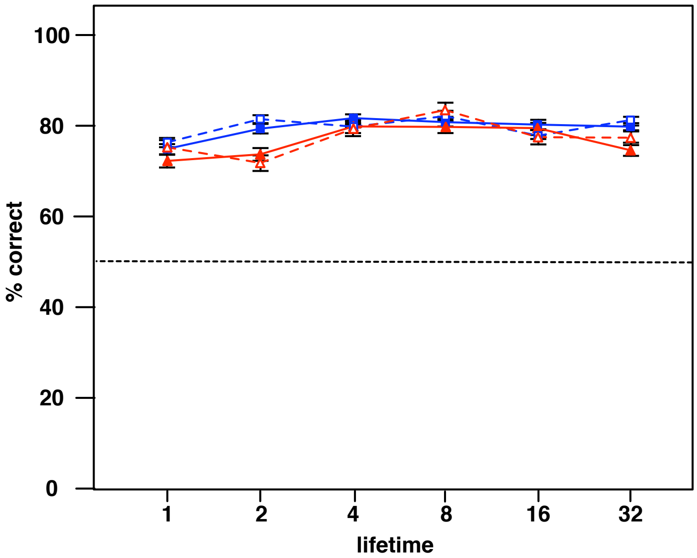

Figure 2.5: Experiment 1B: Averaged percentage of correct responses as a function of point lifetime. Data are split by profile (continuous lines) and half-profile view (dotted lines) and by orthographic (squares) and perspective projection (triangles). Error bars represent the standard error over subjects.

consecutive frames. That is, in the condition without local motion, points were relocated to a new position in every frame, while in the condition with local motion, points were relocated every two frames. To confirm that there is no benefit from local motion information an additional experiment was performed to test a wider range of point lifetimes. The lifetimes varied between 1 and 32 frames. This experiment concentrated on the condition with 128 points per trial and 4 points per frame since performance with 512 points per trial was already almost saturated in the lifetime 1 and 2 conditions, and the 2 points per frame condition with 128 points per trial in most cases indicated a decline in performance for lifetime 2 over lifetime 1. Therefore, chances to see any benefit from higher lifetime would be highest in the 128 points per trial and 4 points per frame conditions.

Seven subjects took part in this experiment, four of them also participated in

Experiment 1, three were new subjects. The seven subjects included two of the authors of the study. Each of the seven subjects conducted three experimental sessions. Each session had 48 different conditions (6 lifetime x 2 play-directions x 2 projections x 2 orientations) with 9 repetitions, and consisted therefore of 432 trials. Stimuli had 128 points per trial and 4 points per frame and were shown in profile and half-profile view. The frontal view will be tested separately in Experiment 3.

The results are displayed in Figure 2.5. Performance remained around 80 percent correct in all conditions. There was no effect of lifetime on performance in any condition. A 3-way repeated measures ANOVA with the factors lifetime, perspective and viewing angle (6 x 2 x 2 design) gave no significant effects or interactions.

2.4.4 Discussion

As explained in the Introduction, perspective projection of the movements of the light points might help to recognize biological motion if the walker is presented in non-profile views. Experiment 1 revealed no differences between the perspective and orthographic projections for any view, indicating that such motion-in-depth cues do not improve the recognition.

Overall performance differed between views. With respect to the half profile view, a possible advantage of perspective motion (lifetime 2 and higher) over non-motion (lifetime 1) might have been expected if local motion contributes to the perception of the 3D structure of the walker. However, increasing the lifetime did not have a positive effect on the performance. This showed that local motion information is not necessary for performing the task, and even does not give any advantage. In contrast, with 128 points per trial there even was a significant decrease from the profile view to the half-profile view, which was independent of all other factors. Discrimination performance for the limited-lifetime walkers otherwise was similar for the profile and the half profile view. Performance for 512 points per trial was as high as for classic walkers in profile and half-profile view.

There was a strong decrease in performance in the frontal view compared to the profile or half-profile view. Even with 512 points per trial the performance was a lot poorer than for the classic walker. For the classic walker there was also a small decrease in performance for the frontal view (t-test; $p = 0.04$). Results from both walker types suggest that the discrimination of walking direction in the frontal view is more difficult than in other views.

2.5 Experiment 2

The first experiment answered the main question in that the results showed no influence of local motion information on the perception of biological motion in perspective projection. The results of the first experiment suggested further that the frontal view is a special case for the forward/backward detection task. Therefore, in the following experiments the problem of the frontal view is further examined. In Experiment 2 it was asked whether the poor performance for limited-lifetime walkers in the frontal view is limited to just the frontal view or whether performance gradually decreases between 40° and 90°. In Experiment 3 benefits from longer point lifetime the frontal view were investigated. In Experiment 4 the role of the foot-points for the frontal view was investigated.

2.5.1 Methods

Limited-lifetime walkers in different viewing angles were presented randomly. The orientations of the walkers varied in ten-degree steps from 40 to 90 degrees. The walkers had 4 points per frame and 512 points with a frame duration of 11.1

ms. Only limited-lifetime walkers with lifetime 1 were used as there was no benefit for higher lifetimes in Experiment 1. All other methods were identical to those of Experiment 1.

2.5.2 Results and Discussion

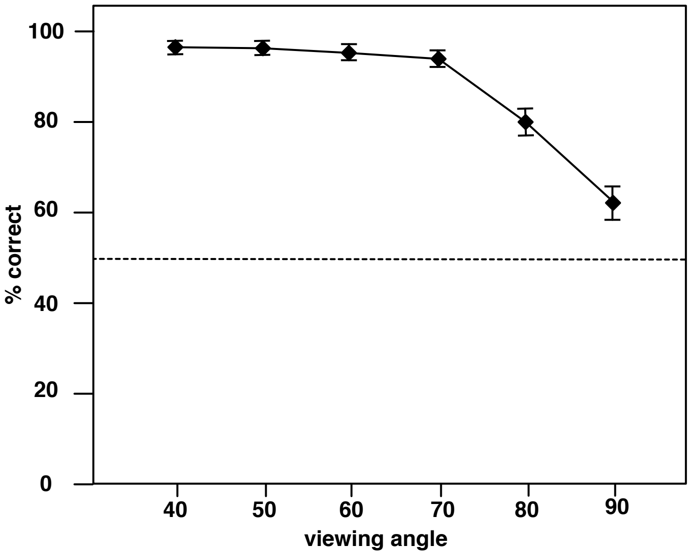

Figure 2.6: Experiment 2: Averaged correct responses as function of the viewing angle. The stimulus was a limited-lifetime walker with lifetime 1, 8 points per frame, and 512 points per trial. The error bars represent the standard error over the individual subjects.

The percentages of correctly recognized walking directions are displayed in Figure 2.6. Performance was about equally good between 40° and 70° and with approximately 96 percent similar to the performance for the profile view in Experiment 1. From 70° to 80° the performance dropped to 80 percent and to 60 percent for 90°. However, performance remained above chance level even in the frontal view (t-test; frontal view $p = 0.02$; all other views $p > 0.0001$).

A repeated measures ANOVA showed a significant influence of viewing angle

(F(5,30) = 65.7 , $p < 0.0001$). Paired comparisons by Scheffé tests revealed no significant differences between 40° and 60° and neither between 60° and 70°. All other comparisons were significant ($p > 0.0003$). Thus, one can conclude that a forward backward discrimination is more difficult in the frontal view than in other views.

2.6 Experiment 3

The two above experiments showed that identifying the walking direction of a limited-lifetime walker was especially difficult in the frontal view. Although there was no difference between lifetime 1 and 2 in Experiment 1, it may be that information that becomes available only at longer point lifetimes is important. Therefore, the question was whether an increase of lifetime beyond two frames could lead to an enhancement in performance in the frontal view. In Experiment 3 the lifetime was gradually increased until the lifetime of the limited-lifetime walker was similar to the lifetime of a classic walker. The limited-lifetime walkers were compared with a classic walker as control condition.

2.6.1 Methods

Limited-lifetime walkers with a lifetime of 1, 4, 8, 16, or 32 frames were used. For the lifetime 32 condition each of the points changed its position on the body only once per trial. The limited-lifetime walkers had 12 points per frame, identical to the classic walker, and 384 points per trial with a frame duration of 44.4 ms (i.e. 8 multiples of the 180 Hz at which the monitor displayed). In addition to the limited-lifetime walkers also a classic walker was presented, which consisted of 12 light-points, displaying the joints of shoulders, elbows, wrists, hips, knees and ankles.

2.6.2 Results

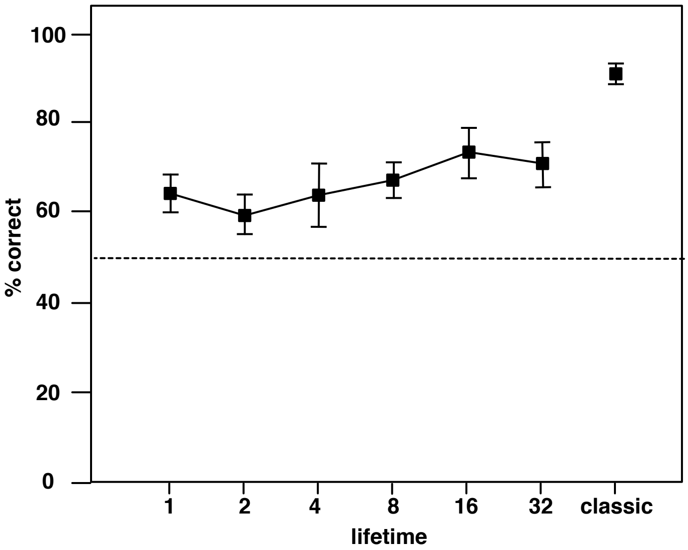

Figure 2.7: Experiment 3: Averaged correct responses in the frontal view as function of point lifetime. The error bars represent the standard error over the individual subjects.

The results are displayed in Figure 2.7. Discrimination performance was high for the classic walker. For the limited-lifetime walker, overall performance was between 60 and 75 percent and never reached the performance level of the classic walker. A one-way repeated measures ANOVA revealed a significant main effect over conditions (F(6,25) = 10.4, $p < 0.0001$). The Scheffé tests revealed significant differences between the classic walker and all limited-lifetime walker conditions ($p < 0.04$), but no differences between the limited-lifetime walker conditions.

2.6.3 Discussion

Increasing the point lifetime yielded to no significant improvement of performance, so there was still a clear difference between lifetime 32 and the classic walker. Thus, the difference between classic and limited-lifetime walker cannot be explained by a difference in lifetime, or local motion information, respectively.

The limited-lifetime walker with a lifetime of 32 frames resembled the classic walker in many aspects. Both possessed similar amounts of structural and local motion information, as the points of the lifetime 32 condition were relocated only once during the trial. However, the trajectories of the light points of the two walkers differ strongly because the points occupied different positions on the limbs.

For the classic walker the knee-, elbow-, hip-, shoulder-, and wrist-points move with approximately sinusoidal velocity profiles. Therefore their trajectories are symmetric in shape and carry practically no information about the walking direction. In contrast, the trajectories of the feet are asymmetric. They show a long backward movement with a slow rise when the foot is lifted, and a quick forward movement with a sharp drop when the foot is put down. This asymmetry of the foot trajectory carries information about the walking. For instance, when walkers facing to the left have to be discriminated from walkers facing to the right, subjects rely very much on the visibility of the feet (Mather et al., 1992). Moreover, even when the other joints of the body are dislocated, the movement of the feet alone supports the perception of walking to the left or walking to the right (Troje & Westhoff, 2006). The information used is the differences in acceleration of the feet in the lift and drop phases (Chang & Troje, 2009). For the limited-lifetime walker, information from the feet is not directly available. Due to the random locations of the points there is not always a point located in the vicinity of a foot. Moreover, due to the relocation of the points, the amount of trajectory information from a single point depended on the lifetime of the point and was less than a walking cycle. Therefore, a comparison of the drop and the lift phases was not always possible.

Although the role of the feet for the discrimination between forward and backward walking has not been studied so far, and other views than the profile view have not been used, it is very plausible that the feet carry most of the information in those cases as well. Experiment 4 was designed to specifically test the role of the feet for walking direction discrimination in the frontal view.

2.7 Experiment 4

To investigate the influence of the information from the foot-points a classic walker was generated on which the foot-points could be shifted toward the knees on the lower leg and thereby shortened the lower leg. The closer the foot-point was shifted to the knee point the less distinct was the asymmetry of the foot trajectory.

2.7.1 Methods

Walkers were generated in which the lowest points was either positioned directly on the ankle, at three quarter of the lower leg, at the half of the lower leg, or at one quarter of the lower leg, and a walker in which the foot-points were entirely omitted. In all other respects the walkers were identical to the classic walker with 12 points. All walkers were presented in the profile view and in the frontal view in blocked conditions.

2.7.2 Results

Figure 2.8 displays the results. Performance was overall lower in the frontal view than in the profile view and decreased with decreasing distance from the lowest point to the knee. For the profile view the performance was high for all inter-joint conditions. Performance was at chance-level when the foot-points were omitted for both profile and frontal view.

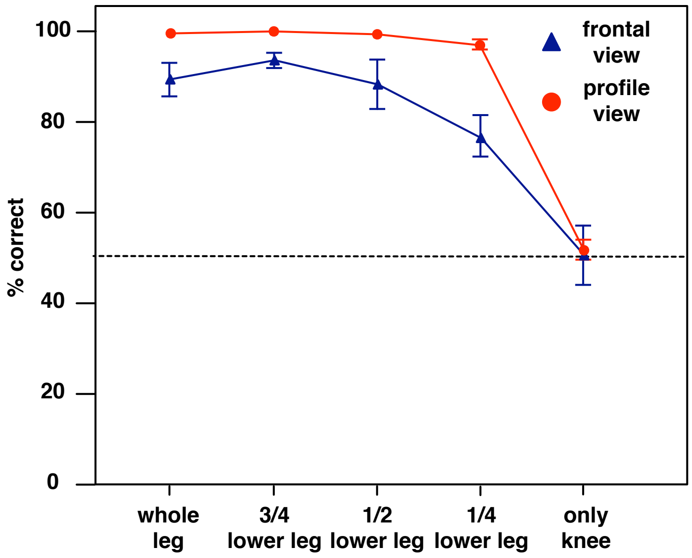

Figure 2.8: Experiment 4: Percentage of correct responses for walkers in which the lowest visible point was on different locations of the lower leg. The error bars give the standard error over subjects.

A two-way repeated measures ANOVA showed a highly significant main effect of the condition foot-point ($F(4,6) = 89.6$, $p < 0.0001$) and a main effect of viewing angle ($F(1,6) = 9.1$, $p = 0.02$). Paired comparisons by Scheffé tests revealed that the significant differences occurred between the condition where the foot-points were presented at one fourth of the leg and the condition where the foot-points were presented at three fourth of the leg ($p < 0.002$) as well as between the condition were only the knee points were shown and every other condition ($p < 0.0001$). Both conditions with missing foot-point did not differ from chance-level (t-test).

2.7.3 Discussion

For both views good performance for most of the inter-joint conditions was found, except for the condition where the lowest points were presented near to the knee-points, at one fourth of the leg. In this condition the subjects performed significantly poorer. For the condition where only knee points were presented the performance was at chance-level.

Therefore, information about the movement of the lower leg is necessary to discriminate walking direction. A similar importance of the foot-points has also been noted for example by Mather et al. (1992), who found that visibility of the trajectory of the feet and wrists are necessary for the discrimination of coherency and facing direction. Simulations of the template-model (Lange & Lappe, 2006) supported Mather's conclusion and suggested this is because the configuration of the extremities carries the most of the information about facing direction. Troje & Westhoff (2006) examined the inversion effect of biological motion and found that inverting only the feet of point-light displays has a much stronger detrimental effect than inverting all points except the feet. Chang & Troje (2009) showed that the comparison of the acceleration of the feet in the lift and drop phases is important.

However, the discrimination performance was above chance level even when the point was presented at different locations between foot and knee. This suggests that it is not the foot itself that is important but rather information about the movement or the configuration of the lower leg. Moreover, in Experiment 2 discrimination was possible for the limited-lifetime walkers in view angles between 0 and 70 even though the trajectory of the foot was not directly available in those stimuli. In this case, information about the movement of the lower leg may be derived from a template analysis of the body configuration similar to that proposed for walkers in profile view (Lange & Lappe, 2006; Lange et al., 2006). The same can be expected for classic walkers with the lowest point placed somewhere on the lower leg since the movement of the leg can be inferred from the motion of a point on the leg in relation to the knee. However, if the lowest point is very close to the knee estimation of the leg configuration is more difficult and errors become more likely.

In the geometric projection of the frontal view, however, the lower legs move almost exclusively vertically. The movement of the foot-point in the classic Johansson walker accurately describes the movement of the leg. The movement of a point somewhere on the lower leg also describes the movement of the lower leg sufficiently enough. In the limited-lifetime walker, however, the position of the point on the leg changes unpredictably. In other words, if the point on the lower leg is high in one frame and low in the next this must not indicate that the leg is lifted but could rather have resulted form the relocation of the point. Therefore, in the limited-lifetime walker the movement of the lower leg cannot be calculated from the positions of the light point on the leg over time. However, this is true only for the frontal view, because in all other orientations the perspective projection of the leg movement has a sideways component. In this case, a sequence of point positions on the leg allows to trace the orientation and movement of the leg over time. Thus, information about the leg configuration is available and supports estimation of the walking direction. Therefore, the failure to discriminate the walking direction of a limited-lifetime walker in the frontal view results from the particular projection properties of the walker in that orientation.

2.8 General Discussion

The use of local motion signals in discriminating biological motion of 3D oriented point-light walkers in profile, half-profile, and frontal views was studied.

Since the frontal view turned out to be a special case I will begin by discussing the profile and half-profile views and return to the frontal view thereafter. The main question was whether speed and depth information derived from the local motion of the light points can aid biological motion recognition of 3D oriented walkers in perspective projection. To answer this question discrimination performance on walkers with limited lifetimes of the point lights was measured. Performance in perspective and orthographic projections between point lifetimes of 1 or more frames was compared. If the participants would use local motion to perceive walking direction one would expect a higher performance for higher lifetimes. Experiment 1 revealed no increase in performance between lifetime 1, 2, or higher. Thus, local motion information is not necessary for performing the walking direction detection task in either the profile view or the half-profile view. Thus, we found no evidence of the necessity of local motion information for performing the walking direction detection task in either the profile view or the half-profile view.

Our stimuli either presented 2 or 4 points per frame. Because of visible persistence the apparent number of simultaneously visible points was higher than the number of physically present points (see methods). An increase in lifetime reduces the apparent number of simultaneously visible points, because the position change of a moving point between two frames is too small to give rise to two different apparent positions. One might argue, therefore, that the shorter the lifetime, the more configural information will be present in the stimulus. One might argue further that a constant level of performance for stimuli with longer lifetime might rely on local motion signals to compensate for the drop in configural information. Two lines of evidence argue against this. First, the performance in various discrimination tasks over a wide a range of lifetimes and frame durations has been shown to depend essentially only on the total number of points that were physically displayed in the stimulus (Beintema et al., 2006). Second, the template model (Lange & Lappe, 2006; Lange et al., 2006) could reproduce the performance data of Beintema & Lappe (2002) on lifetime variation, while simulating visible persistence but without using local motion signals.

As expected from other studies with varying viewing angle, the performance for the classic walker was high in every view angle with only a slight drop in the frontal view. Vanrie et al. (2004) for example used a facing direction task and classic walkers in perspective projection (their Experiment 3). They found no difference in performance between profile and half-profile view, but in the frontal view the performance was only about 85 percent correct. Gender classification tasks (Mather & Murdoch, 1994) and person identification tasks (Troje et al., 2005), on the other hand, indicated an advantage for the frontal view with respect to profile and half-profile view.

The results for the limited-lifetime walker depended stronger on viewing angle. For profile and half-profile views, discrimination performance was similar for the classic walker and for the limited-lifetime walker with 512 points per trial, but for the frontal view performance with the limited-lifetime walker was much lower. The different behavior for the frontal view was investigated in three experiments. The first one (Experiment 2) showed that the difficulty for direction discrimination was restricted to a small range around the frontal view. Experiment 3 showed that longer lifetimes did not change performance, ruling out a direct role for local motion. Experiment 4 revealed that the missing information to perform the task was in the foot-points. Classic walkers with missing foot-points resulted in poor performance in frontal and profile view. Varying the position of the lowest point on the lower leg influenced performance on the limited-lifetime walkers but also on the classic walker. In the limited-lifetime walker, the lowest light point occupies varying positions on the lower leg and may therefore have led to an overall lower performance.

To reconcile the two main results, i.e. the lack of influence of local motion and the importance of the movement of the feet for the walking discrimination, one need to discuss how the movement of the feet (and the rest of the body) can be estimated from the stimuli that were presented. The results of experiment 1 show that local motion detectors are not used. This is consistent with current theoretical frameworks for biological motion recognition such as the template approach (Lange & Lappe, 2006; Lange et al., 2006) or the interactive encoding model (Dittrich, 1999). The interactive encoding model of Dittrich (1999) assumes three different routes by which the trajectory information of the points of an biological motion stimuli is processed further. These routes are connected by so called motion integrators. The first route is strictly based on the analysis of the structural components of human motion to reconstruct 3D body information out of the 2D trajectory information. The second route is linked to the memory system and allows to apply cognitive constraints relating to the human body and its motion trajectories for the 3D reconstruction. The third route relies on visual semantic stored in respect to action categories. Here, in contrast to the template matching, cognitive processes aid the perception process. If low-level motion is lacking, recognition can be enhanced because input signals can be amplified by stored information. Thus, these findings support the interactive encoding model's proposition that the level of processing is highly variable depending on the type of information available to the viewer.

A slightly different approach is provided by template matching (Lange & Lappe, 2006; Lange et al., 2006). Template matching may estimate leg configuration if suitable templates are available. These templates may exists either as multiple 2D view templates with varying orientation or as 3D templates. A computation directly from the trajectory of the foot-point, which would be possible for the classic walker, is not possible for the limited-lifetime walker because the foot-point is unlikely to be visible for an extended amount of time due to point relocation. A computation of the leg trajectory from local motion signals of points on the leg is also unlikely because results in the lifetime 1 condition and the higher lifetime conditions were identical even though local motion is not available in the lifetime 1 condition. This suggests that the trajectory of the leg is derived from template matching analysis. Whether this involves 2D, view-dependent templates or 3D templates is a matter or further research. Template matching processes as proposed by Lange & Lappe (2006) are simple automatic processes, which can take place at an early stage of visual processing.

One can conclude that for discrimination of the walking direction of 3D oriented walkers there is no need for a local motion mechanism. Instead, the use of a local motion independent high-order motion detector based on static templates seems more plausible. The results of the profile and half-profile view show that local motion information does not add essential information to the perception process. However, global mechanisms that are used in other views are useless in the frontal view. Here local mechanisms like the use of local motion information, trajectory information or relative motion of the foot-points with respect to the knee aid perception.

Chapter 3

Action Recognition from Limited Lifetime Biological Motion

Contents

3.1 Abstract

Studies with sequential-position stimuli suggested that biological motion does not require local motion detection. So far, these studies only used walking stimuli. Walking is a very common and easy to recognize action. Most other actions are more complex and slightly different body postures can lead to a different meaning. Therefore, there may be a benefit of local motion information when details of a more complex action than walking have to be recognized. The performance on an action recognition task was compared for classic and sequential position stimuli. The sequential position stimuli were either presented with or without valid local motion information and with different amount of form information. The results show that observers were able to recognize actions from stimuli that contain only sequential position information. Moreover, the recognition of the portrayed action in the sequential position condition without valid local motion information was as good as in the classic-condition. Whereas, for the sequential position condition with valid local motion information there was a clear decrease in performance.

3.2 Introduction

When it comes to human action perception a number of studies have been done with stimuli in the form of line drawings or photographs (e.g., Fiez & Tranel, 1997), or the object and action naming battery by Druks & Masterson, 2000. It has been shown that subjects can distinguish walking and running based on static pictures of stick figures just by the orientation of the tibia (Todd, 1983). These static representations are not equally suitable for all research goals since the dynamic component of the action is not included.

Therefore, a number of researchers have used the point-light technique to investigate action perception. With the classic point-light stimuli it is easy to recognize actions like walking, cycling, climbing or dancing (Johansson, 1973). Observers are able to tell different dance movements apart (Walk & Homan, 1984) and can even accurately estimate the weight of a lifted object or the distance an object is thrown (Runeson & Frykholm, 1981, 1983; Bingham, 1987). Dittrich (1993) used different action types with classic (light attached to joints), inter-joint (light attached between joints), and upside-down presentations. In their study locomotory actions were recognized better and faster than social or instrumental actions. Biological motion was recognized much better and faster when the light-spot displays were presented in the normal orientation rather than upside down, whereas recognition rate was only slightly impaired under the inter-joint condition. Norman et al. (2004) compared the ability to recognize actions from point-light displays in terms of aging. They showed that older observers can also effectively perceive and discriminate among various forms of biological motion. This is true even when the depicted motions of the stimulus elements are either brief and / or fragmentary by the presence of occluding bars. Their results show ability to perceive and discriminate different types of actions from biological motion is robust and is relatively well preserved throughout the latest periods of life. Pollick et al. (2001) showed that subjects even were able to perceive internal stages like fear, anger or tiredness from actions by using point-light displays of knocking and drinking arm movements.

The brain network for the recognition of biological motion includes besides visual areas (Vaina et al., 2001; Grossman et al., 2000; Servos et al., 2002) also structures of the mirror-neuron system (Rizzolatti & Craighero, 2004; Saygin et al., 2004; Michels et al., 2005). The mirror-neuron system is found to be active during action execution as well as during action recognition. The mirror-neuron system is thought to be responsible for the understanding of motor actions (Jeannerod, 1994; Decety et al., 1997; Carey et al., 1997; Rizzolatti & Matelli, 2003), learning of actions by imitation (Jeannerod, 1994) and estimate consequences of other actions (Gallese et al., 1996; Rizzolatti et al., 1996; Kilner et al., 2004). A recent study by Saygin et al. (2004) using point-light actions (walking, walking up stairs, jogging, jumping jacks, throwing, underarm throwing, skipping, stepping up, a high kick into the air, and a lower kick) revealed that the premotor brain regions containing mirror-neurons are also activated in response to point-light human motion.

Studies, which were concerned with the question which information is essential for the perception of biological motion normally use walking stimuli. Walking is an action, which is not only very common to us, but has also some aspects by which recognition is easier than for many other actions. Walking is a symmetric action, i.e. the left half of the body is identical to the right, but 180° out of phase. For judging the facing direction (Troje & Westhoff, 2006) as well as for judging the walking direction the information of the foot points is crucial (see chapter 2). Most other actions are composed of different movement parts and either part is important to recognize the action correctly. A slightly different body posture or arm movement can lead to a different meaning. Therefore, there may be a benefit of local motion information when details of a more complex action than walking have to be recognized.

Kuhlmann & Lappe (2006) investigated the influence of local and global cues by the mean of action recognition from blurred natural scenes. By blurring the scenes the visual cues were modified, particularly the local form and motion information. The results demonstrated that reductions of local form and local motion information by blurring can be compensated by global form change and global motion. Another technique to show biological motion without valid local motion information is the sequential position technique. Until now the ability to recognize actions from sequential position stimuli has not been tested. The sequential position stimuli differ from the classic stimuli mainly in the amount of form information and in the absence of valid local motion information. Many tasks as facing detection or judge the walking direction are easy to perform with sequential position stimuli (Beintema et al., 2002, 2006). The fact that McLeod et al. (1996) found the motion-blind patient LM able to spontaneously identify a wide range of actions (walking, hand-shaking, embracing, standing up, cycling, or a couple dancing) when shown a video of the original Johansson film, is a clear hint that local motion information is not needed for action perception.

In the present chapter an experiment is described in which action recognition was compared for classic and sequential position stimuli. The sequential position stimuli had either a point-lifetime of one frame (no valid local motion) or a point-lifetime of four frames. The stimuli with lifetime 4 differed from the lifetime 1 stimuli mainly by having less form information but valid local information. If we need local motion information to perform action recognition properly it should be more difficult to do the task with the sequential position stimuli than with the classic stimuli. On the other hand if form information is essential there should be a clear difference in performance between the two sequential position stimuli in favor for the lifetime 1 condition.

3.3 Methods

3.3.1 Subjects

Thirty subjects (20-30 years, 18 females) participated in the experiments for course credit. All of them had no prior experience with psychophysical experiments involving biological motion stimuli. The participants were naive to the objective of the experiments.

3.3.2 Stimuli

Stimuli were point-light displays of humans performing an action and consisted of white points (0.15 x 0.15 deg) on a black background. Width and height of the stimulus subtended approximately 5 x 9 degree visual angle. The stimuli were based on the 3D joint positions of real humans recorded using MotionStar Wireless (Ascension Technology Corp., Burlington, USA). The point-light displays were either classic-displays or sequential position displays with a point-lifetime of one frame or sequential position displays with a point-lifetime of four frames. The classic-displays consisted of twelve points located on the major joints of the body (see 3.1). The sequential position displays both had eight points positioned on randomized locations on the body. For the sequential position displays the points were reallocated after one frame for the lifetime 1 condition respectively after four frames for the lifetime 4 condition. To avoid that all the points are reallocated at the same time, the starting phase of the points was randomized.

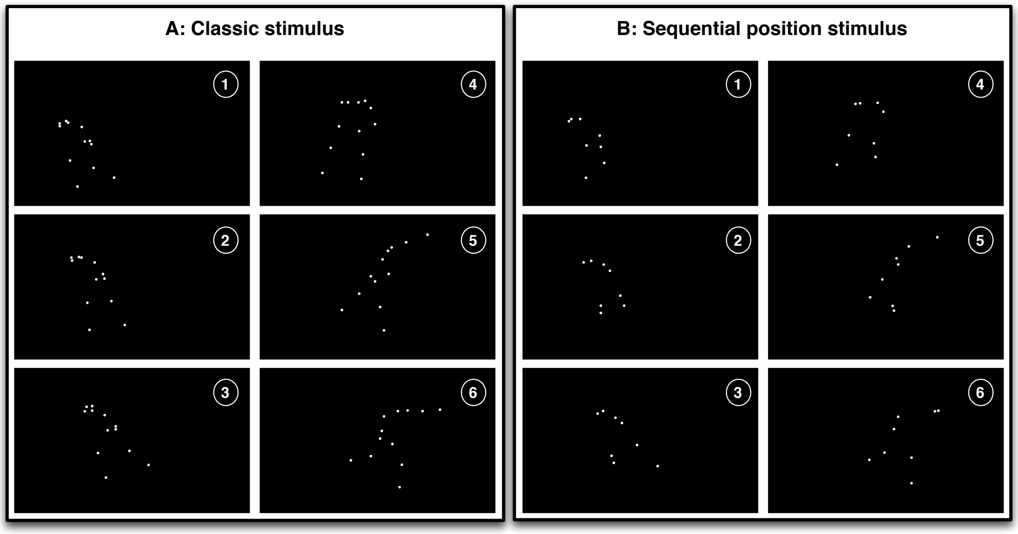

Figure 3.1: Example frames of the action shot-put. Panel A shows frames for a classic stimulus. Panel B shows the same frames for a sequential position stimulus

33 actions were chosen from a pool of actions recorded with the MotionStar Wireless (18 performed by female actors and 15 by male actors). The actions were sport exercises and everyday life activities. The actions were chosen for their good recognizability and their commonness.

Table 3.1: List of used actions

Used actions in alphabetical ordering		
alternate-foot jumps	hula hoop	shot put
boxing	jogging	side kicks
breaststroke	jumping jack	sit down on the floor
cart wheel	knee bending	ski jumps
crawling	marching	soccer shot
crouch jump	one-leg hopping in a circle	stand up from the floor
cycling on the back	one-leg hopping on the spot	tennis serve
discus throwing	pirouette, arms outstretched	two-leg hopping
dance a jig	pirouette, arms to the body	walking on the hands
fencing	push-ups	walking with a walking stick
golf shot	reaching into the air	

3.3.3 Procedure and Experimental Set Up

Stimuli were displayed on an Iiyama Vision Master CRT Monitor (40 x 30 cm, 800 x 600 pixel) at a vertical refresh-rate of 100 Hz. The subjects were seated in a dimly lit room with their eyes about 70 cm in front of the monitor. The point-light displays were presented in the middle of the screen with a fixation point at the center. Point-light displays measured 5 x 9 degrees of visual angles. After each trial the subjects had to press the space bar and an input field appeared. Stimulus duration was between 2 and 7 seconds depending on the complexity of the action. The task was to observe the displayed action and describe the action afterwards as clear as possible. The answers were given by typing the name or the description of the action into the input field.

Each subject conducted one experiment session. In each experiment session all of the 33 different actions were shown to the subject in a randomized order.

3.3.4 Data analysis

The proportion of correct responses was assessed. To test for statistical differences a two-way was conducted repeated measures ANOVA with factors stimuli-type and action. As an a posteriori procedure the Scheffé-test was used with an alpha significance level of 0.05. Error bars in the figures give the standard error of the mean.

3.4 Results

The results are illustrated in figure 3.2. Each subject saw all of the 33 actions. Over all, the subjects had no problems recognizing the displayed actions. In the classic-condition and in the sequential position condition with lifetime 1 the performance was on average over all subjects about 87 percent. In the sequential position condition with lifetime 4 the subjects performed poorer with only 73 percent.

A two-way repeated measures ANOVA revealed significant main effects of stimulus-type (F(2,850) = 18.3, $p < 0.0001$) and no effect of action (F(32,850) = 5.1, $p < 0.0001$), but significant interaction (F(64,850) = 1.1, $p = 0.2677$). The Scheffé test as a posteriori procedure showed no difference among the classic and the sequential position condition with lifetime 1, but significant differences be-

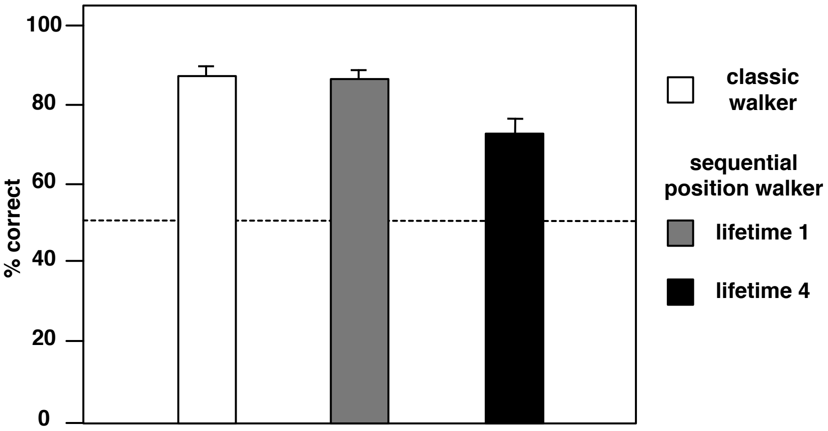

Figure 3.2: Averaged correct responses split by stimulus-type. The white bar shows results for the classic stimuli, grey bar for the sequential position stimuli with lifetime 1 and the black bar the results for the sequential position stimuli with lifetime 4. Error bars represent the standard error over subjects.

tween classic and the sequential position condition with lifetime 4 as well as between the sequential position condition with lifetime 1 and the sequential position condition with lifetime 4.

To see how good one is able to perform with the sequential position stimuli, one subject was shown the sequential position condition with lifetime of one frame, but was asked to describe the action as detailed as possible. Table 3.2 shows a selection of these action descriptions. The subject recognized all but one of the actions correct and was able to give a good and detailed description.

Table 3.2: Examples for detailed action descriptions

action	**subjects action description**	**translation**
discus throwing	Langsam angefangen Anlauf zu nehmen, dabei leicht nach hinten gebeugt und immer das gleiche Bein vorne, schneller geworden, sich gedreht und eine Wurfbewegung gemacht	As start-up first a slow run, while bending slightly backwards, always the same leg in front of the other leg, the run becoming faster, then spinning around followed by a throwing movement
shot put	Strecken nach links, dabei auf einem Bein stehend und ziemlich weit nach links gebeugt, dann aufs rechte Bein gesprungen, Anlauf genommen und Wurfbewegung	Stretching to the left, while standing on one leg, then bending forward far to the left, jumping onto the right leg, gathering momentum followed by a shot-put
golf shot	Frontansicht, gebückte Haltung, Arme leicht schwingend, dann ausgeholt wie beim golfen, zwei mal	Frontal view, bending forward, arms slightly swinging, backward swing to get drive golf shot
jumping on two legs	Nach rechts gehüpft, geschlossene Beine, Arme am Körper angelegt, kurze Sprünge	Jumping to the left, legs straight, arms alongside the body, short jumps
alternate-foot jumps	Von einem Bein aufs andere gehüpft, dabei jeweils ein Bein zur Seite hoch gestreckt, die Arme vor dem Körper ruhig gehalten	Jumping with alternating legs, always one leg stretched to the side, arms stay alongside the body
one-leg hopping in a circle	Auf einem Bein gehüpft, dabei das andere Bein angewinkelt, im Kreis gedreht, Arme am Körper	Jumping on one leg, the other leg is bent, jumping in a circle, arms stay alongside the body
walking with a walking stick	Erst nach links gelaufen, dabei ein Bein angewinkelt, hinkende Bewegung, Drehung und auf gleiche Weise zurück gelaufen	Walking to the left, favoring one leg, limping movement, turning around and walking back the same way

3.5 Discussion

In this study the effectiveness of sequential position stimuli for action recognition tasks was examined. The performance for two different kind of sequential position stimuli was compared with the performance for classic point-light stimuli. The task was to name the action or give a short description. The results show that observers were able to recognize actions from stimuli that contain only sequential position information. Moreover, the recognition of the portrayed action in the sequential position condition with a point-lifetime of only one frame was as good as in the classic-condition. Whereas, for the sequential position condition with a point-lifetime of four frames there was a clear decrease in performance.

To see if one is able to recognize also details in this kind of stimuli, one subject was shown the sequential position condition with lifetime 1, but the task was to describe the actions as detailed as possible. The subjects descriptions were rich in detail and showed how well the sequential position stimuli can be used for action recognition tasks. The answers were over all good descriptions of the movement sequences with details on arm, leg and body posture. This was true even for complex movement sequences like discus throwing, shot put or a golf put (see table 3.2).

Until now action recognition studies using light-point stimuli were done with classic stimuli. One could assume that positioning the points at random locations of the body may make it harder to recognize actions. In classic stimuli the points do not only mark the joints but thereby give also the end-positions of the limbs. For complex movement sequences this may be helpful to determine the exact positioning of the limbs. This is not the case as was already shown by Dittrich (1993). Dittrich compared the ability to name a portrayed action with normal and inter-joint conditions of classic stimuli. He found no difference in performance between the two conditions. The inter-joints condition is a modification of the classic stimuli as the points are placed at the midpoints between the major joints but still follow correct motion paths and therefore provide clear and valid local motion information.

Similar to inter-joint displays for sequential position stimuli the points are not located on the major joints. However, sequential position stimuli differ from the inter-joint displays as the points do not stay on their position but are reallocated after a certain life-time. In the lifetime 1 condition the points are reallocated to a new random position on the body every frame. By doing so, 98 percent of usable local motion information is removed, because an individual point cannot be tracked over frames. However, due to the reallocation more parts of the body are drawn over time and thereby more form information is provided. With longer point-lifetime each point stays on the same position for more frames and therefore will carry more valid image motion information. On the other hand with longer point-lifetime also fewer points are reallocated, resulting in less of the body being drawn out over time. Therefore, the difference between the two sequential position conditions used in this study lies not only in the amount of valid local motion information but also in the amount of form information. The lifetime 1 condition contains not only less local motion information, but also more form information than the lifetime 4 condition.

From our results, one can conclude that local motion information does not aid action recognition more than less form information reduces the ability to recognize the human figure and therefore the ability to recognize the portrayed action. This would support the idea that form analysis plays a dominant role in the perception of biological motion and suggesting that biological motion perception does not rely on local motion. If biological motion perception would benefit from local motion information, one would have expected performance to be better in the lifetime 4 condition compared to the lifetime 1 condition. Rather, there was a decrease in the recognition rate from lifetime 1 to lifetime 4 suggesting an influence of the amount of form information rather than the amount of valid local motion information.

The results of this study support the theory of the existence of a template-matching process. Template matching assumes that the visual system contains a set of templates of the evolving shape of the human body during an action. The global form information of each frame is compared to a sequence of templates to find the best fitting template and the sequence of these templates would lead to the best fitting action (Bobick & Davis, 2001; Lange et al., 2002; Giese & Poggio, 2003).

Chapter 4

Avoiding Local Motion Information by Applying a Counter-Change Technique

Contents

4.1 Abstract

Studies, which used sequential position stimuli, provided strong evidence for the view that biological motion perception is based on form rather than motion perception. However, the sequential position technique has the disadvantage that the proportions of provided form and motion information cannot easily be quantified. On one hand, through visible persistence additional form information is provided. On the other hand, false local motion information can occur. A novel biological motion stimulus was developed, which is similar to the sequential position stimuli in that way that only randomly located points on the walker are shown, but without visible persistence and without false local motion information. A field with static randomly located black and white dots was presented. By moving over the center of a background point, an otherwise invisible walker caused the point's color to flip. As motion perception can only occur if luminance decrements and increments occur at neighboring element locations, this minimized valid local motion information strongly. This random luminance stimulus was tested in a forward/backward detection task. A strong dependency on the number of stimulus-points was found, which supports the idea that biological motion is perceived from a sequence of spatiotemporally sampled forms.

4.2 Introduction

Our visual system is remarkably sensitive to biological motion. In general, the concept of biological motion refers to characteristic movement patterns of living creatures, but is mainly used to describe human motion. The Swedish psychologist Johansson (1973) was the first to show that human observers are able to identify the form of a human person depicted only by point-lights placed on the major joints of a moving actor. Point-light stimuli similar to Johansson's are still the most commonly used stimuli to study the perception of biological motion. In these classic point-light stimuli, each image frame provides a displacement vector to the corresponding point in the previous frame, the new motion vector (velocity and direction) for each point.

The inability of naive observers to recognize the underlying figure out of a static frame led Johansson (1973) to propose that biological motion is perceived by relative movements of the points on the basis of local motion signals. Also many following studies on biological motion perception have suggested or implicitly relied upon the assumption that the perception is processed by means of local motion signals (Cutting et al., 1988; Mather et al., 1992; Neri et al., 1998).

A stimulus, which enables to study the contribution of local motion signals in biological motion perception, was introduced by Beintema & Lappe (2002). The local motion signal was manipulated by limiting the lifetime of points. Additionally, the points were situated anywhere on the four limbs (arms and legs), and assessed to a new random location after a certain number of frames. Such relocations minimized the local motion information (motion vector and trajectory information) carried by each point, without altering the temporal sampling of the sequence. When points are relocated with each single frame, all motion signals consistent with the limb's movement are strongly reduced. Because the information is carried only by the sequence of point positions, this stimulus is referred to as sequential position walker. Beintema & Lappe (2002) found naive subjects able to recognize a walking figure from the sequential position walker just as often as from a classic point-light walker, in which points are continuously visible on the joints. Also, subjects performed in coherence and direction discrimination tasks with sequential position walkers just as well as with classic walkers. Beintema et al. (2006) extended these findings by confirming that this not only applies to the 50-ms frame duration previously tested, but to other frame durations as well. Performance in coherence and forward/backward discrimination was reduced with fewer points per frame and with longer frame duration. They suggested that biological motion perception might be achieved by an analysis of the dynamic form of the human figure. Local motion signals may play only a role in more complicated tasks and are not essential for perception.

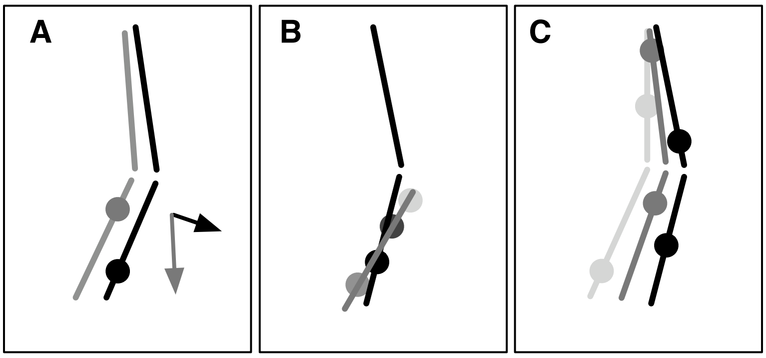

Figure 4.1: Sequential position stimuli can provide additional information. A: false local motion information through false binding B: form from motion information by jumps on one line segment. C: form information through visible persistence; Black points and lines represent the actual frame, gray bars and points represent earlier positions of the leg. Black arrows indicate correct motion path, gray arrows indicate false motion path.

The sequential position technique has the obvious advantage that local motion information is strongly reduced but there are also some problems, which make it difficult to identify the information source. The first problem for sequential position stimuli is, that there might be local motion cues through false binding (see figure 4.1 A). Although local motion information seems not to be necessary for the perception of biological motion per se, local motion detectors may play a role that is nonspecific to biological motion analysis. These false local motion cues may interfere at tasks where local motion otherwise would aid the process.

Another effect that can occur is that line segments are ”drawn” by the jumps of the points. Thereby, form information can be gained through motion information (see figure 4.1 B).

The third problem is the visible persistence of stimulus points. Visual stimuli displayed for a brief interval remain visible for some time after their physical offset (see Coltheart, 1980, for a review). This period of additional visibility is referred to as visible persistence and is estimated to be 100–150 ms for brief stimuli of about 10 ms. It is also known to decrease exponentially with exposure duration. With frame durations smaller than 100 ms older point-positions are still visible. Even though these points do not give the actual and therefore correct position information, they can aid the perception process by giving additional form-information (see figure 4.1 C). This can be helpful for a form-based mechanism like the template matching process proposed by Lange et al. (2002).

Although it is unknown whether visible persistence also applies to the units that process biological motion, it might explain the slight decrease in performance observed with longer lifetime at very short frame durations. To estimate the effect of visible persistence for sequential position stimuli Beintema et al. (2006) assumed that visible persistence causes points to remain visible for 100 ms but does not extend the visibility of points that are displayed longer than that. This effect would, for instance, cause the number of visible points at the 50 ms lifetime to be doubled with respect to the actual number of displayed points per frame. They observed a small effect of lifetime when visible persistence was taken into account but the effect was not significant for any binned number of visible points per trial. Thus, even when visible persistence is taken into account, their data still suggest that local motion plays a minor role.

To avoid these problems a novel biological motion stimulus was developed for this study, which is similar to the sequential position stimuli in that way that only randomly located points on the walker are shown.

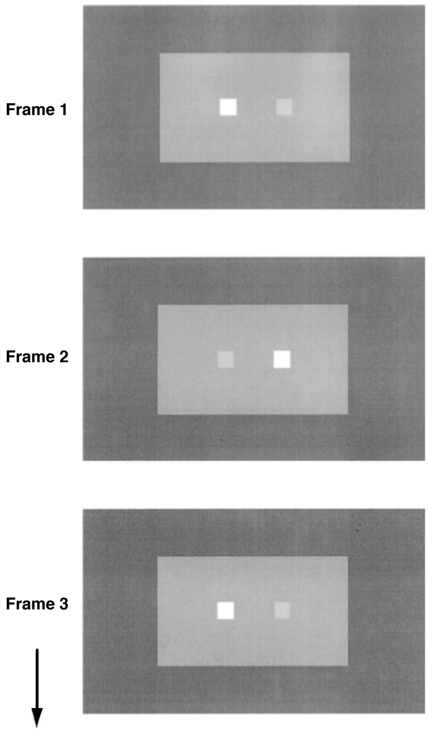

Figure 4.2: Illustration of the generalized apparent motion stimulus, as used in Hock et al.(1997) and Hock et al. (2002).

To produce such a stimulus without visible persistence and without false local motion information a technique from Hock et al. (2002) was adapted. For standard apparent motion only one element is visible at a time. A generalized version of single-element apparent motion was systematically studied by Hock and colleagues (Hock et al., 1997, 2002). They used stimuli composed of two simultaneously visible elements whose luminance alternated between two values (Figure 4.2 shows an illustration of this stimulus). Hock et al. (2002) demonstrated that apparent motion is specified by background-relative, counterchanging luminance. Motion starts where luminance changes toward the background luminance value and ends where luminance changes away from the background luminance. Luminance decrements and increments must occur at both element locations for motion to be perceived. Therefore, in case of co-changing luminance no motion is perceived.

We used this phenomenon to decrease local motion in the random luminance walker. A presentation field consisting of randomly located points on a gray background was displayed, with the color of each point randomly assigned to either black or white. The stimulus was a walker, designed as stick figure. However, the limbs of the walker were not visible, but by moving over the center of a background point, causing the point's color to flip from black to white, respectively from white to black with a certain probability. This avoided successive increments along a common path in order to eliminate motion induced across several frames. Incrementing point luminance and leaving them incremented additionally eliminated motion between pairs of points. There were no moving points, because no points appeared or disappeared. Additionally, the non-changing background points serve to mask the form of the walker.

4.3 Methods

4.3.1 Subjects

Six subjects (20-26 years, all females) participated in the experiments. All of them were psychology students who participated for course credit. They had no experience on psychophysical experiments involving biological motion stimuli. The participants were naive to the objective of the experiments.

4.3.2 Stimuli

The presentation field measured 23 x 11.5 cm and consisted of randomly located points (5 x 5 pixel) on a mean luminance gray background. Each point was presented with at least 10 x 10 pixel distance to the next point. The total number of presented points was either 1200 or 1600. The color of each point was randomly assigned to either black or white that is either minimum or maximum luminance. Therefore, the points had always the same contrast to the background and differed only in their relative luminance to the background. Figure 4.3 A shows an example frame of presentation field with 1200 points, Figure 4.3 B shows an ex-

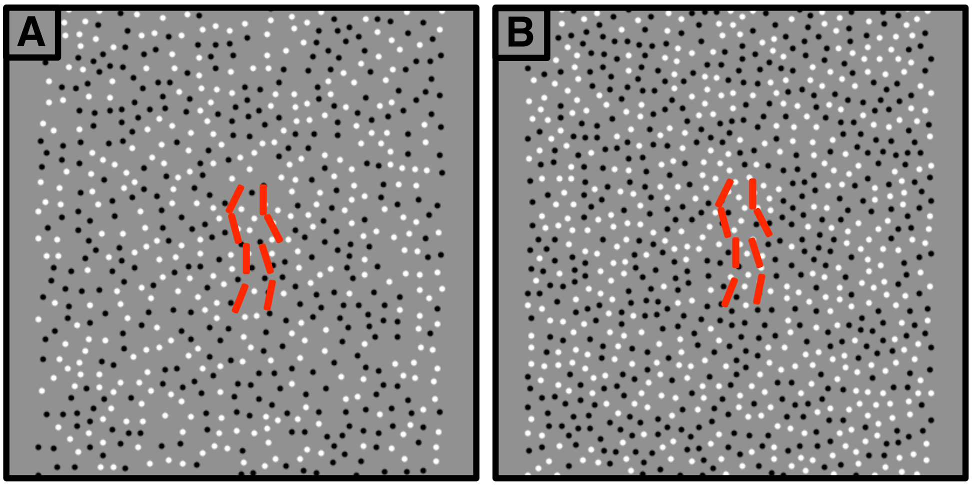

Figure 4.3: Example frames for the different presentation fields. Panel A shows a frame with 1200 points, panel B a frame with 1600 points. The red stick figures mark where the stimulus was presented.

ample frame of a presentation field with 1600 points. The red stick figures mark the stimulus position and were not shown in the actual stimuli. Pretests showed that if the number of presented points is too high, the field of points is perceived as uniform background. Then the points are perceived as belonging to a pattern instead of masking the figure.

Stimuli displayed walking human figures. Width and height of the stimulus subtended approximately 5 x 9 degree visual angle. The stimuli were based on the 3D joint positions of five walking humans (3 male and 2 female) recorded using MotionStar Wireless (Ascension Technology Corp., Burlington, USA). The forward translation was subtracted giving the impression of walking on a treadmill. The walkers were designed without occlusion. Walking speed was normalized so that a complete walking cycle, consisting of two steps, took about 1.39 seconds. The stimulus sequence was either presented in normal (forward walking) or reversed (backward walking) frame order and was shown for four complete walking cycles. The walker started from a random phase in the step-cycle.

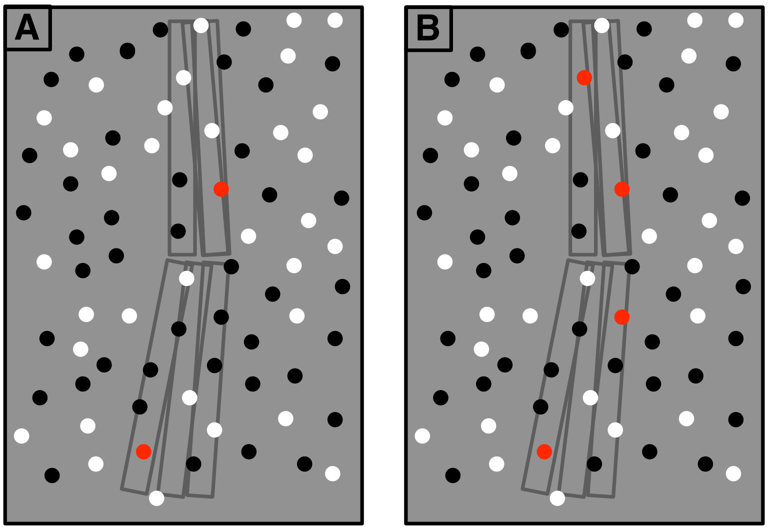

Figure 4.4: Demonstration of color flips for one leg with 10 percent change (A) or 20 percent change (B). The outlined gray bars indicate the leg movement over the background for a time interval of 3 frames. The red points indicate color flips. The Figure represents a cut-out of Figure 4.3 with a 2.5 magnification factor.

The walkers were designed as stick-figures, where each limb had a width of 1 pixel. The limbs of the walkers were not visible, but by moving over the center of a point the point's color flipped from white to black or from black to white with a probability of 10 or 20 percent. A demonstration of the color flips for one leg is shown in Figure 4.4 for the two probability conditions. The images show the location of a segment in three subsequent frames, on a gray background with black and white, stationary, randomly located, points. It shows a cutout of Figure 4.3 with a 2.5 magnification factor. The outlined bar represents the (invisible) limbs, which moves from left to right. Red points indicate color flips.

The two factors presentation field and flip probability influence the number of point flips per frame (walker-points per frame) and thereby the total number of point flips per trial (walker-points per trial). The number of walker-points per frame differs from trial to trial. The same is true for the total number of walker-points per trial. For comparability to other point-light stimuli we also assesses the mean number of walker-points per frame as well as the mean number of walker-points per trial.

4.3.3 Procedure and Experimental Set Up

Stimuli were displayed on an Iiyama Vision Master CRT Monitor (40 x 30 cm, 800 x 600 pixel) at a vertical refresh-rate of 100 Hz. The subjects were seated in a darkened room with their eyes about 60 cm in front of the monitor. They were asked to fixate a red fixation point in the middle of the screen. The walkers were presented in the center of the screen and were always shown for two complete walking cycles. When the walker disappeared the subjects had to press a response key. Thereafter a new trial started and a new walker appeared after 200 ms. The subjects task was to detect the walking direction (forward/backward) of the walker; pressing the ‘up’ (forward) and ‘down’ (backward) arrow keys of the keyboard.

Each subject conducted 10 sessions. In each session five different walkers were shown, with all of them in two walking directions and two orientations. The background consisted of either 1200 or 1600 points and the probability of point flip was either 10 percent or 20 percent. Every combination of these conditions was repeated once, so that one session consisted of 160 trials, which were presented in random order.

4.3.4 Data Analysis

The proportion of correct responses was assessed. For statistical testing repeated measures analysis of variance (significance level = 0.05) on the d' values were conducted. The Scheffé-test was used as a posteriori procedure. For all post hoc test an alpha significance level of 0.05 was used. Error bars in the figures give the standard error of the mean.

4.4 Results

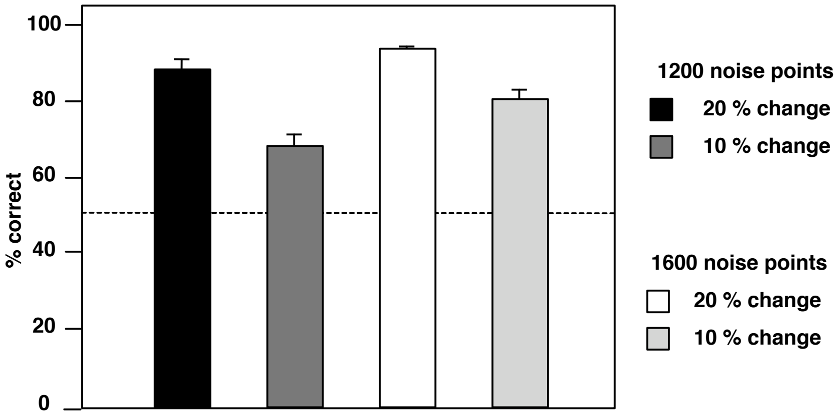

Figure 4.5: Averaged correct responses split by stimulus-type. Error bars represent the standard error over subjects.

Figure 4.5 illustrates the results of the experiment. Both the total number of points in the presentation field and the probability to flip had an influence on performance. With fewer points in the presentation field the performance increased. The same is true for the probability to flip, but the increase is stronger.

A two-way factorial ANOVA was performed on the factors number of background points and fraction change. The ANOVA showed significant main effects of background points (F(1,5) = 102.403, p = 0.0002) as well as of fraction change

(F(1,5) = 165.651, p< 0.0001), and no significant interaction (F(3,5) = 8.475, p < 0.0334). The Scheffé test as a posteriori procedure confirmed the results.

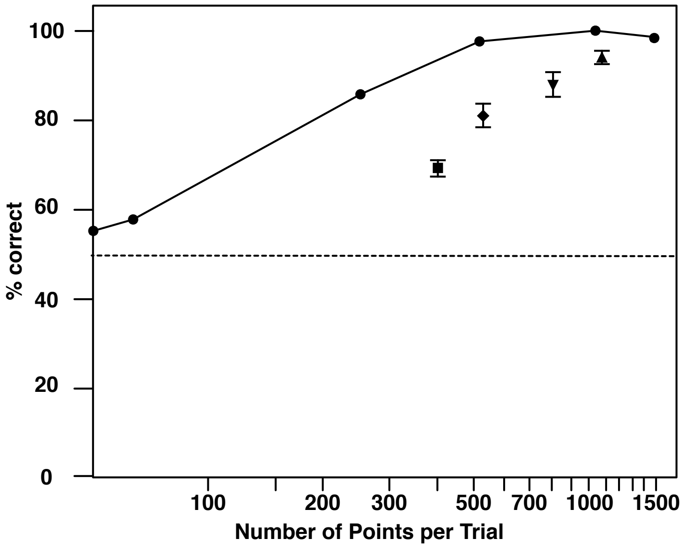

Figure 4.6: Averaged correct responses as a function of total number of points per trial (number of points per frame x number of frames per trial). Square: 1200 points, 10 percent probability for color flip; diamond: 1600 points, 10 triangle: 1200 points, 20 percent probability for color flip; upside-down triangle: 1600 points, 20 percent probability for color flip. For comparison we also included results from Beintema et al. (2006) (shown by circles). Error bars represent the standard error over subjects.

With a 10 percent probability to flip, the number of walker-points per frame was between 0 and 6 points (mean = 5.15) for 1200 points in the presentation field. The number of walker-points per frame was between 0 and 7 points (mean = 5.55) for 1600 points. With a 20 percent probability to flip, the number of walker-points per frame was between 0 and 8 points (mean = 6.85 points) for 1200 points in the presentation field and between 0 and 10 (mean = 8.45) for 1600 points. The mean number of walker-points per trial was about 401 for 10 percent probability to flip and 1200 points in the presentation field, about 531 for

10 percent and 1600 points, about 805 for 20 percent and 1200 points and about 1086 for 20 percent and 1600 points.

For sequential position walkers Beintema et al. (2006) found a dependency on the number of points per trial. To see if a similar dependency can be found for the random luminance stimulus, the results were sorted by the number of walker-points per trial and were plotted together with the results of Beintema et al. (2006) for the forward-backward task (see Figure 4.6). The results show a similar dependency on the number of points per frame as Beintema et al. (2006) had found for the sequential position walkers. However, the performance for the random luminance walker was about factor two poorer than the performance for the sequential position stimuli.

4.5 Discussion

The number of presented points in the presentation field was varied as well as the probability of point flips. This has effects on the walker and on how the background is perceived. First, the number of presented points has an influence on the number of possible walker points, i.e. the more points on the screen the higher the probability for a point to be at the area where the walker moves. Second, the points in the presentation field mask the form of the figure, which is only visible by point flips. With higher number of points in the presentation field the masking increased. The ability to detect biological motion in the mask requires global integration of the points constituting the walker and subsequent segregation of the resulting figure from the background points.

The probability of point flip, i.e. the probability of a point to change its color when the walker moves over it, only influences the walker and not how the background is perceived. If in one frame ten points are in the area of the lower leg, one of these points flips its color for a probability of 10 percent, whereas 2 points flip for a probability of 20 percent.

With higher number of points in the presentation field as well as with higher probability of point flip the performance increased. The increase was slightly stronger for the probability of point flip with 1600 points in the presentation field. This can be explained by the fact that for probability of point flip the number of point flips is doubled and for number of points in the presentation field the increase is only one quarter.

The random luminance stimuli are in some way similar to the sequential position walkers used in the experiments of Beintema & Lappe (2002). In the random luminance walker only the walker moves and the points keep their positions, therefore, random locations on the walker are shown by a luminance change. The parameters walker-points per frame and walker-points per trial are similar to the parameters points per frame and total number of points per trial, which are used for describing the sequential position stimuli (Beintema & Lappe, 2002). As can be seen in Figure 4.6 there is a similar dependency on the number of points per frame as Beintema et al. (2006) had found for the sequential position walkers. However, the performance for the sequential position stimuli is about factor two higher. An explanation for this is that the sequential position walker provides more form information compared to the random luminance walker because of visible persistence and form through motion information due to the ”painting effect” described above. Another factor, which has to be taken into account for explaining the difference, is the masking of the form due to background points in the random luminance stimulus.

There are studies, which also used luminance or contrast changes to disrupt low-level processing for biological motion stimuli, but all of them used stimuli with moving points. Mather et al. (1992) for example found that randomly reversing point contrast seriously impaired direction discrimination. They embedded their figures in a mask of random-noise points whose contrasts were changing randomly between darker and brighter than background from frame to frame. They found poorer performance for stimuli with random point contrast. Therefore, Mather et al. (1992) concluded that low-level motion-detecting processes have a major role in the analysis of biological motion. Ahlström et al. (1997) also used points varying in contrast to test if they generate vivid biological motion. They used black and white points appearing against a gray background and the color of each point was selected randomly for each frame. The task was to judge whether normal or phase-scrambled walkers were shown. Subjects saw twenty frames of the sequence whereas subjects in Mather's experiments saw only eight frames of the sequence (Mather & West, 1993). They found no difference between constant or random contrast. Recently, Aaen-Stockdale et al. (2008) compared the ability to detect first- and second-order biological motion walkers in noise using a random-polarity stimulus similar to that used by Mather et al. (1992) and Ahlström et al. (1997). By mixing first-order and second-order dots within the same stimulus, they found, that when equally visible, first-order noise dots can mask a second-order walker, and vice-versa. In a direction-discrimination task the same pattern for second-order as that obtained with first-order stimuli. These results are consistent with biological motion being processed by a mechanism that is cue-invariant.

The points in the stimuli do not move, but change their luminance and provide thereby position information about locations on the walker. The stimuli gave a vivid impression of a walking human and subjects were able to detect the walking direction. The strong dependency on the amount of form information we found and the fact that motion information in these stimuli is strongly reduced argues for the importance of form information in the perception of biological motion.

A possible way how our brain accomplishes the perception of biological motion by form is the integration of position signals given by the luminance changes in every frame. The integration of these position signals gives the information about the actual posture of the walker. The form-based model of Lange (2006) for the perception of biological motion perception provides a good explanation how our brain could achieve this. The model suggests that each frame is first compared for similarity with templates of whole-body postures. Perception of an action sequence like for example walking is then achieved by integration of the individual postures over time.

Part III

General Discussion

Chapter 5

General Discussion

Despite several decades of research, there is still disagreement about the fundamental mechanisms underlying the perception of biological motion. Biological motion, i.e., the movement of the human figure, may be derived from local motion analysis of the light points or by a global analysis of the changing shape of the body.

Biological motion is a very complex form of movement as already explained in Chapter 1. Stimuli consisting of point-light are well suited for the investigation of biological motion as they give the possibility to exclude or minimize certain features. Biological motion displayed by point-light stimuli contains both form and motion information, but depending on the stimulus, different proportions of the individual information sources are available. In classical point-light stimuli, structural information is reduced, but the points provide local position information and clear local and global motion information. For sequential position stimuli with a limited lifetime of one frame and for random luminance stimuli the points have random positions and change their positions along the limbs unpredictably from frame to frame. These manipulations do not disrupt perception of the body shape, but make it impossible to perceive valid local motion. By presenting sequential position stimuli with lifetimes above one frame it is possible to add valid local motion information and control the amount of it.

The objective of my study was to investigate the role of form and motion information for the perception of biological motion by the mean of different stimuli. Sequential position and random luminance stimuli were used to minimize and control the amount of valid local motion and the amount of form information.

Concerning the role of local motion, we found no evidence that local motion is required for the perception of biological motion. Local motion did not contributes to the perception of the 3D structure of the walker (see Chapter 2). Observers performance was high for judging the walking direction of point-light walkers even when local motion was strongly minimized. This was shown in Chapter 2 for sequential position stimuli and was found as well for random luminance stimuli in Chapter 4. Observers were similar good in an action recognition task with sequential position stimuli as they were with classic stimuli (see Chapter 3). Adding local motion did not improve performance neither in a forward/backward task (see Chapter 2) nor in an action recognition task (see Chapter 3). Concerning the role of form information, we found for both sequential position (see Chapter 2) and random luminance (see Chapter 4) stimuli, a clear dependency on the number of points in the stimulus and therefore on the amount of form information.

The results of the above described experiments argue for global motion perception on the basis of global form perception rather then on the basis of local motion information. How would such a perception process on the basis of form information look like? A good approach is offered by modeling studies. Biological motion perception can be regarded as a template-matching process (Giese & Poggio, 2003; Lee & Wong, 2004; Lange et al., 2006; Lange & Lappe, 2006). According to the model proposed by Giese & Poggio (2003), the perception of biological motion relies on the feedforward analysis of local features (local form and local motion). Local form and local motion signals are thought to proceed in two separate hierarchical pathways, with gradually more complex motion and form features extracted along these pathways. In the ”form path” points are grouped to line segments, limbs and then to complete body representations. In the ”motion path” local motion signals are combined into complex optic-flow patterns. In both paths the resulting patterns are compared to learned templates. However, their ”form path” in contrast to their ”motion-path” was not able to generalize from full-body to point-light stimuli. They concluded that both pathways contribute to the recognition of normal biological movement stimuli, and that point-light stimuli are analyzed predominantly in the dorsal pathway.

A far better explanation for our results is given by the template-matching model suggested by Lange & Lappe (2006) In contrast to the feedforward model of Giese & Poggio (2003) the model of Lange & Lappe (2006) is purely based of form information. The model uses global form templates and their temporal evolution for the analysis of biological motion from point-light stimuli. At the first stage the incoming information is compared for similarities with stored representations of whole-body postures. At the next stage, temporal integration of activity of these template cells leads to the perception of the action sequence. Comparing their model psychophysical data as well as to neuroimaging and single-cell data, they found striking similarities.

Evidence from other psychophysical studies support this view. For example, consistent with the idea that biological motion perception is achieved by a temporal analysis of form information are the results of Beintema & Lappe (2002) and Beintema et al. (2006). They examined the perception of biological motion with sequential position walkers and found that reducing the number of points as well as reducing the frame rate led to poorer performance. Moreover, discrimination performance was independent of the lifetime, demonstrating that local motion is not necessary for the perception of biological motion. Bertenthal & Pinto (1994) also provided evidence for a global form analysis in biological motion perception. They used distracting noise points, with motion trajectories identical to those of the walker points, but different, randomly chosen positions. Despite identical motion signals in stimulus and noise, recognition rates always stayed above chance level. They concluded that biological motion perception results from a global top-down form recognition process, rather than a bottom-up local motion analysis.

There are also case studies in lesion patients (Vaina et al., 1990; McLeod et al., 1996; Vaina et al., 2002) who found motion blind patients able to perceive biological motion. Vaina et al. (1990)) reported a study on a patient with bilateral lesions including visual pathways like MT. The patient had serious problems in tasks including spatial localization and low-level motion tasks, but was able to solve tasks that involving form perception. Furthermore, he had no problems identifying biological motion, unless it was presented in moving noise. In the study of McLeod et al. (1996) investigated a patient with lesions including motion processing areas like MT. The patient was unable to perceive simple motion or to report the direction of a random dot pattern movement, but could identify the movement direction of biological motion. Vaina et al. (2002) described a case study, in which the patient had no problems recognizing biological motion, but had difficulties to integrate local motion signals to a motion percept or to perceive structure-from-motion.

Part IV

Summary

Chapter 6

Summary

Contents

6.1 Summary

This study investigated the influences of form and motion information on the perception of biological motion. For this purpose, different kind of stimuli were used, by which it was possible to investigate the proportions of form and motion information as separate as possible. In chapter 2 and in chapter 3 besides the classic biological motion stimuli, where the individual points are located directly on the major joints, also so called sequential-position stimuli were used. The points of sequential position stimuli do have in contrast to classic stimuli random positions on the body and are relocated to another location on the body after a certain lifetime. In this stimulus the individual points have random positions on the body similar to the sequential position stimuli. However, this is achieved by a different technique. Here the walker is only visible by luminance changes of background points. An otherwise invisible walker moves over a field of static, randomly located black and white points. If the walker moves over a background point, this point flips from black to white, respectively from white to black with a certain probability.

6.1.1 Perception of Limited Lifetime Biological Motion from Different Viewpoints

Chapter 2 was concerned with a series of experiments with classic stimuli and sequential position stimuli with limited lifetime. Performance on forward/backward walking discrimination was compared of walkers in orthographic and perspective projection when view orientations and point lifetime was varied.

Experiment 1 tested whether local motion information can improve performance on a forward/backward discrimination in perspective projection, when the walkers are presented in different view orientations. Orthographic projections in contrast to perspective projections do not contain direct information about the structure and motion in depth. A benefit from this additional information should therefore lead to a better performance in perspective projections. For both cases equally good performance was found for sequential position walkers with limited lifetime in profile and half-profile view and poor performance for the frontal view. A detailed analysis of the individual parameters lifetime, points per frame and points per trial showed similar results for perspective and orthographic projection. Performance was generally better with 512 points per trial than with 128 points per trial. With 512 points per trial the performance for the limited-lifetime walkers was as good as the performance for the classic walker independent of the number of points per frame. Increasing the point lifetime from one to two frames did not have a positive effect on the performance. An additional Experiment confirmed that even for higher point lifetimes there is no benefit from local motion information for profile and half-profile view. The first Experiment showed that discrimination performance became extremely difficult in the frontal view. Therefore, follow-up experiments were performed that tested lifetime, view orientation, and specific information from the feet to investigate the role of the frontal view further.

Experiment 2 tested whether the poor performance for sequential position walkers walkers in the frontal view was limited to just the frontal view or whether performance gradually decreases between 40° and 90°.

The orientations of the walkers varied in ten-degree steps from 40 to 90 degrees. The walkers had 4 points per frame and 512 points per trial and the points had a limited lifetime of one one frame.

Performance was about equally good between 40° and 70°. From 70° to 80°

there was a strong decrease as well as from 80° to 90°. This Experiment confirmed that identifying the walking direction of a sequential position walkers walker was especially difficult in the frontal view.

Experiment 3 investigated the question whether an increase of lifetime beyond two frames could lead to an enhancement in performance in the frontal view. The lifetime was gradually increased until the lifetime of the sequential position walkers walker was similar to the lifetime of a classic walker.

Sequential position walkers walkers with a lifetime of 1, 4, 8, 16, or 32 frames were used. For the lifetime 32 condition each of the points changed its position on the body only once per trial. The sequential position walkers walkers had 12 points per frame, identical to the classic walker.

Increasing the point lifetime yielded to no significant improvement of performance, so there was still a clear difference between the highest lifetime condition and the classic walker. This experiment showed that the difference between classic and sequential position walkers walker, which has been found in the first Experiment could not be explained by a difference in lifetime, or local motion information, respectively.

Experiment 4 was designed to specifically test the role of the feet for walking direction discrimination in the frontal view. A classic walker was generated on which the foot-points could be shifted toward the knees on the lower leg and thereby shortened the lower leg. Five different walkers were tested, either with the foot-points positioned directly on the ankle, at three quarter of the lower leg, at the half of the lower leg, or at one quarter of the lower leg, and one walker in which the foot-points were entirely omitted. All walkers were presented in the profile view and in the frontal view.

For both views good performance for most of the inter-joint conditions was found, except for the condition where the lowest points were presented near to the knee-points, at one fourth of the leg. In this condition the subjects performed significantly poorer. For the condition where only knee points were presented the performance was at chance-level.

These results indicated that the dependence on viewing angle can be explained by the reliance of the forward/backward discrimination on information about the movement of the lower legs, which is difficult to ascertain in the frontal view.

6.1.2 Action Recognition from Limited Lifetime Biological Motion

Chapter 3 investigated the ability to perform action recognition with sequential position stimuli. It has been shown that task like judging the orientation of a walker or deciding the walking direction are easy to perform with these stimuli. As action recognition is a more complicated task the performance for classic and sequential position stimuli was compared in this chapter.

The performance on 33 actions was tested in a recognition task with classic and sequential position stimuli. The classic stimuli had 12 points located on the major joints of the body and the sequential position stimuli had 8 points located on random positions on the body. The sequential position stimuli had either a point-lifetime of one frame (no local motion) or a point-lifetime of four frames. The stimuli with lifetime 4 differed from the lifetime 1 stimuli mainly by having less form information but valid local information.

If local motion information is needed to perform action recognition properly it should be more difficult to do the task with the sequential position stimuli than with the classic stimuli. On the other hand if form information is essential there should be a clear difference in performance between the two sequential position stimuli in favor for the lifetime 1 condition. The results showed clearly that sequential position stimuli are adequate stimuli for action recognition tasks. Observers were good in recognizing actions from stimuli that contain only sequential position information. Moreover, the recognition of the portrayed action in the sequential position condition with a point-lifetime of only one frame was as good as in the classic-condition. Comparing the performance for sequential position stimuli with different point-lifetimes showed that less form information in the stimuli can not be compensated by local motion information.

6.1.3 Avoiding Local Motion Information by Applying a Counter-Change Technique

In Chapter 4 a counter-change technique was used to avoid local motion information in the stimuli. For that an invisible walker moved over a field of random black and white points, causing points, which lay under the walker to flip their color. To investigate the influence of the number of walker points and therefore the influence of the amount of form information, the number of presented points in the presentation field was varied as well as the probability of point flips. The points in the presentation field additionally masked the form of the figure, which was then only visible by point flips.

The presentation field consisted of either 1200 or 1600 randomly placed black or white points. The probability of point flip was either 10 percent or 20 percent. The performance on forward/backward walking discrimination was examined. The results were compared to previous results on a forward/backward discrimination task with sequential position walkers.

The results showed clearly that subjects were able to detect the walking direction in this kind of stimuli. A comparison with data of sequential position stimuli showed a similar strong dependency on the amount of form information. Furthermore, the fact that motion information in these stimuli is strongly reduced argued against the use of local motion information and for the importance of form information in the perception of biological motion.

6.2 Zusammenfassung

In der vorliegenden Arbeit wurden die Einflüsse von Form-und Bewegunsinformation auf die Wahrnehmung von biologischer Bewegung untersucht. Dafür wurden unterschiedliche Stimuli verwendet, durch die es möglich war die Anteile von Form- und Bewegungsinformation möglichst getrennt zu untersuchen. In Kapitel 2 und in Kapitel 3 wurden neben den klassischen Stimuli, bei denen die einzelne Punkte direkt auf den Gelenken platziert sind und ihre Position beibehalten auch sogenannte Sequential-Position Stimuli verwendet. Die Punkte bei Sequential-Position Stimuli haben im Gegensatz zu den Punkten beim klassischen Stimuli randomisierte Positionen auf dem Körper und werden nach einer bestimmten Lebensdauer auf eine neue randomisierte Position auf dem Körper versetzt. In Kapitel 4 wurde ein weiterer Stimulus-Typ eingeführt der sogenannte Random-Luminance Stimulus. Hierbei handelt es sich um einen Stimulus, bei dem die Punkte ähnlich dem Sequential-Position Stimulus auf randomisierten Positionen des Körpers liegen. Dies wird jedoch durch eine andere Technik erreicht. Bei dieser Technik wird der Läufer nur über sie Veränderung der Hintergrundpunkte dargestellt. Ein ansonsten unsichtbarer Läufer bewegt sich hier über ein Feld von statischen, randomisiert verteilten schwarzen und weißen Punkten. Wenn sich der Läufer über eine Hintergrundpunkt bewegt, wechselt dieser mit einer bestimmten Wahrscheinlichkeit von Schwarz zu Weiß beziehungsweise von Weiß zu Schwarz.

6.2.1 Wahrnehmung von Biologischer Bewegung mit Limitierter Punkt-Lebensdauer in Verschiedenen Ansichten

Kapitel 2 beschäftigt sich mit einer Reihe von Experimenten mit klassischen Stimuli und Sequential-Position Stimuli, bei denen die Lebensdauer der Punkte limitiert wurde. Die Leistung in einer Vorwärts-Rückwarts Unterscheidungsaufgabe wurde untersucht, wobei ein Vergleich von Läufern in orthographischer und perspektivischer Darstellung erfolgte. Zusätzlich wurden Ansichten der Läufer und die Lebensdauer der Punkte variiert.

Experiment 1 Mit Läufern in unterschiedlichen Ansichten wurde untersucht ob lokale Bewegungsinformation die Leistung in einer Vorwärts-Rückwarts Unterscheidungsaufgabe verbessern kann, wenn die Sequential-Position Läufer in perspektivischer Projektion dargestellt werden. Die orthographische Projektion enthält im Gegensatz zur perspektivischen Projektion keine Information über Struktur und Bewegung in der Tiefe. Ein Vorteil durch diese zusätzliche Information sollte sich demnach in einer besseren Leistung bei der Darstellung in perspektivischer Projektion zeigen.

In beiden Projektionen wurden für die Sequential-Position Stimuli sowohl für die Profil-Ansicht als auch für die Halb-Profil Ansicht gleich gute Leistungen gefunden, jedoch nicht für die Frontal-Ansicht. Eine detaillierte Analyse der einzelnen Parameter Punkt-Lebensdauer, Punkte pro Einzelbild und Gesamtzahl der Stimuluspunkte zeigte ähnliche Ergebnisse für orthographische und perspektivische Projektion. Die Leistung war generell besser für eine Gesamtzahl von 512 Stimuluspunkten als für eine Gesamtzahl von 128 Stimuluspunkten. Mit einer Gesamtzahl von 512 Stimuluspunkten war die Leistung für Sequential-Position Läufer mit limitierter Punkt-Lebensdauer genauso gut wie die Leistung bei klassischen Läufern. Dieses Ergebnis war unabhängig von der Anzahl der Punkte pro Einzelbild. Eine Erhöhung der Punkt-Lebensdauer von einem auf zwei Einzelbilder hatte keinen positiven Einfluss auf die Leistung. Ein zusätzliches Experiment bestätigte, dass weder in der Profil noch in der Halbprofil-Ansicht höhere Lebensdauern einen Vorteil bringen.

Das erste Experiment zeigte, dass die Vorwärts-Rückwarts Unterscheidungsaufgabe in der Frontalansicht sehr schwer wurde. Deshalb wurden Folgeexperimente durchgeführt, die die Rolle der Frontalansicht weiter untersuchten. Dabei wurden unterschiedliche Punkt-Lebensdauern, unterschiedliche Ansichten und spezifische Information der Fußpunkte untersucht.

Experiment 2 untersuchte ob die niedrige Leistung für Sequential-Position Läufer mit limitiert Punkt-Lebensdauer auf die Frontalansicht begrenzt ist oder ob die Leistung graduell zwischen 40° und 90° abnimmt.

Die Ansichten der Läufer wurden in zehngrad Schritten von 40° bis 90° variert. Die Läufer hatte vier Punkte pro Einzelbild, die Gesamtzahl der Stimuluspunkte betrug 512 und die Lebensdauer der Punkte wurde auf ein Einzelbild gesetzt.

Zwischen 40° und 70° war die Leistung etwa gleich hoch. Zwischen 70° und 80° und zwischen 80° und 90° fiel die Leistung jeweils stark ab. Dieses Experiment bestätigte, dass die Identifizierung der Laufrichtung bei Sequential-Position Läufern mit limitier Lebenszeit in der Frontalansicht besonders schwer ist.

Experiment 3 untersuchte die Frage ob eine Erhöhung der Lebensdauer um mehr als 2 Einzelbilder zu einer Verbesserung der Leistung in der Frontalansicht führt. Die Lebensdauer der Punkte wurde graduell erhöht bis die Lebensdauer ähnlich der eines klassischen Läufers war.

Klassische Läufer und Sequential-Position Läufer mit einer Lebensdauer von 1, 4, 8, 16, oder 32 Einzelbilder wurden verwendet. In der Bedingung mit einer Lebensdauer von 32 Einzelbildern ändert jeder Punkt nur einmal seine Position auf dem Körper. Die Sequential-Position Läufer hatten 12 Punkte pro Einzelbild, identisch zum klassischen Läufer.

Eine Erhöhung der Punkt-Lebenszeit führte zu keiner signifikanten Verbesserung der Leistung, so dass immer noch ein deutlicher Unterschied zwischen der höchsten

Lebensdauer Bedingung und dem klassischen Läufer bestand. Dieses Experiment zeigte, dass der Unterschied zwischen klassischen Läufern und Sequential-Position Läufern nicht durch eine Unterschied in der Lebensdauer der Punkte, beziehungsweise durch einen Unterschied an lokaler Bewegungsinformation erklärt werden kann.

Experiment 4 wurde durchgeführt um besonders die Rolle der Füße für die Erkennung der Laufrichtung in der Frontalansicht zu untersuchen. Ein klassischer Läufer wurde entwickelt, bei dem die Fußpunkte in Richtung der Knie verschoben werden konnten und somit die Länge des Unterschenkels variiert werden konnte. Fünf verschiedene Läufer wurden untersucht. Entweder war der Fußpunkt direkt auf dem Fußgelenk positioniert, auf Dreiviertel des Unterschenkels, auf der Hälfte des Unterschenkels, auf einem Viertel des Unterschenkels oder der Fußpunkt wurde ganz weggelassen. Alle Läufer wurde sowohl in der Profilansicht als auch in der Frontalansicht präsentiert.

Für beide Ansichten wurde eine gute Leistung für die meisten Bedingungen gefunden, bei denen die Punkte zwischen den Gelenken präsentiert wurden. Eine Ausnahme bildeten die Bedingungen, bei denen der unterste Punkt sehr nah am Knie präsentiert wurde, auf einem Viertel des Unterschenkels. In dieser Bedingung war die Leistung der Versuchspersonen signifikant schlechter. Für die Bedingung, bei der nur die Kniepunkte präsentiert wurden war die Leistung auf Zufalls-Niveau.

Diese Ergebnisse zeigen, dass die Ansichtsabhängigkeit bei Sequential-Position Läufern, dadurch erklärt werden kann, dass Vorwärts/Rückwarts Entscheidung von der Information über die Bewegung des Unterschenkels abhängt, die in der Frontalansicht schwer zu deuten ist.

6.2.2 Handlungserkennung bei Biologischer Bewegung mit limitierter Punkt-Lebensdauer

Kapitel 3 untersuchte die Fähigkeit Handlungserkennung mit Sequential-Position Stimuli durchzuführen. Es wurde bereits gezeigt, dass Aufgaben wie das Erkennen der Orientierung der die Entscheidung über die Laufrichtung einfach auszuführen sind mit dieser Art von Stimulus. Da Handlungserkennung eine deutlich kompliziertere Aufgabe ist, wurde in diesem Kapitel die Leistung bei klassischen und bei Sequential-Position Stimuli verglichen.

In einer Handlungserkennungsaufgabe wurden 33 Handlungen mit klassischen und Sequential-Position Stimuli getestet. Die klassischen Stimuli bestanden aus 12 Punkten, die auf den Hauptgelenken platziert wurden. Sequential-Position Stimuli hatten 8 Punkte, die auf randomisierten Positionen auf dem Körper präsentiert wurden. Die Punkte der Sequential-Position Stimuli hatten entweder eine Lebensdauer von einem Einzelbild (keine valide lokale Bewegung) oder eine Lebensdauer von vier Einzelbildern. Die Sequential-Position Stimuli mit einer Punkt-Lebensdauer von vier unterschieden sich von denen mit einer Lebensdauer von einem Einzelbild hauptsächlich dadurch, dass sie weniger Forminformation enthalten dafür aber valide lokale Bewegungsinformation.

Wenn lokale Bewegungsinformation benötigt wird um Handlungserkennung richtig ausführen zu können, sollte die Aufgabe mit Sequential-Position Stimuli schwieriger sein als mit klassischen Stimuli. Wenn andererseits Forminformation essentiell ist, sollte sich ein deutlicher Unterschied zwischen den beiden Sequential-Position Stimuli zugunsten der Bedingung mit einer Lebensdauer von einem Einzelbild zeigen.

Die Ergebnisse zeigen deutlich, dass Sequential-Position Stimuli angemessene Stimuli zur Handlungserkennung sind. Versuchspersonen haben erfolgreich Handlungen in Stimuli erkannt, die nur sequentielle Positionsinformation enthielten.

Außerdem zeigte sich, dass Handlungserkennung bei Sequential-Position Stimuli mit einer Punkt-Lebensdauer von einem Einzelbild genauso gut war wie in der klassischen Bedingung. Ein Vergleich der Sequential-Position Stimuli Bedingungen zeigte zusätzlich, dass weniger Forminformation im Stimulus nicht durch das Vorhandensein von lokaler Bewegungsinformation kompensiert werden kann.

6.2.3 Vermeidung von lokaler Bewegungsinformation durch eine Counter-Change Technik

In Kapitel 4 wurde eine sogenannte Counter-Change Technik verwendet um lokale Bewegungsinformation im Stimulus zu vermeiden. Dazu wurde ein unsichtbarer Läufer über ein Feld von randomisiert verteilten, statischen Punkten schwarzen und weißen Punkten bewegt. Wenn sich Punkte unter dem Läufer befanden, konnten diese mit einer bestimmten Wahrscheinlichkeit ihre Farbe wechseln. Um den Einfluss der Anzahl der Punkte im Läufer und damit den Einfluss der Menge an Forminformation zu untersuchen, wurde zum einen die Anzahl der Hintergrundpunkte variiert und zum anderen die Wahrscheinlichkeit eines Punkte seine Farbe zu wechseln. Da die Hintergrundpunkte zusätzlich die Form der Figur maskierten, war der Läufer nur sichtbar durch die Farbwechsel der Punkte.

Der Hintergrund bestand entweder aus 1200 oder aus 1600 randomisiert verteilten schwarzen oder weißen Punkten. Die Wahrscheinlichkeit dafür, dass ein Punkt seine Farbe wechselt lag bei entweder 10 oder 20 Prozent. Die Leistung in einer Vorwärts/Rückwärts Entscheidung wurde untersucht. Die Ergebnisse wurden mit Ergebnissen für Sequential-Position Stimuli aus früheren Studien verglichen.

Die Ergebnisse zeigen deutlich, dass Versuchspersonen mit dieser Art von Stimulus in der Lage sind die Richtung in die der Läufer geht zu erkennen. Ein Vergleich mit Daten eines Experiments mit Sequential-Position Stimuli zeigte eine vergleichbare Abhängigkeit von der Menge an Forminformation im Stimulus. Außerdem argumentiert die Tatsache, dass Bewegungsinformation in diesen Stimuli stark reduziert ist gegen die Notwendigkeit von lokaler Bewegungsinformation und für die Wichtigkeit von Forminformation bei der Wahrnehmung von biologischer Bewegung.

Part V

Appendices

Appendix A

List of Abbreviations

2D two-dimensional

3D three-dimensional

ANOVA analysis of variance

EBA extrastriate body area

FFA fusiform face area

ISI inter stimulus interval

KO kinetic occipital region

LGN lateral geniculate nuclei

LOC lateral occipital complex

MST middle superior temporal area

MT middle temporal area

OFA occipital face area

STS superior temporal sulcus

STSp posterior part of the superior temporal sulcus

V1 primary visual cortex

V2 secondary visual cortex

V3, V4 and V5 higher visual areas

Appendix B

Curriculum Vitae

Name:	Simone Kuhlmann
Born:	May 14, 1976 in Recklinghausen

1982 - 1986	Basic School, Recklinghausen
1986 - 1995	Pre-University Education, Recklinghausen

10/1995 - 06/2004	Studying Biologie at the Ruhr-Universität Bochum
08/1995 - 06/2004	Graduation Thesis: ”Psychophysische Untersuchungen zur Erkennung biologischer Bewegung aus tiefpassgefilterten Stimuli”

10/2004 - 6/2009	PhD-Student at the Westfälische Wilhelms-Universität Münster, Department of Psychology Supervisor: Prof. Dr. Markus Lappe

Appendix C

Publications

Journal Articles

- Lappe, M., Kuhlmann, S., Oerke, B. and Kaiser, M. (2006). The fate of object features during perisaccadic mislocalization. *Journal of Vision*, 6(11): 1282-1293.
- Kuhlmann, S. and Lappe, M. (2006). Recognition of biological motion from blurred natural scenes. *Perception*, 35(11):1495-1506.
- Kuhlmann, S., de Lussanet M.H.E. and Lappe, M, Perception of limited-lifetime biological motion from different viewpoints. *Journal of Vision*, (in press)

Proceedings

- Kuhlmann, S.; de Lussanet, M. H. E., and Lappe, M. (2007). Perception of three-dimensional biological motion from limited lifetime stimuli. *Perception*, 36 (suppl.), 74.

- Kuhlmann, S. and Lappe, M. (2005). Recognition of biological motion from blurred natural scenes (8. Wahrnehmungskonferenz). *In Bülthoff, H. H.; Mallot, H. A.; Ullrich, R., and Wichmann, F. A., Eds., Proceedings of the 8th Tübinger Perception Conference (TWK), Kirchentellinsfurt. Knirsch.*

- Lappe, M. and Kuhlmann, S. (2003). Perisaccadic compression merges position but preserves stimulus attributes. *Soc. Neurosci. Abstr.*, 29.

- Lappe, M.; Kuhlmann, S.; Oerke, B., and Kaiser, M. (2006). The fate of object features during perisaccadic mislocalization. *Journal of Vision*, 6(11), 1282–1293.

Acknowledgment

Besonders bedanken möchte ich mich bei Herrn Prof. Dr. Markus Lappe, für die Möglichkeit diese Arbeit durchführen zu können. Ich danke ihm für seine Betreuung, seine Ideen und seine Unterstützung während der gesamten Zeit dieser Arbeit. Ich möchte mich auch für die schöne Zeit, die ich in seiner Arbeitsgruppe verbringen durfte, bedanken. Ich habe immer gerne mit ihm zusammen gearbeitet. Für die Bereitschaft diese Arbeit zu begutachten, möchte ich mich natürlich auch bei Herrn Prof. Dr. Norbert Sachser bedanken.

Der gesamten jetzigen und ehemaligen Arbeitsgruppe danke ich für die schöne Zeit und die gute Zusammenarbeit. Allen, die mich mit Rat und Tat – zum Beispiel beim Korrekturlesen dieser Arbeit – unterstüzt haben, vielen vielen Dank. Auch wenn ich diese Dissertation selbständig verfasst habe, so bin ich mir doch im Klaren, dass zu dem Gelingen dieser Arbeit auch andere auf unterschiedlichste Weise beigetragen haben. Während des Schreibens habe ich versucht, dies durch den Plural ”we“ zum Ausdruck zu bringen.

Ich möchte mich bei meinen jetzigen und ehemaligen Mitdoktoranden und meiner Bio-Motion-Truppe für die schöne Zeit innerhalb und auch außerhalb der Uni bedanken. Ganz besonders möchte ich da natürlich Dr. Marc de Lussanet herausheben. Marc hat mich während dieser Arbeit auf so viele Weisen unterstützt. ich möchte ihm für die Programmierung der Stimuli, die Hilfe bei der Statistik und die vielen Ideen und fruchtbaren Diskussionen bedanken, die mir immer

weiter geholfen haben.

Besonderer Dank gilt auch Ursula Husemann, der guten Seele der Arbeitsgruppe. Nicht nur dafür, dass sie mir und auch allen anderen bei bürokratischen und finanziellen Angelegenheiten eine große Hilfe war, sondern vor allem war sie mir in dieser Zeit auch immer eine gute Freundin.

Ich möchte mich auch bei all meinen Freunden bedanken, die mir in dieser Zeit zur Seite standen. Meinen Eltern danke ich für ihr Verständnis, ihre Unterstützung und Hilfe während der gesamten Zeit meines Studium. Ohne Euch hätte ich das nie geschafft. Danke. Großer Dank gilt natürlich meinem Freund Bert, der mich mich immer wieder aufgemuntert hat, wenn es nicht so lief, wie ich es mir vorgestellt hatte.

Bibliography

Aaen-Stockdale, C.; Thompson, B.; Hess, R. F., & Troje, N. F. (2008). Biological motion perception is cue-invariant. *J. Vis.*, *8 (8)*(6), 1–11. (doi:10.1167/8.8.6.).

Aggarwal, J. K. & Cai, Q. (1999). Human motion analysis: a review. *Comp. Vis. Image Understand.*, *73*, 428–440.

Aggarwal, J. K.; Cai, Q.; Liao, W., & Sabata, B. (1998). Nonrigid motion analysis: articulated and elastic motion. *Comp. Vis. Image Understand.*, *70*(2), 142–156.

Ahlström, V.; Blake, R., & Ahlström, U. (1997). Perception of biological motion. *Perception*, *26*(12), 1539–1548.

Allison, T.; Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.*, *4*(7), 267–278.

Baker, C. L. & Braddick, O. (1985). Temporal properties of the short-range process in apparent motion. *Perception*, *14*, 181–192.

Beintema, J. A.; Georg, K., & Lappe, M. (2006). Perception of biological motion from limited lifetime stimuli. *Percept. Psychophys.*, *68*(4), 613–624.

Beintema, J. A. & Lappe, M. (2002). Perception of biological motion without local image motion. *Proc. Nat. Acad. Sci. USA*, *99*(8), 5661–5663. (doi:10.1073pnas.082483699).

Beintema, J. A.; van den Berg, A. V., & Lappe, M. (2002). Receptive field structure of flow detectors for heading perception. In Dietterich, T. G.; Becker, S., & Ghahramani, Z., Eds., *Advances in Neural Information Processing Systems*, volume 14, pages 149–156. MIT Press, Cambridge, MA.

Bertenthal, B. I. & Pinto, J. (1994). Global processing of biological motions. *Psychol. Sci.*, *5*(4), 221–225.

Bingham, G. P. (1987). Kinematic form and scaling: further investigations on the visual perception of lifted weight. *J. Exp. Psychol. Hum. Percept. Perform.*, *13*, 155–177.

Bobick, A. F. & Davis, J. W. (2001). The recognition of human movement using temporal templates. *IEEE-PAMI*, *23*(3), 257–267. (doi:10.1109/34.910878).

Bonda, E.; Petrides, M.; Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J. Neurosci.*, *16*(11), 3737–3744.

Brown, R. W.; Pola, J., & Matin, L. (1974). Visual persistence: effects of flash luminance, duration, and energy. *Vis. Res.*, *14*(4), 295–303.

Bülthoff, I.; Bülthoff, H. H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nat. Neurosci.*, *1*(3), 254–257. (doi:10.1038/699).

Carey, D.; Perrett, D., & Oram, M. (1997). Recognizing, understanding and reproducing actions. In Jeannerod, M. & Grafman, J., Eds., *Handbook of neuropsychology – action and cognition*, pages 111–130, Amsterdam. Elsevier.

Casile, A. & Giese, M. A. (2005). Critical features for the recognition of biological motion. *J. Vis.*, *5*, 1–14. (doi:10.1167/5.4.6).

Chang, D. H. F. & Troje, N. F. (2009). Acceleration carries the local inversion effect in biological motion perception. *J. Vis.*, *9 (1)*(19), 1–17. (doi:10.1167/9.1.19.).

Chatterjee, S.; Freyd, J., & Shiffrar, M. (1996). Configural processing in the perception of apparent biological motion. *J. Exp. Psychol. Hum. Percept. Perform.*, *22*(4), 916–929.

Chen, Z. & Lee, H.-J. (1992). Knowledge-guided visual perception of 3-D human gait from a single image sequence. *IEEE-SMC*, *22*(2), 336–342. (doi:10.1109/21.148408).

Coltheart, M. (1980). Iconic memory and visible persistence. *Percept. Psychophys.*, *27*(3), 183–228.

Cutting, J. E. (1981). Coding theory adapted to gait perception. *J. Exp. Psychol. Hum. Percept. Perform.*, *7*(1), 71–87.

Cutting, J. E. & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bull. Psychonom. Soc.*, *9*, 353–356.

Cutting, J. E.; Moore, C., & Morrison, R. (1988). Masking the motions of human gait. *Percep. Psychophys.*, *44*(4), 339–347.

Decety, J.; Grezes, J.; Costes, N.; Perani, D.; Jeannerod, M.; Procyk, E.; Grassi, F., & Fazio, F. (1997). Brain activity during observation of actions: Influence of action content and subject's strategy. *Brain*, *120*, 1763–1777.

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception*, *22*(1), 15–22. (doi:10.1068/p220015).

Dittrich, W. H. (1999). Seeing biological motion - is there a role for cognitive strategies? In Braffort, A.; Gherbi, R.; Gibet, S., & Richardson, J., Eds., *Gesture-based communication in human-computer interaction*, pages 3–22, Berlin. Springer-Verlag.

Dittrich, W. H.; Troscianko, T.; Lea, S. E. G., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*(6), 727–738.

Downing, P. E.; Jiang, Y.; Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*, 2470–2473.

Druks, J. & Masterson, J. (2000). *An Object and Action Naming Battery*. Hove, East Sussex: Psychology Press.

Fiez, J. & Tranel, D. (1997). Standardized stimuli and procedures for investigating the retrieval of lexical and conceptual knowledge for actions. *Mem. Cognit.*, *25*(4), 543–569.

Gallese, V.; Fadiga, L.; Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119 (Pt 2)*, 593–609.

Gavrila, D. M. (1999). The visual analysis of human movement: a survey. *Comp. Vis. Image Understand.*, *73*, 82–98.

Gazzaniga, M. S.; Ivry, R. B., & Mangun, G. R. (1998). *Cognitive Neuroscience - The Biology of the Mind*. Norton, NY.

Giese, M. A. & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nat. Rev. Neurosci.*, *4*(3), 179–192. (doi:10.1038/nrn1057).

Grossman, E. D. & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vis. Res.*, *41*, 1475–1482.

Grossman, E. D. & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron*, *35*(6), 1167–1175. (doi:10.1016/S0896-6273(02)00897-8).

Grossman, E. D.; Donnelly, M.; Price, R.; Pickens, D.; Morgan, V.; Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *J. Cog. Neurosci.*, *12*(5), 711–720. (doi:10.1162/089892900562417).

Hock, H.; Kogan, K., & Espinoza, J. (1997). Dynamic thresholds for the perception of single-element apparent motion: Bistability from local cooperativity. *Percept. Psychophys.*, *59*, 1077–1088.

Hock, H. S.; Gilroy, L., & Harnett, G. (2002). Counter-changing luminance: A non-fourier, nonattentional basis for the perception of single-element apparent motion. *J. Exp. Psychol. Hum. Percept. Perform.*, *28*(1), 93–112.

Hodgins, J. K.; O'Brien, J. F., & Tumblin, J. (1998). Perception of human motion with different geometric models. In 4, Ed., *IEEE Transactions on Visualization and Computer Graphics*, volume 4, pages 307–316. (doi:10.1109/2945.765325).

Hoffman, D. D. & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biol. Cybern.*, *42*, 195–204.

Jeannerod, M. (1994). The representing brain. neural correlates of motor intention and imagery. *Behav. Brain Sci.*, *17*, 187–245.

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

Kilner, J. M.; Vargas, C.; Duval, S.; Blakemore, S. J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nat. Neurosci.*, *7*(12), 1299–1301. (doi:10.1038/nn1355).

Knoblich, G. & Prinz, W. (2001). Recognition of self-generated actions from kinematic displays of drawing. *J. Exp. Psychol. Hum. Percept. Perform.*, *27*(2), 456–465. (doi:10.1037//0096-1532.27.2.456).

Kozlowski, L. T. & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Percept. Psychophys.*, *21*(6), 575–580.

Kuhlmann, S. & Lappe, M. (2006). Recognition of biological motion from blurred natural scenes. *Perception*, *35*(11), 1495–1506. (doi:10.1068/p5500).

Lange, J. (2006). *Perception of biological motion by form analysis*. PhD thesis, Physik, Universität Marburg, Markus Lappe, Allgemeine Psychologie, Universität Münster.

Lange, J.; Georg, K., & Lappe, M. (2002). Ideal-observer-model and psychophysical experiments on the role of form information biologial motion perception. In Würtz, R. P. & Lappe, M., Eds., *PAI Proceedings in Artificial Intelligence*, volume Dynamic Perception, Workshop of GI section 1.0.4 ”Image Understanding” and the European Networks MUHCI and ECOVISION, pages 109–114, St. Augustin. Infix Verlag.

Lange, J.; Georg, K., & Lappe, M. (2006). Visual perception of biological motion by form: a template-matching analysis. *J. Vis.*, *6*(8), 836–849. (doi:10.1167/6.8.6).

Lange, J. & Lappe, M. (2006). A model of biological motion perception from configural form cues. *J. Neurosci.*, *26*(11), 2894–2906. (doi:10.1523/JNEUROSCI.4915-05.2006).

Lee, J. & Wong, W. (2004). A stochastic model for the detection of coherent motion. *Biol. Cybern.*, *91*, 306–314. (doi:10.1007/s00422-004-0516-0).

Livingstone, M. & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, *240*, 740–749.

Marr, D. (1982). *Vision*. Freeman, San Francisco.

Mather, G. & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proc. R. Soc. Lond. B*, *258*, 273–279.

Mather, G.; Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proc. R. Soc. Lond. B*, *249*, 149–155.

Mather, G. & West, S. (1993). Recognition of animal locomotion from dynamic point-light displays. *Perception*, *22*(7), 759–766.

McLeod, P.; Dittrich, W.; Driver, J.; Perrett, D., & Zihl, J. (1996). Preserved and impaired detection of structure from motion by a 'motion-blind' patient. *Vis. Cogn.*, *3*(4), 363–391. (doi:10.1080/135062896395634).

Michels, L.; Lappe, M., & Vaina, L. M. (2005). Visual areas involved in the perception of human movement from dynamic form analysis. *NeuroReport*, *16*(10), 1037–1041.

Neri, P.; Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature*, *395*, 894–896.

Norman, J. F.; Payton, S. M.; Long, J. R., & Hawkes, L. M. (2004). Aging and the perception of biological motion. *Psychol. Aging*, *19*(1), 219–225. (doi:10.1037/0882-7974.19.1.219).

Oram, M. W. & Perret, D. I. (1994). Responses of anterior superior temporal polysensory area STPa neurons to "biological motion" stimuli. *J. Cog. Neurosci.*, *6*(2), 99–116.

Oram, M. W. & Perret, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area STPa of the macaque monkey. *J. Neurophysiol.*, *76*(1), 109–129.

Peelen, M. V. & Downing, P. E. (2005). Is the extrastriate body area involved in motor actions? *Nat. Neurosci.*, *8*(2), 125. (doi:10.1038/nn0205-125a).

Perrett, D. I. & Oram, M. W. (1998). Visual recognition based on temporal cortex cells: Viewer-centred processing of pattern configuration. *Z. Naturforsch.*, *53*, 518–541.

Pinto, J. & Shiffrar, M. (1999). Subconfigurations of the human form in the perception of biological motion displays. *Acta Psychol.*, *102*, 293–318.

Pollick, F. E.; Fidopiastis, C., & Braden, V. (2001). Recognising the style of spatially exaggerated tennis serves. *Perception*, *30*(3), 323–338. (doi:10.1068/p3064).

Rashid, R. F. (1980). Towards a system for the interpretation of moving lights display. *IEEE Trans. Patt. Anal. Mach. Intell.*, *2*(6), 574–581.

Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system. *Ann. Rev. Neurosci.*, *27*, 169–192. (doi:10.1146/annurev.neuro.27.070203.144230).

Rizzolatti, G.; Fadiga, L.; Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.*, *3*(2), 131–141.

Rizzolatti, G. & Matelli, M. (2003). Two different streams form the dorsal visual system: anatomy and functions. *Exp. Brain Res.*, *153*(2), 146–157. (doi:10.1007/s00221-003-1588-0).

Runeson, S. & Frykholm, G. (1981). Visual perception of lifted weight. *J. Exp. Psychol. Hum. Percept. Perform.*, *7*(4), 733–740.

Runeson, S. & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: expectation, gender recognition, and deceptive intention. *J. Exp. Psychol. Gen.*, *112*(4), 585–615.

Saygin, A. P.; Wilson, S. M.; Hagler, D. J., J.; Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *J. Neurosci.*, *24*(27), 6181–6188. (doi:10.1523/JNEUROSCI.0504-04.2004).

Servos, P.; Osu, R.; Santi, A., & Kawato, M. (2002). The neural substrates of biological motion perception: an fMRI study. *Cereb. Cortex*, *12*, 772–782.

Shiffrar, M. & Freyd, J. (1990). Apparent motion of the human body. *Psychol. Sci.*, *1*(4), 257–264.

Shiffrar, M. & Freyd, J. (1993). Timing and apparent motion path choice with human body photographs. *Psychol. Sci.*, *4*(6), 379–384.

Shiffrar, M.; Lichtey, L., & Chatterjee, S. H. (1997). The perception of biological motion across apertures. *Percept. Psychophys.*, *59*(1), 51–59.

Shipley, T. F. (2003). The effect of object and event orientation on perception of biological motion. *Psychol. Sci.*, *14*(4), 377–380.

Thornton, I. M.; Pinto, J., & Shiffrar, M. (1998). The visual perception of human locomotion. *Cogn. Neuropsychologia*, *15*(6/7/8), 535–552.

Todd, J. T. (1983). Perception of gait. *J. Exp. Psychol. Hum. Percept. Perform.*, *9*(1), 31–42. (doi:10.1037/0096-1523.9.1.31).

Troje, N. & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a ”life detector”? *Curr. Biol.*, *16*, 821–824. (doi:10.1016/j.cub.2006.03.022).

Troje, N. F. (2002). Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. *J. Vis.*, *2*, 371–387. (doi:10:1167/2.5.2).

Troje, N. F.; Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Percep. Psychophys.*, *67*(4), 667–675.

Ullman, S. (1984). Maximizing rigidity: the incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception*, *13*(3), 255–274.

Ungerleider, L. G. & Mishkin, M. (1982). Two cortical visual systems. In Ingle, D. J.; Goodale, M. A., & Mansfield, R. J., Eds., *The Analysis Of Visual Behavior*, pages 549–586, Cambridge, MA. MIT Press.

Vaina, L.; Cowey, A.; LeMay, M.; Bienfang, D., & Kikinis, R. (2002). Visual deficits in a patient with ’kaleidoscopic disintegration of the visual world’. *Eur. J. Neurol.*, *9*(5), 463–477.

Vaina, L. M.; Cowey, A.; Eskew, R. T. J.; LeMay, M., & Kemper, T. (2001). Regional cerebral correlates of global motion perception: evidence from unilateral cerebral brain damage. *Brain*, *124*(Pt 2), 310–321.

Vaina, L. M.; Lemay, M.; Bienfang, D. C.; Choi, A. Y., & Nakayama, K. (1990). Intact biological motion and structure from motion perception in a patient with impaired motion mechanisms: A case study. *Vis. Neurosci.*, *5*, 353–369.

Vaina, L. M.; Solomon, J.; Chowdhury, S.; Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proc. Nat. Acad. Sci. USA*, *98*(20), 11656–11661. (doi:10.1073pnas.191374198).

Vanrie, J.; Dekeyser, M., & Verfaillie, K. (2004). Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception*, *33*(5), 547–560. (doi:10.1068/p5004).

Verfaillie, K. (1993). Orientation-dependent priming effects in the perception of biological motion. *J. Exp. Psychol. Hum. Percept. Perform.*, *19*(5), 993–1013.

Walk, R. D. & Homan, C. P. (1984). Emotion and dance in dynamic light displays. *Bull. Psychonom. Soc.*, *22*(5), 437–440.

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