Perception of biological motion
by form analysis
(Wahrnehmung biologischer Bewegung
durch Formanalyse)

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich Physik

der Philipps-Universität Marburg

vorgelegt von

Joachim Lange

aus

Korbach

Marburg/Lahn, 2006

Contents

List of Abbreviations

EBA	extrastriate body area
F5	premotor cortex
FFA	fusiform face area
fMRI	functional magnetic resonance imaging
IT	inferior temporal area
KO	kinetic occipital region
LGN	lateral geniculate nuclei
MST	medial superior temporal area
MT	middle temporal area
OFA	occipital face area
PET	positron emission tomography
SD	standard deviation
SE	standard error
STPa	anterior superior temporal polysensory area
STS	superior temporal sulcus
TE	temporal area
TEO	temporal occipital area
TMS	transcranial magnetic stimulation
V1	primary visual cortex
V2	visual area 2
V3	visual area 3
V3a	visual area 3a
V4	visual area 4
V5	visual area 5
V5a	visual area 5a

Chapter 1

Introduction

1.1   General introduction

Everyday we see other human individuals. We observe them passively or interact with them by reacting on their actions. For both, the passive and the active actions we need to perceive and analyze the movements of others fast and accurately. The process of analyzing other humans movements is complex for there are various movements, which are subtle. Slightly different movements and gestures can have a different meaning and impact on other individuals. Mostly, recognition of the individual and its action also needs interpretation in a social context. Nevertheless, the perception of humans and their actions is fast and accurately.

Johansson (1973) has shown that humans can perceive the movements of others even if the visual information is degraded to a handful of moving point-lights. Johansson recorded the movements of human actors in the dark with only small point-lights attached to their joints. If human observers viewed the recordings of these point-lights, they immediately recognized the portrayed actions. Even if more than one person was shown the sets of point-lights could easily be disentangled. Later on, other studies showed that this limited visual information is sufficient to recognize friends and the gender of the walker (Cutting and Kozlowski, 1977; Dittrich, 1993).

These findings contradict the intuitive idea that we need detailed information about other individuals to interact with them. Instead, it shows that not as much visual information as available is needed to perceive others and that the our brains may have developed specialized mechanisms which are sensitive and well adapted to this common and often experienced visual information.

Over the last years, several studies have investigated the perception of other human beings. The common approach of experiments is to present a stimulus as an input to the visual system and measure the response of the human observer. From the relationship of stimulus input and output of the visual system (i.e. the response) the experimenter can draw conclusions about the mechanisms of the brain. For this, clear definitions of the visual input (i.e. the stimulus) and output are required to obtain unambiguous models of the visual system of the brain.

1.1.1 Challenging problems

Especially when we investigate the perception of human movements, the problem of clearly defined input and output arises. Human movements are complex, containing many degrees of freedom. Many factors define the stimulus and many aspects can be retrieved. Local features like position and local motion signals define the stimulus, as well as global aspects like the general form, or temporal characteristics like stimulus duration or velocity. The stimulus can have different geometrical features like size or depth information. The stimulus may be three or two dimensional, moving or static.

On a higher level, the stimuli may differ in their semantic meaning. For example, observers can retrieve identity, gender, emotions and intentions from the stimulus. This means that similar basic aspects like form and motion have a different impact in a different context.

Our visual system has to deal with these multi-dimensional aspects. Surprisingly, the study by Johansson (1973) has shown that the visual system can cope with stimuli of humans accurately and fast even if they are ill-defined. Missing depth and reduced structural information does not impoverish the perception noteworthy.

But, how does the brain overcome the missing information? Does the brain hierarchically analyze the information or does it overcome the missing information with a priori known constraints, assumptions, and internal models about the surrounding world? What information of all available is used by the brain? To answer these questions it would be desirable to study specific stimulus information and the output (i.e. the response) of the visual system to this specific feature isolated. Unfortunately, this is impossible because the aspects defining the stimulus are always coupled. Global features like the form of the stimulus cannot be identified without taken local features like position and local motion signals into account. The overall percept of the stimulus cannot be disentangled from these local features and not from interpreting the movement.

It is possible to reduce the influence of specific features like structural information or local motion signals in point-light walkers (see section 1.2.1) but not to omit them.

Computational models can help to understand the mechanisms of the brain particularly in the case of the perception of human movements. They can concentrate on specific aspects of the stimulus and analyze and compute these aspects isolated. We can make assumptions on the internal models that might be used by the brain. By comparing the output of the computational model to experimentally obtained results we can draw theoretical conclusions on the mechanisms the brain may use.

In the following sections I will provide some background about what we know up to now about the perception of human movements. For this, I will focus on the local and global influence of form and motion signals and on the temporal characteristics such as the mechanisms for the fast perception. I will introduce results from psychophysical, electrophysical, imaging studies and computational models.

1.2 Biological Motion

Originally, the term 'biological motion' refers to the stimulus used in the study by Johansson (1973) . It describes the compelling example of the visual system's ability to recover object information from sparse input. Johansson showed that human observers are able to identify the form of a human person depicted only by thirteen point-lights in a fraction of a second. Since then the term 'biological motion' has been used variously. Almost all movements of 'biological' objects, e.g. hands, whole bodies, are referred to as biological motion. Also, the perception of it can be investigated by stimuli, other than point-light stimuli. Yet, point-light walkers are still the most commonly used stimuli. Since originally 'biological motion perception' referred to the perception of point-light stimuli and all experiments in my studies are conducted with this kind of stimulus the expression 'biological motion' denotes point-light displays of walking human persons in my studies.

1.2.1 Point-light walker

In his early study Johansson (1973) not only coined the expression 'biological motion' but also introduced the point-light walker as a stimulus. He used only thirteen light-points on the major joints (i.e. shoulders, elbows, wrists, hips, knees, and ankles) and

the head of an otherwise invisible human body. This way the stimulus minimizes the structural information about the body. The only information, which can be effectively used for a visual analysis, is the sparse local position signals of the light points and their local motion trajectories if the stimulus moves (see section 4.3.2 for details). Several studies have shown that this highly degraded stimulus contains enough information not only to recognize the walker per se, but also to perform more sophisticated tasks like gender recognition (Kozlowski and Cutting, 1977; Mather and Murdoch, 1994) or identification of individuals (Cutting and Kozlowski, 1977; Troje, 2002). Point-light displays of humans attract the attention of infants more than meaningless displays do (Fox and McDaniel, 1982) and children can recognize the form of animals and humans when they view point-light displays (Pavlova et al., 2001). Even cats are able to recognize point-light displays of other cats (Blake, 1993).

While Johansson and others recorded the movements of real actors, Cutting (1978) developed a computer program to generate an artificial walker based on a mathematical algorithm, closely similar to a real walker (for details see Figure 2.3 a). This stimulus was only available for walking, but had the advantage of being easy to modify and replicate.

Beintema and Lappe (2002) presented an adapted version of these point-light walkers to minimize local motion signals without changing structural information. They also used the computer-generated walker by Cutting but changed it in a way that additional to structural information also local motion signals are minimized. To this end, the dots were no longer positioned on the joints but randomly on the limbs (i.e. arms and legs). From frame to frame the dots change their position also randomly. Thus, the local motion signals provided by the dots do not give useful information (see section 2.2 for details). By keeping the dots' position on the limbs, constant for a defined number of frames the contribution of the local motion signals (i.e. lifetime of the dots) could be manipulated. Beintema and Lappe (2002) showed that the recognition rate for this stimulus was only marginally worse than for the classical 'Cutting walker'.

All in all, point-light walkers are suited for the investigation of human movements for the influence of features like structural information and local motion signals can be manipulated or reduced to a near-absent minimum. Especially in the present thesis point-light walkers are helpful since here the emphasis is on the investigation of the roles of form and motion analysis.

1.3   The influence of form and motion

The recognition of a human action from a stimulus is equivalent to the recognition of
the global form of the stimulus. That is, the stimulus as a whole will be recognized as
a human not only specific features like limbs. But, it is unclear from what this global
impression is derived. Divers mechanisms of the brain are possible. The simplest
hypothesis is that in the brain the stimulus is matched as a whole to an existing
internal model of a human. Other possible mechanisms are that the brain gradually
integrates local features such as the local motion signals or the local position signals
to obtain the global impression of a human person.

In this section I will introduce studies, which favored either the approach via motion
or via form analysis. And I will discuss studies that investigated the role of the stimulus'
temporal characteristics. Since motion and form analysis are linked to characteristic
areas of the brain I will start by giving a short overview of the visual pathways.

1.3.1   Visual pathways

When light falls on the eye, this visual information is acquired by the retina. The
visual information is passed via the *lateral geniculate nuclei* (LGN) to the *primary
visual cortex* (V1) where the information arrives in the cortex and is passed to V2.
Beyond V2, the visual cortex is believed to divide in two general streams of information
processing: one for motion and location, the other for color and form. These are known
as the dorsal and ventral streams. Because of the functions they serve, they are also
called the 'where' and 'what' paths (Mishkin et al., 1983).

The dorsal pathway, dedicated to the analysis of motion, runs via V3a to the *middle
temporal area* (MT in monkeys, the human homolog is known as V5) to the *middle
superior temporal area* (MST in monkeys, V5a in humans). Similarly in structure, the
ventral path passes the information via V4 to TE and TEO (known as *inferior temporal
area* (IT) in humans). This pathway computes the form and color of objects but is
insensitive to the object's motion.

Along each serial pathway the receptive fields of the neurons increase, i.e. the visual
fields the neurons respond to, enlarge. Meanwhile, the neurons show an increasing
degree of functional and hierarchical organization and they become more unspecific
for details (Kobatake and Tanaka, 1994). For example, while neurons in V1 and V2
respond selectively to simple lines and their orientations, cells in located in the IT
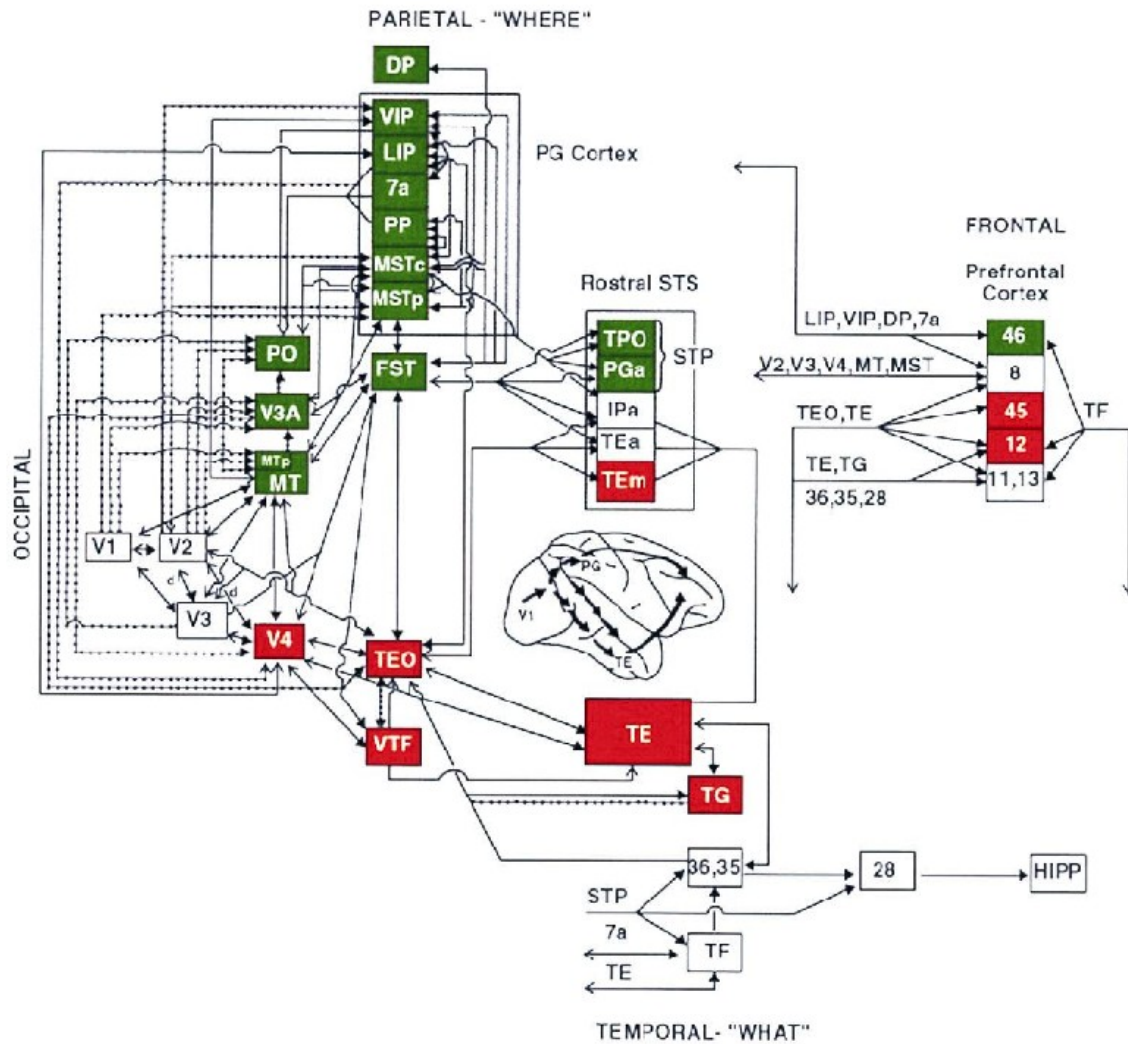
Fig. 1.1: Vision is a complex sense and involves many brain areas to process the visual information. The visual cortex is often divided in two separate processing streams named 'where' (or 'dorsal') for motion analysis and 'what' (or 'ventral') stream responsible for object recognition. Adapted from Ungerleider (1995).

complex are known to respond selectively complex objects and to faces (Rolls, 1992) but not to simple stimuli. Similarly, in the dorsal stream the selectivity for optic flow patterns becomes more specialized the higher the hierarchical organization is.

In the last years, several connections and interactions between the dorsal and the ventral pathways have been found so that its distinct separation does not longer really hold. Nevertheless, the idea of two different paths for the analysis of form and motion is still popular.

1.3.2 Motion analysis

The absence of structural information and the inability of naive observers to recognize a human person out of a static frame inspired Johansson (1973) to ascribe biological motion perception to a phenomenon of motion analysis. Johansson (1973) and Cutting (1981) developed models, which analyze the local motion signals of the point-light stimulus. Furthermore, Cutting (1988) concluded from a masking study that local motion analysis of the limbs is essential for recognizing biological motion. Here, he showed biological motion stimuli embedded in different kinds of noise. His results showed that recognition of the walking direction was worst when the noise consisted of dots with motion trajectories identical to that of the stimulus dots. Mather et al. (1992) presented the stimulus embedded in randomly moving noise dots. Subjects viewed the stimulus frames alternating with a mask consisting of blank frames. Mather et al. varied the duration the mask was presented (60-100 ms). Presenting these blank intervals should interfere local motion detectors. The results should, therefore, show if local motion signals have an influence on the perception of biological motion. Indeed, the results showed that direction discrimination is not possible if blank inter-stimulus frames intermit the stimulus in noise. From this finding Mather et al. concluded that local motion signals are essential for the perception of biological motion. Neri et al. (1998) argued in a similar way. They used biological motion or simple translatory motion as a stimulus and asked subjects to detect the stimulus in noise. The results showed no differences for detection of the two stimuli. Both revealed a linear increase of threshold for increasing stimulus dots. In another experiment, they revealed that performance threshold for discriminating the walking direction of a biological motion stimulus in noise increased non-linearly with an increasing number of stimulus dots. Neri et al. (1998) concluded from the first experiment that the common information of the two stimuli (that is motion) is the driving force for biological motion perception. These biological motion filters are flexibly adapted to the stimulus, as reflected in the non-linearity revealed by the second experiment.

Moreover, fMRI studies showed that viewing a biological motion stimulus activates area MT (Grossman et al., 2000; Vaina et al., 2001). Yet, these studies did not show whether MT activation is essential for perceiving biological motion or just an artifact of perceiving a moving stimulus. In addition, they also showed an involvement of the ventral path, just as well not clearly giving explicit information about the role of

these areas. Ahlstrom et al. (1997) showed that perception of biological motion does not necessarily rely on first-order motion (i.e. luminance changes). Their biological motion stimulus, based on second-order motion (i.e. contrast changes), was still easily detectable.

1.3.3 Form analysis

Psychophysical studies with patients who suffered from strokes argue against the view that global form perception is derived from integrating local motion signals (for details see section 1.5.3). Also, there are psychophysical studies with healthy subjects that argue against an involvement of low-level motion. Shiffrar et al. (1997) investigated biological motion perception in the context of the aperture problem. The aperture problem is known as the phenomenon that detecting the direction of a homogeneous motion becomes locally ambiguous if the motion is perceived through a small hole ('aperture'). Shiffrar et al. (1997) showed line drawing stimuli of biological motion and objects like cars to human observers. The stimuli were visible only through apertures distributed over the monitor. The results showed that only the perception of biological motion stimuli can overcome the aperture problem in contrast to other, non-biological objects. This indicates that the ambiguous local motion information does not interfere with biological motion perception.

Beintema and Lappe (2002) used a novel stimulus which minimized local motion signals but maintained the form information available in the classical point-light walker. Yet, observers' recognition rate was not different from recognition of the classical walker. In addition, they gave direct evidence that increasing available form information increased performance whereas increasing local motion signals does not. Other studies excluded a major role for dynamic symmetry of the limbs (Pinto and Shiffrar, 1999). Pinto and Shiffrar used a variation of the classical biological motion stimulus. They showed that observers still can recognize the stimulus although in their stimulus the common symmetry of the limbs and, thus, the opposing movements of the limbs were missing. A study that used distracting noise dots as a mask for the biological motion stimulus also favored the influence of form information over local motion signals (Bertenthal and Pinto, 1994). Bertenthal et al. showed that noise with the same motion trajectories as the stimulus impoverished the recognition, but recognition rates always stayed above chance level. Thus, the noise is only able to interfere with the

perception on subordinate levels such as limbs but not on the level of global form perception. Despite the noise, the overall structure of the walker can always be recognized in contrast to other less recognizable stimuli such as upside-down walker.

1.3.4 Dynamics

A few studies investigated the influence of the temporal characteristics, i.e. the dynamical change of the human body. Shiffrar and Freyd (1990) and Chatterjee et al. (1996) demonstrated that human form and the interpretation of its action depended strongly on the time interval in which it is presented and on the dynamics of the movement. The study of Shipley (2003) argued that the dynamics of a movement has a stronger influence on the correct recognition than the form per se. He demonstrated this by presenting point-light displays of a walker on his feet and on his hands, respectively. The results indicate that the way the display moves has a stronger influence than the pure form analysis. Similar results are presented in a study that investigated person identification by gait (Troje, 2002). Here, the psychophysical data and the underlying model on the basis of gait analysis gave evidence that form per se is important. However, a more reliable cue for person identification is the way this form behaves in time.

1.4 Characteristics of biological motion perception

We may assume that retinal input is processed in a straightforward way to the visual cortex. This direct perception, which processes the visual information to hierarchically higher areas of the brain is called bottom-up. With more complicated visual stimuli or tasks, it is more unlikely that the perception results only from retinal input. Previous experience, assumptions about the external world and known constraints of the stimulus may influence the visual perception. We are also able to use prior knowledge, expectations, memory and attention to influence our visual perception. These processes, which exert a control function, are called top-down effects. In contrast to a biased bottom-up process, such constructive top-down processes appear when we are able to change our perception of ambiguous figures. Also, when we assign previously unknown objects or group familiar looking objects to a categorical prototype. We may also assume that generally the comparison of the visual information to internal states

of the brain or to memory is more time-consuming than a direct bottom-up process and that both processes may interact at some level.

In the case of biological motion, there are arguments for both, bottom-up (fast perception of the stimulus) as well as top-down (previously unseen movements can be recognized) influences. Thus, a question is to what extend do only bottom-up processes drive the perception and what influence do top-down processes have?

Short-range motion signals (i.e. motion up to a duration of ≈ 100 ms) are considered to involve early processing (Braddick, 1974) and, therefore, are believed to reveal bottom-up processes. Mather et al. (1992) tested the performance of subjects using inter stimulus intervals between 60 ms and 100 ms. They argued that these blank intervals should interfere with early local motion detection but not with high level processes. However, subjects' performance broke down so that Mather et al. concluded that biological motion is a low-level process.

This position has been questioned by Thornton et a. (1998). They showed that observers still correctly recognize biological motion using the set-up of Mather et al. (1992) if only the presentation time of the stimulus is long enough. Chatterjee et al. (1996) reported a similar result. In there study they applied apparent motion stimuli. Apparent motion denotes the illusion of motion when two a series of still pictures is shown (cinema relies on this effect). The results revealed that apparent motion displays of biological motion are perceived better for longer stimulus durations. This argues for a high-level process that operates at longer frame durations. Verfaillie (2000) showed in priming experiments that top-down processes (i.e. prior knowledge of the stimulus) influenced the perception of biological motion. The reaction times for recognizing the walking direction were significantly shorter when the stimulus used for priming had the same orientation as the test stimulus. Two studies by Thornton and others have demonstrated the discordance between the two views. While his first study argued for a high-level process, the second study showed that biological motion could be perceived incidentally using only low-level features. The first study used a 'chimeric' walker, i.e. a point-light walker, derived by superimposing two walkers with opposite walking direction (Thornton et al., 2003). To analyze the visual information only in a straightforward way would yield in an ambiguous percept of two walkers. However, subjects reported to perceive only one walker with a bias for walking to the right. In the other study, Thornton and Vuong (2004) showed that biological motion perception could be achieved in a passive, bottom-up fashion. In all, these studies
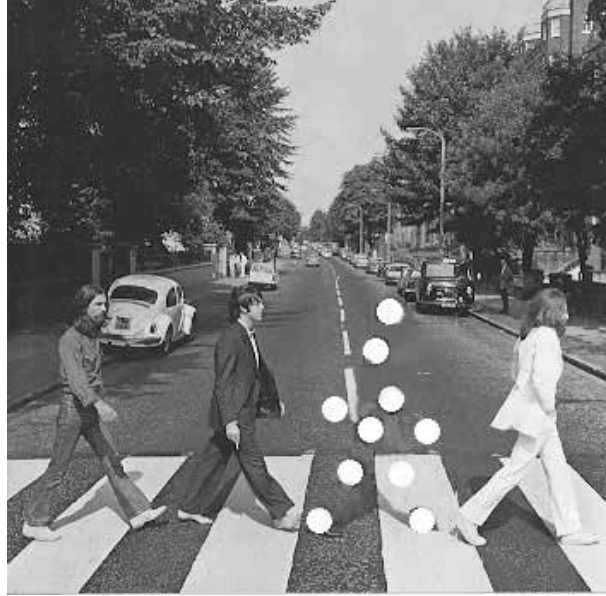
Fig. 1.2: Without a context, naive observers do not perceive a form in a point-light display of a person. Yet, if presented in the correct context, it is quite easy to recognize the human figure. Figure adapted from Thornton et al. (2003).

suggest that biological motion perception uses low-level mechanisms and processes the outputs to high-level stages where they are compared to internal models of perception. These top-down processes modulate the bottom-up information if stimulus durations are long enough.

1.5 Cortical representation of biological motion

The visual cortex is believed to be structured hierarchically and functionally. Along the dorsal and ventral path different areas are assigned to more or less specialized functions. The contribution of single areas to the perception of biological motion can be investigated by means of different methods. I will describe the results in the following sections.

1.5.1 Studies in non-humans

Electrophysiology is an invasive method to examine directly the behavior of a single cell or populations of cells mainly applied in non-human primates. Oram and Perrett (1994) accomplished the first study to investigate cell responses to a biological motion

stimulus.They found direct evidence that a population of cells in the anterior part of the *superior temporal polysensory* area (STPa) respond selectively to specific views of whole human bodies in articulated motion but not to single moving limbs. This selectivity to specific form and direction was also given for point-light displays. In another study, they found cells in STPa which are selectively responsive to the form of the stimulus. Other cells in this area were selective to the motion exclusively, and a third population integrated form and motion and was only active when the stimulus' form, orientation, and direction of movement were compatible (Oram and Perrett, 1996).

These patterns of activation are not found exclusively for moving human bodies because purposeful hand-object actions such as reaching for, picking, tearing and manipulating objects have evoked similar responses (Perrett et al., 1989; Perrett et al., 1990; Jellema et al., 2000). Jellema and Perrett (2003) showed that cells in STPa also show preferences to articulated movements of a body or body parts in comparison with non-meaningful motion. Moreover, cell responses were stronger when the observed posture implied a motion compared to responses to static postures per se.

1.5.2   Studies in humans

The first study to show a direct involvement of the *superior temporal sulcus* (STS), presumably the analog of monkey STPa in the human brain, was a study using *positron emission tomography* (PET) (Bonda et al., 1996). The posterior part of STS and Amygdala specifically showed activation when subjects viewed point-light displays of humans performing different actions.

Most of the studies used the method of *functional magnetic resonance imaging* (fMRI) to investigate the network involved in biological motion perception. Grossman et al. (2000) reported that viewing point-light figures significantly activates STS. In contrast, coherent motion and motion defined by kinetic boundaries activated only MT and the *kinetic occipital region* (KO), STS is also activated when imaging biological motion or when viewing upside-down walker, albeit that activation is less strong than for up-right biological motion displays (Grossman and Blake, 2001). Other studies have confirmed the crucial role of STS. Grezes et al. (2001) showed an anterior-posterior gradient of activation in STS for rigid vs. non-rigid motion (like biological motion). Additionally, they observed activation in the left intraparietal cortex for biological motion displays. Vaina et al. (2001) reported that, besides STS, perceiving biological

motion also specifically involves other areas like KO, parts of the cerebellum and LOC. Grossman and Blake (2002) also reported other areas than STS. They found activation for biological motion compared to scrambled biological motion in areas in the ventral path like the *occipital* and *fusiform face area* (OFA and FFA), similar to the findings of Vaina et al. (2001). The study of Beauchamp et al. (2002) showed that even static pictures of a human body can activate STS and the ventral cortex. When motion is added, the activation can also increase (Beauchamp et al., 2003). Earlier, Downing et al. (2001) reported an area which they called the *extrastriate body area* (EBA) because it showed selective responses to static pictures of human bodies and stick figures. However, the role EBA plays in perceiving point-light displays remains unclear. Downing et al. (2001) observed a stronger activation for biological motion displays than for scrambled versions. Yet, Grossman and Blake (2002) did not confirm this finding. But, Peelen and Downing (2005) reported also significant activation in FFA, involved in perceiving faces, for human bodies shown without a head.
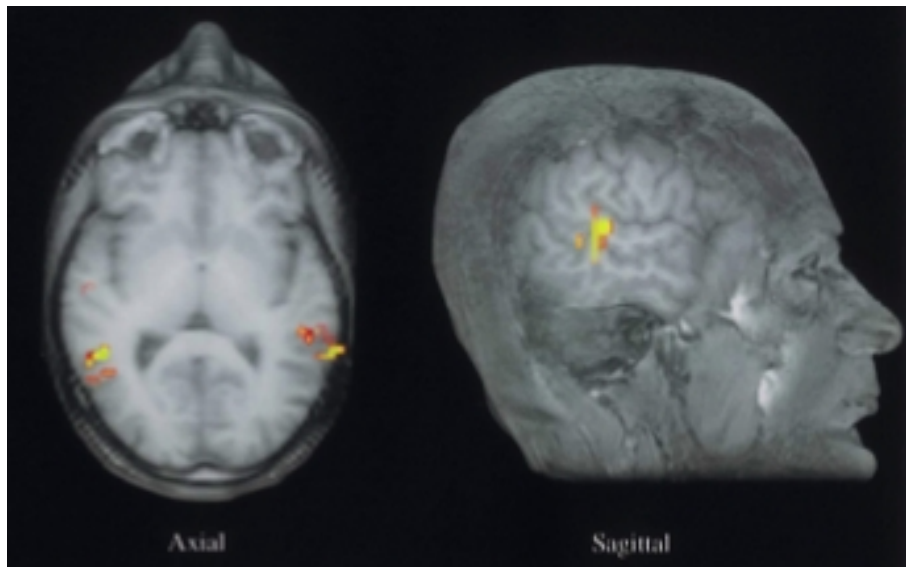
Fig. 1.3: The colored patches show the cortical representation of area STS in the human brain in axial and sagittal view. Adapted from Grossman et al. (2001).

Especially the studies of Oram and Perrett (1994, 1996), Beauchamp et al. (2003), Vaina et al. (2001) and Grossman and Blake (2002) imply that biological motion perception benefits from the motion and may be a form of structure-from-motion perception. Yet, the study of Grossman et al. (2002) revealed an indifferent response in MT to biological motion, while Bradley et al. (1998) identified MT to be the center of structure-from-motion perception.
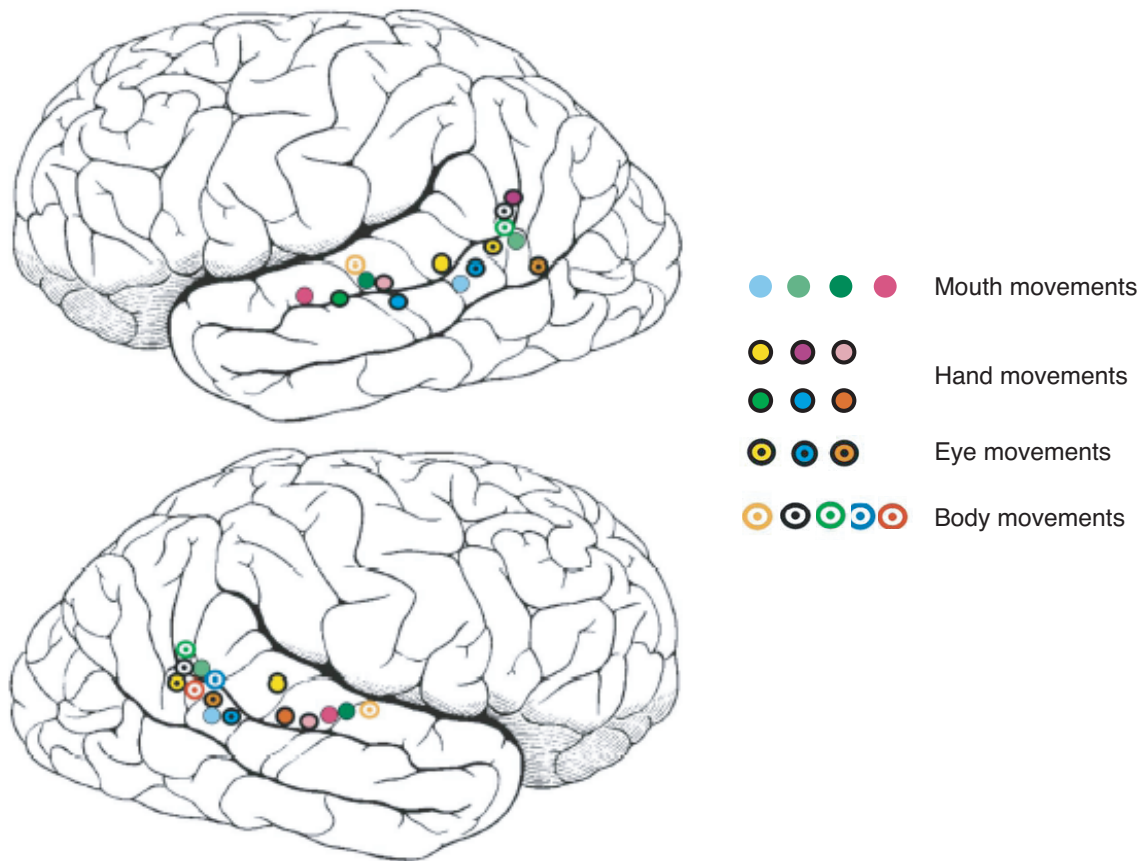
Fig. 1.4: Activation of the STS by different kinds of biological motion stimuli. Top: left hemisphere, Bottom: right hemisphere. Adapted and modified from Allison et al. (2000).

All these studies have in common that they uniformly identify STS as a crucial part of the network. Yet, the role of STS remains not clearly identified. Several studies showed that the activation is not limited to walking persons. This area is also sensitive to viewing movements of the lips, hands, and eyes (Fig. 1.4).

Due to the involvement of the area in actions that involve or require the interaction with other humans, area STS is sometimes referred to as the perceptual locus of 'social cognition' (Allison et al., 2000). For example, the gaze of another human can give important information about the person's intention or emotions. Neurophysiological studies have shown that paying attention to the gaze activates STS while paying attention to other parts of the face does not (Perrett et al., 1985). Yet, if this 'social cognition' system exists, it is a wider network that includes the Amygdala and OFA as well (Adolphs, 2003).

On the other hand, STS has been identified to play a major role in cross-modal

perception. Some studies have shown that audio-visual integration in perceiving visual speech activates STS or enhances the activity, respectively (Sekiyama et al., 2003; Wright et al., 2003). Moreover, visual perception of actions that belong to the motion repertoire of humans activates the premotor cortex in addition to visual areas (Buccino et al., 2001; Buccino et al., 2004). In particular, it has been demonstrated that perception of biological motion point-light displays also activates areas in the premotor cortex (Saygin et al., 2004). Thus, STS may not only be specific for biological motion perception but the primary role of STS may also the integration of inputs from different modalities.

1.5.3 Lesions studies

Another way to examine visual perception is to study patients with deficits, e.g. lesions of parts of the brain due to a stroke. Although it is hard to determine the exact locations of the lesions and nearly impossible to determine the affected areas, the implications on vision can be easily investigated. Vaina et al. (1990) investigated a patient who had bilateral lesions including visual pathways like MT. While he could solve tasks that involve form perception he had severe problems in tasks including spatial localization and low-level motion tasks. In contrast, he had no problems to identify structure-from-motion or biological motion displays, unless the stimulus was presented in moving noise. Similarly, McLeod et al. (1996) reported a study with the patient LM. Previous studies have shown that LM lacked a motion percept after a stroke which destroyed form processing areas like MT (Zihl et al., 1983). In this study, she was additionally confronted with a biological motion stimulus. Although she was unable to perceive simple motion or discriminate the direction of a random dot pattern movement, she was able to solve the same tasks if the stimulus contained biological motion. Also, Vaina et al. (2002) described a case study in which the patient had difficulties to integrate local motion signals to a motion percept. In addition, he could not perceive structure-from-motion but he had no problems to recognize the biological motion stimulus.

Conversely, while these studies introduce patients who had problems in motion tasks but not biological motion, there are also studies which showed that biological motion perception can be impaired despite of intact motion perception. Batelli et al. (2003) studied three patients with lesions in the parietal cortex. Although their ability to

perceive low-level tasks was indifferent to normal subjects they were unable to perceive biological motion stimuli. Batelli et al. argued for deficits in tasks which need attention driven by the parietal cortex. Schenk and Zihl (1997) examined stroke patients with lesions in the parietal cortex. They found that the perception of biological motion may be unimpaired but when a segregation from the background is necessary these patients fail to fulfill the task. Another study revealed that patients can have normal object and motion recognition performances without perceiving a form in a biological motion stimulus (Cowey and Vaina, 2000). Vaina and Gross (2004) studied four patients with brain damages due to strokes. All of them were unable to recognize a walker from a point-light figure. They had normal object recognition rates but were impaired on recognition of objects from degraded incomplete information. Their performances on motion tasks differed but were slightly impaired. Cowey and Vaina (2000) as well as Vaina and Gross (2004) presumed that all patients had a damage in STS and, thus, were unable to integrate the given information to a percept of biological motion.

1.6 Models

The recognition or reconstruction of a human form from displays can be theoretically achieved in different ways. Two main kinds of model approach can be distinguished: models with no a priori knowledge about the human form. Here, motion signals give information about correspondence between dots, thus, I will call them motion models. The other kind, form models, have an implicit model of the human form and match the images to their a priori template. Some of the models are developed solely for computer vision; others try to explain how the human brain copes with such kind of stimuli. I will describe models of both approaches.

1.6.1 Motion models

Parallel to his first description of the amazing capability of humans to perceive complex patterns of movements depicted by point-lights Johansson also presented a model to explain the analysis (Johansson, 1973). The analysis of common motion directions provides information to connect these dots to a rigid element. This analysis is carried on beginning at the center of the body (i.e. hips) to the more distal limbs. This rigidity assumption has been used in the model of Ullman (1984), too. He showed that pure

bottom-up analysis of two-dimensional motion vectors is mathematically insufficient to reconstruct a human form. Therefore, he used the assumption that observers attempt to interpret rigidly moving objects whenever possible.

The assumption of local rigidity to delimit pair-wise connections of the point lights has also been used by others (Webb and Aggarwal, 1982; Hoffman and Flinchbaugh, 1982). They assume that the axis of rotation remains fixed in a two-dimensional plane. These pair-wise connections are then iteratively combined into a hierarchically organized global form.

Especially the studies of Dittrich (1993) and Beintema and Lappe (2002) have challenged the models based on local rigidity. Dittrich (1993) used point-light stimuli in which the dots are not placed on the joints but between them. Although the constraint of rigidity of the joints is no longer present, this stimulus can still be easily recognized. Also, the stimulus developed by Beintema and Lappe (2002) which shows randomly chosen dots on the limbs without any rigidity contradicts the local rigidity approach. Moreover, impoverished recognition of upside-down displays (Sumi, 1984; Dittrich, 1993) and orientation specific recognition (Pavlova and Sokolov, 2000) argues against models that assume local rigidity to reconstruct a human figure.

Giese and Poggio (2003) have presented a neurophysiologically inspired model that exploits the local motion signals in a point-light display. In fact, their model used two distinct bottom-up processes, one for pure motion analysis, the other for pure form analysis. The first one is supposed to simulate biological motion processing in the dorsal path by integrating local motion signals to more and more complex optic flow patterns which are compared to stored templates. Giese and Poggio claim that this approach can account for many experimental data. Instead, their approach for modeling the form analyzing ventral path fails to do so. Here, they just connect nearest dots to lines without any prior knowledge of the form.

In the studies of Bobick and Davis (2001) motion is used to segregate a temporal pattern of a human body from the static background. By computing changes in pixel luminance, the moving part of an image is extracted from the image. This temporal pattern is matched to pre-stored templates to identify the action the body performs.

1.6.2 Form models

Most of the motion based approaches described above are difficult since they need to find correspondences between the single dots. A simpler way is to use pre-defined

models of a human form. Marr and Nishihara (1978) proposed such a form based
model for the recognition of three-dimensional shapes. They suggested static object-
centered representations of shapes by volumetric primitives. This idea has been used
in other models. Hogg (1983) and Rohr (1994) both used these cylindrical primitives
to reconstruct a human body in an image sequence. Marr and Vaina (1982) and Vaina
and Bennour (1985) built up their model on the idea of Marr and Nishihara (1978).
By tracking this shape over time, they provide early examples of the motion-from-
form idea. In their model Chen and Lee (1992) applied stick figures to recover the
three-dimensional configuration of a moving subject according to its two-dimensional
image.

Although the form-based approach has been popular in the construction of artificial
vision system, it did not have much influence on the investigation of biological motion
perception. Yet, it may be a viable route by which the visual system could analyze
and interpret biological motion. Only one model used an approach similar to that from
computer vision to explain psychophysical data. Lee and Wong (2004) used templates
of point-light walkers to investigate the perception of biological motion. As predicted
by other studies (Neri et al., 1998) they found a non-linear relationship between dots
in the stimulus and dots in the distracting noise. Although their model can only
qualitatively explain these data it shows that a template-matching model per se can
account for psychophysical data.

1.7   Objective of this work

Biological motion stimuli are complex stimuli containing motion and form cues and
involving rigid and non-rigid elements. Despite many degrees of freedom humans can
easily recognize the human form in a fraction of a second even if the visual information
is highly degraded. Psychophysical experiments so far emphasize different strategies
depending on the task. Sometimes the results even seem to be contradictory. Imaging
studies like fMRI have revealed that STS plays an important role in the perception
of biological motion. These methods have also shown that a network of other areas
is involved in this process. Yet, they failed to identify other areas involved as non-
ambiguous as STS. Moreover, the role of the determined areas is unclear. That is, it
is only known that this area is somehow involved but not to what extend and what
its role is in this network. Is it essential, redundant or just an artifact which is not

necessarily unique for biological motion? Another way to explore the contribution of certain aspects is by computational models. In common, these models rely on certain assumptions which restrict the common explanatory power. But, computational models can help to investigate the theoretical borders of certain hypothesis. So far, most models for the recognition of biological motion do not encounter mechanisms possibly used by the brain but are developed for computer vision. Only one model relies explicitly on assumptions that are compatible with known brain mechanisms (Giese and Poggio, 2003). They are able to explain some data by local motion analysis but fail to explain many other data by their approach. Thus, it is still not investigated theoretically to what amount the ventral path, that is object recognition, can contribute to the perception of biological motion.

The objective of this thesis is to explore the role of pure form information on the perception of biological motion by means of a computational model and psychophysical experiments. As outlined in section 1.4, there is evidence that top-down processes involving prior knowledge of a human form modulate the perception of biological motion. Therefore, I developed a template matching model which relies solely on form information and that neglects local motion signals. To what amount can this approach explain experimental data? In the first chapter, I apply a model which relies on extracting information from static frames. I compare the models results to previous experimental data and to data from own psychophysical experiments. The second chapter describes a model which relies on neurally plausible assumptions to analyze the form information. I will again compare these results to psychophysical data and in addition to data from an fMRI study. In the last chapter, I will investigate the perception of biological motion in noise. I will test the hypothesis that the perception in noise can also be accomplished by form analysis but involves an additional step of segregation. These studies are fulfilled by the dynamic model described in the second chapter and psychophysical studies.

I will discuss the results and predictions of the models in the face of the also described psychophysical experiments and in the context of other psychophysical studies. I will also discuss the model in the context of fMRI studies. And based on my model and on the predictions and implications following out of it I will present a general model how the perception of biological motion may be implemented in the brain which is compatible with my own results and other findings.

Chapter 2

General methods

My thesis consists of three chapters, which describe a model for the perception of biological motion, and its performance compared to existing and my own psychophysical experiments. Additionally, I compare model simulations to data obtained from experiments using fMRI. Here, I will describe the model in general, the general methods of the psychophysical and fMRI experiments and the stimulus used for both model simulations and experimental sessions. I will describe specific methods in the corresponding chapter.

2.1 The model

The aim of my work was to develop a model, which uses form information exclusively. Therefore, features of object recognition had to be taken into account. Object recognition in general is often thought to be based on learned two-dimensional views of objects rather than on a three-dimensional representation (Bülthoff and Edelman, 1992; Logothetis et al., 1994; Logothetis and Pauls, 1995; Riesenhuber and Poggio, 1999). This has especially been shown for biological motion: destroying the depth-component of a point-light walker has no influence on the perception (Bülthoff et al., 1998). Rather, it has been shown that knowledge of the form is important. If this is given by connecting the dots of a point-light walker to a stick figure rotation of the walker around its vertical axis is interpreted as changes of the two-dimensional structure. In contrast, if nearest dots are connected so as to give no structural information, the rotation is interpreted correctly (Sinha and Poggio, 1996).

I adopted this hypothesis and employed a template-matching model in which I

assumed a library of static two-dimensional postures of a walking human person. These templates were matched to incoming frames of the stimulus.

2.1.1 The model's templates

To obtain the templates used by the model I recorded the limb movements of nine real human persons (5 male, age 20-29). They walked normally on a wooden catwalk with sensors attached to the main parts of their bodies. A motion tracking system (MotionStar, Ascension) recorded the trajectories of these sensors. I will describe this system and the template extraction from these recordings in detail in this section.

Motion tracking system

The motion tracking system consists of two cubes which generate a magnetic field, 20 sensors which record the actual position and orientation in this magnetic field, a transmitter in a backpack which transmits the signals to a computer which records and potentially monitors the sensor recordings. The two cubes were positioned on a platform such that the middle of each cube was in 1.5 m height. The cubes had a distance of 1.2 m from each other, thus generating a virtually usable magnetic field of about 3x7x3 m^3 (Fig 2.1).

Two cubes using three dipoles independently generated the magnetic field in the three Cartesian directions. Up to 20 sensors detected the magnetic field strength with a frequency of 86 Hz. Cables connected the sensors to a transmitter, which the subjects could carry by in a backpack. Each sensor recorded 12 data: three Cartesian coordinates, which indicated the relative position to one of the cubes and a 3x3 matrix, which represented the spatial orientation of the sensor. The data were transmitted wirelessly to a computer so that the subjects could move freely. A computer eventually recorded the data from the sensors and saved them for subsequent analysis.

Since magnetic fields interact with metal or electricity the homogeny of the field can be easily disturbed. Thus, these disturbances tamper with the sensor recordings. Therefore, recordings nearby the wall had to be discarded. Moreover, disturbances by metal in the floor were eliminated by a wooden catwalk 60 cm above the floor. This reduced the area in which the recordings could take place to 1.5x7x2.4 m^3 (Fig 2.1).
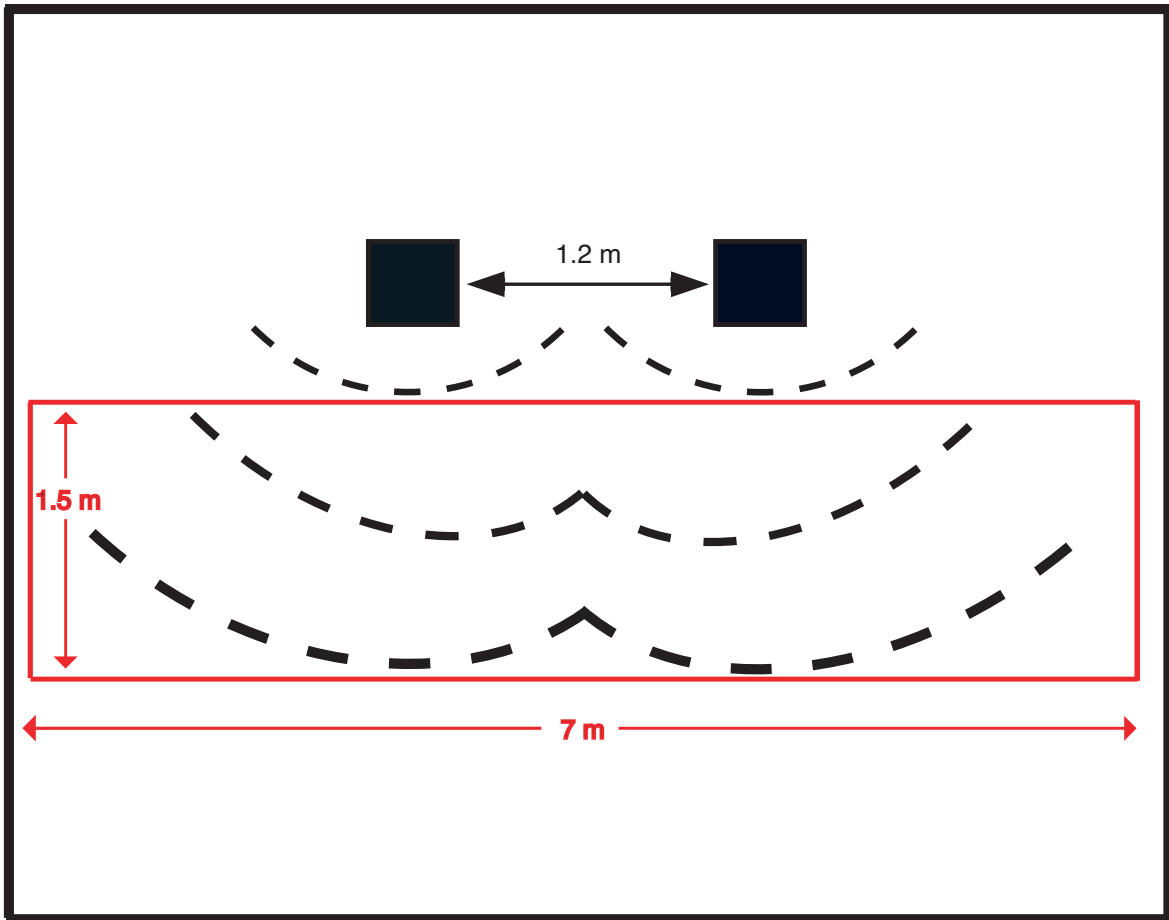
Fig. 2.1: Sketch of the MotionTracker environment. The black squares indicate the two cubes, which generate the magnetic field, the red lines the outlines of the catwalk. The dashed black lines indicate erroneous recordings obtained if straight lines parallel to the two cubes were expected. These systematic errors due to metal and electricity had to be corrected by an additional program.

Data evaluation and template construction

For the template construction I used only 14 of the available 20 sensors. They were attached to the subjects' head, shoulders, elbows, wrists, hips, knees, and ankles using hook and loop fastener (Fig. 2.2).

Subjects walked normally on the catwalk while the motion tracking system recorded the sensor data. For the evaluation, only the position coordinates were extracted and the orientation matrix was discarded. Due to the disturbances by the electricity supply in the walls only maximal three step cycles provided reliable data. Out of these, I extracted one step cycle for the template construction.

Data analysis revealed a non-negligible systematic disturbance of the recordings (Fig
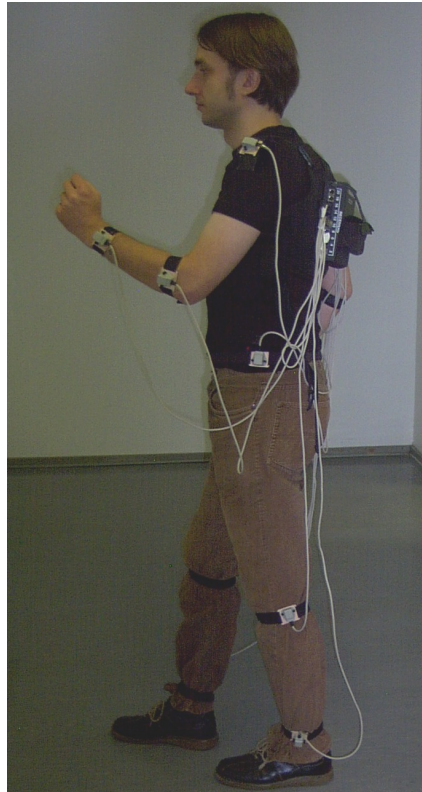
Fig. 2.2: The set-up for recording the human movements which were used to construct the model templates. Sensors, which recorded the magnetic field strength, were attached to the subject's major joints. The sensors were connected to a backpack that transmitted the recordings to a computer.

2.1). Therefore, I developed an additional program written in Mathematica (Version 4, WolframResearch) to correct for these deviations. Analogous to the stimulus I used in the experiments (see section 2.2), all translatory elements of a person's walking sequence were extracted. Thus, the person seems to walk on a treadmill.

Since the postures of the subjects at the beginning and at the end of the step cycle were actually not identical I superimposed single frames of the starting and the ending phase of a cycle with different weights. When the walking sequence is looped this provided the impression of a smooth continuous motion.

All walkers were normalized for height, and the walking sequence was subdivided into 100 temporally equidistant frames additionally providing normalization for speed. In general, the model used these single static frames as its internal templates of a walking human person. The exact mechanisms the model relied on and how it solved the tasks I will explain in the corresponding chapters.
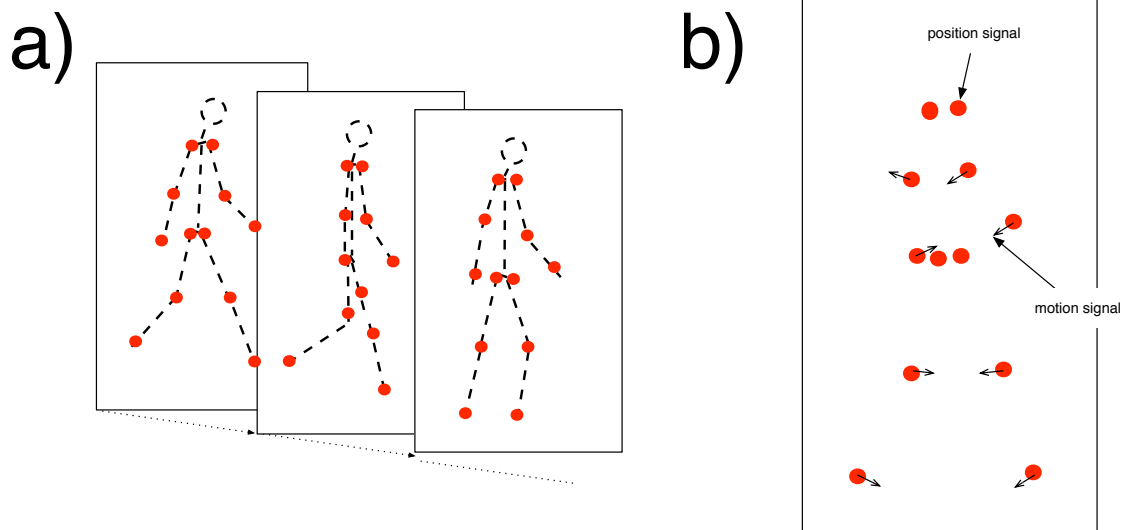
Fig. 2.3: Illustration of the 'classical walker' similar to the one used by Johansson (1973). a) Three frames of a walking cycle. The dashed lines are only for illustration. In the real stimulus only the red dots are presented. The dots are always located on the joints. b) The information available in this stimulus consists of local position signals of the dots and of local motion signals if the stimulus is set into motion. The small arrows indicate the motion signals provided by the single dots.

2.2   Stimulus

For the biological motion stimulus I used an algorithm adapted from J.E. Cutting that mimicked the movements of a human walker. Twelve point-lights appear on the major joints (i.e. shoulders, elbows, wrists, hips, knees, ankles) of the otherwise invisible body and produce smooth trajectories when the stimulus is in motion. The stimulus is projected on the two-dimensional monitor plain and all net translatory movement components were eliminated, giving the impression of a person walking on a treadmill. In this stimulus, local position signals of the single dots and local motion signals are coupled (Fig. 2.3).

Since I want to investigate the influence of form information in the absence of local motion signals, a decoupling of both signals is desired. Therefore, I manipulated the stimulus such that the single dots did not keep a constant position on the joints but were positioned randomly on the limbs and change their position every single frame by jumping to a new, randomly chosen position on the limbs (Beintema and Lappe, 2002) (Fig. 2.4 a). This way the single dots still provide local position signals, but
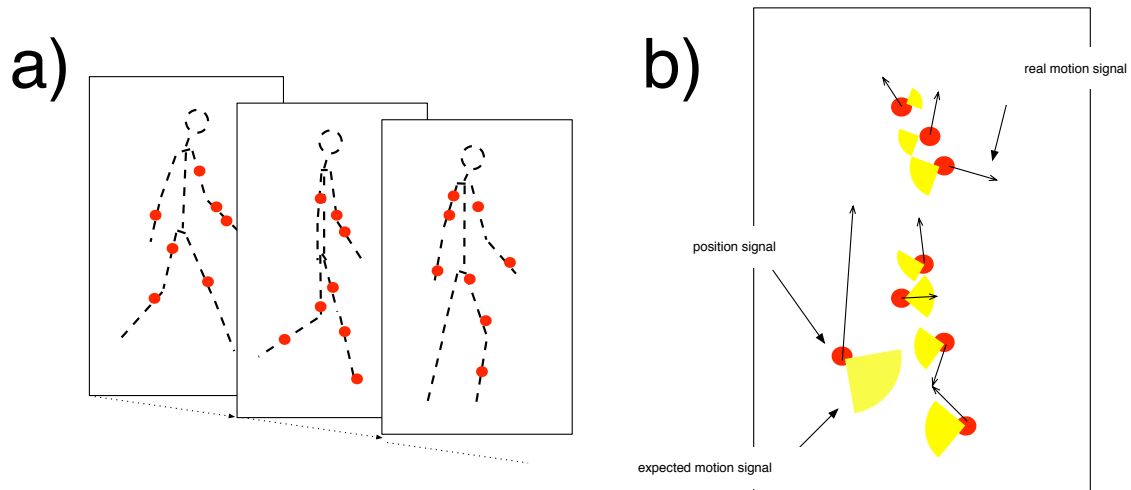
Fig. 2.4: Illustration of the stimulus I used in this study. a) Three frames of a walking cycle. The dashed lines are only for illustration. In the real stimulus only the red dots are presented. In each frame the dots have a new, randomly chosen position somewhere on the limbs. b) In this stimulus the dots still provide local position signals. However, the local motion signals (indicated by arrows), calculated for the transition from frame 2 to frame 3 in a), give erroneous information even if a cone of 10 degrees around the expected motion vector is assumed (indicated by yellow sectors of a circle).

the local motion trajectories between close-by dots in consecutive frames are scattered more or less randomly. Beintema and Lappe (2002) have shown that only 2% of all motion signals are usable because they are within a 10 degree cone around the expected trajectory if the dots would not change their position (Fig. 2.4 b). The stimulus was presented without a head because this part of the body is always in the same position and thus would give static and therefore unwanted position cues.

This stimulus is also suited to manipulate the amount of local motion signals. By increasing the number of frames, the amount of useful local motion signals increases. Thus, this stimulus is suited to control the amount of local position signals by varying the number of dots per frame and additionally to control the amount of local motion signals (Fig. 2.5).

I used this stimulus for the model simulations as well as for the psychophysical experiments.
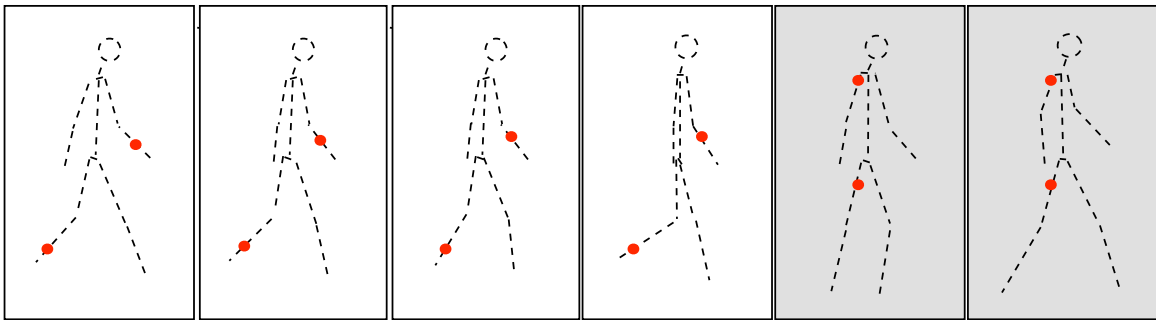
Fig. 2.5: Illustration of a stimulus with 2 dots per frame and a lifetime of 4 frames. This is illustrated by the 4 frames (white background) in which the dots keep their position on the limbs constant. This time they produce motion trajectories and, thus, local motion signals. After 4 frames they jump to a new position where they will stay for another 4 frames (grey background).

2.3 Tasks

To test the model I applied three different tasks and compared the model's results to the performance of human observers in the same tasks. Here, I will only describe the general tasks. Special tasks and the way the model deals with these tasks I will describe in the corresponding chapters.

2.3.1 Direction task

The stimulus depicted a walking person facing and walking to the right or its mirror image to the left (Fig. 2.6). Human observers and the model had to discriminate the walking direction.
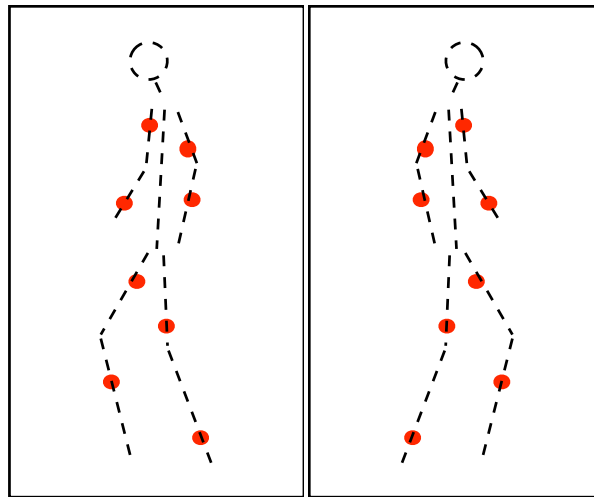
Fig. 2.6: Two single frames of the stimuli used in the direction task. The left frame depicts a walker facing and walking to the left, the right frame shows its mirror image, a walker to the right. The dashed lines are only for demonstration and not shown in the real stimulus.

2.3.2 Coherence task

Here, human observers and model had to decide whether the upper and lower part of the body were facing and moving in the same (coherent) or opposite (incoherent) direction (Fig. 2.7).

2.3.3 Forward/backward task

In this task, the model's and human observers' challenge was to decide whether the walker was moving in forward or backward direction. Both conditions consisted of
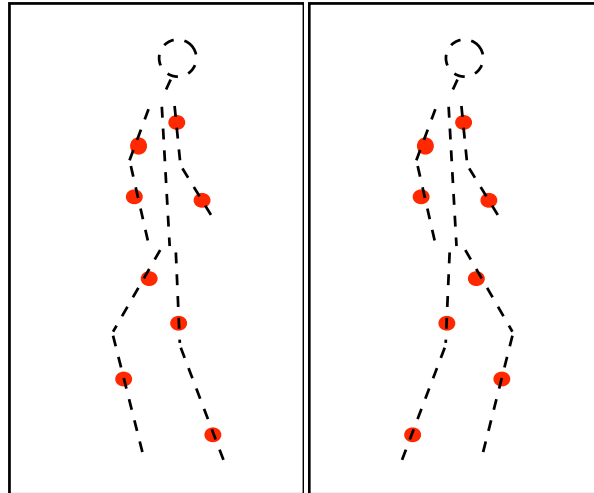
Fig. 2.7: Two single frames of the stimuli used in the coherence task. The left frame depicts an incoherent walker (upper and lower part of the body in opposite directions), the right frame shows a coherent walker (upper and lower part of the body in the same direction). The dashed lines are only for demonstration and not shown in the real stimulus.

identical frames. The sequence of frames was shown either in correct order, giving the impression of a walker moving forward, or in reversed order in which case the walker appeared to move backwards (Fig. 2.8).
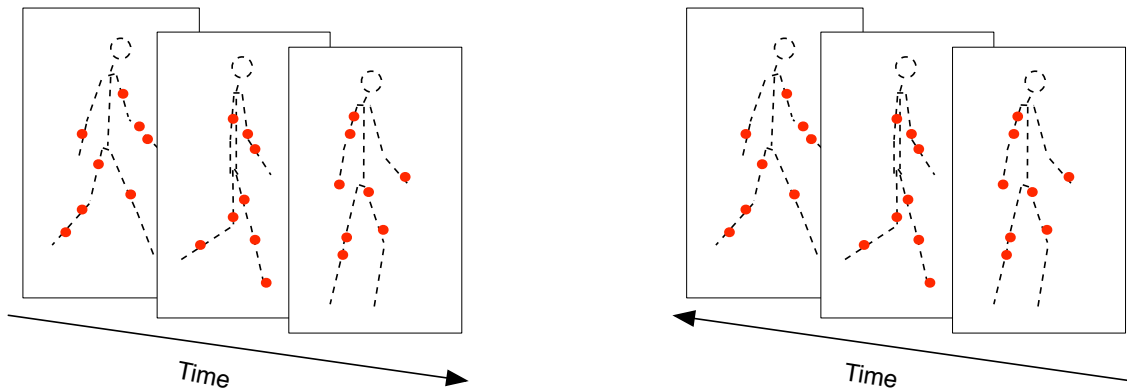
Fig. 2.8: The left side illustrates a walking sequence with the frames shown in correct order (forward movement). On the right side, the identical frames are shown in reversed order. This gives the impression of a backwards moving walker. The dashed lines are only for demonstration and not shown in the real stimulus.

2.4 Psychophysical experiments

To draw conclusions about the explanatory power of the model its results are compared to psychophysical data. If possible I use experiments already conducted, most of them by Beintema, Georg, and Lappe (Beintema and Lappe, 2002; Beintema et al., 2005; Georg, 2002). If the experimental data do not exist I will perform the psychophysical data myself. Here, I will describe the methods of my own experiments and, since they slightly differ, briefly the methods used by Beintema, Georg, and Lappe.

2.4.1 My own experiments

Stimuli were generated on *Power Mac G4* (*Apple Computers*) and presented on monitors (*Iiyama, Vision master 505*) with a resolution of 1280 x 1024 pixels and a display size of 30 cm x 40 cm. The monitor refresh rate was 100 Hz. If not indicated otherwise, a single stimulus frame was presented for 50 ms (5 monitor frames). A total stimulus consisted of a full step cycle, which lasted for 1.6 s and comprised 32 frames. The stimulus covered a field of 5° x 10° and consisted of white dots (5x5 pixels) on a black background. Trials were presented in random order and the stimulus position was in the middle of the monitor with a randomly chosen offset to avoid spatial cues from familiar positions. Starting phase was randomized throughout the trials.

In each experiment 4-6 subjects participated. They were members of the department and their age ranged from 24 to 35. All subjects had normal or corrected to normal vision and experience with biological motion stimuli but were naive to the goal of the study. They sat in a dark room and viewed the stimulus binocularly in approximately 60 cm distance from the monitor (head and eye position was not controlled explicitly). Subjects could move their eyes freely. After each stimulus, subjects indicated their answer by pressing a key on the keyboard in front of them without feedback. Then the next stimulus started.

2.4.2 Other experiments (Beintema, Georg, and Lappe)

The psychophysical experiments conducted by Beintema, Georg, and Lappe differed only in minor details from my own described above. I will shortly report the differences. The experiments were conducted on monitors with a refresh rate of 75 or 85 Hz. Stimulus frames were presented for 4 or 5 monitor frames, respectively, resulting in an

effective presentation duration of 52-55 ms. 3 to 10 subjects participated and viewed the stimulus from 45 or 60-70 cm distance from the monitor. For details see (Beintema and Lappe, 2002; Georg, 2002; Beintema et al., 2005).

2.5   fMRI experiments

In chapter 4 I will compare model predictions to fMRI data. Therefore, I will briefly describe the methods used in this experiment. For details see (Michels et al., 2005). 4 subjects participated in this study. They viewed the stimuli supine in a 1.5 T scanner. Stimuli were either the classical moving walker (CWm) from Cutting's algorithm (Cutting, 1978), a static frame of this stimulus (CWs), the stimulus mainly used in my studies and described above (SWm), or the same stimulus, but remaining in one posture for the whole duration, while the dots are changing their position on the body frame by frame (SWs). The control condition consisted of a set of static dot, which covered approximately the same area as the stimulus.

The study recorded the percent signal change of the stimuli compared to the control condition.

Chapter 3

Visual perception of biological motion by form: a template-matching analysis

3.1 Abstract

Biological motion perception is referred to as the ability to recognize a moving human figure from no more than a few moving point-lights. Such point-light stimuli contain limited form information about the shape of the body and local image motion signals from the moving points. The contributions of form and motion to the vivid perception of point-light displays are subject to controversy in the discussion. While some studies claim that local motion signals are critical, others emphasize the role of global form cues. Here, we present a template-matching approach to investigate the role of global form analysis. We used a template-matching method that ideally derives biological motion exclusively from form information. The algorithm used static postures monitored from walking humans as stored templates. We compared the simulation results to psychophysical experiments with the commonly used point-light walker and a variant point-light walker with near-absent local motion signals. The common result in all experiments was a high correlation between simulation results and psychophysical data. The results show that the limited form information in point-light stimuli might be sufficient to perceive biological motion. We suggest that it is possible for humans to extract the sparse form information in point-light walkers and to use it to perceive biological motion by integrating dynamic form information over time.

3.2   Introduction

Perceiving human movements is a complex task for the visual system since human movements contain many degrees of freedom and involve both rigid and non-rigid elements. Yet, nave human observers readily recognize moving human figures and their complex actions within fractions of a second. This is true even if the stimulus is degraded to only twelve point-lights attached to the joints on the body (Johansson, 1973). This striking phenomenon is referred to as perception of biological motion.

Biological motion contains different kinds of motion and form information (Fig. 3.1). Each light-point changes position over time and thus provides apparent motion signals. We call these the local or image motion signals. The instantaneous positions of all light points at any time provide structural information about the momentary posture of the body. Although this information is only weak in a single snapshot of a human body, temporal integration of the instantaneous position signals over a sequence of postures may provide increased structural information. We call this the global form information. Changes of the structural information of the body posture over time also provide motion information. In this paper, this is referred to as global motion information (Fig. 3.1).

The perceptual origin of global motion impressions is still an issue of discussion. Beintema and Lappe (2002) investigated whether normal observers can perceive biological motion in the absence of image motion. They developed a stimulus, which consisted of a fixed number of dots spread randomly over the skeleton of a human figure. The dots were reallocated to a new position every n-th frame. For n=1, the position was changed for each frame, thus, minimizing useful local image motion information in the stimulus. By varying n, the contribution of local image motion signals could be manipulated (see our section Methods/Stimulus for details). Spontaneous recognition of this new stimulus by nave observers was similar to that of the classical Johansson stimulus. In various discrimination experiments, Beintema and Lappe (2002, 2005) investigated more precisely the role of form information and image motion signals. They manipulated the amount of form information by changing the number of simultaneously visible dots. The results revealed a clear relationship between available form information and discrimination performance of the subjects. Adding local motion signals, on the other hand, did not improve the subjects performance and, in fact, their performance deteriorated marginally. Beintema and Lappe suggested that biological
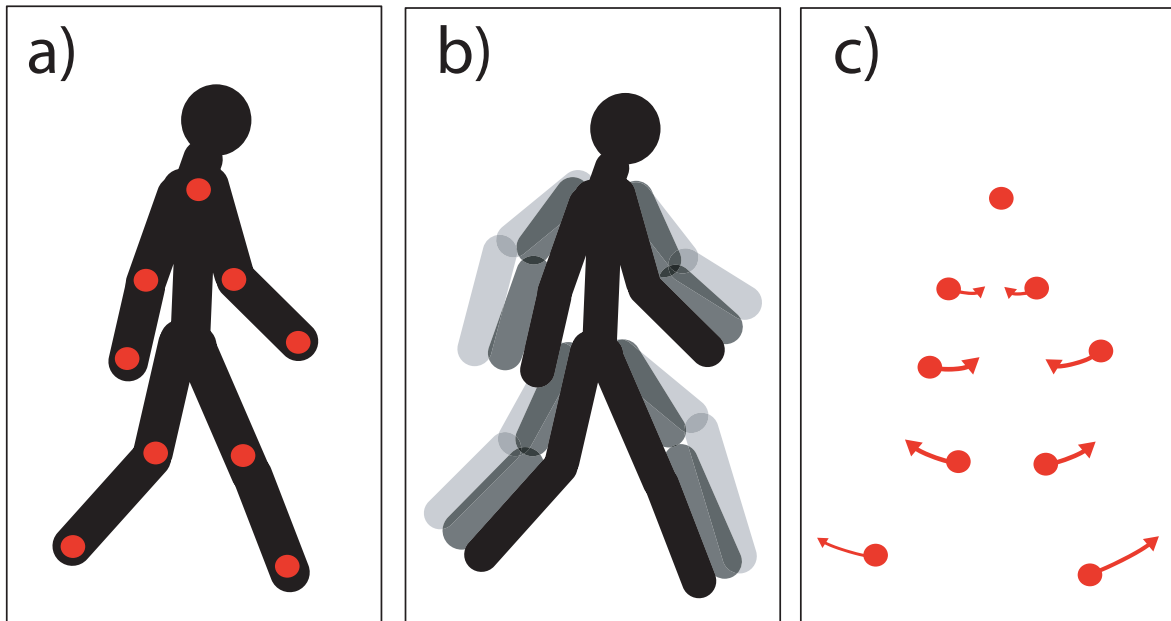
Fig. 3.1: a) The shape of a human figure contains global (illustrated by the black sketch) and local (illustrated by the red dots) features. b) The impression of a walking human person may occur from the integration of the global shape over time (differently shaded figures) or c) by integrating the local image motion signals (illustrated by arrows).

motion perception might be achieved by an analysis of the dynamic form of the human figure and that image motion signals have a supporting character in more complicated tasks and are not essential for perception.

The importance of form cues for biological motion perception has also been demonstrated in earlier studies. Chatterjee et al. (1996) studied the perception of apparent motion from sequential full-body images and found a higher level of usage for biomechanically consistent motion paths compared to impossible motion paths. This motion percept relates to the global motion of the body and overwrites local apparent motion signals when there is a conflict between the two. In another study Shiffrar et al. (1997) report an orientation-specific recognition of biological motion through apertures while other objects could not be identified in this manner. Both studies support a role of global form mechanisms for biological motion perception. Because they used line drawings or full body photographs, the question remains open whether global form analysis can also explain biological motion perception from point-light stimuli.

Bertenthal and Pinto (1994) investigated the importance of form for the recognition of point-light biological motion. Using masks comprising dots with trajectories

identical to those of the walker itself but with different, randomly chosen positions, they concluded that biological motion perception results from a global top-down form recognition process, rather than a bottom-up local motion analysis. This conclusion was challenged by Giese and Poggio (2003), who proposed that a hierarchical bottom-up process using only local motion signals combined with an attention process could account for the results. Neri et al. (1998) claimed that the perception of biological motion in the presence of noise is driven mainly by the integration of local motion signals.

Studies that emphasized the contribution of local motion signals often argue that the information from a single static picture of a point light walker does not allow a naive observer to perceive a walking human figure. Spontaneous biological motion perception occurs only in an animated sequence (Johansson, 1973). Therefore, most studies on biological motion perception have suggested or implicitly relied upon the assumption that the perception is processed by means of local image motion signals (Johansson, 1973; Cutting, 1981; Mather et al., 1992; Neri et al., 1998). However, while a single static frame is insufficient to recognize a walker, biological motion perception might also be derived from temporal integration of the sparse form information in each frame.

Computational studies have also emphasized the role of local motion signals. Giese and Poggio (2003) proposed a model, which analyzed form and motion cues separately. Their model accounts for a variety of experimental results purely by using the extracted local motion signals. In contrast, the form-analyzing pathway did not reveal selectivity for biological motion stimuli. Based on Giese and Poggios approach, Casile and Giese (2005) developed a model, which relied on the local motion signals in the stimulus. This model contained detectors of local motion signals that move in opposing direction. Casile and Giese computed the amount of opponent motion signals in the stimulus proposed by Beintema and Lappe (2002) and developed a new artificial stimulus with the same amount of opponent motion signals. From the approximate similarities between the two stimuli and the corresponding model simulations, Casile and Giese claimed that these opposing local motion signals might act as a critical feature in biological motion perception. This debate clearly reveals the controversy relating to which processes are necessary for perceiving biological motion as opposed to those, which are supplementary in nature.

While several studies investigated the contribution of local motion signals, in this

study, our objective was to investigate quantitatively the contribution of global form information. We present a simple model based on template-matching, which relies on form analysis only and completely ignores any image motion signals. We investigated how much form information is available from point-light walkers and whether this information could contribute to tasks that use point-light walkers as a stimulus. By comparing the performance of the model to both the psychophysical results described above and to the additional experiments reported below, we assessed quantitatively the contribution of form information. Among the many and often complicated characteristics of biological motion, we will focus on basic and often used low-level discrimination tasks. We chose these tasks on the one hand because they are simple and allow a straightforward quantitative comparison, and on the other hand because we believe that restricting the scope of the model is advisable for an early investigation. For a similar reason we concentrated on stimuli without masking noise. Beintema and Lappe (2002) have argued that biological motion recognition within noise may involve not only the perception of the biological motion stimulus per se, but also the segmentation of the figure from the background, which could be a different process. The relationship between our model and the masking studies will be considered in the Discussion.

3.3 Methods

3.3.1 Stimuli

Stimuli were computer-generated two-dimensional point-light walkers (Cutting, 1978). All translatory movement components were eliminated giving the impression of a person walking on a treadmill.

In the classical case, point-lights appear on the major joints of the body and produce smooth trajectories when the stimulus is in motion. We manipulated these stimuli such that the single dots did not keep a constant position on the body, but rather changed position each frame by jumping to a new, randomly selected position on the limbs (Beintema and Lappe, 2002). This way, we minimized local motion signals or selectively manipulated them by varying the lifetime of the dots (number of frames, respectively) before the jump takes place.

3.3.2 Tasks

Following the psychophysical studies to be simulated, we used three different tasks to compare the model to psychophysical data:

Direction task

In this task human observers and the model had to decide whether the walker was facing and moving to the right or to the left.

Coherence task

Here, model and human observers had to decide whether the upper and lower part of the body were facing and moving in the same (coherent) or opposite (incoherent) direction. This task was used previously by Mather et al. (1992). For the model this decision was similar to the direction task with the difference that the upper and lower halves of the stimulus were initially treated separately. After making a direction decision as described for the direction task for each half separately, the model decided whether both halves were walking in the same or in different direction.

Forward/backward task

In this task, the models and human observers' challenge was to decide whether the walker was moving in forward or backward direction (Beintema et al., 2005). Both conditions comprised identical frames. The sequence of frames was shown either in correct order, giving the impression of a walker moving forward, or in reversed order. In this case, the walker appeared to move backwards.

3.3.3 Templates

We attached sensors to the major joints of the bodies (i.e. shoulders, elbows, wrists, hips, knees, ankles) of nine human persons (five male) and recorded their walking movements on a catwalk with a body tracking system (MotionStar, Ascension). Since we used only two-dimensional stimuli we omitted the depth component of the movement patterns. We spatio-temporally averaged (Giese and Poggio, 2000) the individual walking patterns and connected the dots in a biological appropriate way resulting in a stick figure model of a mean walker. We subdivided one step-cycle of this walking

pattern into 100 temporally equidistant frames. The model used these frames as its templates of a common walking human person (Fig. 3.2).

3.3.4   Template-matching analysis

The template-matching analysis was achieved by a frame-by-frame template-matching algorithm, which evaluates the distances between the templates and the stimulus frames (Fig. 3.2).
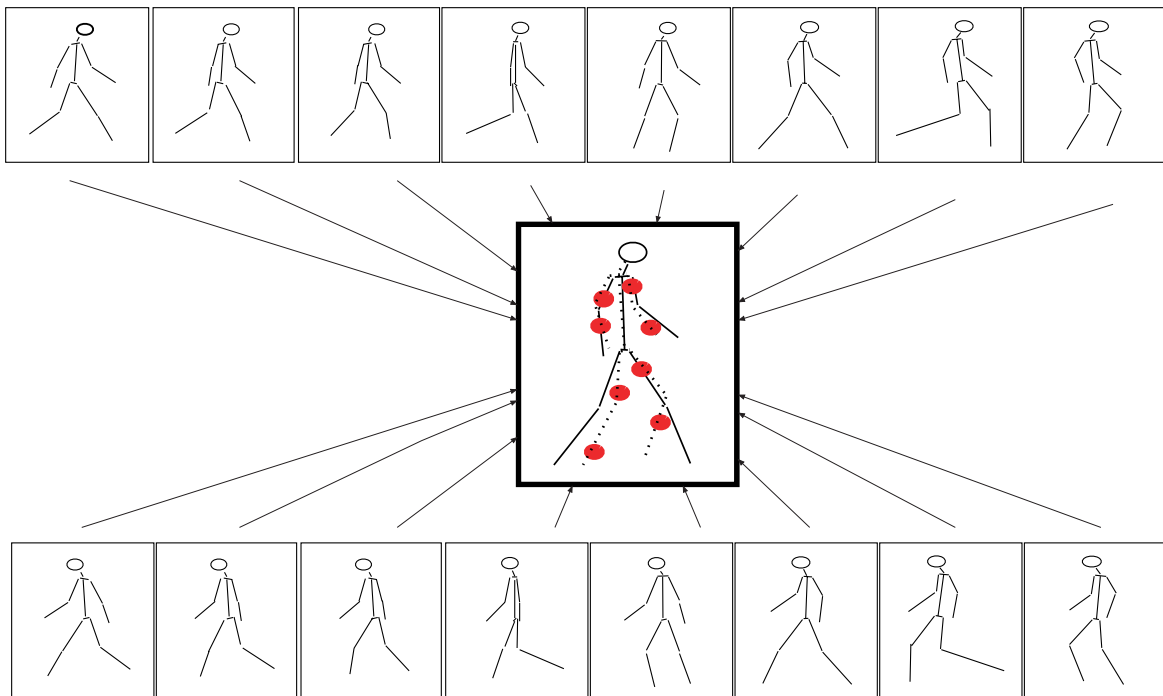
Fig. 3.2: One frame of the stimulus (filled circles, the dashed lines are only for clarification and are not shown in the real stimulus), which is matched to a set of templates of a walker moving and facing to the right and a walker moving and facing to the left (stick figure, solid lines). The match depends on distance measurements between stimulus dots and the template, indicated by lines.

In stage 1, the model uses a library of size-normalized static template frames with known coordinates x_p^T. The set of template points x_p^T comprises not only the joint positions of the template but all points on the limbs of the template walker. The input to stage 1 is the coordinates x_i^S of the stimulus dots of a given frame S. The model computes the distances d^{S,T_L} and d^{S,T_R} between a given stimulus frame S and each of the templates T_L (templates for walking to the left (L)) and T_R (templates for walking to the right (R)) by calculating the minimum Euclidian distance between each of the

stimulus dots x_i^S and all locations x_p^T on each template frame, and adding all single distances up. This procedure is done independently for each set of templates d^{S,T_L} and d^{S,T_R} without adding any internal or external noise:

$$d^{S,T_L} = \sum_{i=1}^{n} \min_p(|x_i^S - x_p^{T_L}|)$$

$$d^{S,T_L} = \sum_{i=1}^{n} \min_p(|x_i^S - x_p^{T_L}|)$$

(3.1)

with n: number of stimulus dots.

Stimulus dots were not restricted to a specific limb nor were the number of dots per limb restricted. After summing these minimum Euclidian distances of all dots in each frame, the frame with the shortest total distance was selected from the set of template frames. This choice was based on a winner-take-all principle. For a given stimulus frame S the best matching templates are determined by finding within each template set T_L and T_R the templates with the minimum distances d^{S,T_L} and d^{S,T_R}. This matching procedure is done independently within each template set T_L and T_R:

$$d_{min}^{S,T_L} = \min_{T_L}(d^{S,T_L}) = d^{S,T_L^{S,min}}$$

$$d_{min}^{S,T_R} = \min_{T_R}(d^{S,T_R}) = d^{S,T_R^{S,min}}$$

(3.2)

$T_L^{S_{min}}$ and $T_R^{S_{min}}$ determine the template frames within the template set for walking to the left (T_L) and for walking to the right (T_R) that match the stimulus frame S best. $d^{S,T_L^{S_{min}}}$ denotes the distance between the given stimulus frame S and the best matching template frame $T_L^{S_{min}}$ for walking to the left and $d^{S,T_R^{S,min}}$ denotes the distance between the given stimulus frame S and the best matching template frame $T_R^{S,min}$ for walking to the right. The models decision criterion at stage 1 (c_S^1) to discriminate the stimulus walking direction in a single stimulus frame S is based on the minimum distance measure $d_{min}^{S,T_{L,R}}$:

$$c_S^1 = 1 \quad for \quad d_{min}^{S,T_L} < d_{min}^{S,T_R} \quad and \quad c_S^1 = -1 \quad otherwise$$

(3.3)

For $c_S^1 = 1$ the model decides in favor of walking to the left, for $c_S^1 = -1$, the model decides in favor of walking to the right. Note, that the criterion is always well-defined because in the model is always $d_{min}^{S,T_L} \neq d_{min}^{S,T_R}$.

A trial consists of N stimulus frames. Each frame is evaluated independently by the above described computation. At the end of a trial the model computes an overall decision criterion at stage 1 c^1 by averaging all single decision criterions c_S^1:

$$c^1 = \frac{\sum_{S=1}^{N} c_S^1}{N} \tag{3.4}$$

For $c^1 > 0$ the model decides in favor of walking to the left, for $c^1 < 0$ it decides in favor of walking to the right. For the rare case of $c^1 = 0$, the model randomly decides in favor of left or right. This procedure is applied to each of the 100 trials of a simulation run and the proportion of correct decisions is expressed as percentage correct.

In the following stage 2, the model evaluates the temporal order of the best matching templates $T_L^{S_{min}}$ and $T_R^{S_{min}}$ for all stimulus frames S. The template frames are ordered depending on their temporal position in the entire walking sequence from 1 to t. For two consecutive stimulus frames S and $S+1$ the decision criterion in stage 2 $(c_{S,S+1}^2)$ is:

$$c_{S,S+1}^2 = 1 \quad for \quad T_{L,R}^{S,min} \geq T_{L,R}^{S+1,min} \tag{3.5}$$
$$c_{S,S+1}^2 = -1 \quad for \quad T_{L,R}^{S,min} \leq T_{L,R}^{S+1,min}$$

If the two consecutive frames are recognized by the model as temporally ascending or equal, $c_{S,S+1}^2 = 1$, if they are descending or equal, $c_{S,S+1}^2 = -1$. In case the selected template frames for the stimulus frames S and $S+1$ are from different template sets (e.g. S from leftward oriented walkers and $S+1$ from rightward oriented walkers) $c_{S,S+1}^2 = 0$. An entire trial consists of N stimulus frames. This leads to a time series TS with N-1 entries for the $c_{S,S+1}^2$. An overall decision criterion after one trial for a forward (c_f^2) and a backward (c_b^2) movement is achieved by applying two functions F_f and F_b on the series TS:

$$c_f^2 = F_f(TS) \tag{3.6}$$
$$c_b^2 = F_b(TS)$$

F_f finds chains of consecutive entries of 1 and determines the length of the longest chain; F_b finds chains of consecutive -1 and determines length of the longest chain of 1 values. The model decides in favor of a forward movement if

$$c_f^2 > c_b^2 \qquad (3.7)$$

and for a backward movement if

$$c_f^2 < c_b^2 \qquad (3.8)$$

For $c_f^2 = c_b^2$ the model randomly decides in favor of forward or backward movement. This procedure is applied to each of the 100 trials of a simulation run and the proportion of correct decisions is expressed as percentage correct.

During the simulations, all stimulus properties like trial duration, stimulus size, and stimulus position were identical to the conditions used in the psychophysical tasks. Starting phase of the walking cycle of the stimulus was randomized over trials.

In comparing psychophysical to computational data, we needed to account for the phenomenon of visible persistence (Coltheart, 1980) Visible persistence refers to the fact that light-points presented to an observer for a time period shorter than 100 ms are perceived for as long as 100 ms, while dots shown for longer periods are perceived for the time they are actually presented.

In psychophysical experiments subjects reported to see more points on the screen than were presented in any single frame. Quantitative analysis showed that in accordance with the literature reviewed by Coltheart, at 50 ms frame duration subjects perceived about twice as much dots than are really shown (Beintema et al., 2005). We, therefore, feel that visible persistence is part of the process of interpreting these stimuli and consequently needs to be implemented in the template-matching analysis. We adapted the model to this effect as simple as possible: to include the effect of visible persistence by overlapping the dots in a stimulus frame with the dots of the preceding frame if the presentation duration of the frame was less than 100ms. The model uses a view-based approach that treats size and position of the stimulus as constant. We believe that these assumptions, especially knowledge of height and position, are appropriate for a template-matching model when a discrimination stimulus is presented in isolation as in the experiments that we modeled. The model does not use any adjustable parameters which could be fitted to the psychophysical data. The model stages were chosen to be as simple and intuitive as possible. In the simulations, we compared the models decisions in each processing stage to the stimulus properties

and determined the percentage of correct decisions within 100 trials, each containing a full walking cycle of the stimulus. Note that the stimulus is a computer generated artificial walker whereas the templates were obtained from recordings of actual human walkers. Therefore, the stimulus will never exactly match any of the templates. Thus, the model is not expected to yield recognition rates of 100%. We believe this is an appropriate comparison to the psychophysical task in which this same computer generated walker was presented to human observers. If, as we predict, human observers use templates of body postures then it is likely that these templates are also learned from observing real human walking. We compared the models recognition rates to data from psychophysical experiments or psychophysical data obtained from other studies with the same tasks and stimuli as used in the mode simulations.

3.3.5   Experimental methods

In addition to already existing psychophysical studies (Beintema and Lappe, 2002; Beintema et al., 2005; Casile and Giese, 2005; Mather et al., 1992), we conducted further experiments and compared the data to simulation results. These experiments followed the procedures described in Beintema and Lappe (2002) and Beintema et al. (2005). The stimuli in all experiments were presented on monitors with a resolution of 1280x1024 pixels and a display size of 30 cm x 40 cm. The duration of a single frame was 52-55 ms (which means that a stimulus frame was presented for 4 or 5 monitor frames, depending on the display's refresh rate). The subjects were seated 60 cm in front of the monitor and viewed the stimulus binocularly. The stimulus covered a field of 5 x 10 and consisted of white dots (5x5 pixels) on a black background. Trials were presented in random order and the stimulus position had a randomly chosen offset. At the end of each trial subjects had to indicate their decision by pressing on of two buttons. Each experiment contained 30 trials. Data represent the average over 2-6 subjects depending on the task.

3.4   Results

We intended to study the importance of form information. We used the psychophysical data from the studies by Beintema and Lappe (2002) and Beintema et al. (2005) and simulated these tasks with our model. We presented one step-cycle of the stimulus

with a varying amount of form information by presenting different numbers of dots per frame (1, 2, 4 or 8). The frame duration and the lifetime of the dots were kept constant (52 ms = 1 frame). We first simulated a direction discrimination task in which the walker stimulus imitated a full step cycle comprising 40 single frames (see Fig. 3.3a for results). We secondly simulated a coherence task (results see Fig. 3.3b).
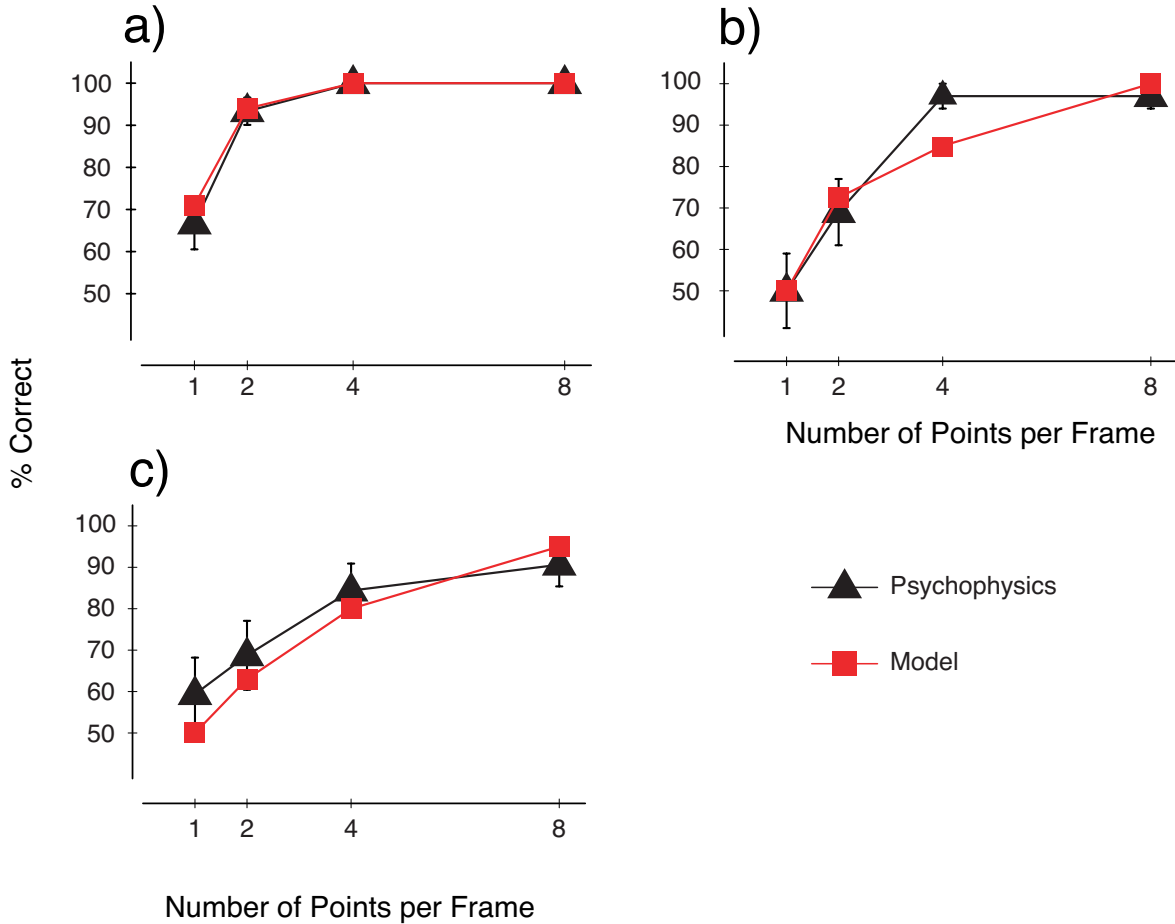
Fig. 3.3: Percentage of correct answers as a function of the number of points shown per frame for (a) the direction task, (b) the coherence task, and (c) the forward/backward task for model and human observers. Psychophysical data are adapted from Beintema and Lappe (2002) and Beintema et al. (2005) and are shown as Mean ± SE.

Statistical analysis revealed a significant influence of number of points per frame (ANOVA with repeated measures, $F(3,6) = 10.36, p < 0.01$) in the direction task. Logarithmic regression revealed that the psychophysical data in the direction task increased with increasing number of stimulus dots ($r^2 = .65, F(46) = 85.3, b_0 = .63, b_1 = .20, p < 0.01$). In order to quantify the model simulations, we compared the model data and the mean value of the psychophysical data (Tab. 3.1).

Table 3.1: The table shows the parameters for a quantitative comparison between model data and the psychophysical data for the different tasks (one-sample t-test, $\alpha = .05$). n.d. denotes t- and p-values that are not defined because the standard error of the psychophysical data is equal 0 due to ceiling effects. Psychophysical data are taken from Beintema and Lappe (2002) and Beintema et al. (2005).

Task	No. of points	Data Model [%]	Mean Psycho-physics [%]	SE [%]	Lower confidence limit [%]	Upper confidence limit [%]	df	t-value	p-value
Direction	1	71	66.7	6.1	54.9	79.1	2	0.71	> .25
	2	94	93.3	3.2	86.6	99.3	2	0.21	> .40
	4	100	100	0	100	100	2	n.d.	n.d.
	8	100	100	0	100	100	2	n.d.	n.d.
Coherence	1	50	50	9	32.2	67.8	1	0	1
	2	73	69	8	53.2	84.8	1	0.44	> .35
	4	85	97	3	91.1	100	1	4.0	> .05
	8	100	97	3	91.1	100	1	1.0	> .25
Forward/ Backward	1	50	59.4	8.8	42.0	76.8	1	1.06	> .20
	2	63	68.8	8.3	52.3	85.2	1	0.69	> .30
	4	80	84.4	6.5	71.5	97.3	1	0.67	> .30
	8	95	90.6	5.2	80.3	100	1	0.84	> .25

When we tested the influence of number of points in the coherence task, the psychophysical data of Beintema et al. (2005) revealed only a trend for an influence of number of points per frame (ANOVA with repeated measures, $F(3, 3) = 4.51, p = 0.13$). Since the qualitative behavior of the data is very similar to the data in the direction task (which showed a significant influence of number of points), we assume that the non-significant effect is probably due to the small number of subjects. Nevertheless, logarithmic regression revealed that the psychophysical data in the coherence task increased with increasing number of stimulus dots ($r^2 = .68, F(6) = 13.0, b_0 = .53, b_1 = .24, p < 0.05$).

In order to quantify the model simulations, we compared the model data and the mean value of the psychophysical data (Tab. 3.1). Some t- and p-values could not been calculated because of the ceiling effects in the direction task. All model data were located in the range defined by the lower and upper confidence limits around the mean of the psychophysical data. Thus, the t-values and p-values never reached statistical significance.

The model was able to solve both, the direction and coherence task solely on the basis of static form information. Therefore, both tasks rely on form analysis and do

not necessarily depend on the perception of the temporal pattern of walking. In the forward/backward task introduced by Beintema et al. (2005), recognition of the stimulus depends on temporal integration. Both stimuli, forward and backward walking, consisted of the same individual frames; the only difference was the order in which they were shown. In normal frame order the impression of a forward moving walker occurred, while in reversed order the impression of a backward moving walker occurred.

Similar to the coherence task, statistical analysis on the influence of number of points revealed only a trend (ANOVA with repeated measures,$F(3, 3) = 6.57, p = 0.08$), presumably due to the small number of subjects. Nevertheless, logarithmic regression revealed that the psychophysical data in the forward/backward task increased with increasing number of stimulus dots ($r^2 = .77, F(6) = 20.5, b_0 = .59, b_1 = .16, p < 0.01$). In order to quantify the model simulations, we compared the model data and the mean value of the psychophysical data (Tab. 3.1).

As shown in Fig. 3.3c and Tab. 3.1, the results of the model again simulated the human performance accurately. All model data were located in the range defined by the lower and upper confidence limits around the mean of the psychophysical data. Thus, the t-values and p-values never reached statistical significance.

The total amount of form information in one stimulus may be varied in two ways. Either for fixed stimulus duration the number of points shown per frame can be varied, or the number of points is kept constant and the duration of stimulus presentation is varied. As shown by Beintema et al. (2005), these two conditions can be used interchangeable. Next, we therefore intended to test how performance depends on presentation duration.

We conducted a new psychophysical experiment. We used a direction task. Six subjects participated. The lifetime of single dots (1 frame) and the frame duration (52 ms) were kept constant. Stimulus duration was varied from 100 ms to 1.6 s (2 to 40 frames, respectively) in pseudo-randomized manner. In blocked trials either 2, 4, or 8 stimulus dots per frame were presented. The subjects had to decide the direction of the walker after each trial and indicate their decision with a button press, whereupon the next trial started. The results are shown in Fig. 3.4.

Statistical analysis revealed a significant influence of trial duration (two-way ANOVA with repeated measures, $F(4, 20) = 49.9, p < 0.01$), number of points per frame ($F(2, 10) = 193.3, p < 0.01$) and the interaction of number of points and trial du-
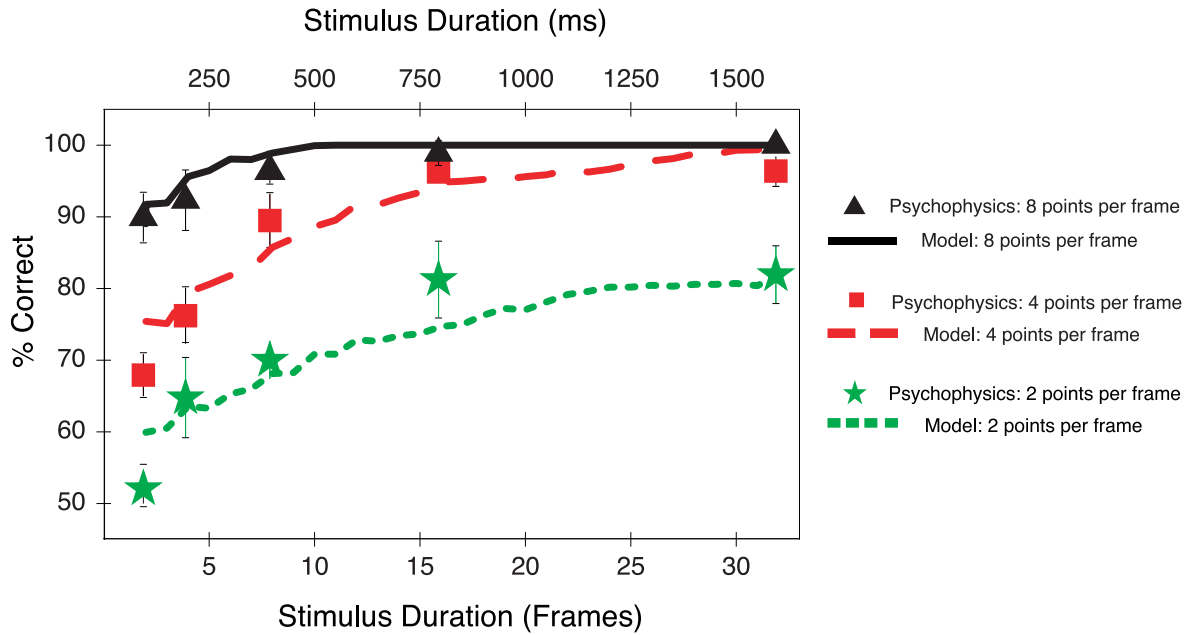
Fig. 3.4: Percentage of correct answers as function of the total stimulus duration for 2, 4, and 8 points per frame in the direction task. Psychophysical data are shown as Mean $\pm$ SE.

ration ($F(8, 40) = 5.0, p < 0.01$). Furthermore, logarithmic regression revealed that the psychophysical data in the direction task increased with increasing number of stimulus dots (two points per frame: $r^2 = .59, F(34) = 49.4, b_0 = .52, b_1 = .09, p < 0.01$, four points per frame: $r^2 = .77, F(34) = 112.0, b_0 = .57, b_1 = .13, p < 0.01$, eight points per frame: $r^2 = .57, F(34) = 45.9, b_0 = .76, b_1 = .08, p < 0.01$). In order to quantify the model simulations, we compared the model data and the mean value of the psychophysical data (Tab. 3.2).

At each point in time the human observers revealed qualitatively and quantitatively the same behavior as the template-matching analysis. The analogy between both curves reveals that the total amount of available form information (accumulating with increasing stimulus duration) influences the percentage of correct answers predominantly.

3.4.1 Local motion signals

We next looked at the influence of local motion signals on the perception of biological motion. To compare model simulations to psychophysical data, we adapted the data of Beintema and Lappe (2002) and Beintema et al. (2005). Beintema and Lappe (2002) added local motion signals to the stimulus by varying the duration (number of frames)

Table 3.2: The table shows the parameters for a quantitative comparison between model data and the psychophysical data for the direction task depending on the number of points per frame and trial duration (one-sample t-test, $\alpha = .05$).

No. of points	Trial duration [ms]	Data Model [%]	Mean Psycho- physics [%]	SE [%]	Lower confidence limit [%]	Upper confidence limit [%]	df	t-value	p-value
2	100	59	52.1	2.1	46.3	57.3	5	1.53	$> .05$
	200	63	64.8	5.2	51.4	78.2	5	0.09	$> .45$
	400	68	70.0	2.4	63.8	76.2	5	0.31	$> .40$
	800	74	81.3	4.3	70.3	92.4	5	0.61	$> .25$
	1600	82	81.8	3.4	73.1	90.6	5	0.82	$> .20$
4	100	75	67.9	2.6	61.1	74.6	5	1.12	$> .15$
	200	79	76.0	3.4	67.1	84.9	5	0.39	$> .35$
	400	86	89.5	3.2	81.3	97.7	5	0.86	$> .20$
	800	95	96.3	1.5	92.5	100.0	5	0.39	$> .35$
	1600	100	96.5	1.8	91.8	100.0	5	0.80	$> .20$
8	100	92	90.0	2.9	82.5	97.5	5	0.26	$> .45$
	200	96	92.2	3.5	83.3	100.0	5	0.39	$> .35$
	400	99	96.5	1.8	91.9	100.0	5	0.59	$> .25$
	800	100	99.0	0.6	97.4	100.0	5	0.64	$> .25$
	1600	100	99.8	0.2	99.4	100.0	5	1.6	$> .05$

for which a dot kept its position on the limb before being extinguished and relocated to a different position. If a dot remained at a specific limb position for several frames, its spatio-temporal profile allows estimating local image motion. Since the model, however, relies on form information only, it does not evaluate this local motion signal. If human observers do take advantage of local motion signals the answers of model and humans should differ, therefore. The difference should become more obvious with prolonged lifetime of the dots.

The simulations were conducted with the direction task to examine whether local motion signals would improve performance in general, and with the forward/backward-task in which local motion signals should provide the most useful additional information (see Fig. 3.5a-d). In addition to varying the lifetime of the dots we also varied the number of dots per frame (Fig. 3.5e). Human data in the direction task were taken from Beintema and Lappe (2002). Data in the forward/backward task were taken from Beintema et al. (2005). Statistical analysis of the data revealed an effect of number of points (two-way ANOVA with repeated measures, $F(3,6) = 102.5, p < 0.01$) and on lifetime ($F(3,6) = 5.0, p < 0.05$).

The model matched the performance of human subjects even though it uses form

Table 3.3: The table shows the parameters for a quantitative comparison between model data and the psychophysical data for the direction task depending on the number of points per frame and lifetime of the dots (one-sample t-test, $\alpha = .05$). n.d. denotes t- and p-values that are not defined because the standard error of the psychophysical data is equal 0 due to ceiling effects. * indicates statistically significant differences between model and psychophysical data. Psychophysical data are taken from Beintema and Lappe (2002).

Task/ No. of points	Life-time [ms]	Data Model [%]	Mean Psycho-physics [%]	SE [%]	Lower confidence limit [%]	Upper confidence limit [%]	df	t-value	p-value
Direction 1	50	72	66.7	9.3	26.7	100.0	2	0.67	> .25
	100	58	58.3	10.9	11.3	100.0	2	0.20	> .40
	200	50	65.0	2.8	52.3	77.4	2	5.0	< .05*
	400	56	46.7	6.0	20.8	72.5	2	1.67	> .10
Direction 2	50	92	93.3	4.4	74.3	100.0	2	0.25	> .40
	100	74	88.3	1.7	81.2	95.5	2	6.5	< .05*
	200	76	63.3	6.0	37.5	89.2	2	2.17	> .05
	400	68	73.3	6.7	44.6	100.0	2	0.71	> .25
Direction 4	50	100	100	0.0	100.0	100.0	2	n.d.	n.d.
	100	85	100	0.0	100.0	100.0	2	n.d.	n.d.
	200	84	90	5.7	65.2	100.0	2	1.0	> .20
	400	76	91.7	6.0	65.8	100.0	2	0.76	> .20
Direction 8	50	100	100.0	0.0	100.0	100.0	2	n.d.	n.d.
	100	98	100	0.0	100.0	100.0	2	n.d.	n.d.
	200	94	98.3	1.7	91.2	100.0	2	2	> .20
	400	96	100	0.0	100.0	100.0	2	n.d.	n.d.
Forw/ Back 8	50	98	97.5	2.0	93.9	100.0	2	0.25	> .40
	100	94	100	0.0	100.0	100.0	2	n.d.	n.d.
	200	94	96.3	2.0	92.7	100.0	2	1.13	> .15
	400	96	96.3	2.0	92.7	100.0	2	0.13	> .45

information only. Only two model data points are statistically different from the corresponding psychophysical data. The data of Beintema and Lappe show that the subjects' performance does not increase but remains constant or even decreases with prolonged lifetime. The model is able to reproduce the qualitative and quantitative behavior of the psychophysical data (Tab. 3.3).

Prolonging lifetime results in an increase of local motion signals. Therefore, an increase of correct answers with prolonging lifetime should be expected if the perception relies on local motion signals. There was no increase of humans' performance, however. Rather, the data revealed a slight decrease as lifetime increased. This result would be expected if form information were dominant over local motion signals. With longer

lifetime each dot remains on a fixed position on the body for a longer time. Thus, the sampling rate of the form of the body is reduced. Also, the number of dots effectively perceived due to visible persistence decreases when lifetime is prolonged. In the model this is equivalent to decreasing the number of effectively used dots. The model does not use any local motion signals but fits the human's performance well.
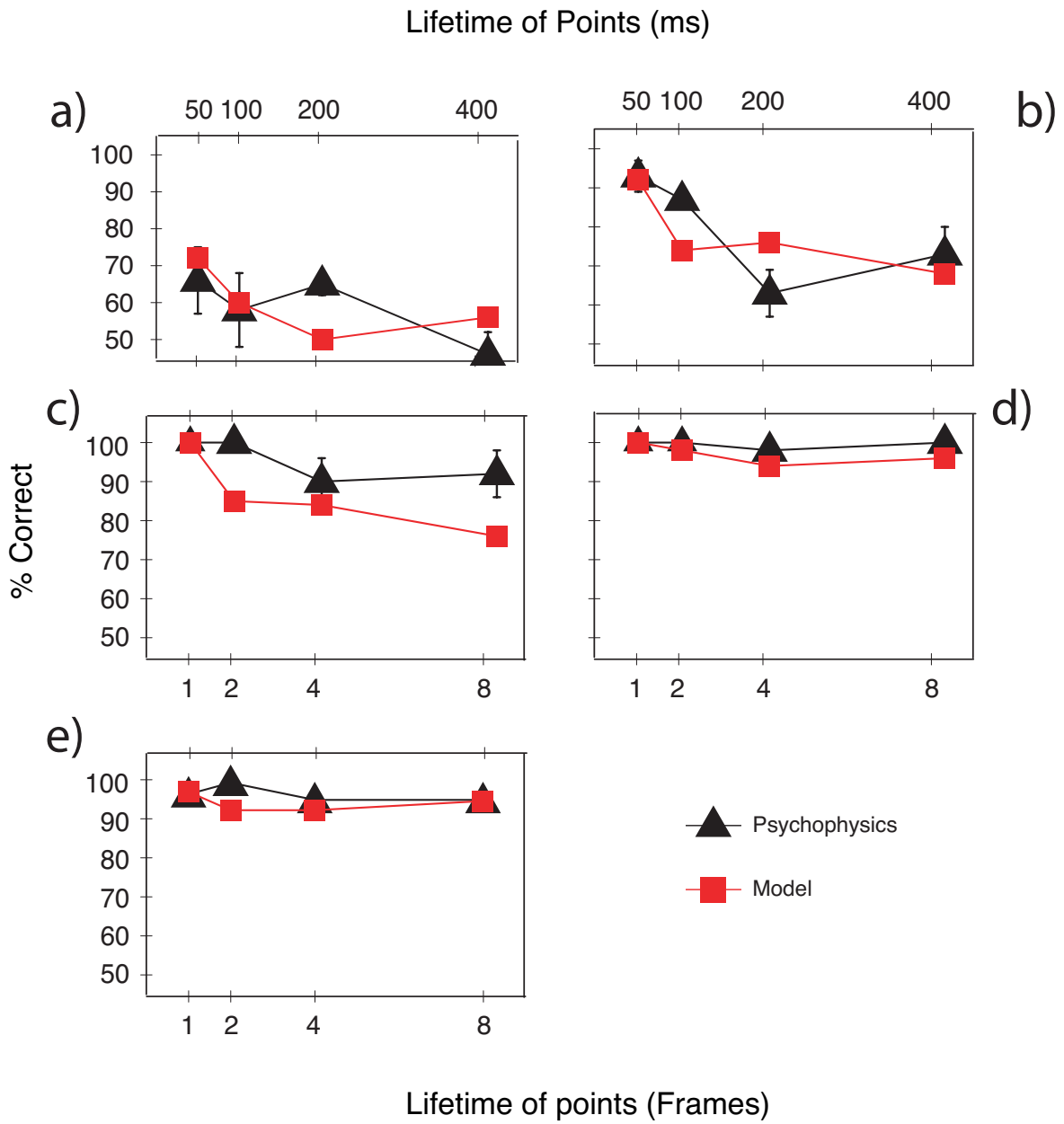
Fig. 3.5: Percentage of correct answers as a function of the lifetime of single points in the direction task for (a) 1 point per frame, (b) 2, (c) 4, and (d) 8 points per frame and in the (e) forward/backward task for 8 points per frame. Psychophysical data (Figure a-d were adapted from Beintema and Lappe, 2002; Figure e was adapted from Beintema et al., 2005) are shown as Mean ± SE.

3.4.2 Other walkers

We tested discrimination tasks using the classical point-light walker introduced by Johansson (1973). Fig. 3.6 shows the results of the simulation of an experiment by Mather et al. (1992). They presented the classical walker in a direction discrimination task in conditions in which specific dots were omitted from the walker. In the first condition all dots were shown while in the other conditions four dots were removed, either shoulder and hips, or elbow and knees, or wrists and ankles. Particularly the omission of wrists and ankles had a deteriorating effect on perception.
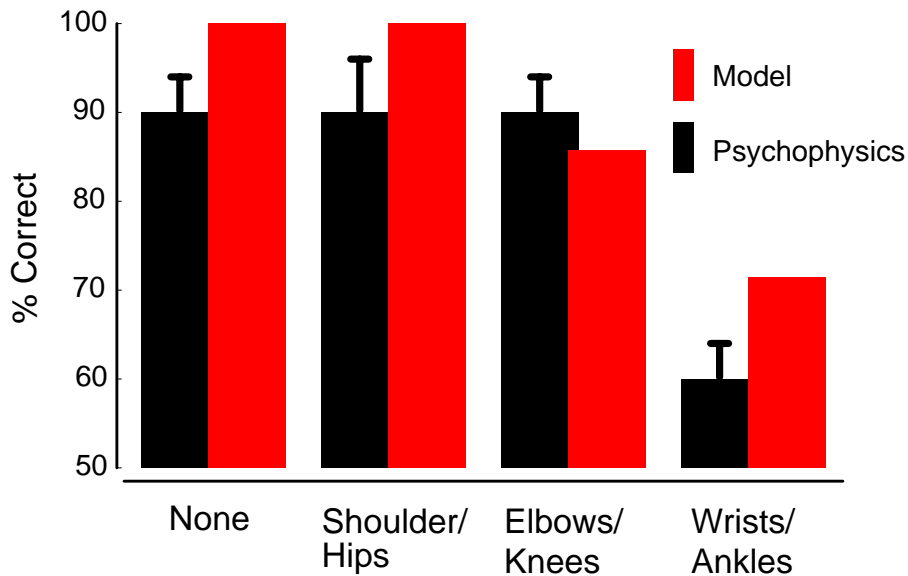
Fig. 3.6: Percentage of correct answers as a function of the dots omitted. Psychophysical data (adapted from Mather et al., 1992) are shown as Mean ± SE.

The models results revealed the same dependence on visible dots as the psychophysical data of Mather et al. Leaving out the ankles and wrist impaired the perception, elbows and knees showed little effect while omitting the shoulder and hips had no influence at all. Statistical analysis revealed that all model data are located within the confidence limits of the corresponding psychophysical data (Tab. 3.4).

Mather et al. concluded that the feet and hands are most important because they follow the longest trajectories, hence providing most motion information. Our simulations revealed similar results and confirm that distal dots are most important. Since the model uses only form information, we conclude that the reason why wrists and

Table 3.4: The table shows the parameters for a quantitative comparison between model data and the psychophysical data for the direction task depending on the number of points per frame and trial duration (one-sample t-test, $\alpha = .05$). Psychophysical data are adapted from Mather et al. (1992)

Dots omitted	Data Model	Mean Psycho-physics	SE	Lower confidence limit	Upper confidence limit	df	t-value	p-value
None	100	90	5	80	100	5	2.0	> .05
Shoulder/Hips	100	90	6	80	100	5	1.6	> .05
Elbows/Knees	85	90	5	75	95	5	1.0	> .20
Wrists/ankles	71	60	5	50	70	5	2.2	> .05

ankles are more important than other proximal dots, is because they offer the most reliable spatial information about the posture of the walker.

Casile and Giese (2005) argued against the idea that the form information in the point-light walkers with strongly degraded local motion information is sufficient to explain psychophysical data by Beintema and Lappe (2002). Casile and Giese proposed a model that used only opponent local motion features to model the psychophysical data of Beintema and Lappe (2002) and for a newly developed stimulus.

A new stimulus (critical feature stimulus, CFS) proposed by Casile and Giese was similar to the stimulus used by Beintema and Lappe in terms of local motion signals but it was also degraded in terms of positional information. The stimulus consisted of four regions. In two of these regions, roughly corresponding to the position of hands and feet, dots move with a sinusoidal horizontal component and a random vertical component. The other two regions contain dots that move completely random. The spatial arrangement of the four regions was derived from the spatial arrangement of a human person walking to the right or to the left. Casile and Giese's psychophysical data with the CFS-stimulus are similar to the psychophysical results by Beintema and Lappe (2002). Beintema and Lappe reported a counterintuitive decline of recognition rates for prolonged lifetimes of the stimulus dots (Fig. 3.5). Although Casile and Giese were unable to replicate this decline of recognition rates, they claimed that the remaining sparse local image motion information acted as a critical feature for the recognition.

We used this CFS-Walker to test the available form information in the CFS-Walker
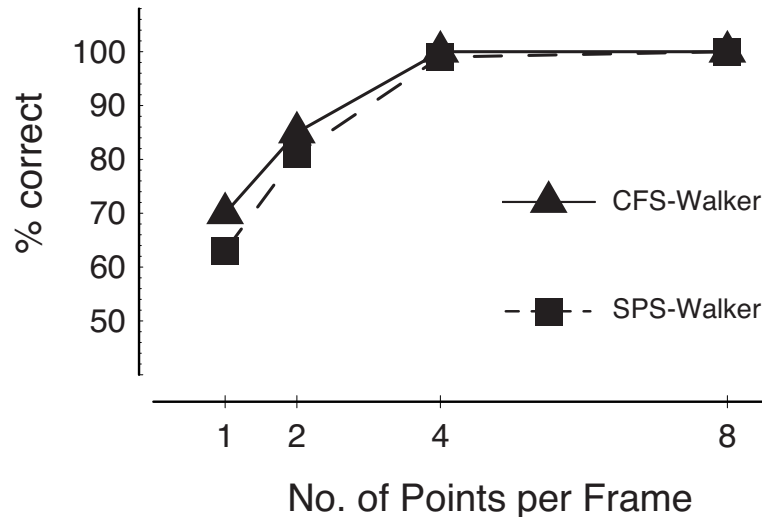
Fig. 3.7: Comparison of the models performance in a direction task for the stimulus used in this study (Sequential Position Stimulus (SPS)) and the stimulus proposed by Casile and Giese (2005) (CFS-Walker)

stimulus with our template-matching analysis. Recognition rates for the CFS-Walker in our template-matching model were similar to recognition rates of the SPS-Walker. Both stimuli revealed the same dependency on points per frame. These simulations showed that the positional information in the CFS walker is still comparable to the information in the SPS walker and that this information is sufficient to explain the psychophysical data (Fig. 3.7).

As described above, the walker consisted of four regions, each with a specific spatial offset from the vertical axes. If the there would be no spatial displacement of the four regions at all, a stimulus facing to the left would be identical to a stimulus facing to the right. Thus, the discrimination task would be unsolvable and the spatial displacements are essential to observe results different from chance level. Therefore, it is these spatial displacements that act as the critical feature rather than the opposing motion signals.

Although the CFS-Walker matched the form of a human body only very roughly, the models performance was similar to the SPS walker. At a first glance this result seems surprising because we compare a walking stimulus to an artificial stimulus with coarse position information approximately matching the original stimulus. But in the SPS-Walker which is mainly used in the present study, the stimulus dots also almost never exactly match the template. Actually, the SPS-Walker also represents an artificial

computer stimulus. Therefore, the model identified the SPS as well as the CFS walker as a noisy walking stimulus. The simulation results obtained with both the CFS walker and the psychophysical results by Casile and Giese indicate that the visual system is very robust against noise.

3.5   Discussion

In this paper, we addressed the question of possible mechanisms underlying the perception of biological motion. Global biological motion, i.e. the motion of the human figure, may be derived from local image motion analysis of the light points or from structural information from the changing shape of the body.

In our study, we investigated quantitatively the contribution of global form to discrimination tasks with point-light walkers. By assuming a library of static postures of a walking human person, we developed a model based on template-matching, which uses only sequential posture information rather than local motion signals. In this way, we could quantify the amount of global form signals available in the depicted point-light displays in the absence of local image motion signals. Global motion information will be derived from the change of these postures over time rather than from local motion signals.

We presented three different experiments and compared the results of the template-matching analysis to the psychophysical data. We varied the number of visible points per frame, the stimulus duration, and the amount of local motion signals. The comparison revealed a strong dependence on available form information in the template-matching model similar to data from previous psychophysical studies (Beintema and Lappe, 2002; Beintema et al., 2005). These similarities were consistent when the total form information was not varied within one single frame but instead in the overall information mediated by stimulus duration.

Beintema and Lappe also reported a counterintuitive decline of recognition rates for prolonged lifetime of the stimulus dots. Casile and Giese (2005) proposed a model that was supposed to explain these psychophysical data by exploiting local motion features. However, the model predicted a slight increase of recognition rates for a prolonged lifetime of stimulus dots. In contrast to Casile and Giese, our simulation results based on global form showed a decline for a longer lifetime of stimulus dots similar to the psychophysical data. In accordance with Beintema et al. (2005), we

suggest that this decline is due to a decreasing form sampling in combination with visible persistence, rather than the addition of local motion signals. For a detailed analysis of the relationship of recognition rates and visible persistence see Beintema et al. (2005).

In another experiment, Casile and Giese introduced a new artificial stimulus, called critical feature stimulus (CFS). They intended to show that this stimulus does not contain any information but opponent local motion features. They claimed these opposing motion vectors to be critical features that are essential to recognize biological motion. Our simulations showed that the CFS walker still contained coarse global form information. This information is strongly degraded but still sufficient to solve the applied task. We suggest that the critical feature in this stimulus is global form rather than local motion signals.

Several studies examined the perception of biological motion in masking experiments, i.e. the stimulus is shown in a number of distracting noise dots. Beintema and Lappe (2002) have argued that these tasks comprise both, the perception of the biological motion stimulus per se but also the segmentation of the stimulus from the background, which could be a separate process. Within the current analysis, we did not simulate noise experiments since we wanted to keep the model simple and confined to the biological motion task without an additional segmentation stage. Such a separate segmentation stage, however, might not be needed. Lee and Wong (2004) recently presented a template model for the recognition of biological motion that is similar in spirit to our approach but uses point-light templates rather than stick-figure templates. They showed that a form-based template-matching approach could also account for perception of biological motion in noise with results similar to psychophysical data (Neri et al., 1998). In addition, since their model did not included segmentation by local image motion, it should also work for more complicated noise patterns. Although more work would be needed to confirm this hypothesis for our approach, we suggest that our template-matching would be able to reveal similar results.

Our computational approach suggests that perception of biological motion is possible from matching the sparse stimulus frames to (dynamic) form-templates and integrating this information over time. The concept that learned global prototypes underlie the interpretation of perceived body structures in a top-down process relates back to initial ideas by Marr and Nishihara (1978). The general idea is supported by other psychophysical experiments. Sinha and Poggio (1996) connected the dots on the ma-

jor joints of a human body so that it showed the line drawing of a person. If this rigid structure was rotated about its vertical axis, it was seen as a walking figure. However, if the wrong joints were connected (so that the figure did not represent a human body), the rotation was correctly interpreted. Sinha and Poggio argued that the visual system interprets the human figure in terms of how the human structure is expected to change. Our study supports this idea of a form-based top-down process mediating the perception of biological motion. The simulations reveal that Sinha and Poggios idea can be extended from static line drawings to moving point-light figures. Even if a single frame did not provide enough information to recognize the human figure, the succession of point-light images was sufficient. The simulations do not exclude that human observers can use available local image motion signals if they are useful.

Our approach derives global motion information from an analysis of the changing shape of the figure rather than from local motion detectors. As such, it bears some relationship to feature-based motion systems as suggested by Cavanagh (1992) and Lu and Sperling (1995). From a computational viewpoint, the advantage of such feature-based motion systems over lower-level, energy based motion systems is particularly high for biological motion recognition because in contrast to rigid object motion, biological motion relies inevitably on form information due to the large number of degrees of freedom in the non-rigid motion of the body.

Perception of biological motion is not just a single phenomenon. The perception of biological motion comprises of a rich palette of different aspects such as action recognition (Johansson, 1973; Dittrich, 1993; Pollick et al., 2001), gender discrimination (Troje, 2002; Pollick et al., 2002; Troje et al., 2005), and identification of identity (Cutting and Kozlowski, 1977; Loula et al., 2005). It has been shown that humans can use different information to judge movements depending on the task (Pollick et al., 2001; Troje, 2002) and that the influence of bottom-up and top-down processing, and attention differs among tasks (Thornton et al., 1998; Thornton et al., 2002; Thornton and Vuong, 2004). In the present study we have focused for simplicity on straightforward discrimination tasks. These simple tasks can be solved by a global form analysis in the absence of local motion signals. One may now ask: which of the more complex aspects of biological motion perception are local and which ones are global? Which ones require motion per se and which ones are based on structural cues? In principle, a template model such as ours may be sufficient to also discriminate action, gender, or identity provided that the appropriate templates are available. The model arrives

at a description of the temporal structure of the body posture change over time and thus may also discriminate actions and use dynamic cues (Troje, 2002), even if they are not derived from local motion analysis. Whether this is truly sufficient would have to be investigated in further studies, however, it is also likely that among the many aspects of biological motion there are some that benefit from additional motion signals. However, for the task we studied here these local motions signals do not form a critical feature for biological motion.

3.6   Conclusion

In this chapter, we investigated the perception of biological motion by form analysis. We were able to show that the results of the model compared to psychophysical data showed a high correlation. We conclude that the approach of template-matching model is a suitable method to analyze the form information of a point-light walker and, therefore, to investigate the perception of biological motion by form analysis. The model in these experiments is based on approaches from computer vision. Thus, this approach can only hypothetically explain how the perception of biological motion is implemented in the primate brain. Moreover, this approach cannot explain the influence of frame duration and the dynamics in the stimulus. While this has a strong influence on humans' perception (Parish et al., 1990; Beintema et al., 2005) the model cannot deal with this. These two important questions we will study in the next chapter.

Chapter 4

A model of biological motion perception from configural form cues

4.1 Abstract

Biological motion perception is the compelling ability of the visual system to perceive complex human movements effortlessly and within a fraction of a second. Recent neuroimaging and neurophysiological studies have revealed that the visual perception of biological motion activates a widespread network of brain areas. The superior temporal sulcus (STS) has a crucial role within this network. The role of other areas are less clear.

We present a computational model based on neurally plausible assumptions to elucidate the contributions of motion and form signals to biological motion perception and the computations in the underlying brain network. The model simulates receptive fields for images of the static human body, as found by neuroimaging studies, and temporally integrates their responses by leaky-integrator neurons. This configuration-based approach model for biological motion perception reveals a high correlation to data obtained by neurophysiological, neuroimaging, and psychophysical studies.

4.2 Introduction

The visual motion generated by human actors are complex because the body comprises many degrees of freedom. Despite the complexity and diversity of the visual stimulus, humans can easily recognize the movements and gestures of others.

Many studies that investigated the perception of human movement used point-lights walker stimuli (Johansson, 1973). These stimuli consist of twelve point-lights that are attached to the joints of an otherwise invisible human body. Point-light walkers allow to investigate the impact of the different features of a walking human figure. Generally, these features can be divided in local and global features of motion and form and the dynamics of global motion and form.

Global motion can theoretically be derived from a suitable integration of local motion signals of the trajectories of the point-lights over time (Webb and Aggarwal, 1982; Hoffman and Flinchbaugh, 1982; Giese and Poggio, 2003). Alternatively, the visual system may analyze the global form that is sparsely available in the stimulus at each point in time. Although this information is insufficient to recognize a walker from a single frame (Johansson, 1973), temporal integration of the sparse form information may allow the identification of a walker (Chen and Lee, 1992; Beintema and Lappe, 2002; Beintema et al., 2005).

The superior temporal sulcus (STS) has often been implied in the perception of biological motion (Oram and Perrett, 1996; Bonda et al., 1996; Puce et al., 1998; Grossman et al., 2000; Vaina et al., 2001; Beauchamp et al., 2002; Santi et al., 2003; Thompson et al., 2005). Because it receives input from from and motion areas it is in a prime location to integrate form and motion processing (Oram and Perrett, 1996; Vaina et al., 2001; Beauchamp et al., 2002). The role of other brain areas is less clear. Some studies found selective activation of the middle temporal gyrus (MT) (Vaina et al., 2001; Ptito et al., 2003) and the kinetic occipital area (KO) (Vaina et al., 2001; Santi et al., 2003), which are believed to process local motion signals. Other studies reported activation of these areas not different to a control stimulus that presented dots with identical motion signals, but a randomized spatial arrangement that did not depict a human figure (Grossman et al., 2000; Downing et al., 2001). The extrastriate body area (EBA) is activated by static images of the human body (Downing et al., 2001). Area EBA is also activated by point-light walkers, but the activation does not differ from that evoked by the control stimulus mentioned above (Downing et al., 2001;

Grossman and Blake, 2002). Thus, while the role of the STS in biological motion recognition is undisputed, the contribution of signals feeding into the STS is currently not clear.

We are particularly interested in the possible contribution of form processing to biological motion recognition and present a model for its perception. The model is based on global, configural form information only and uses neurally plausible assumptions. We compared the model's performance to data from fMRI, neurophysiological, and psychophysical studies. The results demonstrate that perception of biological motion, even from point-light walkers, can be achieved by the analysis of global form recognition over time.

Background and motivation of the model

Classical point-lights walkers were introduced by Johansson (1973) and comprise twelve point-lights attached to the joints of an otherwise invisible human body. Point-light stimuli limit information about the walker's body structure. The visible points provide information about the joint positions, but the connections between them are absent. A single static picture of a point-light walker is insufficient to induce the percept of a human figure in naive observers (Johansson, 1973). When the stimulus is in motion, the individual dots provide fully correct motion signals. Therefore, many studies have concluded that biological motion perception is derived from local motion signals. For instance, Mather et al. (1992) presented a point-light walker embedded in randomly moving noise dots. Subjects viewed the stimulus frames that alternated with a mask consisting of blank frames. The duration of the mask was varied (60-100 ms). Direction discrimination was not possible if blank inter-stimulus frames intermit the stimulus in noise. Mather et al. concluded that local motion detectors which are disturbed by the blank frames are essential to recognize biological motion. Neri et al. (1998) argued in a similar way. They used biological motion or simple translatory motion as a stimulus and asked subjects to detect the stimulus in noise. The results showed no differences for detection of the two stimuli. Both revealed a linear increase of threshold for increasing stimulus dots. Performance threshold for discriminating the walking direction of a biological motion stimulus in noise, however, increased non-linearly with the number of stimulus dots. Neri et al. (1998) concluded from the first experiment that the common information of the two stimuli (that is motion) is the driving force for biological motion perception. These biological motion filters are flexibly adapted

to the stimulus, as reflected in the non-linearity revealed by the second experiment.

Early computational considerations also focused on local motion signals. Johansson (1973) and Cutting (1981) hierarchically reconstructed the human figure from common pendular movements of neighbored dots. The recent computational model of Giese and Poggio (2003) integrates local motion signals and local form signals in independent processing pathways to reconstructed templates of human motion. Only the motion processing pathway of this model was able to reconstruct a human body from point-light walkers.

The reliance on local motion signals is called into question by some observations in neurological patients. Vaina et al. (1990) studied a patient with bilateral lesions including area MT. This patient had severe difficulties in low-level motion integration tasks but no problems identifying biological motion displays. McLeod et al. (1996) reported that patient LM, who lacked all motion perception after a stroke (Zihl et al., 1983) was able to recognize action from point light biological motion stimuli. Her ability to see biological motion was lost, however, when the stimulus was embedded in noise. Vaina et al. (2002) described a patient that had difficulties to integrate local motion signals into a coherent motion percept or to perceive structure-from-motion but could recognize point light biological motion. These three cases demonstrate that biological motion perception is possible even when general motion analysis is impaired.

To study biological motion perception in the absence of local motion signals in healthy observers Beintema and Lappe (2002) developed point-light walkers in which point lifetime was limited to a single animation frame. In these stimuli, 98% of the local motion information is removed. Yet, naive observes readily recognized a walking human figure from these stimuli. Moreover, when observers had to identify the orientation of the walking figure, the addition of local motion signals by increasing the lifetime of the point lights did not aid performance. Beintema, Georg and Lappe (2005) showed similar results for a different biological motion task, namely the discrimination of forward from backward walking. Also in this task, which clearly involves the global motion direction of the figure, local motion signals did not contribute to task performance. Beintema et al. suggested that biological motion perception here was driven by the analysis of the variation of the form of the figure over time. These results prompted us to develop a model of biological motion perception from global form analysis.

Shiffrar et al. (1997) earlier emphasized the importance of global form analysis for interpreting biological motion. They presented stick-figures of walking humans

seen through apertures. Despite the ambiguous motion signals through the apertures subjects recognized the human figure easily. Heptulla Chatterjee et al. (1996) showed that form information in biological motion can override local motion signals. They presented a two photograph series of human movements and asked subjects to report the apparent motion path. Subjects reported the biomechanically consistent path rather than the shortest path, which would be reported if subjects used only local apparent motion signals.

Bertenthal and Pinto (1994) provided further evidence for an involvement of global form analysis in biological motion perception. They presented point-light walkers surrounded by noise dots. The motion trajectories of the noise dots were identical to those of the walker dots; only the global spatial configuration was different for walker and noise. Despite the identical motion signals in the noise, subjects could still recognize the walking figure. Bertenthal and Pinto argued that 'the perception of structure in a point-light walker does not require the prior detection of individual features or local relations'.

The above studies indicate that form analysis of the human body is involved in biological motion recognition. However, it is also clear that biological motion perception is a special function that goes beyond simple form (or motion) analysis. For instance, findings in patients have shown that biological motion perception can be impaired despite intact motion and form perception. Batelli et al. (2003) studied three patients with lesions in the parietal cortex. Although their ability in low-level motion tasks was normal, they were unable to perceive biological motion. Batelli et al. explained this with deficits in attention allocation. Schenk and Zihl (1997) examined stroke patients with lesions in the parietal cortex. In some patients the perception of biological motion as an isolated stimulus was possible but became impossible when a segregation from the background was necessary. Another study revealed that patients can have normal object and motion recognition performances without perceiving a form in a biological motion stimulus (Cowey and Vaina, 2000). Vaina and Gross (2004) studied four patients with brain damage due to strokes. All of them were unable to recognize a walker from a point-light figure. They had normal object recognition rates and only partial motion deficits but were impaired on recognition of objects from degraded incomplete information. These patients had damage to STS and were presumably unable to integrate the given information to a percept of biological motion. The specific impairment of biological motion recognition despite intact from and motion processing argues for a

separate integration stage in which signals that support biological motion analysis are integrated to achieve the percept.

Recent fMRI studies provided more insight into the neural correlates of biological motion perception. These studies almost uniformly report activation of STS when subjects viewed biological motion displays (Puce et al., 1998; Grossman et al., 2000; Vaina et al., 2001; Beauchamp et al., 2002; Santi et al., 2003; Peuskens et al., 2005; Thompson et al., 2005). STS gets input from both motion and form processing areas. FMRI studies reported selective activation of motion sensitive areas KO and MT (Vaina et al., 2001; Santi et al., 2003; Peuskens et al., 2005), while other studies found that the activation of MT and KO is not specific to biological motion (Grossman et al., 2000; Downing et al., 2001). Grossman et al. (2005) reported that transcranial magnetic stimulation to knock-out MT activity did not influence the perception of biological motion whereas TMS over STS impaired the perception of biological motion. Other studies found selective activation in form areas such as the fusiform gyrus or the occipital face area (Vaina et al., 2001; Grossman and Blake, 2002; Peelen and Downing, 2005b; Michels et al., 2005). Beauchamp (2003) showed that point-light displays of human actions activate the ventral temporal cortex although this activation is less strong than for whole body displays. Michels et al. (2005) used different biological motion stimuli which varied in the amount of available motion and form information. Activation levels in areas sensitive to processing static human form depended strongly on the amount of structural information in the stimuli but not on local motion signals. This suggests that form processing areas are recruited for biological motion perception.

Specific form processing areas and the STS are also driven by static images of the human body (Beauchamp et al., 2002). These activations are increased when motion is added (Beauchamp et al., 2003). The extrastriate body area (EBA) shows selective responses to static pictures of human bodies and of stick figures (Downing et al., 2001). However, the role of EBA in perceiving point-light displays remains unclear. Downing et al. (2001) observed a stronger activation for biological motion displays than for scrambled non-human figures with identical motion signals. They attributed this signal increase to the engagement of attention driven by the presence of a body configuration. This finding was not confirmed by Grossman and Blake (2002). They found that EBA responds to biological motion stimuli but also to the scrambled controls. Peelen and Downing (2005) reported significant activation in FFA for human bodies shown without a head.

Thompson et al. (2005) presented displays of walking mannequins which were either intact or with the limbs and torso scrambled. Stimuli were either completely visible or partially occluded. Activation in STS was always greater for the intact walkers than for the scrambled walkers regardless of whether parts of the body were occluded or not. Thompson et al. concluded that processing of biological motion in STS is driven by configural processing of the walking stimulus rather than tracking the movement of individual limbs. This provides means to process biological motion even in the case of occlusion.

4.3   Methods

4.3.1   The model

From the above studies, we can, for the purpose of our model, derive three assumptions. First, biological motion may be inferred from form analysis without local motion processing. Second, form analysis in some areas of the ventral stream is selective for the shape of the static human body. Third, biological motion perception is a specialized process that combines the analysis of the global form of the human body with its global motion. Our model follows these assumptions. It uses form sensitivity and processes biological motion in two stages, a static and a dynamic form stage. Leaky integrator neurons of the second stage dynamically integrate the output of neural template cells from the first stage. These template cells are formed by Gaussian response functions that simulated receptive fields for human bodies.

Fig. 4.1 shows a schematic overview of the model. We assume a library of upright static template cells of human walkers which are implemented in a view-based template-matching approach. The three-dimensional configuration of a walking human body is represented by a collection of two-dimensional postures. We assume that this view-based approach is invariant to size and position of the perceived object, similar to the properties of neurons in higher areas of the ventral stream (Logothetis et al., 1995b; Tanaka 1996; Riesenhuber and Poggio, 1999).

We generated the template cells from recordings of the movements of nine human walkers (age 20-29, 5 male). The individuals walked normally on a catwalk with sensors attached to their major joints (head, shoulders, elbows, wrists, hips, knees, ankles) while a motion tracking system (Ascension MotionStar, USA) recorded their

movements at 95 Hz sampling rate. To reduce noise we filtered the tracking data by averaging three successive data points of each sensor. If necessary, additional data points of the walking sequence were obtained by interpolation between the filtered recording data. Then, each of the nine walking sequences was divided in temporally equal intervals to obtain a set of 50 sequential body configurations for each walker. The recorded joint positions for each configuration were connected in the anatomically correct way to obtain stick-figures of a common walking sequence. These stick figures formed the basis of the body template cells of the model. Each such body template cell is selective for a particular body posture. The cells response to a biological motion stimulus is derived from the total of the responses to the individual stimulus dots. The response to a dot near a particular position on the body is assumed to be maximal if the dot is located on the body and drops off with a gaussian function of distance of the dot to the nearest point on the body (Fig. 4.1). Because our study is intended to investigate the contribution of global form, our model decidedly uses a top-down approach. The model treats the body as a global figure without explicitly taking into account local stimulus features (orientation, motion). This is different from earlier models, which combine local features hierarchically into a percept of a human body (Johansson, 1973; Cutting, 1978; Giese and Poggio, 2003).

We used two different sets of template cells: one for a walker oriented and moving to the right and one for a walker oriented and moving to the left. Differential activity within those two sets is used for decisions in the discrimination tasks we describe below. In each set, the nine different walkers redundantly represented each of the 50 static postures for a total of 450 templates.

The model consists of two stages: a first stage for the analysis of the form (posture) of the walker and a second stage for the analysis of the global motion (postural change) of the walker (Fig. 4.1). Our choice of different stages for these tasks is partly motivated by the above mentioned fMRI studies, which showed different selectivities for static and moving human bodies, and in part by differences observed between biological motion tasks. For instance, Vaina et al. (2001) showed that identical displays of biological motion may activate different brain regions depending on the task. When the subjects had to discriminate between the shape of the walking pattern and a scrambled control stimulus, different regions were activated than for judging the overall motion direction of the dots. Results of Beintema et al. (2005) also suggested a task-specific analysis of biological motion stimuli. When subjects were asked to identify the direction in

which a point-light walker faced (left or right), they mainly used information about the shape of the figure. When asked to discriminate between forward and backward walking point-light figures, subjects used also information about the global motion of the stimulus. These results argue for a task-dependent analysis of a biological motion stimulus as implemented in the different stages of the model.

Stage 1

At the onset of stimulation the first stimulus frame is present in stage 1. This frame is compared to the templates of each of the template cells. Each dot of the stimulus frame contributes to the cells response weighted by the distance to the nearest part of the body. Each cell sums the responses for all single dots to obtain an overall response measure to this stimulus frame(Eq. 4.1).

$$F_{tc} = \sum_{i=1}^{n} e^{-\frac{(\mu_{tc} - p_i)^2}{2 * \sigma^2}} \tag{4.1}$$

where F_{tc} denotes the output of the template cell tc. The outputs of the template cells were obtained by weighting the shortest distance between a stimulus dot and a limb of the template with a Gaussian function. p_i gives the position of the stimulus dot i and μ_{tc} denotes the limb position in the template cell with the shortest distance to the stimulus dot. σ is the width of the template cells' receptive field that is defined by the Gaussian weights.

  This template-matching procedure is done independently for both sets of template cells. A winner-takes-all mechanism selects the maximum output within each set and feds it into a leaky-integrator (Eq. 4.2). The template-matching procedure is repeated for each stimulus frame independently of the preceding one, and the maximum outputs of both sets are fed into two leaky integrators. The activities $u_{1,2}$ of the integrators are computed from

$$\tau \dot{u}_{1,2}(t) = -u_{1,2}(t) + i_{1,2} + w_+ * f(u_{1,2}(t))) - w_- * f(u_{2,1}(t))) \tag{4.2}$$

where $\tau = 10$ ms, $u_{1,2}$ denotes the activities in the decision stage 1 for the two sets of templates and $i_{1,2}$ denotes the bottom-up inputs from both sets of template cells to the decision stage 1 as defined by the maximum outputs of the template cells in Eq. 4.1:

$$i_{1,2} = \max_{tc}(F(t)_{tc})_{1,2}$$

The lateral interaction between the two integrators is given by $f(u_{1,2}(t))$ with f a sigmoid function that integrates the state of the two integrators:

$$f(u_{1,2}(t)) = \frac{1}{1 + e^k}$$

with

$$k = -2 * \frac{u_{1,2}(t) - \max_t(u_{1,2}(t))}{\max_t(u_{1,2}(t))}$$

In Eq. 4.2, lateral interaction is weighted by w_+ and w_- which denote the weights for lateral excitation and inhibition between the states $u_{1,2}$.

The activities $u_{1,2}$ provide a decision criterion for a left/right discrimination in stage 1. The maximum activity over the total trial duration of both kinds of template cells is taken for a decision of the model. The excitatory and inhibitory weights w_+ and w_- are free parameters of the model that will be fixed in a single simulation later(see 4.3.2).

Stage 2

The model in stage 1 does not explicitly consider the temporal order of the stimulus frames. This is implemented in stage 2. We assume that the recognition of one frame influences the expectation of the next frame:

$$\tau \dot{v}_{1,2}(t) = -v_{1,2}(t) + w_{n,m} * u(t) \tag{4.3}$$

τ : 10 ms, $v_{1,2}$ denote the activities in the 'decision' stage 2 for the possible responses 1,2 and u is the bottom-up input from stage 1. $w_{n,m}$ weights the difference between selected frame n and previously selected frame m (Fig. 4.1). This function should be asymmetric and non-linear. We chose :

$$w_{n,m} = \cos(\frac{n-m}{c})^{\{a,b\}}$$

with a for n-m $\leq$ 0, b otherwise
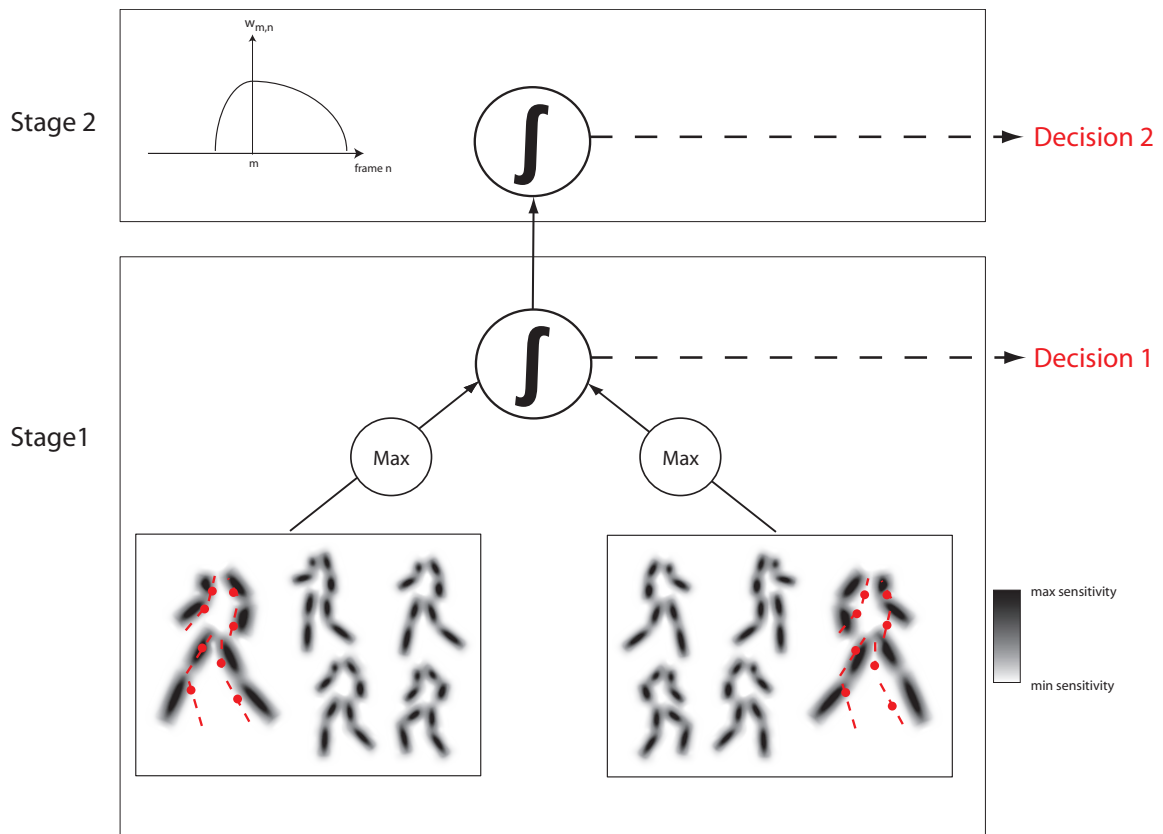
Fig. 4.1: Illustration of the model. Stage 1 consists of template cells for whole body postures. Two sets of template cells represent walking to the left and walking to the right, respectively. Each set comprises static postures with Gaussian response functions indicated by the shaded areas. The red dots indicate the stimulus that is compared to the template by each template cell at each time step of stimulation. In each set, a winner-takes-all process selects the cell with maximum output at each time step. From the temporal integration of these maximum responses in the two sets a decision about stimulus orientation is achieved as a first stage of analysis. This stage of analysis does not considering the temporal order of the stimulus postures. Analysis of temporal order is deferred to stage 2 and provides information for a decision on global motion aspects of the stimulus. This analysis is achieved by weighting the temporal differences between two consecutively selected frames (m,n) by the function $w_{m,n}$. This function non-linearly weights adjacent frames more strongly than frames which are temporally separate. In addition, the function is asymmetric for it weights frames that are in the cell's preferred direction more strongly than frames that are in temporally opposite direction (see illustration of $w_{m,n}$ above for a selected frame m and as a function of the subsequently selected frame n).

4.3.2 Experimental methods

We tested the model by comparing its results to neurophysiological, fMRI and psychophysical data from the literature. In doing so, we adapted the stimulus settings of the corresponding experiments. We conducted additional psychophysical experiments to test model predictions. Here, we used similar experimental settings as described by Beintema and Lappe (2002) and Beintema et al. (2005). In the following, we provide a brief description of the stimulus and the procedure. Further information can be obtained from those publications.

Tasks

The experiments involved either of two discrimination tasks: a direction task or a forward/backward task. In the direction task, the walker was presented either facing to the left or to the right. The subject had to report the walkers facing orientation. In the forward/backward task, the walker was presented in left or right orientation and either with a normal forward gait or in backwards motion, in which case the frames of the walking sequence were displayed in reverse order. The subjects had to report whether the stimulus walked forward or backward.

Stimulus

The stimulus was generated by a computer program and imitated the movements of a walking human (Cutting, 1978). For the model simulations it is important to realize that the stimulus never exactly matches any of the nine recorded templates of real walkers, as it presumably does not exactly match the motions of a real walker to a human observer. In the original program by Cutting, the human body was depicted by light-points attached to the major joints of an otherwise invisible body. Beintema et al. modified this stimulus such that it consisted of a variable number of points (1-8) each with a randomly chosen position on the limbs. Each point was relocated to a new randomly chosen position on the limbs after every single frame of the animation sequence. This stimulus allows to study the perceptual mechanisms of biological motion in conditions with near-absent local motion signals, thus, focussing on the role of form information. We used this stimulus in the experiments described below. The number of dots present in each stimulus frame (1-8), the duration of presentation of each stimulus frame (10-200 ms), and the lifetime of each dot (1-8 frames) are parameters that

influence the amount of form and global motion present in the stimulus (Beintema and Lappe, 2002; Beintema et al., 2005). The parameters used in each of our experiments are described in the respective section.

Experimental procedure

Stimuli were presented on a monitor with a resolution of 1280x1024 pixel and a display size of 30 cm x 40 cm. The monitor refresh rate was 100 Hz. Unless indicated otherwise, a single stimulus frame was presented for a duration of 50 ms (5 monitor frames) and a total trial lasted for 1.6 s, i.e. one walking cycle.

The stimulus covered a field of 5° x 10° and consisted of white dots (5x5 pixels) on a black background. Trials were presented in random order, and the stimulus position had a randomly chosen spatial offset to avoid spatial cues.

Between 4 and 5 subjects (2 female) participated in each experiment. They were between 26 and 35 years of age and had normal or corrected to normal vision. All subjects were student or members of the department and experienced in psychophysical experiments. Subjects were seated 60 cm in front of the monitor and viewed the stimulus binocularly. Subjects had to indicate their decision in the respective discrimination task by pressing one of two buttons in front of them after the stimulus presentation.

Simulation procedure

We compared the models performance to existing data and to data obtained in new experiments. For existing data we mimicked the stimuli described in the corresponding study. For new psychophysical experiments, we used identical stimuli for model simulations and experimental tasks.

Each simulation run consisted of 150 trials with stimuli with randomly chosen starting phases in the walking cycle. The model computed activation levels for these stimuli in stage 1 ($u_{1,2}(t)$) and in stage 2 ($v_{1,2}(t)$). At each model stage we compared activation levels for both possible decisions (left/right, forward/backward) and used them for the decision in the respective perceptual tasks on a trial-by-trial basis. We then calculated the proportion of correct answers over all trials.

To simulate physiological experiments, we compared the activity in the stages of the model induced by the stimulus with that induced by a respective control. We compared model predictions to fMRI data by normalizing the activation levels of the

model's stages to the signal changes for the stimuli obtained by fMRI data for identical stimuli.

Parameter fits

The model stage 1 contains two adjustable parameters, namely the excitatory and inhibitory weights, w_+ and w_- (Eq. 4.2). To estimate the values of these parameters, we conducted a psychophysical experiment and fitted the model to the psychophysical data (Fig. 4.2). The obtained fit was then used for all further simulations in this study.

Since model stage 1 is concerned with form analysis the experiment focussed on stimulus properties that influence form information. First, we manipulated the number of dots per stimulus frame (2-8) to examine the influence of form information per stimulus frame. Second, we varied the form information per trial by varying the stimulus duration (100-1600 ms).

Subjects were asked to report the orientation (left or right) of the walker. The model solved the task by matching the stimulus frames either to template cells for a walker oriented to the right or to template cells for a walker oriented to the left. We varied the free parameters so that the model simulations fitted optimally to the psychophysical data for the condition of 8 points per frame. The parameters ($w_+ = 6.8; w_- = 4.0$) were then fixed for all experiments and simulations reported in this study.

We also tested whether the choice of fit data influenced the model. Fitting results for other conditions with 2 or 4 dots per stimulus frame resulted in the same parameter set. Thus, the results do not rely on the kind of fitting or the data we chose for fitting.

Fig. 4.2 displays observer percent correct and model simulation results for 8, 4, and 2 dots per frame. The observer data for 8 points per frame were used to adjust the weights of the model. Data from the 4 and 2 dot conditions provide an estimate of how well the parameter fit generalizes. The data reveals a clear relationship between form information and performance of the human observers (Fig. 4.2). Our form-based model matches these data for all parameters (form per frame/overall form).

Stage 2 contains one free parameter ($w_{n,m}$) which determines the model's expected frame order. To estimate the value of this parameter we used a forward/backward discrimination task with 8 dots per stimulus frame and varied the amount of form information by varying the total trial duration between 100 and 1600 ms. Unlike the direction task, the forward/backward task cannot be solved solely by spatial analysis

Fig. 4.2: Psychophysical data used to fix the free parameters of stage 1 and evaluate the model's optimum fit. Shown are correct responses as a function of total trial duration for 2, 4, and 8 points per stimulus frame in a direction task. Psychophysical data are presented as mean $\pm$ SE

(Beintema et al., 2005). Since the order of the selected frames has to be taken into account the temporal integration in stage 2 is crucial.

Human subjects were asked to discriminate between a walker moving forward or backward. The model solved the forward/backward task by analyzing, in stage 2, the temporal order of the template cells which were most strongly activated in stage 1 by the sequential stimulus frames. The model used the outputs in stage 2 for an expected forward movement compared to an expected backward movement as the decision criterion to solve the task.

The results from human observers (Fig. 4.3) were used to fit the free parameter of stage 2 of the model ($w_{n,m}$ in Eq. 4.3). This best-fitting weighting function ($w_{n,m} = \cos(\frac{n-m}{9.6})^{\{2400,50\}}$) was then used for all simulations reported in this paper.
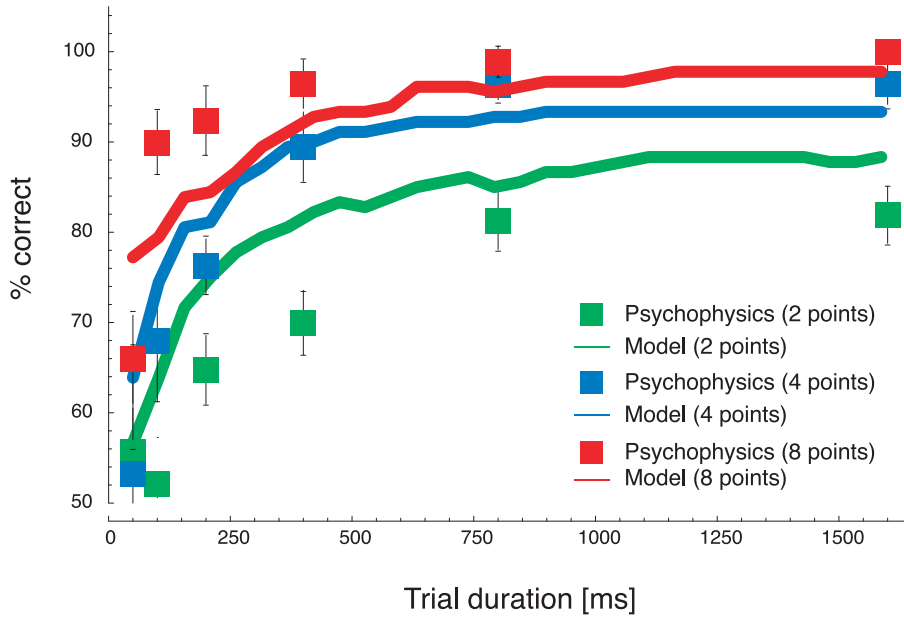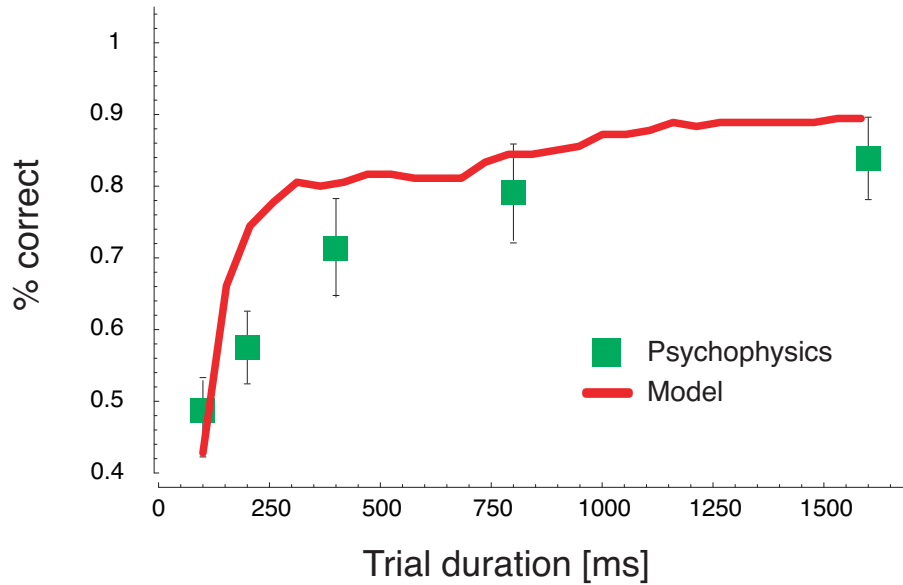
Fig. 4.3: Psychophysical data used to fix the free parameter of stage 2 and evaluate the model's optimum fit. Shown are correct responses as a function of total trial duration for 8 points per stimulus frame in a forward/backward task. Psychophysical data are presented as mean ± SE

4.4   Results

4.4.1   Direction task

Beintema and Lappe (2002) and Beintema et al. (2005) asked subjects to discriminate between a walker facing to the right and a walker facing to the left. They manipulated the number of stimulus dots, the duration of each trial, and the amount of motion signals. Beintema and Lappe (2002) demonstrated that the subjects' recognition rates depended strongly on the available form information rather than on motion signals. Furthermore, the results of Beintema et al. (2005) showed that the recognition rates depend on the number of dots per frame (i.e. form information per frame) and on trial duration (i.e. overall form information per trial). However, across all experiments the recognition rates were constant if the product of trial duration and number of dots was constant, i.e., when the total number of stimulus dots presented during the trial was constant. The model is generally consistent with this because its recognition rates critically depended on the number of stimulus dots presented during the trial (cf. Fig. 4.2). However, the model relies on spatio-temporal integration of form information. This predicts that performance should also depend on the speed with

which new information is acquired. Therefore, we conducted an experiment in which we manipulated the information rate of the stimulus by varying the duration each frame was displayed. Total stimulus duration and walking speed were kept constant. For long frame duration, therefore, the walker remained in one static posture for some time and then changed its posture in a large step to another posture. For short frame durations, the walking sequence was sampled rapidly and appeared smooth. Thus, dynamic sampling of the walking sequence is different for different frame durations as the dynamical change from one displayed posture to the next postures varies.

We presented 4 or 8 points per frame and varied frame duration from 10-200 ms. Fig. 4.4 shows the results for human observers separately for the 4 dots-per-frame and the 8 dots-per-frame condition. Both graphs show a significant influence of frame duration (ANOVA with repeated measures, 4 points: $F(6, 24) = 36.3, p < 0.01$, 8 points: $F(6, 24) = 10.9, p < 0.01$). Linear regression revealed that in both conditions the subjects' performance decreases (4 points: $r^2 = 0.76, F(33) = 103.7, b_0 = 1.04, b_1 = -0.2, p < 0.01$, 8 points: $r^2 = 0.54, F(33) = 38.6, b_0 = 1.03, b_1 = -0.07, p < 0.01$).

The model solves the task by matching the stimulus frames to the template cells for walking to the right and to the template cells for walking to the left and integrating the outputs dynamically. For prolonged frame duration, fewer frames are available within the integration time of the leaky integrator. Therefore, the model performance decreases. The decrease of model performance replicates the psychophysical data qualitatively. For prolonged frame durations, recognition rates drop in a similar way as in the psychophysical data. The model also replicates the stronger and earlier drop of recognition rate for 4 dots per frame. However, a quantitative comparison between model and psychophysical data revealed significant differences for some data points in the condition 4 points (Tab. 4.1). For short frame durations, on the one hand, subjects made only few mistakes due to ceiling effects. This results in small standard errors for the psychophysical data so that only small differences (in percentage correct) for the model result in significant deviations. For frame durations of 90 ms and more, none of the subjects yielded 100% correct recognition rates. Therefore, if ceiling effects are ruled out, the model's performance was within the confidence limits of the psychophysical data. This holds true for 8 points per frame and for 4 points per frame, except for a frame duration of 200 ms when the model data are marginally outside the confidence limits. On the other hand, there might be a shift of about 20 ms on the x-axis between model and psychophysical data. Thus, the significant differences may reflect that the

model's performance drops too early. Nevertheless, the model's dependence on frame duration is qualitatively the same as for the psychophysical data. This supports our hypothesis that form information is integrated over a fixed temporal period. Thus, the results of the direction discrimination task reveal a dependence of recognition rates on form information per frame, form information per trial, and form information per time period.
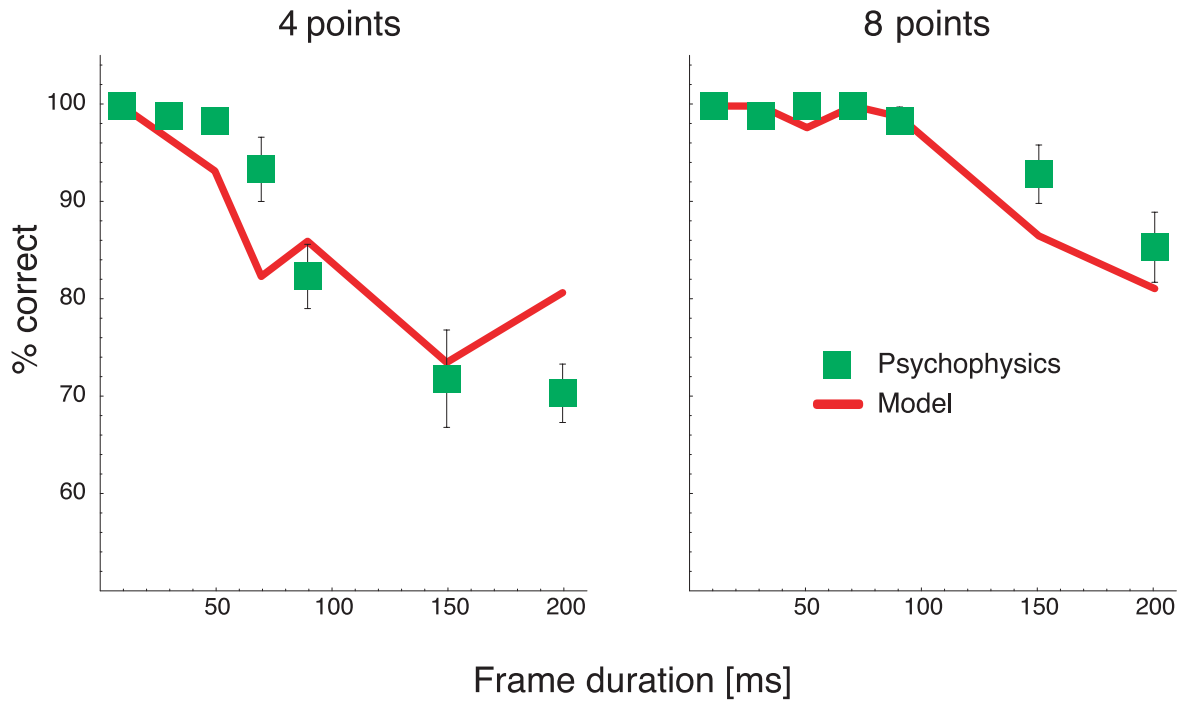
Fig. 4.4: Correct responses as a function of the duration of a single frame for 4 (left) and 8 points (right) per stimulus frame in a direction task. Psychophysical data are presented as mean ± SE.

Table 4.1: The table shows the parameters for a quantitative comparison between model data and the psychophysical data in the direction task for 4 and 8 points per frame depending on frame duration (one-sample t-test, $\alpha = .05$). n.d. denotes t- and p-values that are not defined because the standard error of the psychophysical data is equal 0 due to ceiling effects. * indicates statistically significant differences between model and psychophysical data.

No. of points	Frame duration [ms]	Data Model [%]	Mean Psycho-physics [%]	SE [%]	Lower confidence limit [%]	Upper confidence limit [%]	df	t-value	p-value
4 points	10	100	100.0	0	100	100	4	n.d.	n.d.
	30	97	99.0	0.2	97.8	100.0	4	3.3	$< .05^*$
	50	93	98.5	1.0	95.9	100.0	4	5.2	$< .01^*$
	70	83	93.5	3.3	86.9	100.0	4	4.4	$< .01^*$
	90	86	82.5	3.3	73.5	92.1	4	1.12	$> .10$
	150	74	72.2	5.0	58.3	86.1	4	0.2	$> .40$
	200	81	70.8	2.9	62.7	78.9	4	3.0	$< .05^*$
8 points	10	100	100.0	0	100.0	100.0	4	n.d.	n.d.
	30	100	99.0	0.4	97.8	100.0	4	2.5	$> .05$
	50	98	100.0	0	100.0	100.0	4	n.d.	n.d.
	70	100	100.0	0	100.0	100.0	4	n.d.	n.d.
	90	99	98.4	1.4	94.6	100.0	4	0.36	$> .35$
	150	87	93.4	2.9	85.2	100.0	4	2.0	$> .05$
	200	81	85.8	3.6	75.8	95.8	4	1.25	$> .10$

4.4.2   Forward/backward task

The model simulations and psychophysical results in the direction task provided evidence that the recognition rates depend on form information per time period. This task may be performed using only form information (Beintema et al., 2005). Beintema et al. (2005) therefore introduced the forward/backward discrimination task which cannot be solved solely by spatial analysis. In this task, both stimuli (moving forward and moving backward) comprised exactly the same set of frames. In one condition the frames were presented in forward moving temporal order, in the other condition in reversed temporal order. Thus, the temporal properties of the stimulus are crucial. Beintema et al. investigated the influence of different parameters in this task. We used the same task and compared model predictions for this task to the data from Beintema et al. (2005) and to additional psychophysical experiments reported below. We varied the dynamic behavior in two ways. First, by keeping the total trial duration constant and varying the duration a single frame is presented (variation of frame duration). Second, by manipulating dynamic behavior by varying the walking speed. Here, the number of frames per trial was kept constant and trial duration varied with the duration each frame was presented.

For the first simulation, we adapted psychophysical data of Beintema et al. (2005). A stimulus with 8 points per frame was presented and the duration of a stimulus frame was varied between 30-200 ms. Total stimulus duration was kept constant and always contained one full step cycle. Thus, for longer frame durations the number of frames was reduced and the change of posture between frames was increased to keep the walking speed constant. The observers' task was to discriminate between a walker moving forward and a walker moving backward.

The data are shown in Fig. 4.5 (no isi). For smoother presentations (shorter frame durations), the walking direction was recognized more easily. For longer frame durations, the walking direction was harder to discriminate, at 200 ms frame duration just above chance level. The model showed the same behavior as the human observers in that study.

To further rule out a contribution of local motion detectors to the recognition process Beintema et al. (2005) repeated this experiment with a blank inter-stimulus-interval (isi) between stimulus frames similar to Mather et al. (1992). Above, each frame was presented for only 20 ms and the remaining time of the frame duration was

filled with a blank frame. The psychophysical data for the 'isi'-condition (Fig. 4.5 (isi)) was not different from that of the no-isi condition (Beintema et al., 2005). We compared the model's behavior also to these psychophysical data. In principle, the model should be unaffected by the intermitted blank intervals since it does not rely on local motion signals and thus should not be influenced by the manipulation of local motion signals in the isi condition. However, for the model there is a difference between the two conditions in the reduced presentation time of stimulus frames in the isi condition. The psychophysical data showed in both conditions a significant dependence on frame duration(ANOVA with repeated measures, isi: $F(5,6) = 4.43, p < 0.05$, no isi: $F(5,6) = 7.99, p < 0.05$). A trendtest revealed that in both conditions performance decreases linearly with prolonged lifetime (isi: $F(1,6) = 37.4, p < 0.01$, no isi: $F(1,6) = 73.7, p < 0.01$). During the intermitted blank interval activation in the template cells drops. The results in Fig. 4.5 (isi) show similarity between model and human observers also in the isi condition. Statistical analysis revealed that also quantitatively model and psychophysical data coincide (Tab. 4.2). Only in the no-isi condition one data point showed significant differences between model and psychophysics (frame duration 50 ms). However, this may due to the fact that subjects' performance was at ceiling level so that made less errors. This results in a lower standard error and, thus, in confidence intervals that do not coincide with the model data. Moreover, in other experiments that used the forward/backward task (e.g. task 'velocity' below), subjects revealed lower recognition rates, so that we do not ascribe to much weight on this difference. In agreement with the psychophysical data, the model did not show any differences between the isi- and the no-isi conditions. These results imply that the forward/backward task can be solved by spatio-temporal analysis of form information. The model simulations fit the psychophysical data in the absence of local motion analysis.

Next, we tested the influence of walking speed on the model behavior and compared the results to psychophysical data in a new experiment. We presented a stimulus with 8 points per frame. We kept the number of frames constant (32 frames) and varied the presentation duration of each stimulus frame (20-200 ms). This resulted in slow or fast walking speed. Thus, the overall form information per trial (number of dots per frame x number of frames) kept constant, but the amount of information within a certain temporal integration period and the total duration of the stimulus varied.

Fig. 4.6 shows the results. The human observers reveal a maximum in recognition
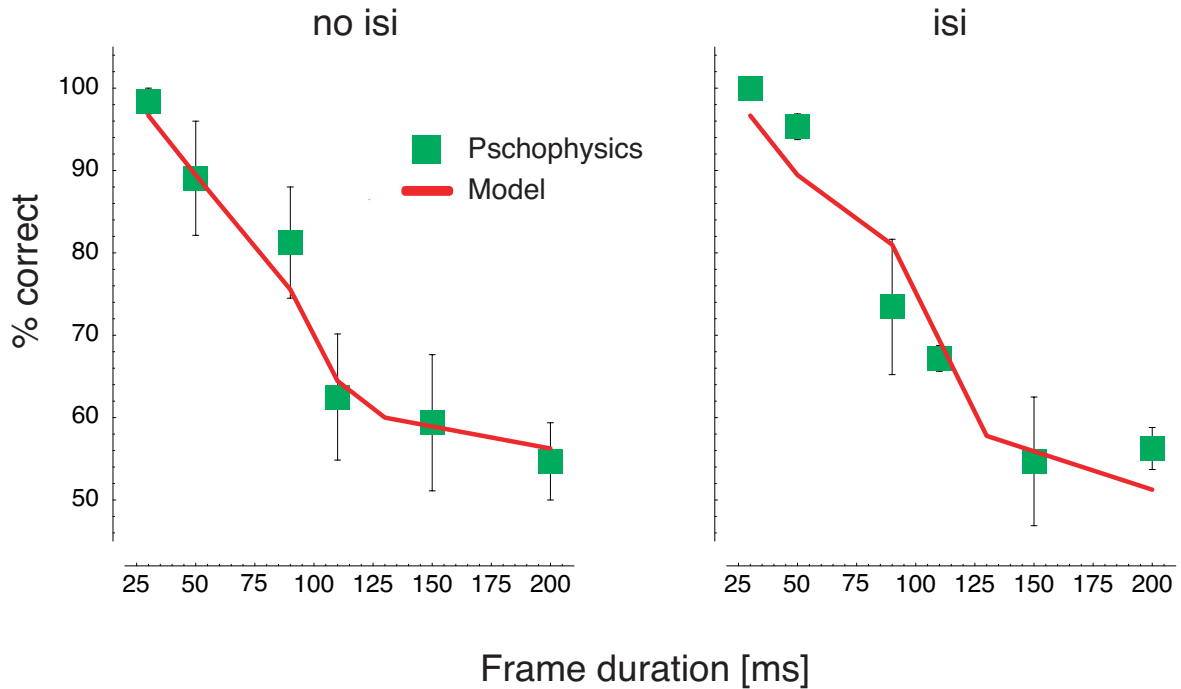
Fig. 4.5: Correct responses as a function of frame duration for 8 points per stimulus frame in the forward/backward task. Frames are presented for the indicated duration (no-isi condition, left) or for 20 ms followed by a blank period for the remainder of the frame duration (isi condition, right). Psychophysical data are presented as mean ± SE. Psychophysical data are adapted from Beintema et al. (2005).

rate for normal walking speed. Statistical analysis revealed a significant influence of velocity (ANOVA with repeated measures, $F(6, 24) = 5.93, p < 0.01$). Recognition rates decrease for higher and lower walking speeds. Comparison of the recognition rates for canonical walking speed (velocity 1) and the other velocities revealed a significant decrease for the two lowest walking speed (velocity 3: $F = 12.20, p < 0.05$, velocity 4: $F = 66.30, p < 0.01$). For higher walking speeds, recognition rates decreased, however, their was only a statistical trend, which was not significant ($F = 5.60, p = 0.077$). The differences are not unexpected because there is a certain preferred speed associated with a particular walking pattern (Giese and Lappe, 2002).

The model simulations also show a decline in performance as the walking speed becomes different from the canonical walking speed. Activation levels of the template cells cannot reach their maximum level if presentation times are short. For long presentation times, the outputs of the template cells will no longer be integrated effectively due to the limited integration period.

Table 4.2: The table shows the parameters for a quantitative comparison between model data and the psychophysical data in the forward/backward task for the isi and no isi conditions depending on frame duration (one-sample t-test, $\alpha = .05$). n.d. denotes t- and p-values that are not defined because the standard error of the psychophysical data is equal 0 due to ceiling effects. * indicates statistically significant differences between model and psychophysical data.

Condition	Frame duration [ms]	Data Model [%]	Mean Psycho- physics [%]	SE [%]	Lower confidence limit [%]	Upper confidence limit [%]	df	t-value	p-value
no isi	30	97	98.4	1.6	94.0	100.0	2	0.89	> .20
	50	89	89.1	6.9	70.0	100.0	2	0.09	> .45
	90	76	81.3	6.8	62.5	100.0	2	0.89	> .25
	110	64	62.5	7.7	41.2	83.8	2	0.19	> .40
	150	60	59.4	6.3	41.9	76.9	2	0.10	> .45
	200	56	54.7	4.7	41.7	67.7	2	0.26	> .40
isi	30	97	100.0	0.0	100.0	100.0	2	n.d.	n.d.
	50	89	95.3	1.6	90.5	99.4	2	4.04	< .05 *
	90	81	73.4	8.2	50.7	96.1	2	0.92	> .20
	110	69	67.2	1.6	62.8	71.6	2	1.13	> .20
	150	57	54.7	7.8	33.1	76.3	2	0.30	> .40
	200	51	56.3	2.6	49.1	63.5	2	1.98	> .05

Table 4.3: The table shows the parameters for a quantitative comparison between model data and the psychophysical data in the forward/backward task depending on the velocity of stimulus presentation (one-sample t-test, $\alpha = .05$). n.d. denotes t- and p-values that are not defined because the standard error of the psychophysical data is equal 0 due to ceiling effects. * indicates statistically significant differences between model and psychophysical data.

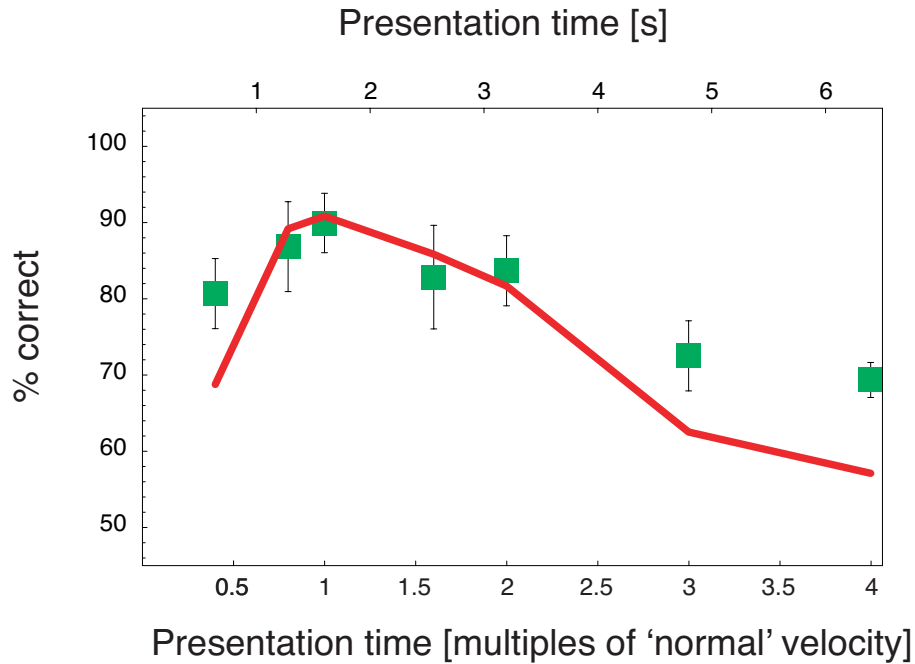| Stimulus duration [s]	Data Model [%]	Mean Psycho- physics [%]	SE [%]	Lower confidence limit [%]	Upper confidence limit [%]	df	t-value	p-value
0.64	69	80.4	4.6	67.5	93.3	4	2.48	> .05
1.12	89	87.0	5.9	70.7	100.0	4	0.34	> .35
1.60	91	91.4	3.9	80.1	100.0	4	0.10	> .45
2.56	86	83.0	6.8	64.1	100.0	4	0.44	> .35
3.20	82	83.6	4.6	70.1	96.4	4	0.35	> .35
4.80	63	72.8	4.6	60.1	85.5	4	2.13	> .05
6.40	57	69.4	2.3	63.0	75.7	4	5.4	< .05*

Fig. 4.6: Correct responses as a function of the walker's velocity for 8 points per stimulus frame in the forward/backward task. Psychophysical data are presented as mean $\pm$ SE

4.4.3 Discrimination in noise

So far, we have investigated the perception of biological motion with an isolated stimulus. Neri et al. (1998) reported a remarkable efficiency of human observers in the temporal integration of biological motion in noise. They presented a point-light walker with a variable number (1-6) of simultaneously visible dots located on the joints of the walker. The dots kept this joint-position for two frames before disappearing and relocating to a new joint location. Therefore, each dot provided useful local motion signals for two frames. The walker was embedded in a random noise mask of dots that changed position in every frame. This stimulus was presented on one side of a fixation dot. The other side displayed the same noise dots plus the number of dots of the walker in random position. Human observers had first to determine in a 2AFC-task the correct presentation side of the stimulus. After correct detection, they had to discriminate between the walking directions of the stimulus, or, in another condition, the coherence of the stimulus. In this coherence task (Mather et al., 1992), the upper and lower parts of the stimulus were shown either in the same (coherent) or in opposite (incoherent) direction. Subjects had to decide whether the stimulus was coherent or not. Neri et al. determined noise thresholds for 75% correct recognition rates in the

detection and the discrimination task. In agreement with previous results (Barlow, 1997), they found that the relationship between number of stimulus dots and number of noise dots is linear for the detection of the biological motion stimulus. In the case of discriminating walking direction, the relationship was non-linear, featuring a more rapid increase of performance with increasing number of stimulus dots.

We simulated the discrimination performance of the model for a stimulus surrounded by random dot noise. We adapted the stimulus of Neri et al. (1998) such that 1 to 6 dots per frame were presented simultaneously on the major joints of the body. They moved on this position for two frames before they were redrawn on a new joint. For a fixed number of stimulus dots, we varied the number of noise dots within a window of 6 by 4.5 times the size of the stimulus. Model simulations were run with these stimuli in the direction and coherence tasks. For the coherence task, the model applied the same steps as in the discrimination task, but separately for the upper and lower parts of the body. The templates were, therefore, subdivided into templates for the upper body (arms) and the lower body (legs). This resulted in two final decisions of the model, one for each body part. Comparing these two decisions resulted in the overall decision whether the walker was coherent or incoherent. We fitted the levels of correct response to a sigmoid function and determined the threshold for 75% correct responses. These values were plotted in a log-log-diagram and slopes of linear regression were determined analogously to Neri et al. (Fig. 4.7).

Linear regression revealed a slope steeper than 1 (independent t-test, direction task: $t(4) = 5.19, p < 0.01$, coherence task: $t(4) = 3.95, p < 0.05$) consistent with the human data. The slope for discrimination of the walking direction is 3.18 ± 0.42 ($r^2 = .95, t(4) = 7.65, p < 0.01$). This is in the range of the two subjects in the study of Neri et al., which had slopes of 2.55 and 4.23. In accordance with Neri et al., the slope in the coherence task was even steeper (4.12 ± 0.79, $r^2 = .90, t(4) = 5.19, p < 0.05$). This is similar to the value Neri et al. obtained from one subject (4.48). We conclude that the template matching approach is able to reproduce the spatial integration properties for discrimination in noise.

A similar conclusion was given by Lee and Wong (2004), who proposed a template matching algorithm based on the distance of dots to the joints not the body segments. The neurally plausible approach of our model is able to replicate the psychophysical data also quantitatively. Cutting et al.(1988) investigated the efficiency of various noise masks on the perception of point-light displays. Detection rates de-
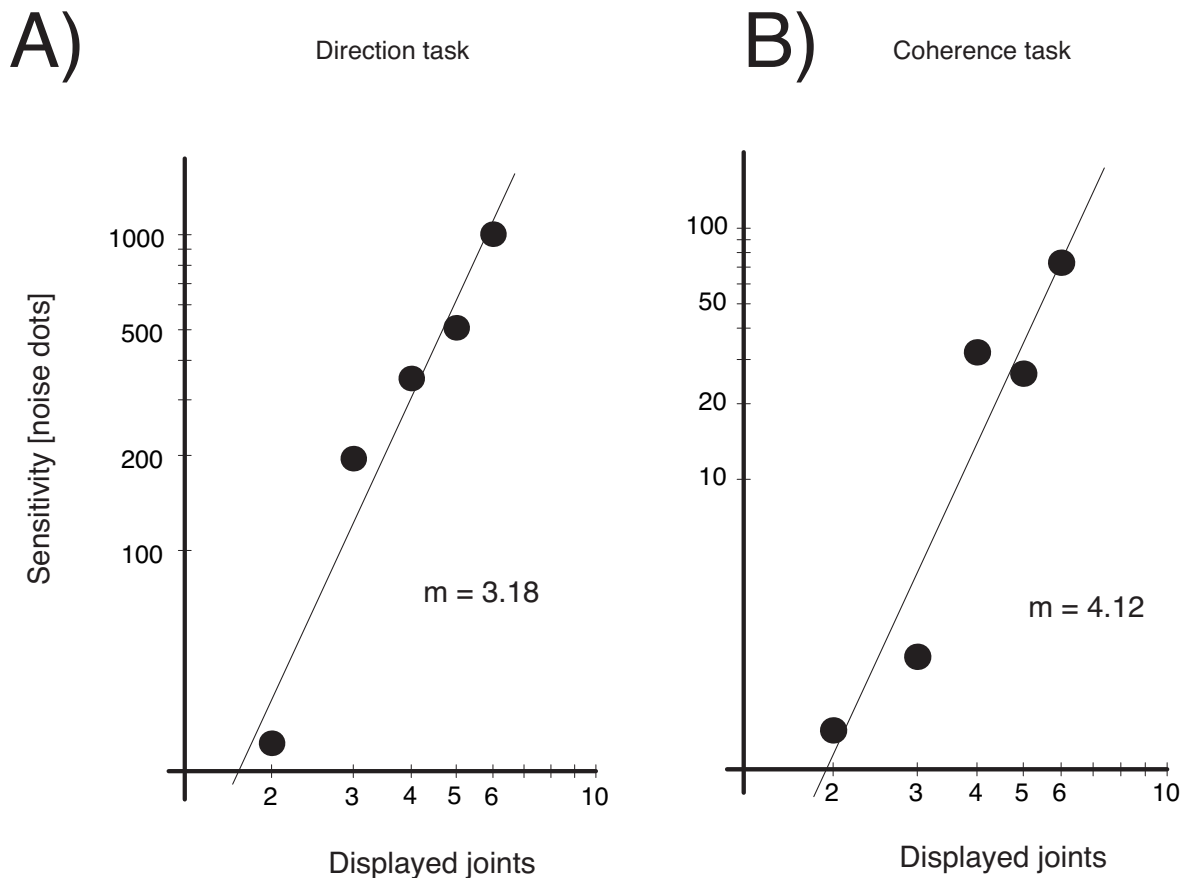
Fig. 4.7: Point-light walkers were presented embedded in a variable number of noise dots. The figure shows the model's thresholds for 75% correct discrimination between the walker's directions as a function of the number of points per stimulus frame. Results for A) Direction and B) Coherence task (see text for details on this task). m denotes the slope of the linear regression of the data.

creased if stimulus dots and noise dots revealed identical motion trajectories.Cutting et al. proposed that the observer's performance included at least two parts: a 'filtering task that ignored about 5/6 of the display area' and a second 'organizational task'. Bertenthal and Pinto (1994) showed that even in noise dots with motion trajectories identical to the motion trajectories of the stimulus dots the global structure of the stimulus is preserved. These results indicate that segmentation and solving the task do not necessarily rely on the same information. Neri et al. (1998) showed that detection threshold increased with local motion information. Our model showed that the recognition process can be explained by global form analysis. In agreement with Cutting et al. (1988) we, therefore, suggest that perception of biological motion in noise

comprises a pre-operating segmentation process and a recognition process fulfilled by template-matching. The segmentation process may be supported by form cues if the density of the stimulus dots is higher than the density of the noise dots. Also, motion signals may help to segment the stimulus from the background even when they are not needed for the recognition process itself (Beintema and Lappe, 2002).

4.4.4 Neuronal activities

Studies in humans and non-human primates suggest a specialized network for the visual perception of biological motion. This network comprises areas of the visual system(Bonda et al., 1996; Oram and Perrett, 1996; Puce et al., 1998; Grossman et al., 2000; Vaina et al., 2001; Beauchamp et al., 2002; Santi et al., 2003; Thompson et al., 2005) and the mirror-neuron system (Buccino et al., 2001; Saygin et al., 2004; Sakreida et al., 2005). The interrelations between these areas and the specific role of each area are not fully understood. Electrophysiological studies in non-human primates have found neurons in the superior temporal polysensory area (STP) selective for biological motion (Oram and Perrett, 1994; Oram and Perrett, 1996). In humans, the presumably homologue of monkey area STP, the superior temporal sulcus (STS), has been linked to biological motion in PET studies (Bonda et al., 1996) and in fMRI studies (Grossman et al., 2000; Vaina et al., 2001). STS receives input from both form and motion processing areas but the functional involvement of these connections in biological motion perception is not known. In this section we compare simulations of model cells to data from monkey area STP.

Stage 1 of the model consists of two types of cells that encode either walking to the right or to the left (Fig. 4.1). Fig. 4.8 (top, left) shows response rates of the two stage 1 cells over time after the stimulus is applied. For the one cell (grey line), the stimulus is in preferred direction. For the other cell (black line) the stimulus is in opposite direction. The activity for both types of cells shows an initial rapid increase. After about 50ms, the cell with the non-preferred direction becomes suppressed by inhibitory interactions from the cell with the preferred directions. Both cell responses settle on these respective asymptotic response levels for as long as the stimulus is present.

Fig. 4.8 (top, right) shows the responses of two cells from stage 2 of the model. In this case, the stimulus presented forward walking. One cell (grey line) was selective for forward walking. The other cell (black line) was selective for backward walking.

The cells show the same qualitative behavior as cells from stage 1 but activity in the non-preferred direction decreases more slowly than in stage 1. Also, the differences between both cell types are smaller for stage 2 than for stage 1.

Oram and Perrett (1996) recorded the responses of neurons in STPa of the macaque monkey when the monkey viewed real walking humans. In one condition, they recorded spike intensity from cells that discriminate between the directions the walking body faces (Fig. 4.8, bottom left). This task corresponds to the direction task in our model. In another condition they recorded cells while the walker was facing in the cell's preferred direction and walked either forward or backward (Fig. 4.8, bottom right). This is similar to the forward/backward task used in our model simulations.

In the direction task, the model simulations show the same behavior as the cell recordings. Both, stage 1 and stage 2, show a rapid increase of activity for the preferred stimulus. Closer examination reveals that the more rapid decrease in stage 1 matches the electrophysiological data even better than the simulations in model stage 2.

In the model the two types of cells can discriminate walking direction within 100-200 ms. This is in accordance with the results of Oram and Perrett (1996) who showed that neurons respond selectively to biological motion stimuli with specific form and orientation from 119 ms after stimulus onset.

The forward/backward task comprises both form recognition and global motion analysis. The model analyzes global motion in stage 2. Therefore, we compare electrophysiological data for this forward/backward walking only to stage 2 predictions (Fig. 4.8 (top, right). Here, too, the model shows a rapid increase as the neuronal data. Also, it shows weaker activity for the non-preferred walking direction, similar to the electrophysiological data.
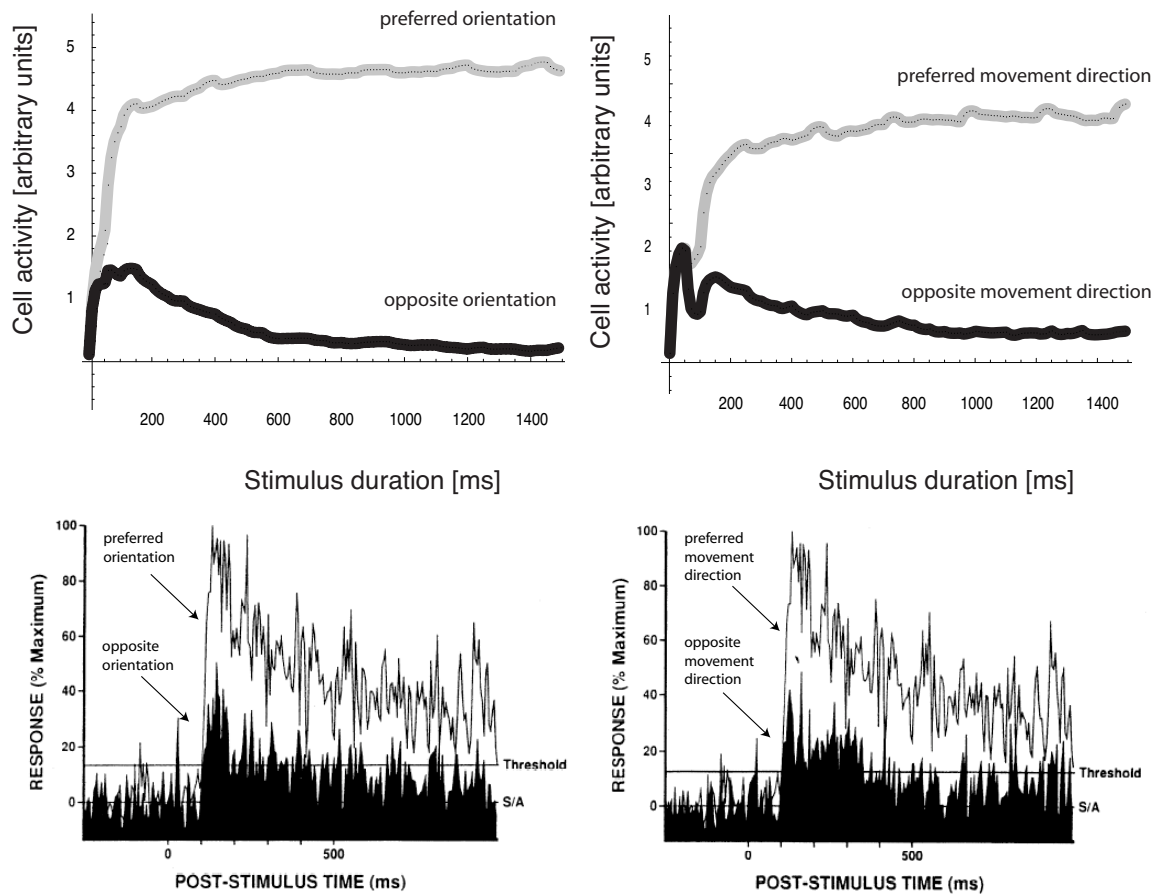
Fig. 4.8: Top left: Simulated activity of the two cells of model stage 1 that encode forward walking while the walker faces opposite orientations. The light gray line shows the cell with preference for the stimulus orientation, the black line the cell with preference for the opposite orientation. The stimulus was a leftward oriented walker with 8 points per frame. Top right: Simulated activity in model stage 2 for a stimulus facing to the left and moving forward. The light gray line shows the cell with the preference for forward movement, the black line the cell with the preference for backward movement. Both activities are simulated for 150 different trials. Bottom left: Single cell recordings from a neuron that responds selectively to a particular combination of body orientation and walking direction (e.g. a walker oriented to the left and moving forward), while a forward moving stimulus was presented either with the preferred orientation or with an orientation that was opposite to the preferred orientation. Bottom right: Single cell recordings from a neuron that responds selectively to particular combination of body orientation and walking direction. Stimuli were presented with a fixed body orientation but opposite walking directions. Figures taken with permission from Oram and Perrett (1996).

4.4.5   Functional MRI data

We are interested in the contributions of STS and form processing areas to the perception of biological motion. STS activity was found in many fMRI studies (Oram and Perrett, 1996; Puce et al., 1998; Grossman et al., 2000; Vaina et al., 2001; Beauchamp et al., 2002; Santi et al., 2003; Thompson et al., 2005). The contribution of form processing areas to biological motion perception is less clear. Because our model is based on form analysis and does not use local motion signals, we will focus in this section on contributions from form processing areas. Activation of form processing areas has been found in many fMRI studies of biological motion analysis (Vaina et al., 2001; Downing et al., 2001; Ptito et al., 2003; Beauchamp et al., 2003; Grossman and Blake, 2002; Santi et al., 2003; Michels et al., 2005; Thompson et al., 2005).Point-light walkers activate area FFA (Vaina et al., 2001; Grossman and Blake, 2002; Santi et al., 2003; Michels et al., 2005) which is believed to process form information. Point-light walkers also activate area EBA (Downing et al., 2001; Michels et al., 2005), which is activated by static images of human figures (Downing et al., 2001). However, the activation in EBA is not different from that of control stimuli that did not depict a human form (Downing et al., 2001; Grossman and Blake, 2002).

The model uses static template cells in stage 1 which are combined for temporal order analysis in stage 2. Possible neural correlates may be areas EBA or FFA, which are sensitive to static postures of human bodies (Downing et al., 2001; Peelen and Downing, 2005b) for stage 1 and area STS, which is sensitive to the global motion of a point light walker (Grossman et al., 2000; Vaina et al., 2001) for stage 2. We computed model predictions of activation levels at the two model stages for different kinds of stimuli and compared the results to experimental data reported for these areas.

Grossman and Blake (2002) recorded fMRI BOLD responses to a stimulus that consisted of 12 point-lights that depicted a human walker and they compared then to BOLD responses to a scrambled control stimulus. In the scrambled control stimulus, the 12 dots had the same motion trajectories as in the biological motion stimulus but the starting positions of the dots were randomized. Thus, the motion path of any dot is consistent with one of the walker dots while the spatial structure of the stimulus is destroyed so that it does not resemble the human form any longer. Grossman and Blake measured BOLD activity in EBA and STS as subjects viewed the biological motion or the scrambled stimulus. They found a slight and non-significant increase

of activation for biological motion over control in EBA and a strong and significant difference in STS.

We simulated the experiments with the same stimuli as Grossman and Blake and compared the outputs of the model's stages 1 and 2 to the results for EBA and STS (Fig. 4.9). We normalized model data to the maximum of the signal change in the fMRI data. The results of the model simulations are in accordance with the fMRI data. A significant difference between biological motion and control is found only in stage 2 ($t = 4.5, p < 0.01$, independent t-test). Moreover, stage 2 activity also matches quantitatively the differences in STS (Fig. 4.9).
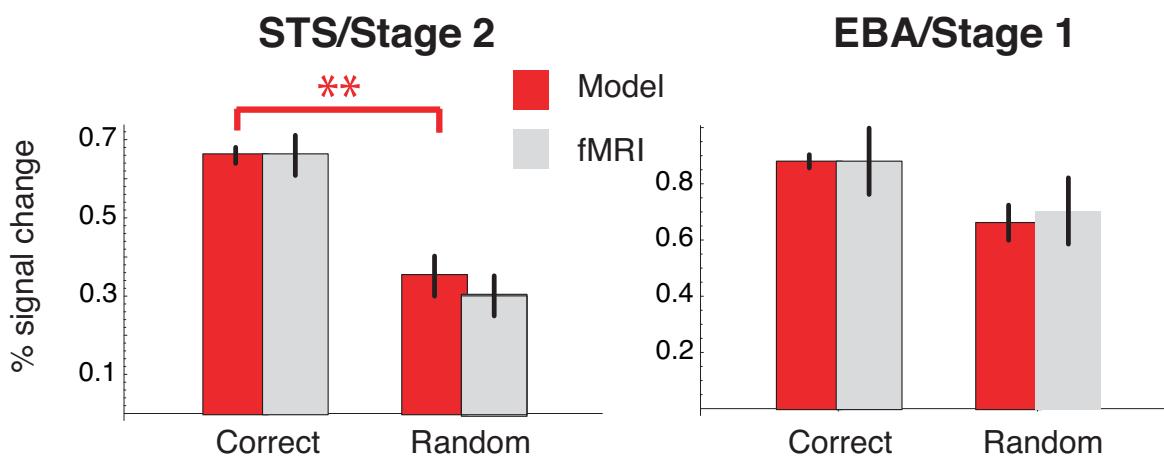

Fig. 4.9: Simulated activity in model stages 1 and 2 compared to fMRI data obtained from EBA and STS for correct and random presentation of the stimulus. Results are presented as mean ± SE. fMRI data are adapted from Grossman and Blake (2002). ∗∗ indicates significant difference at $p < 0.01$

In an earlier study, Grossman and Blake (2001) explored the orientation specificity of biological motion in the fMRI BOLD signal. It is well known that perception of biological motion is impaired for upside-down walkers (Sumi, 1984; Pavlova and Sokolov, 2000; Grossman and Blake, 2001; Grossman et al., 2005; but see Shipley, 2003). Grossman and Blake (2001) recorded neural activity in STS when subjects viewed a canonical (upright) point-light walker or an inverted (upside-down) display of this figure. They found that the response to an inverted walker was approximately half that of an upright walker. We presented both stimuli (upright and inverted) to the model and analyzed the output in stage 2 (Fig. 4.10). In accordance with the results of Grossman and Blake (2001), inverted walkers evoked significantly lower responses than upright walkers in the model (independent t-test, $t = 3.8, p < 0.01$). The reason for

this is that in our model only template cells for upright walkers exist. These template cells match an upright point-light walker more accurately than an inverted point-light walker, consistent with the common explanation for viewer-centered orientation specificity of biological motion perception (Reed et al., 2003; Troje, 2003). Yet, even the poor matches between the upside-down stimuli and the upright templates did elicit some activity in the model as they did in STS. According to the model, this residual activity is still consistent with the assumption of upright only templates. Recent transcranial magnetic stimulation (TMS) results indeed suggest that STS encodes only upright biological walkers but not inverted walkers. Applying TMS over STS impaired the perception of upright biological motion but not the perception of inverted biological motion (Grossman et al., 2005).
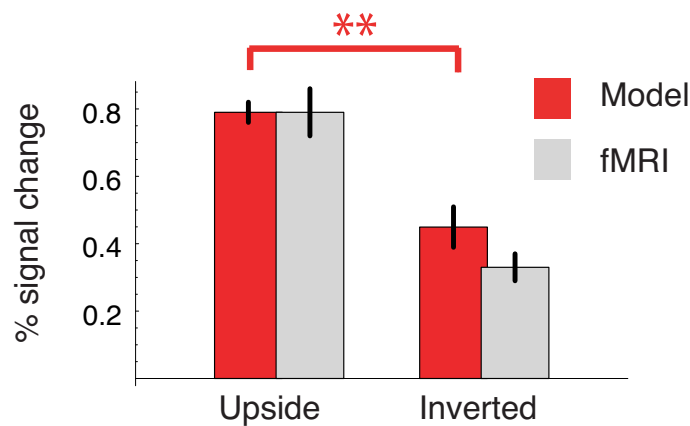
Fig. 4.10: Simulated activity in model stage 2 compared to fMRI data obtained from STS for upside and inverted presentation of the stimulus. Results are presented as mean $\pm$ SE. fMRI data are adapted from Grossman and Blake (2001). $**$ indicates significant difference at $p < 0.01$

In the experiments above, we simulated data from studies that used classical Johansson point-light walkers with light points attached to the major joints. Additionally, we compared model simulations to results of an fMRI study that used the stimulus of Beintema and Lappe and that compared walking to static stimuli (Michels et al., 2005). This fMRI study applied four different stimuli: a) a classical walker computed by the algorithm of Cutting (1978) with the dots on the stimulus' joints ('Classical Walker' moving), b) a static posture of this stimulus presented for the same duration as the moving stimulus ('Classical Walker' static), c) a walking figure with dots reallocating on the skeleton each frame ('SFL-Walker' moving, see stimulus section for details), and

d) a static posture of the stimulus used in c) with the dots changing their position on the static posture frame by frame ('SFL-Walker' static). These four stimuli were also applied in model simulation.

Fig. 4.11 shows the comparison of activity in model stage 1 to fMRI data obtained from EBA and FFA. Fig. 4.12 shows the results from model simulations in stage 2 compared to fMRI data obtained from STS.

Statistical analysis for stage 1 revealed that the model predicts increased activity for both 'SFL-Walker' conditions (moving and static) compared to both 'Classical Walker' conditions (independent t-test, Classic Walker static vs. SFL-Walker moving: $t = 3.5, p < 0.01$, Classic Walker static vs. SFL-Walker static: $t = 2.6, p < 0.05$, Classic Walker moving vs. SFL-Walker moving: $t = 4.8, p < 0.01$, Classic Walker moving vs. SFL-Walker static: $t = 3.9, p < 0.01$). Michels et al. (2005) reported the same statistical differences between the single conditions for both, EBA and FFA. Only in EBA and FFA the 'SFL-Walker' conditions evoked higher activity than for any 'Classical Walker' condition.

The comparison between model and fMRI data within each condition revealed a significantly higher activity in EBA for the 'SFL-Walker moving' condition compared to the model prediction (independent t-test, $t = 4.3, p < 0.01$). Model predictions and fMRI data differed for the 'Classical Walker static'. Here, the model predicts a disproportionate activity for the static condition compared to the fMRI data (independent t-test, Model vs. EBA: $t = 5.1, p < 0.01$, Model vs. FFA: $t = 7.6, p < 0.01$). In the 'classic Walker static' condition a single frame of the stimulus is displayed for the entire trial duration whereas in the three other conditions stimulus dot positions are refreshed every 50 ms. It is likely that fMRI activation in the 'classic Walker static' condition is lower because it is not sustained over the entire trial due to adaptation of the BOLD signal. Such adaptation is not present in the model and, therefore, the model response is larger.

The comparison of stage 2 to fMRI data revealed that among all areas the best correlation is observed between stage 2 and STS. Here, the model predicts that activity for the 'Classical Walker static' condition is significantly lower than for all other conditions (independent t-test, 'Classic Walker static' vs. 'Classic Walker moving': $t = 3.9, p < 0.01$, 'Classic Walker static' vs. 'SFL-Walker moving': $t = 3.6, p < 0.01$, 'Classic Walker static' vs. 'SFL-Walker moving': $t = 6.3, p < 0.01$). The same results were reported by Michels et al. (2005) for STS. In addition, the model shows a

significantly increased activity for the 'SFL-Walker static' condition compared to the 'Classical Walker moving' condition (independent t-test, $t = 2.2, p < 0.05$). Michels et al. (2005) observed only a trend which did not reach statistical significance ($p < 0.09$).

The comparison of model and fMRI data within each condition revealed that the model overestimates activity for the 'Classical Walker static' condition also in stage 2 (independent t-test, $t = 9.1, p < 0.01$). As specified above, this discrepancy can be explained by adaptation of the BOLD signal which is missing in the model.

In summary, the model reveals a high correlation of its stage 1 to EBA and FFA/OFA with a slightly better match for FFA/OFA than for EBA. For stage 2, we found a high correlation between model and STS.

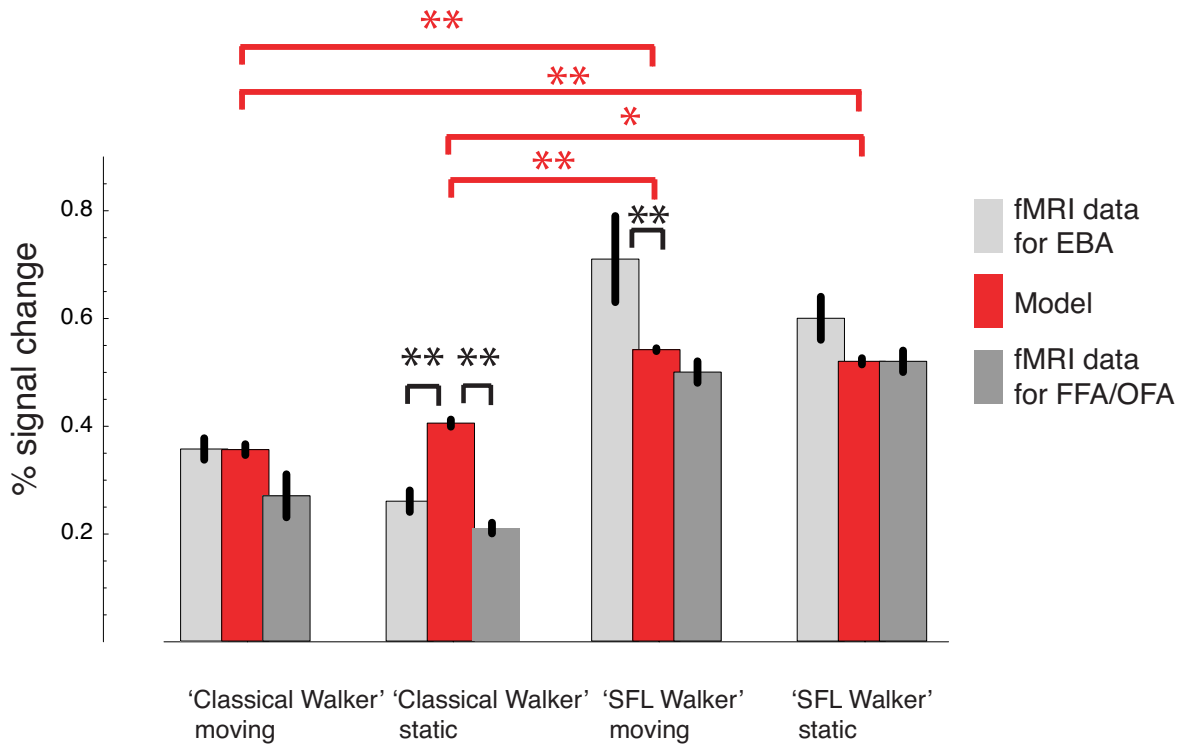Fig. 4.11: Comparison of simulated activities in model stage 1 to fMRI data obtained from area EBA and the FFA/OFA complex. fMRI data are adapted from Michels et al. (2005). Red lines and $*$ indicate differences between model simulations for the different conditions and black lines and $*$ between model simulations and fMRI data within one condition. $**$ indicates significant difference at $p < 0.01$ and $*$ at $p < 0.05$
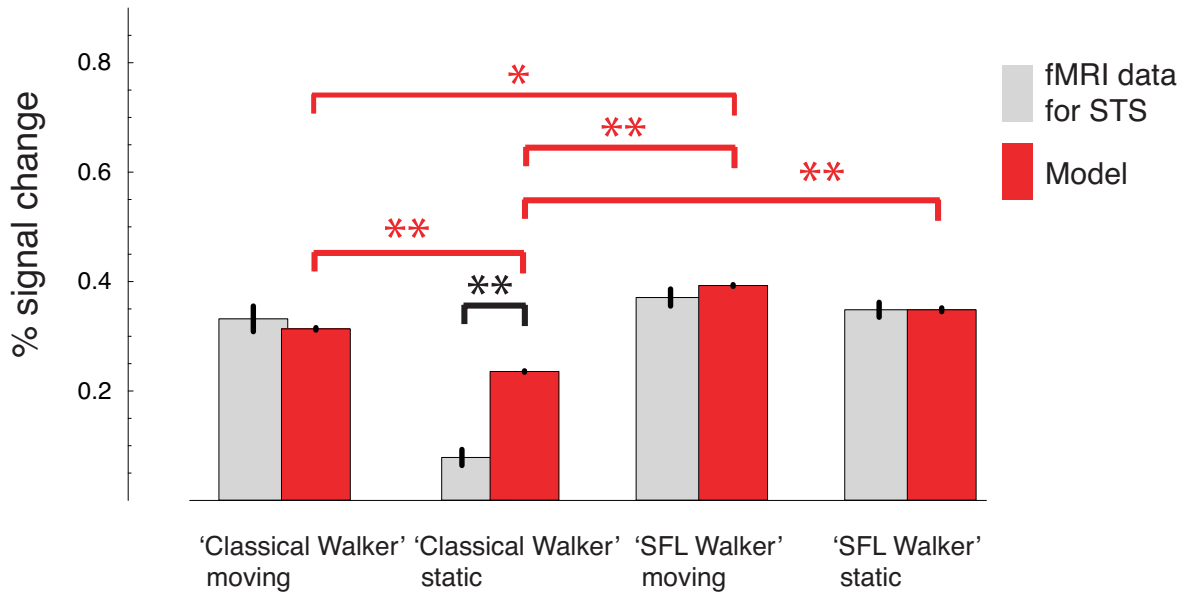
Fig. 4.12: Comparison of simulated activities in model stage 2 to fMRI data from area STS (Michels et al., 2005). Red lines and $*$ indicate differences between model simulations for the different conditions and black lines and $*$ between model simulations and fMRI data within one condition. $**$ indicates significant difference at $p < 0.01$ and $*$ at $p < 0.05$

4.5 Discussion

4.5.1 Biological motion perception from dynamic form

We developed a neurally plausible model of biological motion perception that dynamically integrates the activity of template cells of static postures of the human body. The first stage of the model analyzes only the form information in each sequential frame of the stimulus without knowledge of the temporal order. Local as well as global motion analysis is excluded. The second stage performs global motion analysis by explicitly analyzing the temporal order of the selected frames. The first stage stands for pure form analysis as in the task of direction discrimination. In experiments using this direction task we varied the contribution of form information and the influence of global motion. All data could be accurately replicated by the model solely exploiting form information. This indicates that direction discrimination tasks do not necessarily need global motion information.

In the forward/backward discrimination task, global motion analysis had to be involved because the frames and their available form information were identical. Only their temporal order differed. The model computes this global motion in stage 2 by

analyzing the frame order based on comparing the current most active template with an intrinsic expectancy. We again varied the amount of form information and the dynamics by changing stimulus duration and velocity. This expectancy combined with the form information transferred from stage 1 accounted for all data that used the forward/backward task.

Our model approach is consistent with perceptual investigations that showed that a global analysis underlies the perception of human motion from line drawings or whole-body photographs (Shiffrar and Freyd, 1990; Shiffrar et al., 1997; Chatterjee et al., 1996). These studies uniformly stressed the importance of orientation and form cues for biological motion perception. We extend these conclusions to point-light walkers. We found that the form information in a single frame is not enough information to solve biological motion tasks, but temporal integration within an appropriate time window can provide the required information. The model could also account for psychophysical experiments conduced in interfering noise (Neri et al., 1998). We regard a form-based template-matching model as a possible explanation for the differences observed for discrimination of translatory and biological motion. This supports the conclusion of Neri et al. who considered 'very sensitive, but flexible, mechanisms' as an explanation for their findings.

4.5.2 The cortical network for biological motion analysis

Our form-based model was inspired by various findings from neurological patients that suffered from the loss of motion perception but could see biological motion (Vaina et al., 1990; McLeod et al., 1996; Vaina et al., 2002). The recent surge in functional imaging studies of biological motion allows us to draw comparisons of the model to parts of the cortical network of biological motion perception. Among this network, STS is believed to be crucially involved in the perception of biological motion since it has been found in imaging (Bonda et al., 1996; Puce et al., 1998; Grossman et al., 2000; Vaina et al., 2001; Beauchamp et al., 2002; Thompson et al., 2005) and electrophysiological investigations (Oram and Perrett, 1994; Oram and Perrett, 1996), and it has been functionally implied from lesion studies (Cowey and Vaina, 2000; Vaina and Gross, 2004).

The data produced by our template-matching model predicted activity in areas sensitive to static postures of human bodies. Comparison of the model predictions to fMRI data from area EBA showed a high correlation between model and experimental

data. Moreover, the model predicted that the activation did not differ from activity for the scrambled control as reported by fMRI studies (Downing et al., 2001; Grossman and Blake, 2002). Yet, the model's neural implementation is not restricted to EBA, since Peelen and Downing (2005) also found selective activity for human bodies in the mid-fusiform gyrus. Comparison of model predictions for different types of walker stimuli to fMRI results (Michels et al., 2005) revealed high similarities to activities in EBA and the FFA/OFA complex. This supports our idea that this step of the model analysis may be implemented in form processing areas. The model simulations suggest that EBA, or other areas sensitive to static postures, can be involved in the network of biological motion perception even without showing selectivity for point-light walker vs. scrambled stimuli. Therefore, we suggest that areas like EBA or FFA are candidates for the neural implementation of model stage 1. The results imply that tasks like the direction task which does not necessarily involve global motion analysis (Beintema et al., 2005) can be solved by form analysis in this area.

We compared the model simulations of neural activity in stage 2 to fMRI studies of STS activation (Grossman and Blake, 2001; Grossman and Blake, 2002; Michels et al., 2005). The results imply that the role of STS differs essentially from EBA or other areas processing form information. While scrambled control stimuli activated EBA and model stage 1 comparably to correct walkers, STS and stage 2 are able to discriminate between scrambled and correct stimuli. From our model simulations we hypothesize that the additional global motion which is necessary to solve forward/backward tasks is processed in STS and that the impression of global motion can be derived from the dynamic change of static postures, processed in EBA or FFA. That is, in contrast to EBA, STS involves global motion analysis.

The model also predicts that the spatial structure of the stimulus has a strong influence on activity in STS. This is consistent with Thompson et al. (2005) who found that STS activation is driven by the spatial configuration of the stimulus. The hypothesis is also supported in fMRI studies by Grossman and Blake (2002) for spatially scrambled walkers and by Grossman and Blake (2001) for inverted walkers. Also, Hirai and Hiraki (2005) revealed that the amplitude of event related potentials elicited by point-light biological motion is mainly depending on the spatial structure of the walker rather than by the temporal structure of the dot movement. Temporal structures would be useful for local motion detectors whereas the spatial configuration is useful if the stimulus is mainly processed by global form analysis.

The STS also receives input from motion sensitive areas of the brain and features general motion sensitivity (Ungerleider and Desimone, 1986; Boussaoud et al., 1990). Thus, it is conceivable that low-level motion signals contribute to biological motion processing, even though our model would not seem to require them. In the literature, there is surprisingly little direct evidence that low-level motion signals contribute to biological motion perception (see Beintema et al. 2005 for a discussion). Inactivation of motion processing area MT does not interfere with biological motion perception (Grossman et al., 2005). Some studies revealed selective activation of area KO when biological motion is compared to scrambled control stimuli (Vaina et al., 2001; Santi et al., 2003). A pathway through area KO may allow residual motion perception in patients with lesions of MT (Casile and Giese, 2005). However, other studies reported that KO showed no selectivity to biological motion vs. these control stimuli (Grossman et al., 2000).

4.5.3   Other computational studies

Only a few computational studies have investigated the influence of dynamics and form information on the perception of biological motion. Troje (2002) proposed a model to identify the gender of a walking person from different viewing angles applying a principal component analysis. The results revealed that omitting information about the spatial structure of the walker by averaging over all walker stimuli corrupts performance. Yet, leaving out the dynamical component decreases recognition rates even more strongly. We report the same findings for our simpler discrimination tasks: omitting structural information leads to a decrease of performance and even to a stronger decrease if information about the dynamics is decreased.

Lee and Wong (2004) proposed a template-matching model similar to ours, but they used point-light displays as templates instead of stick figures. Their model could also account for the non-linear relationship between number of stimulus dots and number of noise dots reported by Neri et al. (1998). Our model provides an improvement as it also quantitatively replicates the psychophysical results.

Similar to our approach, Giese and Poggio (2003) assumed snapshots of human walking as the basis for a temporal order analysis. These snapshots are presumably implemented in STS and selectively activated by different human motion patterns. The snapshot neurons get input from motion processing pathways via areas MT and

KO and also from form processing areas. In both cases, however, the information is extracted from the stimulus in a local-to-global bottom-up fashion. For the case of point light stimuli, Giese and Poggio proposed that only the motion analyzing areas are able to lead to the reconstruction because the stimulus is devoid of local form cues such as local orientation. Our model shows that the template matching can be achieved by appropriate form analysis in a top-down process. A template-matching method can also explain why areas like EBA and FFA do not show selectivity for point-light walkers vs. their scrambled versions.

4.6 Conclusion

The model described in this chapter used the template-matching method of chapter 3 but modulated the model in a neurally plausible way. For this, we applied the commonly used leaky-integrator approach. I was able to integrate the dynamic component in the form-based model for the perception of biological motion. The results show that the dynamic template-matching model covers results from psychophysical studies. Furthermore, the model is able to simulate neural activity and, thus, simulate data from neurophysiological studies.

However, it is still difficult to compare model results to other studies. The reason for this is to be found in the experimental set-up. Most studies investigated the perception of biological motion in noise. As sown in section 4.4.3 the model is in general able to handle the noise. Yet, there are still some questions left with regard to the perception in noise. For example, Neri et al. (1998) have demonstrated the differences between detection and discrimination of biological motion. But, the reason for these differences is unclear. Moreover, the role of the model and its assumptions and predictions transferred to the perception in noise remains to be investigated. These issues, we will investigate in the next chapter in detail.

Chapter 5

The perception of biological motion in noise

5.1 Abstract

The perception of a stimulus can be impaired when the stimulus is presented in the context of a masking pattern. Since the perception comprises its detection and recognition, it is unclear which of these processes is masked by the additional noise. Psychophysical studies suggested that the processes of detection and recognition may use different sources of information.

The objective of this chapter was to investigate the processes of detection and recognition for biological motion by using a detection and a discrimination task separately. Additionally, we intended to investigate the sources of information used in these processes. In doing so, we intended to gain closer insights in the perception of biological motion in noise. We presented a biological motion stimulus in noise and varied in two tasks the amount of local motion signals in the stimulus. We asked human observers to discriminate the walking direction of the stimulus. And we simulated the experiments with the form-based template-matching model of Chapter 4.

Model simulations and psychophysical data replicated data from psychophysical studies and showed that these results are independent of local motion signals. However, detection of biological motion in noise may benefit from available local motion signals.

We conclude that a template-matching process may be an adequate explanation for the results of biological motion perception in noise. We suggest two distinct processes for visual analysis of a stimulus in noise: one mechanism for detection and segregation

of the stimulus from the background, which may benefit from local motion signals. Subsequently, an analysis of the stimulus' walking direction, which can be explained by a template-matching process.

5.2 Introduction

Early studies on biological motion perception have shown that humans can easily recognize actions portrayed by point-light displays (Johansson, 1973). Since then several studies have investigated the perception of biological motion when it is masked. That is, a human body depicted by only a handful dots is presented in a cloud of distracting noise dots which make it harder or nearly impossible to perceive the stimulus any more. However, the findings are not unambiguous. Cutting et al. (1988) examined the performance of human observers when the stimulus was presented in different kinds of noise patterns. The results revealed that the most effective masking is when the noise dots possessed the same trajectories as the stimulus dots but had randomly chosen positions on the screen, which no longer represent a human figure. Cutting et al. concluded that local parts like the limbs are masked by this noise pattern and, thus, the perceptual organization of the moving limbs is impeded. Bertenthal and Pinto (1994) investigated the perception of biological motion in noise with similar masking elements. They found that recognition and discrimination of point-light walkers are made more difficult by masking, but recognition and discrimination rates remain significantly above chance for upright but not for upside-down walkers. In contrast to Cutting et al. (1988), they conclude that global form analyses which is not significantly disturbed by moving distractors accounts for the perception of biological motion even in noise. Yet, Giese and Poggio (2003) claim to explain these results by a model solely exploiting local motion signals in addition with an enhancement by attention.

Neri et al. (1998) as well as Beintema and Lappe (2002) also investigated the perception of biological motion in noise. Both used stimuli with a 'limited lifetime' technique. The stimulus used by Neri et al. consisted of dots containing a two-frame apparent motion signal. They found that an increase of stimulus dots resulted in a linear increase of detecting the stimulus but in a non-linear increase of discriminating the walking direction. From this, they concluded that the perception of biological motion essentially involves analysis of local motion signals and that the detectors for biological motion are flexible but sensitive, adjusting to the conditions optimally. Beintema and

Lappe (2002) used a stimulus in which they were able to vary the amount of local motion signals by changing the lifetime of the dots. They found that without noise an increment of local motion signals resulted in a slight decrease of correct discrimination of walking direction. On the other hand, when presented in random noise they found a clear increase of the threshold for discrimination of walking direction for increased lifetime of the dots. They concluded that the role of local motion signals is to segregate the stimulus from the noise but that it is not essential for biological motion perception per se. Some clinical case studies support this interpretation. Vaina et al. (1990) reported a stroke patient with lesions in motion processing areas who had serious problems associated with motion perception. However, he had no problems to perceive a biological motion stimulus unless the stimulus was presented in noise. The study of McLeod et al. (1996) reports similar results. Also, Schenk and Zihl (1997) reported about patients with brain damage due to strokes. Two of these patients had no problems in solving motion or biological motion tasks. But, if the biological motion stimulus was presented in random noise subjects could no longer recognize it. These clinical case studies indicate that the mechanisms for perceiving biological motion in noise differ from the mechanisms of perceiving the stimulus without noise.

While the previous chapters have examined the perception of biological motion without noise, the aim of this study is to investigate the perception of biological motion embedded in noise. In particular, we want to investigate the contribution and role of local motion and global form signals and the differences between the two tasks (with and without noise). Especially the two controversial hypothesis of Beintema and Lappe (2002) and Neri et al. (1998) about the role of local motion signals should be investigated.

5.3 Methods

5.3.1 Stimulus and noise

We used the same stimulus as described by Beintema and Lappe (2002). It was a computer generated stimulus depicting the walking movements of a human body (Cutting, 1978). The dots' positions were chosen randomly on the stimulus' limbs. The dots stayed on this position for a fixed number of frames before choosing a new, again randomized, position on the limbs. This denotes the lifetimes of the dots which are

synonymous to the amount of local motion contribution. In the first experiments (detection experiment) we used a variation of this stimulus as a distractor. Based on the original stimulus (target stimulus) we displaced the dots randomly within the range the stimulus could theoretically cover. Thus, the distractor served as a density control with the same local motion signals (same dot lifetime) but with a different global form than the target stimulus. In a second part of this experiment we set the lifetime of the distractor dots to one, i.e. the dots rearranged every single frame independently from the one before. Noise dots were generated by giving each dot of a walker stimulus a randomly chosen offset within a window of 24 cm x 24 cm. Noise dots could have a lifetime of 1 or 2 frames. In the first case this is identical to scattering the noise dots randomly for every single frame. In the case of a lifetime of 2 frames, each dot has the same trajectory as a dot on the walker's limbs for two frames.

5.3.2 Psychophysical experiments

Five subjects participated in the experiments including the author. All subjects were experienced with biological motion stimuli but naive to the goal of the study (except the author). The subjects sat in a dark room and viewed the stimuli binocularly in a distance of 60 cm from the monitor. The experiments were presented on monitors with a size of 30 cm x 40 cm with a resolution of 1280 x 1024 pixel. Stimuli consisted of white dots (5x5 pixels) on a black background and were 8 cm x 4 cm in size. Noise dots were scattered in a window of 24 cm x 24 cm. We applied two main experiments, one focusing on the detection of biological motion stimuli, the other focusing on the discrimination of walking direction.

Experiment 1

A red fixation point was presented in the middle of the monitor. On the one side of the fixation point (3 cm left or right from fixation point) we presented a biological motion stimulus (target) and on the other side (also 3 cm from fixation point) a distractor stimulus. We shifted the positions of both, target and distractor, with a small, randomly chosen offset (0-1 cm) to avoid cues from familiar positions. Target and distractor were embedded in noise within a 24 cm x 24 cm window. For each, target and noise, we applied two conditions: a lifetime of one and a lifetime of two, resulting in four conditions. In one experiment we always set the lifetime of the distractor's dots to one,

in the other experiment the distractor dots had the same lifetime as the target dots. The number of target and distractor dots was always set to eight while the number of noise dots was a variable. Four subjects participated in this experiment. We asked them first to indicate on which side the target was present and then to discriminate the walking direction (left or right) of the target by pressing a key.

Experiment 2

Here, a stimulus was presented embedded in noise (window of 24 cm x 24 cm). The number of stimulus dots varied in blocks from 2 to 8 dots. Two subjects participated. They had to indicate the walking direction of the stimulus by pressing a key. In each block, we applied a 2-up 1-down staircase to determine the noise threshold for discriminating the walking direction correctly. Analogous to Neri et al. (1998) these thresholds were plotted in a log-log-plot and fitted by linear regression.

5.3.3   The model

We simulated the results of experiment 2 with a form-based model, described in detail in chapter 4. Here, we describe the main aspects of the model briefly. The model uses static stick-figure templates of walking persons. A single stimulus frame is matched to the templates and the distances between stimulus dot and template are calculated weighted by Gaussian response functions. We used one set of templates for walking to the left and one symmetrical set for walking to the right. Each set's maximum response of the matching procedure is forwarded to a layer which dynamically integrates both assemblies' responses. This layer provides a decision measure, which set of templates represents the stimulus frame best.

5.4   Results

The present study consists of two subsections. In the first experiment, we studied the detection of biological motion and the discrimination of the stimulus' walking direction by psychophysical means. In the second experiment, we tested hypotheses developed from the first experiment by means of psychophysical experiments and model simulations.

5.4.1   Experiment 1

This experiment is subdivided in two separate experiments: one investigating the detection of the stimulus (detection experiment), the other the human performance in discriminating the walking direction of the stimulus (discrimination experiment).

detection experiment

Neri et al. (1998) showed that the threshold for detecting a biological motion stimulus in noise increases linearly as a function of number of dots per stimulus frame. They also showed that this linear relationship holds for detecting a simple translatory dot pattern. Since in both stimuli only local motion signals increase in the same way, Neri et al. conclude that local motion signals contribute to the perception in translatory motion as well as in biological motion. On the other hand, Beintema and Lappe (2002) concluded from their experiments that the additional local motion signals only contribute to segregate the stimulus from the background.

To test both hypotheses, we conducted two experiments in which we presented both a target and a distractor simultaneously. Both tasks had in common that the global forms of target and distractor differ, but that they had equal density:

EXP1: Distractor variable

> In the first experiment, target and distractor were variable but always equal in terms of their local motion signals, that is, the dots' lifetime. Since the differences to EXP2 are to be found in the distractor characteristics, we denote this task as 'Distractor variable'.

EXP2: Distractor constant

> In the second experiment, the lifetime of the target's dots was still variable, whereas we set the lifetime of the distractor's dots to a constant lifetime of 1 frame. Therefore, we denote this task as 'Distractor constant'.

If local motion signals are an essential part of perceiving biological motion as several studies imply (e.g. Cutting, 1988; Neri et al., 1998; Giese and Poggio, 2003) the detection of the target should be widely unaffected by the kind of distractor. Targets including local motion signals should be more easily detectable. On the other hand, if

the perception is a global process of form perception and local motion signals are used for segregation as proposed by other studies (e.g. Bertenthal and Pinto, 1994; Schenk and Zihl, 1997; Beintema and Lappe, 2002) than stimuli including local motion signals should be more easily perceived if they differ from the distractor and from the noise in terms of local motion contingent (Fig. 5.1).
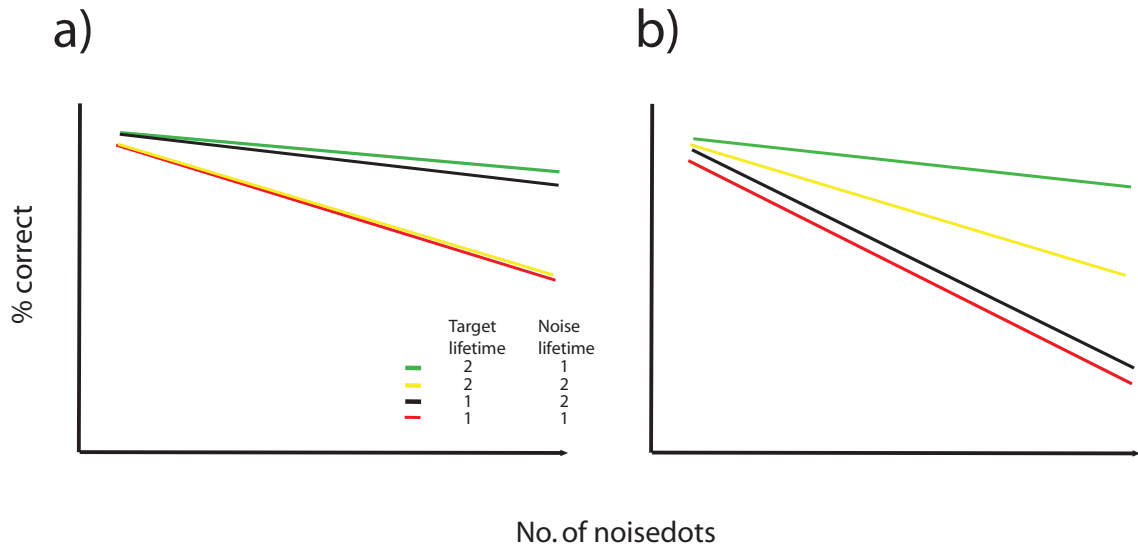

Fig. 5.1: Hypothetical results for *EXP1: Distractor variable*, if it is important that stimulus and noise differ in terms of local motion as predicted by Beintema and Lappe (2002) a) or if it is important that the stimulus contains local motion signals as stated by Neri et al. (1998) b).

Fig. 5.2 presents the data of *EXP1: Distractor variable* (target and distractor include always the same amount of local motion signals and the same amount of dots per frame, only the spatial configuration of the dots varies). Statistical analysis revealed a significant influence of the number of noise dots (two-way ANOVA with repeated measures, $F(4, 12) = 21.33, p < 0.01$) but not of condition ($F(3, 9) = 2.19, p = .16$). If we assume that conditions reveal differences only for high number of noise dots, still no significant differences could be observed. Even the condition containing local motion does not increase performance.

Fig. 5.3 shows the results when we compared conditions of *EXP 1* and *EXP 2*. Here, we only compared conditions in which the distractor had a lifetime of 1, lifetime of target and noise was variable. Statistical analysis revealed a significant influence of the number of noise dots (two-way ANOVA with repeated measures, $F(3, 9) = 16.78, p < 0.01$) but not of condition ($F(3, 9) = 0.63, p = .62$). If we assume that conditions reveal differences only for high number of noise dots, still no significant differences could be observed. Even the condition containing local motion does not increase performance.

Fig. 5.4 shows the results when we compared conditions of *EXP 1* and *EXP 2*. Here, we only compared conditions in which the distractor had a lifetime of 2, lifetime of target and noise was variable. Statistical analysis revealed a significant influence of the number of noise dots ($F(3, 9) = 17.80, p < 0.01$) but not of condition (two-way ANOVA with repeated measures, $F(3, 9) = 2.41, p = .13$). If we assume that conditions reveal differences only for high number of noise dots, a significant influence of condition could be observed for high number (>1000) of noise dots ($F(3, 9) = 4.33, p < 0.05$) but no influence of noise dots ($F(1, 3) = 2.73, p = 0.20$). Scheffé's Post-hoc test on the factor condition revealed that the subjects' performance in the condition 'target: lifetime 2, noise lifetime 1, distractor lifetime 1' was significantly higher than in the condition 'target: lifetime 2, noise lifetime 1, distractor lifetime 2'.

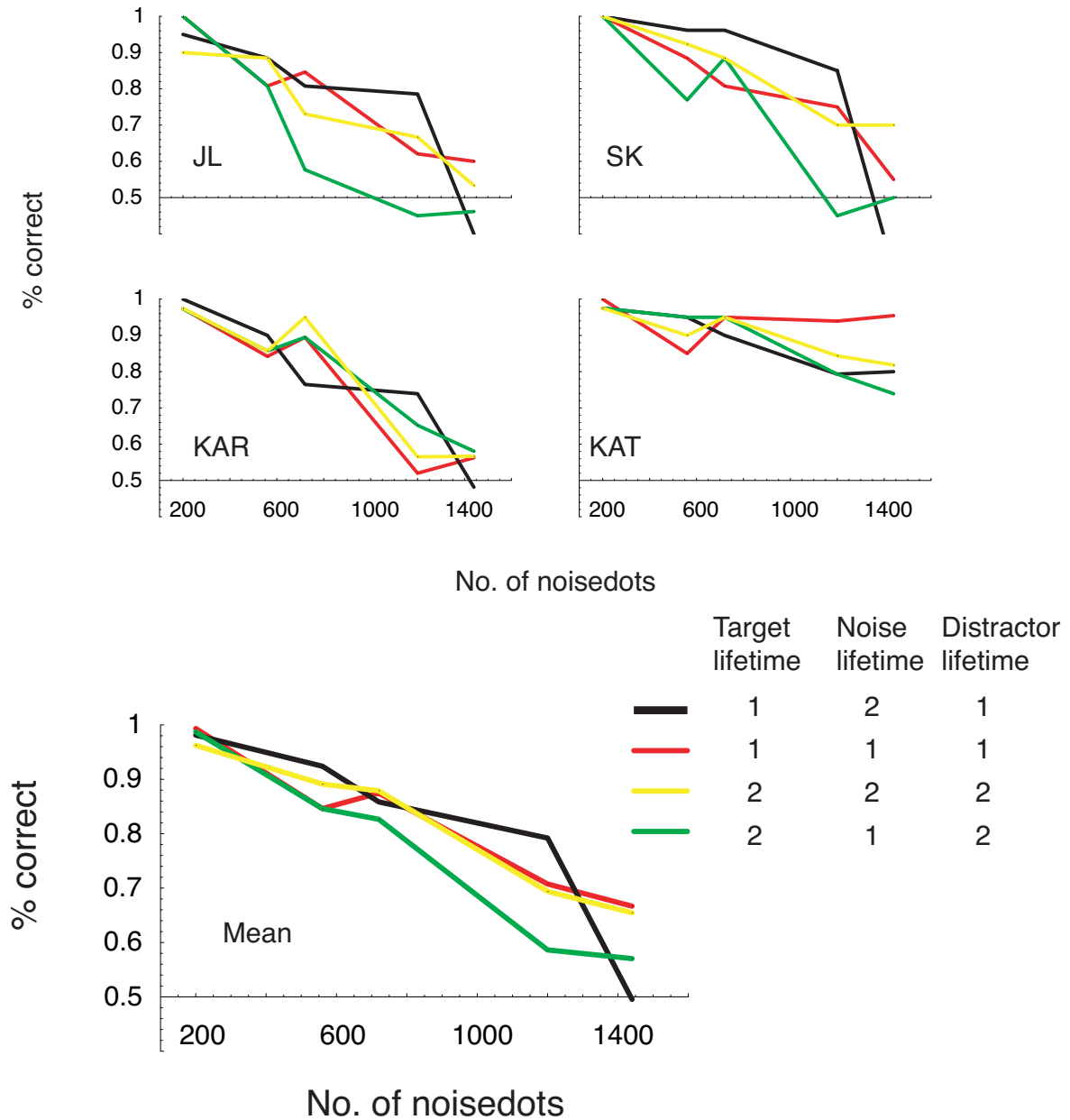# EXP1: Distractor variable

detection experiment

Fig. 5.2: Results of *EXP1: Distractor variable* in the detection task. Target and distractor have always the same lifetime. Results are presented as the fraction of correct answers as function of noise dots for 4 subjects and the mean over all subjects.

**Comparison of EXP1: Distractor variable and
EXP2: Distractor constant**

detection experiment for a target dot lifetime of 1 frame

Fig. 5.3: Comparison of the results of *EXP1* and *EXP2* in the detection task. Target and distractor have always the same lifetime of 1 frame. Results are presented as the fraction of correct answers as function of noise dots for 4 subjects and the mean over all subjects. Solid lines and dashed lines represent data with same conditions but from different experiments and are theoretically expected to be identical for same colors. Therefore, this task serves as a control.

Comparison of EXP1: Distractor variable and
EXP2: Distractor constant

detection experiment for a target dot lifetime of 2 frames

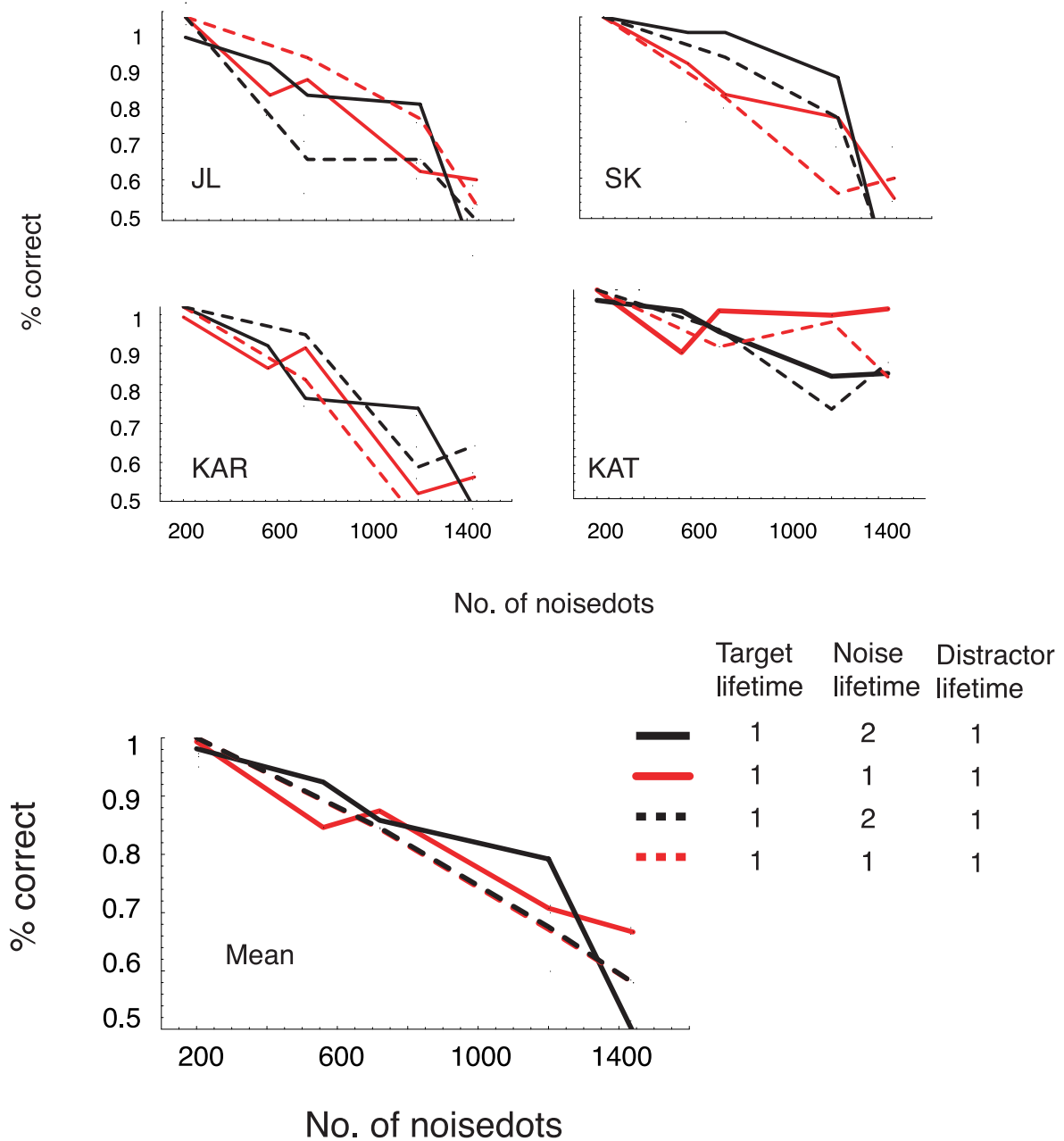Fig. 5.4: Comparison of the results of *EXP1* and *EXP2* in the detection task. Target and distractor are identical in the case of solid lines (lifetime of 2) but differ in the case of dashed lines (distractor: lifetime 1, target: lifetime 2). Results are presented as the fraction of correct answers as function of noise dots for 4 subjects and the mean over all subjects. * indicates significant differences between the two conditions for number of noise dots > 1000.

discrimination experiment

After detecting the stimulus we asked subjects to discriminate the walking direction of the walker. Fig. 5.5 presents the data of *EXP1: Distractor variable* (target and distractor include always the same amount of local motion signals and the same amount of dots per frame, only the spatial configuration of the dots varies). Statistical analysis revealed a significant influence of the number of noise dots (two-way ANOVA with repeated measures,, $F(4, 12) = 17.42, p < 0.01$) but not of condition ($F(3, 9) = 0.63, p = .61$). If we assume that conditions reveal differences only for high number of noise dots, still no significant differences could be observed. Even the condition containing local motion does not increase performance.

Fig. 5.6 shows the results when we compared conditions of *EXP 1* and *EXP 2*. Here, we only compared conditions in which the distractor had a lifetime of 1, lifetime of target and noise was variable. Statistical analysis revealed a significant influence of the number of noise dots (two-way ANOVA with repeated measures, $F(3, 9) = 16.59, p < 0.01$) but not of condition ($F(3, 9) = 0.91, p = .47$). If we assume that conditions reveal differences only for high number of noise dots, still no significant differences could be observed. Even the condition containing local motion does not increase performance.

Fig. 5.7 shows the results when we compared conditions of *EXP 1* and *EXP 2*. Here, we only compared conditions in which the distractor had a lifetime of 2, lifetime of target and noise was variable. Statistical analysis revealed a significant influence of the number of noise dots (two-way ANOVA with repeated measures, $F(3, 9) = 19.56, p < 0.01$) but not of condition ($F(3, 9) = 0.37, p = .78$). If we assume that conditions reveal differences only for high number of noise dots (similar to th detection task above), still no significant influence of condition could be observed for high number (>1000) of noise dots ($F(3, 9) = 1.37, p = 0.31$) and no influence of noise dots ($F(1, 3) = 41, p = 0.57$).

In contrast to the detection experiment subjects did not benefit from additional local motion signals in the discrimination experiment, when we compared *EXP1* and *EXP2* for a lifetime of 2 frames in the target dots (Fig. 5.7). No significant advantage for any condition could be observed in any experiment.

EXP1: Distractor variable

discrimination experiment
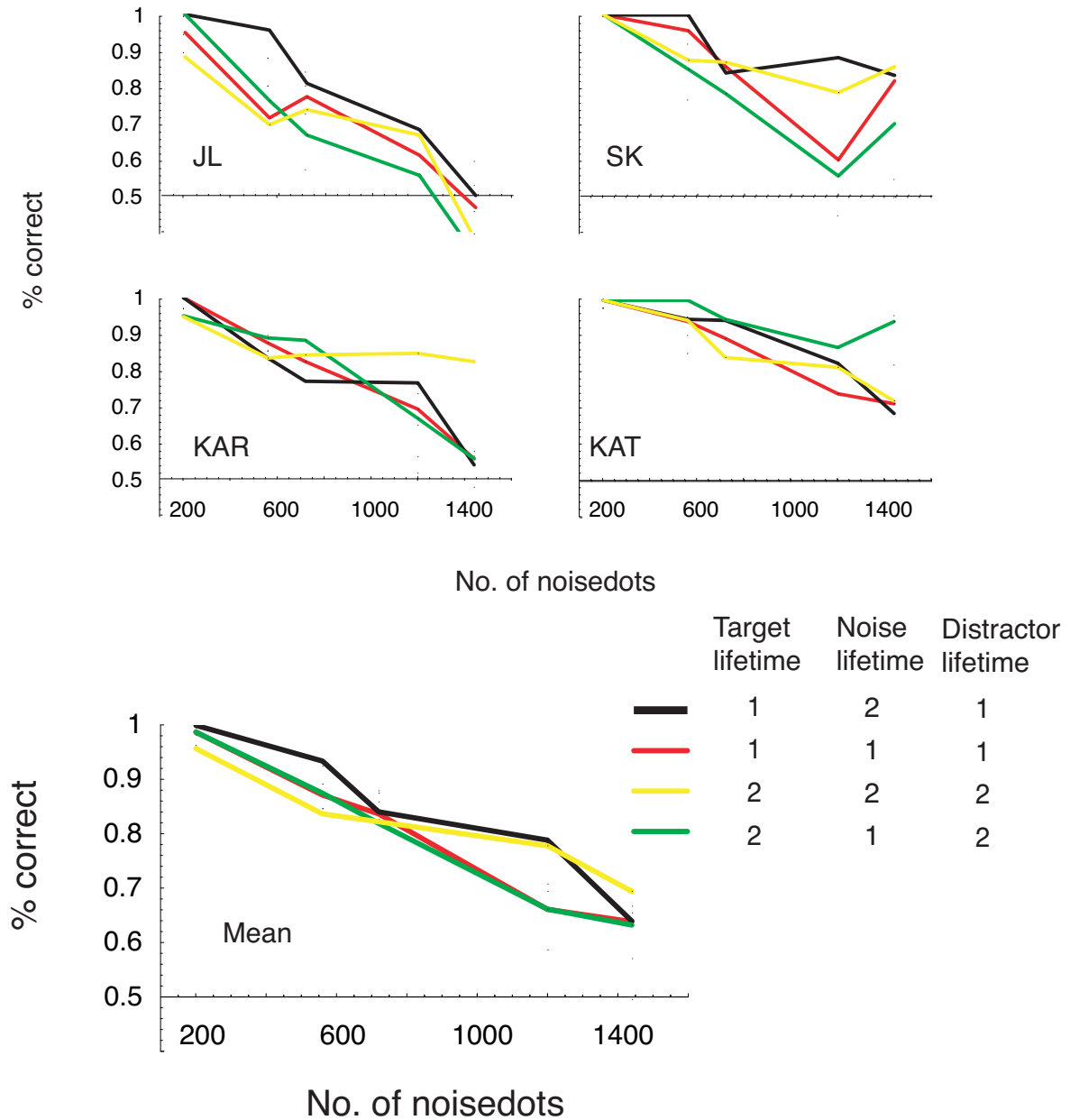
Fig. 5.5: Results of *EXP1: Distractor variable* in the discrimination experiment. Target and distractor have always the same lifetime. Results are presented as the fraction of correct answers as function of noise dots for 4 subjects and the mean over all subjects.

Comparison of EXP1: Distractor variable and EXP2: Distractor constant

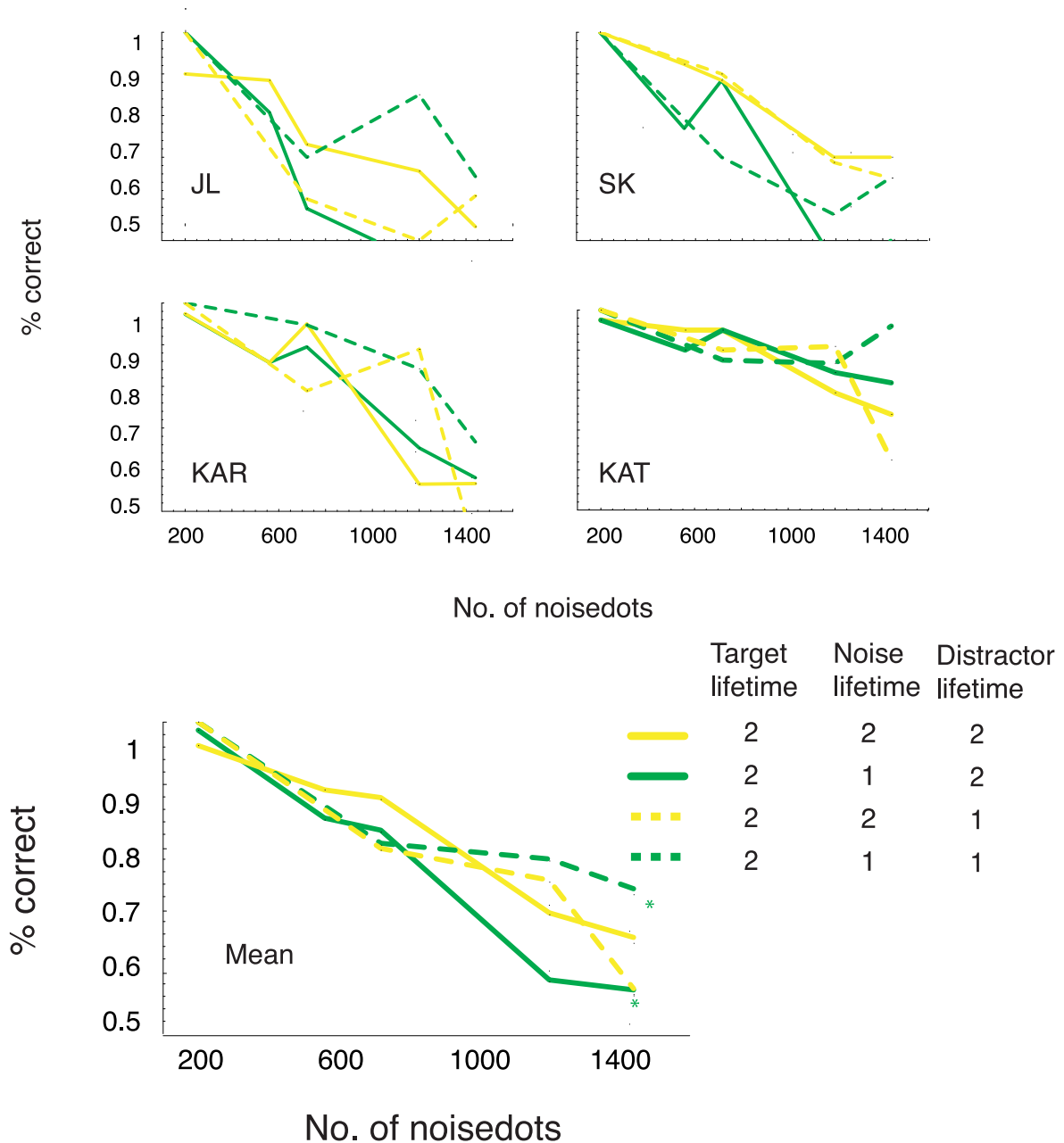discrimination experiment for a target dot lifetime of 1 frame
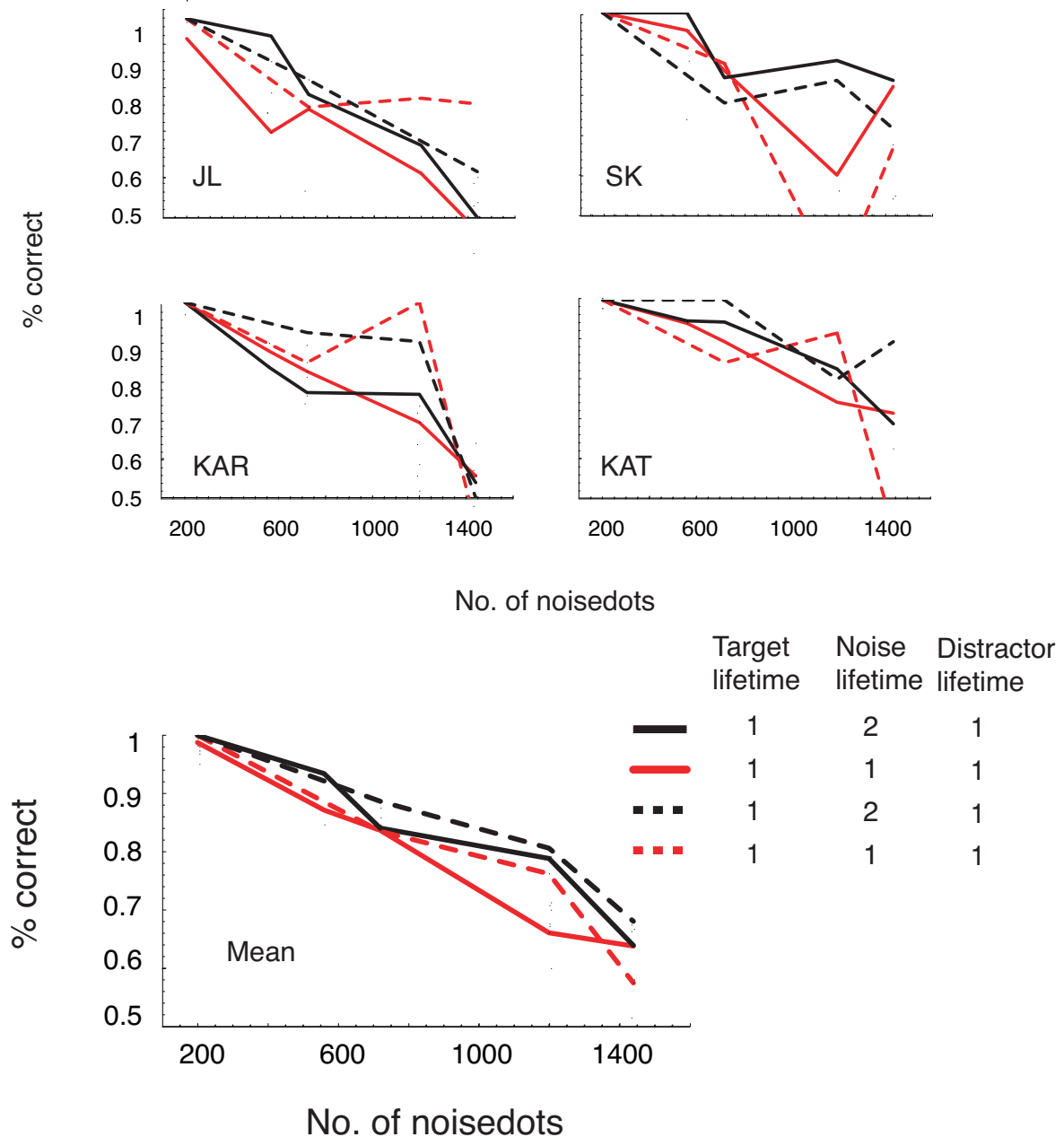
Fig. 5.6: Comparison of the results of *EXP1* and *EXP2* in the discrimination experiment. Target and distractor have always the same lifetime of 1. Results are presented as the fraction of correct answers as function of noise dots for 4 subjects and the mean over all subjects. Solid lines and dashed lines represent data with same conditions but from different experiments and are theoretically expected to be identical for same colors. Therefore, this task serves as a control.

Comparison of EXP1: Distractor variable and EXP2: Distractor constant

discrimination experiment for a target dot lifetime of 2 frames

Fig. 5.7: Comparison of the results of *EXP1* and *EXP2* in the discrimination task. Target and distractor are identical in the case of solid lines (lifetime of 2) but differ in the case of dashed lin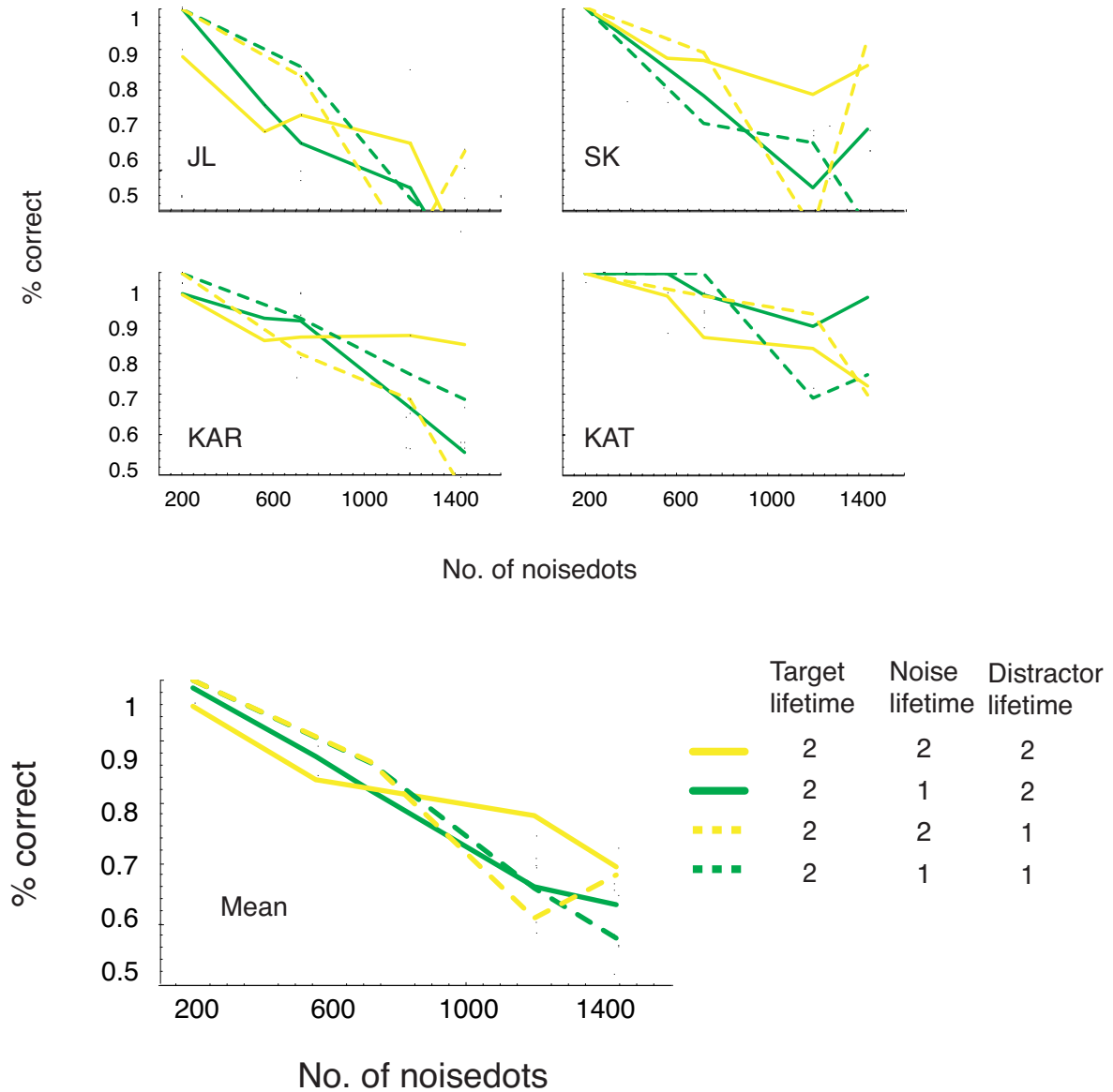es (distractor: lifetime 1, target: lifetime 2) . Results are presented as the fraction of correct answers as function of noise dots for 4 subjects and the mean over all subjects.

5.4.2   Discussion of Experiment 1

The results of the detection experiment do not strengthen the hypothesis of Neri et al. (1998) nor do they clearly show results as expected from the hypothesis of Beintema and Lappe (2002).

If subjects were to look for the target but not for the distractor certain conditions would make the target more easily detectable; that is, conditions which are typical for a biological motion stimulus but not for the scrambled distractor, e.g. local motion signals. However, no differences could be observed in the detection experiment of *EXP1* (target and distractor contain always the same amount of local motion). The reason may be that subjects do not try to detect the target stimulus in noise but rather seem to balance which stimulus is more likely to be a biological motion stimulus. And as the distractor changes the same way the target does, subjects are not able to use any additional information with increasing noise. This result shows for all conditions and, therefore, no condition gets an advantage for increasing noise. The most useful information seems to be structural information. The condition, which contains due to visible persistence the most structural information and the highest signal-to-noise ratio, reveals the highest recognition rates for a given number of noise dots. In contrast, the condition with the lowest signal-to-noise ratio shows the lowest recognitoin rates.

For the discrimination task of *EXP1*, the same behavior can be observed. Again, an influence of noise dots but not of condition could be observed.

In *EXP2* target and distractor could also differ in the amount of local motion signals. Now, subjects benefit from the additional local motion signals in the detection task. Here, they are able to discriminate between target and distractor more easily. This advantage is only obvious for high noise thresholds. Possibly, for a high number of noise dots the other detection mechanisms like structure analysis are masked by the poor signal-to-noise ratio. It is important to point out that the only condition that shows a different behavior is, when the target conditions differ from the conditions of both, distractor and noise. This is another argument that subjects do not pay attention to the target alone, but search for differences between target and distractor. If the additional information (local motion) is also masked by noise, this advantage vanishes.

After detecting the correct side we asked subjects to discriminate the walking direction of the target. Here, we could not observe any differences between the conditions of *EXP1* and *EXP2*. We conclude, if the stimulus is identified correctly none of these conditions does give useful information to discriminate the walking direction. Even

the condition of additional local motion information in the target, which improves the ability to detect the stimulus, does not help to discriminate the walking direction. This also argues in favor of the hypothesis that subjects do not try to identify the target but moreover try to find differences between target and distractor.

5.4.3   Experiment 2

The results of the detection task in experiment 1 implicate that local motion signals are used to segregate the stimulus' location within the noise area if the signal-to-noise ratio is low. Yet, these factors do not influence discrimination of walking direction. Neri et al. (1998) showed that discriminating the walking direction differs from detecting biological motion. They showed that noise threshold for detecting a biological motion stimulus increases linearly with the number of stimulus dots, whereas discrimination of the walking direction does not increase linearly. Other studies (Lee and Wong, 2004; see also section 4.4.3) have shown that a template matching process can account for this process. Here, we test further implications from these studies.

We wondered what the differences between detection and discrimination are. And what are the mechanisms of both? Is there a difference between discriminating the stimulus' walking direction with and without noise?

Neri et al. (1998) determined the threshold for discriminating the walking direction of a point-light walker with a lifetime of 2 frames. They revealed a non-linear relationship between number of noise dots and number of stimulus dots, which reflects in a slope steeper than 1 if the thresholds are plotted in a log-log plot. Our model as presented in chapter 4 has replicated the results of Neri et al. (1998) (section 4.4.3). For the model this non-linear relationship between stimulus and noise dots should always be the same independent of stimulus lifetime. If discrimination is, in contrast to detection, driven by global form analysis accomplished by form-based template-matching then performance of subjects should behave the same way independently of the dots' lifetime in these experiments similar to the expected model results.

To test these hypotheses we conducted an experiment in which human observers and the model had to discriminate the walking direction of a point-light display in noise. We varied the number of dots per frame and determined the noise threshold for discriminating the walking direction to 75% correct. These thresholds were plotted in a log-log-plot and fitted by a linear function.
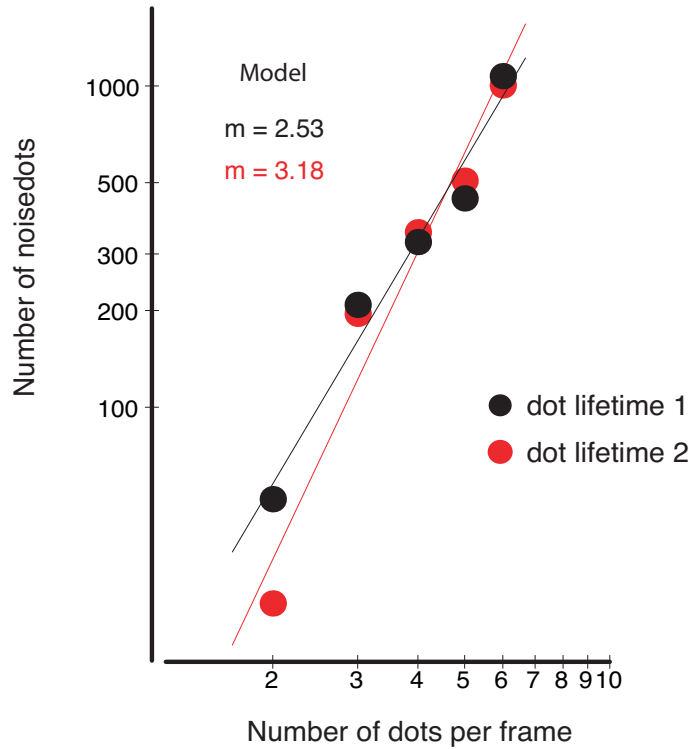
Fig. 5.8: Log-Log-Plot of the model simulations. Noise thresholds for discriminating the walking direction of the stimulus in noise plotted as a function of dots per stimulus frame. Black dots show the results for a stimulus lifetime of 1 and the red dots for a lifetime of 2. m denotes the slope of the linear fits.

The results of the model simulation are presented in Fig 5.8. Both conditions (dot lifetime of 1 and 2 frames) reveal a much steeper slope of the fit than 1 (lifetime of 1: 2.53 ± 0.28 ($r^2 = .97, t(4) = 9.13, p < 0.01$), lifetime of 2: 3.18 ± 0.42 $r^2 = .95, t(4) = 7.65, p < 0.01$). And both conditions show no significant differences ($t(8) = 0.78, p > 0.20$). We obtained very similar results in the psychophysical experiments. Both subjects reveal an increase much steeper than 1 in the condition lifetime 1 (JL: 3.52 ± 0.52 ($r^2 = .90, t(6) = 6.73, p < 0.01$), LM: 3.87 ± 0.41 ($r^2 = .95, t(6) = 9.45, p < 0.01$)) and in the condition lifetime 2 (JL: 4.32 ± 0.54 ($r^2 = .93, t(6) = 7.95, p < 0.01$), LM: 3.88 ± 0.44 ($r^2 = .94, t(4) = 8.82, p < 0.01$)). And, as predicted by the model, both conditions reveal in each subject very similar slopes that are not significantly different (JL: $t(14) = 0.79, p > 0.20$, LM: ($t(14) = 0.01, p > 0.90$))(Fig 5.9).
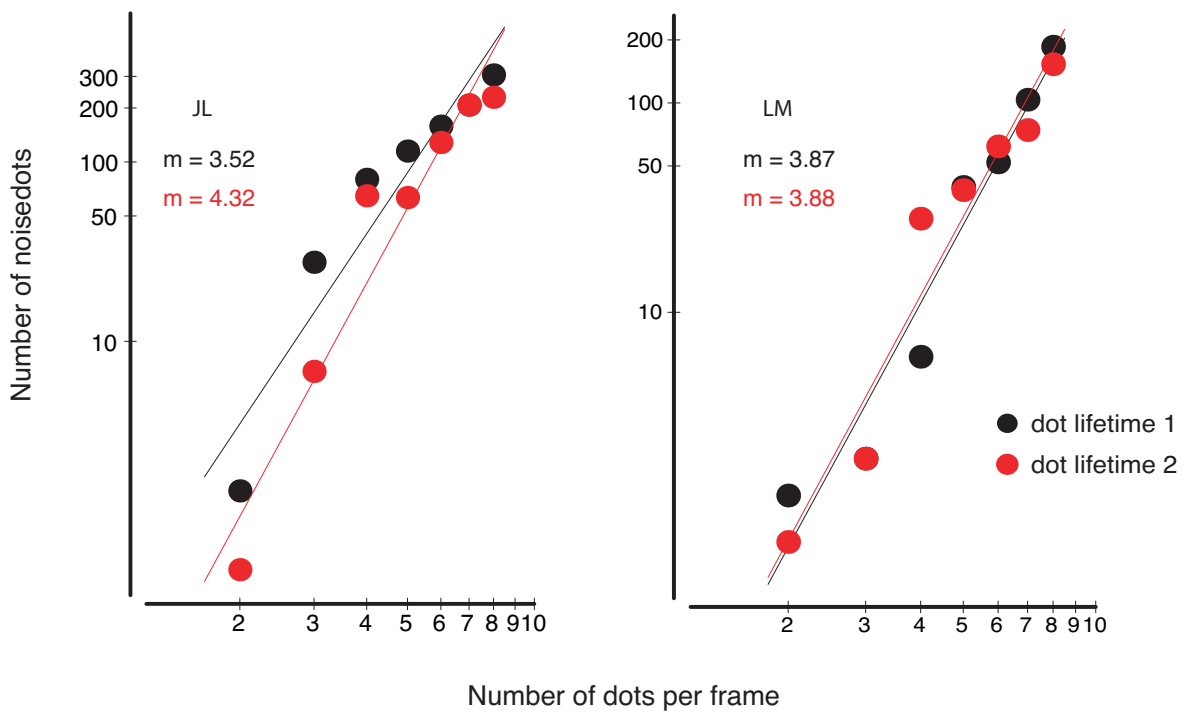
Fig. 5.9: Log-Log-Plot of the psychophysical experiments. Noise thresholds for discriminating the walking direction of the stimulus in noise for two subjects plotted as a function of dots per stimulus frame. Black dots show the results for a stimulus lifetime of 1 and the red dots for a lifetime of 2. m denotes the slope of the linear fits.

5.4.4 Discussion of Experiment 2

In their experiment Neri et al. (1998) showed that discriminating the walking direction of a biological motion stimulus in noise had a non-linear relationship between number of dots in the stimulus and number of noise dots. They observed a slope of 2.55 and 4.23, respectively, in a log-log-plot for their two subjects. Their stimulus dots contained a lifetime of 2 frames and Neri et al. argued that this local motion information, which increases with the number of dots, is the reason for their findings.

The hypothesis to test in this experiment was whether these data really depend on the dots' lifetime. Previous experiments (section 4.4.3) have shown that a form-based template-matching model can replicate the results of Neri et al. without using the local motion information. In the present experiment, we have shown that for a stimulus with a dots' lifetime of 1 frame the same non-linear relationship can be observed. The slope for dots' lifetime of 1 frame is statistically significant larger than 1 as it would be expected if there were a linear relationship between the number of dots in the stimulus and in the noise. But, no statistical significant differences could be found for the slopes of stimuli with lifetime 1 and with lifetime 2. The model's predictions that the findings of Neri et al. are independent of stimulus lifetime were tested in psychophysical experiments. The slopes of the fits for lifetime 1 and lifetime 2 are statistically not significant different for both subjects. And, they are of the same magnitude as reported by Neri et al. and predicted from the model.

Concluding, the model simulations as well as the psychophysical experiments reveal that the non-linear relationship reported by Neri et al. does not depend on local motion signals. We oberserved the same results when using a stimulus which does not contain local motion information. The model, which does not rely on local motion but form information, replicates these results. We have shown that template-matching approach can explain the results. In principle, Beintema and Lappe (2002) have already revealed that performance of human subjects in a discrimination task improves non-linearly with an increasing number of dots. Indeed, they presented their stimulus without noise and this way performance reached 100% very quickly. Due to this ceiling effect, this non-linearity is not clearly observable from their data.

5.5   Discussion

The goal of experiment 1 was to investigate the detection of a biological motion stimulus in noise by psychophysical means. Subjects had to decide on which side of a fixation point the stimulus was. In different conditions we varied the amount of local motion signals in target, distractor and noise. For a low density of noise dots the experiments revealed no advantage for any condition. Only for a high number of noise dots (> 1000) one conditioned differed strongly from the others. This was the condition in which the target contained local motion signals but distractor and noise did not. The reason may be that subjects did not tend to detect the stimulus in noise, but rather they try to detect differences between target, distractor, and noise. Evidence suggests that for high signal-to-noise thresholds subjects can differentiate target and distractor easily. Therefore, they do not detect the stimulus in any particular condition better. For high noise density this detecting mechanism (probably form analysis) is masked. Then a different mechanism must be used, if it exists. That mechanism seems to be finding differences between target dots and the other dots (distractor and noise). The results for noise dots about the threshold of 1000 dots support this hypothesis. The only condition which shows an advantage is when local motion signals are in the target but not in the others. If this result is not due to the differences but it would indicate that local motion signals are essential for detecting biological motion than this advantage of should always be obvious. Instead, only when the signal-to-noise ratio falls below a threshold, so that form information is masked by the noise, there is an advantage of local motion signals.

However, this advantage does not hold for discriminating the walking direction. Here, local motion signals do not give useful information for discriminating at any condition. Nevertheless, the results of Neri et al (1998) would implicate that local motion signals should contribute in this experiment, too.

In experiment 2, we explicitly tested the role of local motion signals in discriminating the walking direction when the stimulus is presented in noise. We used a form-based template-matching model. We were able to show that the results of Neri et al. do not rely on local motion signals, but can also be explained by template-matching. Moreover, the model predicted the same results for dot lifetimes of 1 frame or 2 frames. That means, the results should be unaffected by local motion signals. We could confirm this prediction by showing that subjects' recognition was always the same, irrespective

of the local motion contribution. We obtained in both conditions (dot lifetime of 1 or 2 frames) very similar results like that in the experiment conducted by Neri et al. This shows that discrimination of walking direction and local motion signals are uncorrelated.

All in all, we conclude that detection and discrimination of waking direction are widely unaffected by local motion signals. These local motion signals show an influence only for segregation processes if other more useful information (e.g. form recognition) is masked, e.g. by noise. This should result in a linear relationship between number of stimulus dots and number of noise dots as predicted by Barlow and Tripathy (1997) and shown by Neri et al. (1998) who studied the perception at threshold. Yet, the non-linearity for walking direction discrimination does not rely on local motion signals. We could show by a model and by psychophysical experiments that walking discrimination in noise does not benefit from local motion signals.

We suggest that detecting biological motion and discriminating its walking direction are two distinct processes which are not necessarily coupled. To detect biological motion in noise local motion signals may be important to segregate the stimulus from the noise. If the location of the stimulus is correctly detected the second process which is the real perception process relying on form information takes place. The results suggest that a template-matching process may fulfill this form analysis.

The results are in line with clinical case studies. Mcleod et al. (1996) and Schenk and Zihl (1997) independently reported patients which were able to detect biological motion unless the stimulus was presented in noise. The results of this chapter suggest that in these patients the stage necessary to segregate the stimulus from the noise is impaired. Since all of these patients had difficulty in perceiving common motion signals we believe that the damage of motion processing areas like MT prevented them from segregating the stimulus from the noise. If the stimulus were shown without noise this step was unnecessary and the still intact form analysis could solve the task.

5.6 Conclusion

In this chapter I investigated the perception of biological motion when the stimulus is embedded in noise. The results of the psychophysical study could not confirm the hypothesis that biological motion perception is driven by two distinct processes of segmentation and recognition. These results can mainly be ascribed to difficulties in the

experimental setup. The task put the subjects in a position to use other information than would be necessary to draw conclusions about the exact perceptual process. Anyway, the results show a trend that detection of a biological motion stimulus may benefit from local motion signals. On the other hand, the experiments replicated the findings from studies investigating biological motion perception without noise: local motion signals do not provide useful information to discriminate the walking direction of the walker. The results of the simulation study revealed that the dynamic model is able to reproduce psychophysical data even in a noisy surrounding. Therefore, this chapter provides evidence that discrimination of walking direction is unaffected by local motion signals. The role of a potential pre-operating segmentation process remains to be investigated.

Chapter 6

General discussion

Humans and other species interact with their external world. For this, it is important to perceive and analyze the environment in order to act or react on certain situations. Especially the interaction with other human individuals has a major role in our everyday lives. Persons and their actions involve rigid and non-rigid elements, which are constantly changing. Therefore, these processes are much more complex than perceiving other rigid objects but, in contrast, they are easily and accurately recognized. Even if the information about the human individual is highly degraded, we still have no problems to perceive the person and interpret the actions.

In the classical point-light walkers, structural information about the body is reduced to a minimum. The visible points only provide local position information, but the connections between them are absent and can only be virtually reconstructed when the walker moves. This local motion information can additionally be reduced, but subjects can still be recognize the walker as good as the classical one (Beintema and Lappe, 2002).

Stimuli consisting of point-light walker are suited to investigate the influence of particular features like structural information, local position signals and local motion signals because the other features can be excluded or at least their influence can be minimized.

The objective of my study was to investigate the role of global form information. I applied a computational model to examine theoretically the influence of form in the absence of other information. In psychophysical experiments I tried to exclude or manipulate the contribution of local and global motion signals. Comparison of model and psychophysical results should give further insights how the brain might accomplish

to perceive biological motion.

The computational model formed the main part of my work. I developed a template-matching model, which used form information exclusively and ignored local motion signals. In this chapter, I will discuss the basis of the main assumptions made in the model. Further on, I will discuss the results of model and psychophysical experiments in an overall context and I will discuss the implications of the results. Concluding, I will suggest a hypothesis of how the brain might cope the perception of biological motion.

6.1 Are the model's assumptions reliable?

As described in section 1.6, hypothetically, there are several ways to computationally implement a model of biological motion perception. Giese and Poggio (2003) developed a model, which used either motion or form information. Their approach to implement form recognition was simply based on connecting nearest dots to lines. They compared these stick figures to pre-learned templates. Normally, the lines failed to match the templates so that this approach to exploit form information was unable to recognize a walker. On the other hand, some fMRI studies have demonstrated that the form processing ventral path contributes to the recognition of biological motion (Vaina et al., 2001; Grossman and Blake, 2002; Peelen and Downing, 2005b; Michels et al., 2005).

The approach for form recognition used by Giese and Poggio is based on other models, which performed successfully in general object recognition. Therefore, the results by Giese and Poggio imply that our brains process visual information about biological motion differently than information about common rigid objects. Also, Sinha and Poggio (1996) have demonstrated the differences of biological motion perception compared to the perception of rigid objects. They showed that the same set of dots is interpreted differently depending on how the dots were connected by lines. In one experiment, the dots were connected in a way that showed a stick figure of a human person. If this pattern was rotated around its vertical axes, human observers perceived the display as a walking human figure. If the dots were connected to their next neighbor, the rotation was perceived correctly. Thus, the top-down influence in biological motion perception constrains observers perception so that they probably do not connect the dots the way suggested by Giese and Poggio (2003).

The idea of top-down processes has been also described by Bar (2003) for object

recognition. He proposed that top-down projections from the prefrontal cortex to IT are used to facilitate object recognition in the case of sparse visual information. These top-down processes may help to narrow the set of possible objects that might be perceived in a scene. Similar processes as proposed by Bar might not only help to recognize blurred objects but might also help if the visual information of the stimulus is otherwise degraded, e.g. in the case of point-light walkers.

Therefore, a top-down influenced process seems reasonable. I used a template-matching approach that fulfills the idea of a top-down process. Several studies in computer vision used different kinds of template-matching successfully. As outlined in the Introduction (see section 1.6), these models used 2-dimensional (Chen and Lee, 1992) or 3-dimensional templates (Marr and Nishihara, 1978; Hogg, 1983; Rohr, 1994). Not until recently, this template-matching process has been used exclusively in computer vision to track a human figure in an image. But this approach has not been used to explain the processing of biological motion in the human brain. Lee and Wong (2004) used a template-matching model and compared the results to psychophysical data. In fact, their model failed to explain the psychophysical data fully and quantitatively, but Lee and Wong could demonstrate that a template-matching approach is qualitatively able to explain human recognition rates in psychophysical tasks.

The studies mentioned above and their results argued in favor of a (top-down) template-matching approach to model the ventral paths contribution to biological motion perception. Yet, there are also arguments against this method. For example, the stimuli and the templates somehow have to be normalized to size to match independent of stimulus size. Anyway, even neurons in area V4 show activity indifferent to the size of the stimulus. And, as cells specifically responding to biological motion are apparently located in areas located hierarchically after V4, these neurons should be independent of stimulus size.

From the results of Chapter 3, I conclude that using form information alone can explain the data obtained from psychophysical results. Thus, Chapter 3 reveals that a template-matching approach can explain humans' behavior and is well suited to investigate the mechanisms involved in the perception of biological motion.

In chapter 4 I evolved this template-matching model. The model described in this chapter and also used in chapter 5 was based on neural assumptions. Here, the model's templates consisted of Gaussian response functions centered on the limbs of the templates. The shape of the templates' response function involved a free parameter. This

free parameter was fitted to match the human recognition rates for a single static frame. The fitting procedure revealed that recognition rates of the model increased for broadened response functions. This seemed to be contra-intuitive because broader response functions result in more templates, which are involved in the matching procedure, for the correct as well for the incorrect templates. This should lead to a more indifferent response of the model for a given stimulus frame. But, the results revealed the opposite behavior. Accuracy of the model's responses to a stimulus frame increased with broader response functions. This finding is in accordance with electrophysiological studies. Kobatake and Tanaka (1994) revealed that for higher brain areas in the ventral path the neurons' tuning curves get broader in higher brain areas. However, an ensemble of neurons in higher brain areas shows a high selectivity for objects.

Recent imaging studies have found possible neural correlates for the model's assumption of a library of stored static postures. First, Downing et al. (2001) identified an area (EBA) sensitive to static whole-body or stick figure pictures of human persons. Moreover, Peelen and Downing (2005b) also found a region located in the ventral pathway that showed a high sensitivity for static human bodies. This region was known before to be responsible for processing faces and, thus, named FFA. But, as the study of Peelen and Downing showed, this region also processes human forms and, therefore, it may represent the neural correlate in the ventral path for the layer 1 of the model used in chapter 4 and 5.

I achieved neural plausibility for the model by assuming Gaussian response functions for the templates. That is, dots farer away from the template contribute less to the perception than dots near to the template. This method has a strong influence when the stimulus is presented in noise. Without noise no differences could be observed to the results obtained with the ideal-observer model of chapter 3. In the dynamical, neurally plausible model of chapter 4 and 5, the more outside the dots are the more improbable it is that they belong to the stimulus so that this model is able to cope with noise. When the stimulus is presented in noise the results of model and psychophysical data reveal a good agreement. This indicates that even in noise a template-matching approach can account for data.

Secondly, I achieved neural plausibility by treating the templates no longer as static templates, which do not interact with each other. In Chapter 4 and 5 the static templates are combined by the common and in other studies often used approach of leaky integrators. That is, now the templates interact by inhibiting and exciting each

other. The leaky-integrator method involves a temporal parameter, which was set to a constant, known from object recognition models. This leaky integrator improved the model as it integrates the analysis of temporal characteristics in the form analysis approach. The temporal parameters in this leaky integrator model were achieved from studies on static object recognition (Hamker, 2004) and the sampling or integration window was assumed from psychophysical data (see Coltheart (1980) for a review).

6.2   Features affecting/involved in biological motion perception

Until now, the contributions of form and motion processing areas for viewing a point-light walker are not fully clear. Is the impression of a human person derived from early local motion detectors analyzing and integrating local motion information? Or does our visual system integrate local or global position signal to a general concept of a walker?

The role of these factors has been emphasized differently by psychophysical studies and fMRI studies. While some studies relied on local motion analysis (Giese and Poggio, 2003; Casile and Giese, 2005; Peuskens et al., 2005), others center the role of form information (Bertenthal and Pinto, 1994; Shiffrar et al., 1997; Beintema and Lappe, 2002).

Several fMRI studies have shown that a network of areas in the visual system is active while subjects view a biological motion stimulus. This network involves form and also motion processing areas (Beauchamp et al., 2002; Beauchamp et al., 2003; Grossman et al., 2000; Vaina et al., 2001). A critical area within this network is STS, which gets input from the dorsal as well as from the ventral path.

fMRI is a method, which measures the blood flow in the brain ventricles. This is relative slow signal (about 16 s) and, thus, it is too long to reliably resolve neural events that are separated by only a few milliseconds. Thus, fMRI is no appropriate method to investigate the interactions between the different areas.

In the following sections I will discuss the results of my studies in the context of the different potential contributions and the corresponding brain areas.

6.2.1   Local position and form information and the role of the ventral path for the perception of biological motion

Several studies on biological motion perception claim that in point-light displays like that used by Johansson (1973) the structural information are decreased to a (non-usable) minimum. Therefore these studies suggest that motion is the only useful information in these stimuli. Maybe because of this argument, only a few models have tried to explain behavioral and psychophysical data by a form-based model.

Beintema and Lappe (2002) introduced a new stimulus. This stimulus was based on a limited-lifetime technique. That is, the local motion information of the single stimulus dots could be manipulated or even reduced to near-absent minimum. In the latter case subjects may use only the sparse form information of the stimulus. The results of this psychophysical study implied that form information may be sufficient.

In chapter 3 I developed an ideal-observer-model to test quantitatively the hypothesis that form information is sufficient. I was able to show that point-light walkers still contain enough form information to solve discrimination tasks solely upon this information. The model's behavior and the good accordance to human performance in several tasks imply that humans may rely on the same information the model used.

The model presented in chapter 4 used solely form information, too, but in a neurally inspired approach. The model is divided in three layers. The theoretical assumptions and comparison of the model's results to fMRI data assign the layers 1 and 2 to form processing areas and, thus, to simulating the ventral path. The results show that layers 1 and 2 are sufficient to solve tasks like discriminating the walking direction. Thus the model indicates that pure form extraction is sufficient to explain data from many psychophysical experiments.

Beauchamp (2003) showed that point-light displays of human actions activate the ventral temporal cortex although this activation is less strong than for whole body displays. The computational model used in chapter 4 naturally predicts an increased neural activity for increased form information. My model showed that this quantitative discrepancy between whole body and point-light displays could be explained by the impoverished form information in the point-light walker. This hypothesis is additionally supported by the study of Michels et al. (2005). Michels et al. showed that activation in areas sensitive to processing static human form increased with increasing structural information. This study used different kinds of stimuli with different amount of form

information. These fMRI data were well replicated by my model simulations. This shows that the data by Michels et al. (2005) as well as Beauchamp et al. (2003) can be explained by the use of form information.

The different clinical case studies with patients suffering from brain injuries already described in detail in the introduction (section 1.5.3) can be explained using the assumptions and implications of my model. LM's disability to perceive common motion was explained by her stroke, which largely damaged MT (Zihl et al., 1983; McLeod et al., 1996). Her still available ability to perceive biological motion can be explained by my model: Her intact form processing areas of the ventral path and the intact STS used the global form information to solve this task. Therefore, my model would naturally predict these at first sight surprising abilities of LM. In another case study with LM her ability to read speech from moving faces was tested (Campbell et al., 1997). The results showed that this ability was severely impaired in LM while she still could read speech from static pictures. Although she could perceive moving biological motion stimuli she was unable to recognize the meaning of moving faces. This shows that the two mechanisms of biological motion perception and speechreading are distinct although they share the common characteristics of a global static shape that moves. Similarly, patient AF was still able to perceive biological motion despite impaired motion perception (Vaina et al., 1990). Again, my model can explain this finding: exploiting the sparse form information by the still intact ventral path enables AF to perceive the stimulus.

Some fMRI studies used static stimuli of human persons instead of a walking person (Downing et al., 2001; Peelen and Downing, 2005b). While Downing et al. (2001) claimed to have identified an area they called EBA, sensitive exclusively to biological motion, Peelen and Downing (2005) also found significantly increased activation in FFA for these stimuli. FFA is located in the ventral path and, thus, gives evidence that biological motion can activate form-processing areas. Activation of FFA was also found in the study of Vaina et al. (2001) when subjects viewed moving point-light walker. In contrast, EBA is located dorsally of MT. At some level both areas even overlap. Kourtzi and Kanwisher (2000) showed that motion defined areas like static pictures, which imply a motion of the object, can also activate MT. In that, EBA and MT bear some similarities. Both are located in the dorsal region, which is usually assigned to the perception of motion. But, both areas also show activation to static pictures of humans, which is even stronger if the pictures imply some kind of

motion. A reason for these findings may be that the recognition of a walking human person initializes a back-propagation from STS to MT to evoke a (pre-) impression of a moving stimulus (Kourtzi and Kanwisher, 2000). Also, in a TMS study, Urgesi et al. (2004) revealed that EBA is essentially involved in visual processing of a human body and its parts except the face, respectively. Stimulation of EBA by TMS significantly prolonged subjects' reaction times when they discriminated non-facial-body parts but not discriminated objects or faces.

In an fMRI study, Grossman and Blake (2002) found activation of MT as well as activation of areas located in the ventral pathway. They hypothesized that a structure-from-motion process is responsible for these results and suggested a projection from motion processing areas to form processing areas in the ventral pathway. Yet, it is not clear whether the suggested interaction between MT and FFA is necessary to explain their data. My model showed that the ventral path and FFA in particular could be activated by form information alone. This result does not need to rely on interactions to areas of the dorsal stream.

6.2.2   Local motion information and the role of the dorsal path for the perception of biological motion

I developed my model to quantitatively study the contribution of form information in the absence of local motion signals. Therefore, only indirect and, thus, limited conclusions about the role of local motion signals can be drawn.

The first models for a theoretical framework for biological motion perception relied on analysis of local motion signals (Johansson, 1973; Cutting, 1981). Although some psychophysical studies argue against these models (Sumi, 1984; Dittrich, 1993; Beintema and Lappe, 2002) the assumption of local motion signals as a critical feature is still popular. Giese and Poggio (2003) proposed a model, which integrates local motion information to complex optic flow fields. Especially the study by Casile and Giese (2005) that was based on the model of Giese and Poggio (2003) claimed to explain the results of Beintema and Lappe (2002) purely by local motion extraction. However, their model did not explain the decrease of recognition rates for human subjects when local motion signals (lifetime of the dots) increased. As expected by local motion analysis they predicted a slight increase instead of a decrease. Instead, my model replicated this decrease of recognition rates due to the effects of visible persistence.

This implies that a lower spatial sampling of the human body is responsible for the data observed and local motion signals do not improve recognition. Moreover, Casile and Giese developed a new stimulus that mimicked a walking human person by merely coarse position and motion information. Casile and Giese proposed that the motion information were similar to the motion information in the stimulus by Beintema and Lappe. In contrast, the position information was even coarser. Following their model results, Casile and Giese claimed that the more or less intact local motion signals were sufficient to explain the data by Beintema and Lappe.

In Chapter 3 I could show that even the data by Casile and Giese can be explained by form analysis. Moreover, the results showed that opposing motion vectors are not the critical feature for biological motion perception but the spatial asymmetry of the stimulus. These findings are in accordance with other psychophysical studies. Pinto and Shiffrar (1997) examined which subconfigurations of the human body are important for the recognition of biological motion. For this, they investigated the recognition rates for different stimuli. Their results showed that the stimuli could easily be recognized despite missing opposing motion vectors. Admittedly, the results showed a slight decrease compared to other body configurations. Yet, this result can be explained by the diminished structural information in these stimuli.

Some studies investigated biological motion when the stimulus is presented in randomly moving dots that are noise to the visual system. Here, increasing the contribution of local motion signals improved the perception of biological motion (Neri et al., 1998; Beintema and Lappe, 2002). In contrast to their results, which examined biological motion without noise dots, Beintema and Lappe (2002) reported a positive correlation between local motion signals and detection threshold for the biological motion stimulus. Similar results were reported by Neri et al. (1998). Neri et al. showed that biological motion and simple translatory motion reveal the same linear correlation between number of stimulus dots and detection threshold. This linear correlation disappeared when human observers had to discriminate the walking direction of the biological motion stimulus.

In chapter 4 and especially in chapter 5 I could show that the increase of detection threshold in the study of Neri et al. (1998) is not related to the increase of local motion signals but to the overall increase of form information. The role of local motion signals may be to help to detect the stimulus in noise as hypothesized by Beintema and Lappe (2002). For low noise density the stimulus can still be detected due to the denser spatial

arrangement of the stimulus. But, if the noise density increases it becomes harder to detect the stimulus solely on this basis. Motion signals in the stimulus may help to segregate the stimulus from the noisy background. If it the location of the stimulus is detected correctly, the same processes take place as without noise. In the chapters 4 and 5 I applied the template-matching model to the stimulus plus additional noise. The results revealed the same relationship between number of noise dots and number of stimulus dots as reported in the study of Neri et al. (1998). These results imply that even in noise a solely form-based template-matching approach is a reliable explanation for the psychophysical data. Therefore, I suggest a two-staged process for biological motion perception in noise: the first step to segregate the stimulus from the noise. This can be fulfilled by density or local motion cues. The second process does not differ from biological motion perception without noise: form analysis by template-matching.

In studies using fMRI the same problems occur for motion analysis as described above for form analysis. Yet, in biological motion perception a further obstacle occurs. Point-light walkers always contain motion signals. This is true for the classical walker by Cutting (1978) as well as for the modified walkers by Neri et al. (1998) and Beintema and Lappe (2002). This motion information can be useful local motion signals or just random motion signals that form flicker to the visual system. While in each case there will be activation of motion processing areas like MT, this activation is not necessarily related to usable information (Michels et al., 2005). Therefore, it is complicated to draw conclusions from fMRI studies about the involvement of motion processing areas like MT. Moreover, studies have shown activation of MT when subjects viewed static pictures. Therefore, back-propagation processes from STS to MT cannot be excluded when viewing biological motion stimuli.

Beauchamp et al. (2002) found evidence that motion signals modulate activity of STS when whole body pictures are presented. Subjects viewed photographs of human actions either static or in motion. The results showed an increased neural activity in MT and STS for the moving stimuli but not in the ventral cortex. This indicates that motion signals enhance the activity in STS. Paradoxically, the increase in STS is much stronger than in MT. This argues against an increase, which is solely due to the additional motion signals. But, these results give indications that global motion signals modulate STS activation in the presence of biological motion stimuli.

The results of Beauchamp et al. (2002) on the one hand and the findings of my study and that of Beintema and Lappe (2002) are at a first glance contradictory. Beauchamp

et al. (2002) revealed that neural activity increases with additional motion signals but Beintema and Lappe (2002) demonstrated that perceptual performance is not influenced by additional local motion signals. A reason may be that an increase of local motion signals in the stimuli of Beintema and Lappe (2002) automatically leads to a decrease of form information due to visible persistence and thus this masks the influence of local motion signals. The results of my model simulations argue against this view. They show that the decrease of recognition rates observed by Beintema and Lappe is indeed due to visible persistence. The model results show the same behavior the same behavior as the psychophysical data. This suggests that human observers behave in general like the model: they do not rely on local motion signals if form information is sufficient. Therefore, it is unlikely that the influence of local motion signals is masked in that experiments.

The role of local motion signals may be of a supporting character. That is, with local motion signals the threshold for detecting may be reached faster. This should result in a faster detection for stimuli involving additional local motion signals. But this does not lead necessarily to a more accurate perception (Beintema and Lappe, 2002).

The clinical case studies mentioned above (section 6.2.1) argue against a critical role of local motion signals. Patients with deficits in local motion perception still could perceive biological motion (Vaina et al., 1990; McLeod et al., 1996; Vaina et al., 2002). Other clinical case studies showed that biological motion perception could be impaired although general motion perception is still fairly intact (Cowey and Vaina, 2000; Batelli et al., 2003).

Biological motion has often been described as a structure-from-motion phenomenon. Another well known structure-from-motion display is random-dot display placed on the otherwise invisible surface of cylinder. If the dots start to move with the same trajectories as if they were fixed to the cylinder's surface human observers quickly recognize the rotating cylinder. The perception only occurs if the dots move coherently. These findings are similar to the perception of biological motion. Bradley et al. (1998) identified area MT and Anderson and Siegel (2005) STP, the monkey's homologue of area STS, to be critically involved in structure-from-motion perception. Vaina et al. (2002) reported a case study of a patient who suffered from a stroke. Patient BC was unable to perceive a rotating cylinder but reported without hesitation to see a walking human person in the Johansson point-light walker. This is an indication

that classically structure-from-motion and biological motion perception does not share common paths for processing. Biological motion perception does not occur to require the dorsal pathway via MT like the perception of the well-known rotating cylinder.

My studies give additional evidence for this view. The results suggest that biological motion is not perceived like structure-from-motion but more likely like motion-from-structure. The impression of motion occurs from the change of structure, thus I suggest rather calling biological motion a motion-from-structure phenomenon.

6.2.3 Global motion information and the role of STS for the perception of biological motion

The ideal-observer model described in chapter 3 is able to solve a forward/backward task by taking the frame order into account. Similar, the model described in chapter 4 interprets a walking person as a sequence of static 'objects'. Integrating this different postural information over time results in a perception of a walking human person. In both model assumptions, the impression of global motion of the walker can be derived solely by form information and its postural changes. In chapter 4 I could assign the assumption of this layer 3 due to its theoretical idea and due to the comparison to neural activity to area STS. The layer 3 of the model described in chapter 4 shows the same schema of activity for different types of static and moving walkers as demonstrated in an fMRI study (Grossman and Blake, 2002; Michels et al., 2005).

Beauchamp (2003) demonstrated that moving point-light displays activate the ventral cortex not as strong as full body movies of the same action. This difference between the animated sequences of full body and point-light display decreases in STS. Michels et al. (2005) reported similar results. Instead of a full body display they used the walker by Beintema and Lappe (2002) and compared it to the classical walker from Cutting (1978). They found similar results: The differences between the display with less form information (Cutting walker) and more form information (Beintema and Lappe) are larger in FFA than in STS. Beauchamp et al. (2003) argued that 'visual motion' is necessary to explain these results. However, my model, which does not use local motion information, shows the same results in its simulation of FFA and STS. Since the model integrates form information over time (the dynamical change of the single postures) the 'visual motion' claimed by Beauchamp et al. (2003) would be denoted 'global motion information derived from dynamic changes of postures' in my model's view.

The change of the temporal characteristics of walker has a strong influence on the perception. Especially in the forward/backward task the change of frame duration or stimulus velocity has a strong influence on human subject's behavior. The model replicates this behavior accurately. The change of these dynamical characteristics are directly linked to the perception and analysis of global form and motion. This gives evidence that global motion perception in the case of biological motion can be derived from the dynamical change of static form information. The hypotheses about the neural implementation of the model's layers and findings in fMRI studies suggest that the global motion analysis of biological motion is related to STS.

These findings bear some similarities to results about Glass-patterns. Glass-patterns are constructed by randomly placing dot-pairs, which depict a global form. There is no coherent local motion in a moving sequence of such patterns, only a permanently change of the global form. Anyway, human observers perceive global motion in these patterns. Krekelberg et al. (2003) identified cells in area STS to respond to the global perception of Glass-patterns. Moreover, they found an influence of the dynamics of the Glass-patterns and that this global percept may interact with local motion signals but does not necessarily have to. Additionally, they found that also form areas are involved in the network to process this percept. In that, their findings about this classical motion-from-structure (or as they called it: implied motion) show many parallels to the perception of biological motion and to the way my model interprets its analysis. This indicates that biological motion may potentially be processed by our visual system as proposed by Krekelberg et al. (2003) for Glass-patterns.

Patient AL in the study of Cowey and Vaina (2000) could recognize static pictures of human actions but he was unable to perceive moving pantomimes. Based on my model, the explanation for this behavior is that the ventral path for recognition of static action is still intact. STS integrates the static form information and, thus, is responsible for a global motion perception. Therefore, STS is necessary to interpret the action presented in a moving stimulus. An impairment of STS prevented AL from recognizing a biological motion stimulus. Intact ventral path but damaged STS have also been proposed as an explanation by Cowey and Vaina due to the MRI scans of the damaged brain of AL.

Besides the characteristics of STS mentioned above there are many more qualities ascribed to STS. Schultz et al. (2005) denoted STS as a 'region for processing movements characteristics that characterize living beings rather than simply responding to

the presence of living beings'. Others denote STS as an integrating area for recombining the dorsal and ventral stream of the visual system (Vaina et al., 2001; Beauchamp et al., 2002) or for combining different senses like vision and hearing (Sekiyama et al., 2003; Wright et al., 2003; Beauchamp et al., 2004). I will discuss the implications of these hypotheses on my work in the general model of biological motion perception and on future studies in the next sections.

6.2.4 Temporal information and the role of the cerebellum in the perception of biological motion

I revealed that the temporal characteristics of the stimulus have a critical influence on its perception. But, this influence appeared mainly for a forward/backward task when the impression of a walker occurs. In the direction task, there is only an influence when the overall information per trial is strongly decreased due to strongly reduced stimulus duration. That is, the temporal characteristics seem to play a role when the correct frame order in the stimulus is of importance.

There is evidence that the cerebellum contains internal models of motor actions (see Wolpert et al. (1998) for a review). These internal models may be a part of a predictive system that stores assumptions about sensory consequences and body kinematics and, therefore, monitors correspondences between the momentary and future states (Blakemore and Decety, 2001). Especially, in a PET study Sakai et al. (2002) measured brain activity during a learning task. They reported activation of cerebellum when the subjects learned finger movements with a fixed timing. If subjects learned to use a special finger in a random timing, the activation of cerebellum was missing. Moreover, there is evidence for a more general role of the cerebellum to be associated with tasks that require precise timing (see review by Ivry and Spencer (2004)).

Such a (pre-learned) timing of a sequence is used by the model in the analysis of the forward/backward. A change of this implemented knowledge of the frame order (i.e. the correct temporal order) causes a decrease of recognition rates of the model as well as in human subjects. Thus, the a priori knowledge of the correct sequence, which is used by the model, may be settled in the cerebellum.

In a clinical study, Jokisch et al. (2005) investigated the perceptual performance of patients who suffered from a cerebellar stroke. They found that the patients' ability to discriminate the direction of coherent motion in noise was impaired. In contrast, their

ability to detect a point-light walker in noise was unaffected. Jokisch et al. concluded that an intact cerebellar function is not necessary to perceive biological motion. They put forward the idea that the ventral pathway compensates for the cerebellar deficits or that cerebellar functions are even not necessary at all.

A possible explanation might also be that intact cerebellar function is not necessarily needed to detect the global structure of a human figure. The Cerebellum might only bee needed if the temporal pattern of the stimulus is important. The patients in the study of Jokisch et al. always viewed the walker in frontal view and for short time intervals (200 ms). Therefore, the temporal order gave little information and, thus, may negligible. However, in the coherent motion discrimination task, the dots moved in left or right direction on the monitor display. That is, the temporal order of the succeeding frames was important and, thus, the impaired Cerebellum affected the perceptual performance.

Some fMRI studies revealed activity of the cerebellum when a biological motion stimulus was presented (Grossman et al., 2000; Vaina et al., 2001). Yet, the interpretation differs among the authors. Grossman and Blake (2000) argued that the cerebellum's involvement in motion tasks and in planning motor activity is responsible for the activation. Also, Vaina et al. (2001) found activation of the lateral cerebellum. In both studies the stimulus was presented in sagittal view. That way, the results are in accordance with the interpretation given above for the results by Jokisch et al. (2005) because in sagittal view the temporal order is more important than in frontal. Thus, the Cerebellum might be more involved in this perceptual task. Nonetheless, Vaina et al. (2001) revealed that cerebellar activation depended on the task. If they presented an identical stimulus but now subjects had to discriminate only the direction in which the single dots move, they did not find cerebellar activation. Vaina et al. ascribed the observed activation of the Cerebellum to visual-spatial attention, which is only attended if subjects had to perceive a walker but not if they had to judge a moving direction.

6.3 A general hypothetical model for the perception of biological motion

6.3.1 Non-visual information for the perception of biological motion

In the previous section, I have discussed the influences of aspects I could investigate in my study. I discussed the possible influence of brain areas, namely of the visual system and the cerebellum. Yet, there are indications that other areas of the brain are also involved in the process of perceiving biological motion. These brain areas do not have a direct influence on the model presented in this thesis. But, they may have an indirect influence, e.g. on the assumptions underlying the model. Mainly, these areas have an influence on the general model on perception of biological motion. Therefore, I will briefly discuss studies, which reported other brain areas than the ones discussed so far.

Buccino et al. (2001) reported activation of premotor areas when subjects observed actions fulfilled with mouth, hand or foot. In another study they proposed that 'when the motor representation of the observed action is activated, the observer gains knowledge of the observed action in a 'personal' perspective, while this perspective is lacking when there is no motor activation' (Buccino et al., 2004). Saygin et al. (2004) also reported activation of premotor areas when viewing biological motion point-light displays. Neurons in monkey area F5 are known as mirror neurons (Rizzolatti et al., 1996; Rizzolatti et al., 2002). That means the neurons in the monkey homolog to human premotor cortex fire when the monkey executes an action as well as when the monkey only observes the same action. Analogous to these neurons that react on grasping movements, neurons encoding perception of biological motion may act in a similar 'mirror' fashion.

This may indicate that the visual system interacts with the premotor areas when we perceive biological motion displays. A reason for this may be to survey whether the observed scene belongs to the own motor repertoire and thus represents a (meaningful) action (Rizzolatti and Arbib, 1998; Jacobs and Shiffrar, 2005). From this, support for the model's assumption of templates arises. The interaction between the two distinct visual and motor areas may have direct influences on the templates stored in the visual system. This hypothesis gets additional support by findings that executed actions

without visual feedback have a direct influence on visual areas, namely EBA (Astafiev et al., 2004; Peelen and Downing, 2005a; Astafiev et al., 2005).

Schubotz and von Cramon (2004) proposed a different explanation for activation of premotor areas. In an fMRI study they found that the premotor cortex is not only activated by sequential biological motion stimuli. Regardless whether biological or abstract objects were presented premotor cortex was activated as long as the temporal order of the stimulus had to be processed. Thus, premotor cortex may be involved in the processing of sequentially structured events as for example a walking biological motion stimulus..

Neurons in the motorcortex M1 are usually assigned to be the last cortical step before the execution of limb actions like arm movements. Similar to the neurons in the premotor cortex (Schubotz and von Cramon, 2004), Lu and Ashe (2005) found that neurons in M1 are also involved in anticipatory planning of movements. A hypothetical idea may also represent mirror neurons, which do not only fire upon executed actions but also upon observed actions. Therefore, the knowledge these neurons contain about a future, sequential action may support the models assumption of a priori stored knowledge about the sequence of the biological motion stimulus.

Another area reported frequently in studies investigation biological motion perception is the amygdala. As a part of the limbic system the amygdala is traditionally assigned to emotional and reward-associated actions. Bonda et al. (1996) observed an increased activation in this area for body and limb movements. Ptito et al. (2003) also report significant activation of amygdala and cerebellum. Supported by the view that the amygdala is involved in the network for social perception and cognition (for an overview see Allison et al., 2000) these studies associated activation of the amygdala biological motion with the social meaning of the recognized stimulus.

In addition, there are also studies that showed that other modalities have an influence on biological motion perception. Few studies investigated the influence of auditory signals on biological motion perception. For example, Bidet-Caulet et al. (2005) showed that auditory information about a walker in the absence of visual information is able to activate areas in STS usually known to be activated by visual information alone. In other studies, it has been shown that STS comprises neurons that are activated by visual, auditory, or bimodal information (Beauchamp et al., 2004). Nonetheless, little is known about a specific correlation of the auditory system and biological motion perception. For this reason, I will not discuss the role of auditory signals in detail.

6.3.2   A model for biological motion perception

The sections before have shown that the visual perception of biological motion activates a widespread network, which involves many regions of the brain. Among these areas are predominantly parts of the visual and motor system. But also areas like cerebellum and amygdala are frequently cited. In this concluding section I will try to connect these single findings and combine them with the findings of my thesis to develop a hypothetical model of the perception of biological motion.

Several studies have assigned STS to be crucially involved in the perception of biological motion (Oram and Perrett, 1994; Bonda et al., 1996; Grossman et al., 2000; Vaina et al., 2001; Beauchamp et al., 2002). Thus, a role of STS might be the integration of form and motion which are generally believed to be processed in distinct pathways after the visual cortex' separation after V2 (Mishkin et al., 1983).

STS has also been reported to form a possible linkage between visual and motor related actions (Buccino et al., 2001; Iacoboni et al., 2001). Iacoboni et al. showed that a region located in area STS is activated by both, visual perception of a hand movement and the execution of the same movement without visual feedback. Thus, STS' role may be the part where the visual and the motor system interact to obtain a stable view of the world.

Psychophysical evidence that visual perception is mediated by the motor system has been given by Loula et al. (2005). Loula et al. asked human observers to identify the identity of an actor (self, friend, stranger) presented as a point-light display in different actions. If the discrimination was based solely on visual experience than observers should perceive their friends most easily. On the other, enhanced sensitivity to one's own actions supports the view that one's own motor experience influences the visual perception, because it is the own action that should be implemented in the motor system rather than in the visual system. Friends are more often seen and, thus, the recognizing them more easily should favor visual experience over activity related to the motor system. The results that observers can best recognize their own movements supports the hypothesis that each observer's motor system contributes to the visual analysis of human movements.

Therefore, the importance of STS may be that it forms the junction of the cortical network involved in the perception of biological motion. That is, it combines motion and form pathway as well as the motor cortex and amygdala (Allison et al., 2000).

It is well known for face perception that there are distinct networks for recognizing individuals and for recognizing emotions from faces (Adolphs, 2003). Heberlein et al. (2004) studied a group of patients who had lesions in different brain areas. From the deficits the patients showed Heberlein et al. suggested dissociation between recognition of individuals and emotions similar to face perception. Here, one area that is involved in the network for the recognition of individuals is FFA.

Based on my model studies and the comparison of its results to fMRI data, I suggest that the recognition of individuals from biological motion displays occurs in FFA or EBA. Chan et al. (2004) were unable to find evidence that EBA is involved in the recognition of individuals. Therefore, I hypothesize that FFA, similar to the perception of faces, is involved in the process of perception of biological motion. Especially for the recognition of individuals, FFA may play an important role. The close relationship of this hypothesis to what is already known from studies investigating face perception may lead to new ways to test implications of my study. The role of FFA may be tested by comparing results from face perception and biological motion perception, especially in studies with patients suffering from an inability to recognize individuals by their face (prosopagnosia). If face and biological motion perception share common pathways especially via the ventral path, these patients should also reveal impairments in specific biological motion tasks.

In combination with neurons in the premotor cortex and the somatosensoric cortex, which are also reported to be activated by viewing biological motion (Schubotz and von Cramon, 2004; Sakreida et al., 2005), the frontal cortex may hold mirror neurons for biological motion. These neurons respond to biological motion for action as well as for visual perception. Thus, action and observation of biological motion may influence each other via these mirror neuron system.

Fig. 6.1 summarizes the results discussed in this chapter in a sketch of a hypothetical network engaged in the process of perceiving biological motion.
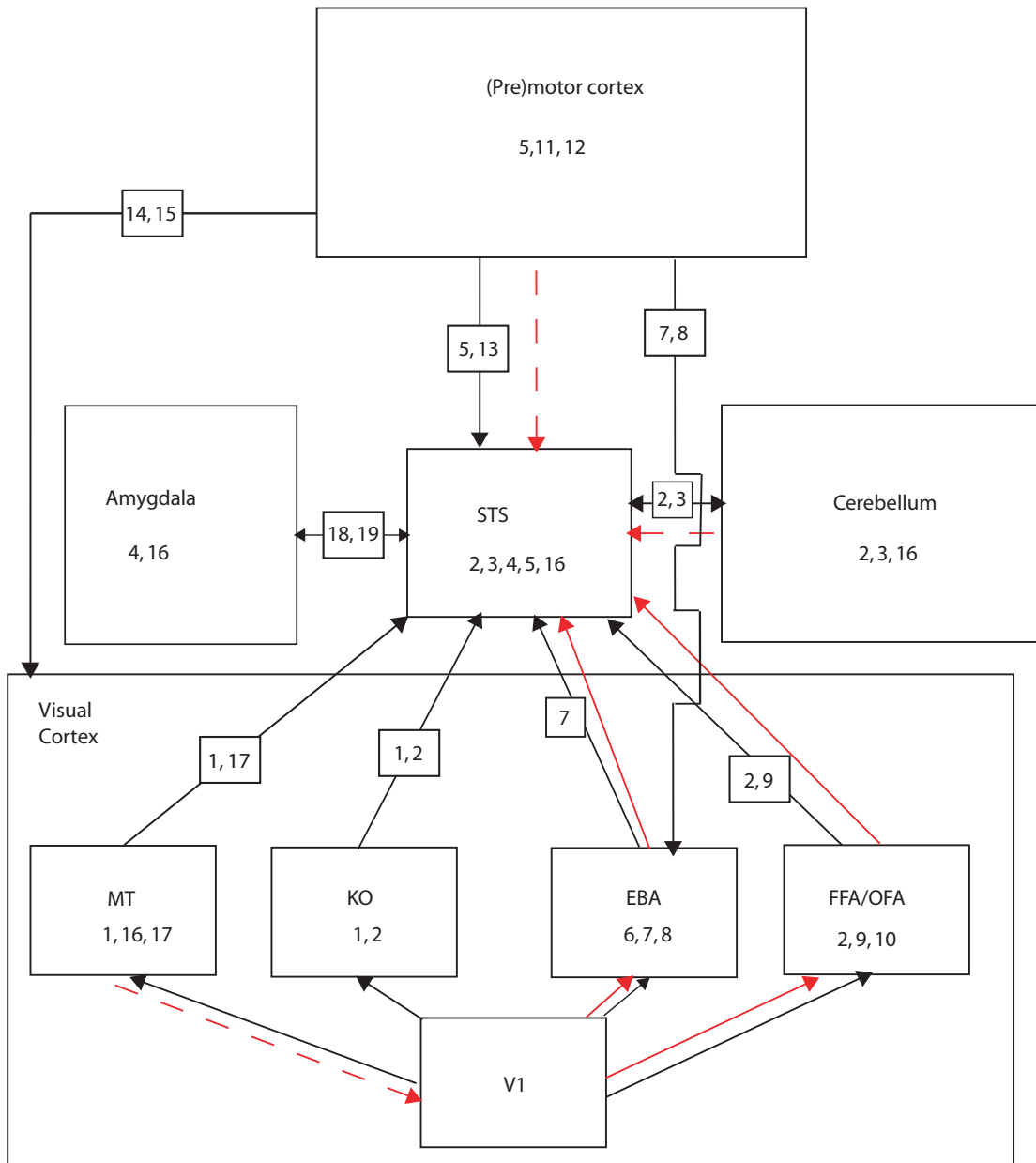
1: Giese & Poggio, 2003	6: Downing et al., 2001	11: Saygin et al., 2003	16: Ptito et al., 2003
2: Vaina et al., 2001	7: Astafiev et al., 2004	12: Buccino et al. , 2004	17: Peuskens et al., 2005
3: Grossman et al., 2000	8: Peelen and Downing, 2005a	13: Keysers and Perret, 2004	18: Allison et al., 2000
4: Bonda et al., 1996	9: Grossman and Blake, 2002	14: Loula et al., 2005	19: Adolphs, 2003
5: Buccino et al., 2001	10: Peelen and Downing, 2005b	15: Jacobs and Shiffrar, 2005	

Fig. 6.1: An illustration of hypothetical network for biological motion perception. Black boxes and numbers illustrate brain areas found to be activated by biological motion stimuli. The numbers indicate the references to these studies. Red solid lines indicate areas and connections suggested by the results presented in this thesis. Red dashed lines indicate potential connections between areas. There is no evidence for a direct connection among these areas but they potentially explain or confirm model results and interpretations. Black solid lines indicate connections among the areas suggested by other studies.

Chapter 7

Summary and conclusions

7.1 Summary

Detection of other living beings' movements is a fundamental property of the human visual system. Viewing their movements, categorizing their actions, and interpreting social behaviors like gestures constitutes a framework of our everyday lives. These observed actions are complex and differences among them are rather subtle. However, humans recognize these actions without major efforts and without being aware of the complexity of the observed tasks. In point-light walkers, the visual information about the human body is reduced to only a handful point-lights placed on the major joints of the otherwise invisible body. But even this sparse information does not effectively reduce humans' abilities to perceive the performed actions.

Neurophysiological and neuroimaging studies suggested that the movement of the human body is represented in specific brain areas. Nonetheless, the underlying network is still issue of controversial discussion. To investigate the role of form information, I developed a model and conducted psychophysical experiments using point-light walkers. My thesis comprises three main chapters. In this concluding chapter, I will summarize the results of the computational simulations and the psychophysical experiments followed by some general conclusions.

7.1.1 Template-matching approach

In Chapter 3 I investigated whether form information in point-light walkers is sufficient to explain results from psychophysical studies. I developed a model that ideally extracts

the available form information frame-by-frame and compared the results to different psychophysical studies. The results of the chapter were:

- Point-light walkers contain sufficient form information to explain discrimination tasks in a noise-free environment

- The results hold for both, a walker with virtually no local motion signals but also for the classical walker containing form and motion information. This indicates that the results do not essentially depend on the stimulus.

- Regardless whether the stimulus comprises local motion signals, all psychophysical results match with the computational simulations although the model did not use local motion signals.

- Visible persistence plays a major role as it influences the effectively visible point-lights and, thus, the available form information.

- The stimuli need to contain only rough information about a human figure.

7.1.2   Neurally plausible approach

In Chapter 4 I investigated whether the results of Chapter 3 could be transferred to a neurally plausible approach. Instead of static templates that were used in the model, here I used dynamic templates, which simulated complex dynamic receptive fields of whole human bodies. Additionally, I conducted psychophysical experiments. I compared the model simulations to these experiments and to the results from other psychophysical studies as well as to neuroimaging and neurophysiological studies. The results can be summarized as:

- All described psychophysical experiments can be explained by exploiting form information in a neurally plausible way.

- Psychophysical experiments and model simulations show a positive correlation between recognition rates and available form information.

- Even if the stimulus is embedded in noise, the perceptual performances of humans in discrimination tasks can be explained by pure form analysis.

- Recognition rates depend strongly on the dynamics of the stimulus presentation.

- The model simulations suggest dissociation between tasks that require discriminating walking direction and tasks that require to perceive a human figure.

- Discrimination of walking direction can be solved solely on the basis of form information and does not necessarily need a global perception of a walking human figure.

- Temporal parameters like the fast neural response times can be explained by the form-analyzing template-matching model.

- The different layers of the model correlate with brain areas, which are reported to be involved in the perception of biological motion.

- Static form processing areas may be essentially involved in the process of biological motion perception.

- Non-selectivity of specific areas for biological motion vs. scrambled biological motion does not mean that this area is not involved in biological motion perception.

7.1.3   Biological motion perception in noise

Several studies have shown that the perception of biological motion differs depending whether the stimulus is presented in noise or isolated. In Chapter 5 I investigated in particular the perception of biological motion when the stimulus is embedded in a distracting noisy environment. I used the neurally plausible model of Chapter 4 and psychophysical studies to gain insight into how the visual system might deal with biological motion in noise. The results were:

- The psychophysical experiments revealed that the detection of biological motion may benefit from local motion signals. However, detection only improves if the stimulus differs from noise and potential distractors in the amount of local motion signals.

- In contrast, discrimination of walking direction does not improve with additional local motion signals.

- Detection and discrimination may be separate processes fulfilled by possibly separate mechanisms.

- Linear relationships between stimulus and noise dots for detection can be explained by the use of local motion signals.

- Non-linear relationships in discrimination tasks can be explained by a form-based template-matching process.

7.2 General conclusions

I have presented a model that relied solely on form information. This approach allowed to quantitatively investigate the role of form information on the perception of biological motion. This template-matching algorithm ould be implemented in a neurally plausible model.

A widely accepted theory claims that in point-light walkers, form information is decreased to a non-usable minimum and, thus, the perception of biological motion is driven by the analysis of motion signals. In my study, I could show that point-light walker indeed contain useful form information. Moreover, I could show that temporal integration of this information is sufficient to explain results from psychophysical, neurophysiological, and neuroimaging studies. In opposition to the standard models of biological motion perception, I could also show that all results can be explained without the analysis of local motion signals.

This thesis was mainly dedicated to the analysis of point-light stimuli that depicted walking in sagittal view. Moreover, the psychophysical studies were restricted to discrimination tasks. Therefore, it can only be speculated if local motion signals are unnecessary, at all. I have tested the model with a range of different tasks and also in noise. In doing so, I could show that form information is sufficient to solve all applied tasks. The range of experiments indicates that the results are not only restricted to the tasks used. Yet, I cannot exclude that in more difficult tasks, the visual system may benefit from local motion signals.

A major problem of most template-matching algorithms is that the results and interpretations are restricted to the set of templates that were used. Because of this, I could mainly use discrimination tasks, since here the set of templates is defined. Therefore, I cannot draw direct conclusion whether the detection of biological motion relies on the same processes as the detection and recognition of the stimulus if subjects are unaware of what the point-lights depict. The discrimination tasks that used a

forward/backward task indicate that the global impression of a walker and its walking direction may also arise from template-matching processes.

The generalization of my model's results on other kinds of actions and on other task like detection and recognition in general may be interesting to explore further by modeling and testing.

7.3 Zusammenfassung

Zu den Aufgaben des menschlichen visuellen Systems gehört die Wahrnehmung der Bewegung anderer Lebewesen. In unserem alltäglichen Leben sind wir ständig damit konfrontiert, die Bewegungen von Mitmenschen wahrzunehmen, zu analysieren und Bewegungen und Gesten zu interpretieren. Menschliche Bewegungen sind in der Regel sehr komplex und Bewegungen mit unterschiedlicher sozialer Bedeutung unterscheiden sich manchmal nur in Nuancen voneinander. Trotz dieser offensichtlichen Schwierigkeiten bewältigt das menschliche Gehirn die täglichen Aufgaben scheinbar so leicht, dass uns die Komplexität der Aufgaben meist gar nicht bewusst wird.

In so genannten Lichtpunkt-Läufern ist die visuelle Information über den gesehenen Körper stark reduziert. In diesen Reizen sind nur etwa ein Dutzend Lichtpunkte sichtbar, die auf den Gelenken eines laufenden Menschen platziert sind. Trotz dieser stark reduzierten Information kann unser visuelles System die dargestellten Handlungen weiterhin schnell und nahezu problemlos erkennen.

In verschiedenen neurophysiologischen und bildgebenden Studien konnten Gehirnareale identifiziert werden, die speziell die Bewegungen eines menschlichen Körpers kodieren. Wie dieses Netzwerk von Gehirnarealen miteinander verknüpft ist, ist weiterhin Gegenstand der Forschung. In meiner Arbeit sollte speziell die Rolle von Forminformationen untersucht werden. Zu diesem Zweck entwickelte ich ein Modell und führte psychophysische Experimente mit Hilfe von Lichtpunkt-Läufern durch. In diesem abschließenden Kapitel fasse ich die Ergebnisse der Experimente und Computersimulationen der drei Hauptkapitel meiner Arbeit zusammen, gefolgt von einigen abschließenden Schlussfolgerungen.

7.3.1 Template-matching Modell

Im 3. Kapitel der Arbeit habe ich untersucht, ob die in einem Lichtpunkt-Läufer enthaltende Forminformation ausreicht, um Ergebnisse psychophysischer Studien zu erklären.

Ich habe ein Computermodell entwickelt, das statische Einzelschablonen aus dem Lauf-
zyklus einer gehenden Person benutzt. Durch Vergleich zwischen den Schablonen und
jedem Einzelbild des Stimulus wird die im Stimulus vorhandene Forminformation in
idealer Weise analysiert, d.h. ohne lokale Bwegungsinformationen zu berücksichtigen.
Die Ergebnisse dieser Simulationen wurden mit den Ergebnissen aus psychophysischen
Experimenten verglichen. Die Ergebnisse dieses Kapitels lassen sich wie folgt zusam-
menfassen:

- Lichtpunkt-Läufer enthalten genügend Forminformationen, um Ergebnisse aus
  psychophysischen Experimenten zu erklären, in denen zwischen zwei Alternativen
  in rauschfreier Umgebung unterschieden werden musste.

- Diese Ergebnisse sind unabhängig davon, ob der Lichtpunkt-Läufer lokale Bewe-
  gungssignale enthält oder nicht. Dies zeigt, dass die Ergebnisse nicht essentiell
  von der Art des präsentierten Reizes abhängen.

- Obwohl die Ergebnisse des nur auf Formanalyse beruhenden Modells unabhängig
  von lokaler Bewegungsinformation sind, zeigen Modell und psychophysische Ex-
  perimente die gleichen Übereinstimmungen, selbst wenn der Reiz lokale Bewe-
  gungssignale enthält.

- 'Visible persistence' hat einen großen Einfluss auf die Anzahl effektiv sichtbarer
  Lichtpunkte. Die Ergebnisse zeigen, dass die Anzahl der durch 'visible peristence'
  effektiv wahrgenommen Punkte entscheidender ist als die Anzahl der Lichtpunkte
  pro Einzelbild eines Reizes.

- Für die beobachteten Erkennungsraten ist eine grobe Kenntnis über die Körper-
  struktur ausreichend.

7.3.2 Neuronal plausibler Ansatz

Im 4. Kapitel meiner Arbeit habe ich untersucht, ob sich die Erkenntnisse aus Kapitel
3 auf ein neuronal plausibles Modell übertragen lassen. Anstelle der statischen Körper-
schablonen die dem Modell in Kapitel 3 zu Grunde lagen, benutze ich dynamische
Schablonen, die komplexe rezeptive Felder für Körper simulieren sollten. Des weiteren
habe ich psychophysische Experimente durchgeführt. Die Ergebnisse der Modellsimula-
tionen wurden anschließend mit den Ergebnissen aus diesen und anderen psychophysi-
schen Experimenten verglichen. Zusätzlich wurden mit Hilfe des Modells Simulationen

neurophysiologischer und bildgebender Daten durchgeführt. Die Ergebnisse lassen sich folgendermaßen zusammenfassen:

- Alle untersuchten psychophysischen Experimente lassen sich mit der neuronal plausiblen Art der Formanalyse erklären.

- Darüber hinaus zeigen Modellsimulationen und psychophysische Experimente eine hohe Korrelation zwischen Erkennungsrate und verfügbarer Forminformation.

- Die oben erwähnten Ergebnisse lassen sich mit diesem Modell auch auf Experimente ausweiten, in denen der Reiz von Rauschpunkten umgeben ist.

- Erkennungsraten hängen stark von den zeitlich-dynamischen Eigenschaften der Reizpräsentation ab.

- Anhand des Modells lassen sich Studien über biologische Bewegungswahrnehmung in zwei separate Gebiete aufteilen: Experimente, bei denen der Läufer erkannt werden muss und Experimente bei denen z.B. Bewegungsrichtungen unterschieden werden müssen.

- Ist bekannt, dass es sich bei dem Reiz um eine menschliche Figur handelt und besteht die Aufgabe dann darin, zwischen zwei Antwortmöglichkeiten zu unterscheiden, so kann diese Aufgabe anhand von Forminformationen gelöst werden. Der Eindruck eines Läufers muss dabei nicht zwangsläufig vorhanden sein.

- Die schnellen Antwortzeiten einzelner Neuronen können durch das Form analysierende Schablonen-Modell erklärt werden.

- Die unterschiedlichen Ebenen des Modells korrelieren mit Gehirnarealen, die zu dem durch biologische Bewegung aktivierten Netzwerk gehören.

- Gehirnareale, die auf Formanalyse statischer Objekte spezialisiert sind, haben entscheidenden Anteil an der Wahrnehmung biologischer Bewegung.

- Diese Areale zeigen allerdings keine Selektivität für biologische Bewegung verglichen mit bestimmten Kontrollreizen. Das dynamische Modell impliziert diese Nichtselektivität, zeigt allerdings, dass diese Areale trotzdem am Wahrnehmungsprozess von biologischer Bewegung beteiligt sind.

7.3.3   Wahrnehmung biologischer Bewegung im Rauschen

Es ist in mehreren Studien gezeigt worden, dass die Wahrnehmung biologischer Bewegung abhängig davon ist, ob der Reiz isoliert oder umgeben von Rauschpunkten präsentiert wird. Im 5. Kapitel habe ich speziell untersucht, wie biologische Bewegung in einer verrauschten Umgebung wahrgenommen wird. Hierfür habe ich das neuronal plausible Modell aus Kapitel 4 benutzt und zusätzliche Simulationen und psychophysische Experimente durchgeführt.

Die Ergebnisse dieses Kapitels sind:

- Die psychophysischen Experimente zeigten, dass die Erkennung des Läufers im Rauschen von zusätzlichen Bewegungssignalen profitiert, allerdings nur, wenn die vorhandenen Rauschpunkte und Distraktoren keine Bewegungssignale enthalten.

- Die Erkennungsraten in Aufgaben, in denen Bewegungsrichtungen unterschieden werden sollen, werden jedoch nicht von Bewegungssignalen beeinflusst.

- Die Erkennung und Unterscheidung von Bewegungsrichtungen scheinen unterschiedliche Prozesse zu sein.

- Der beobachtete lineare Zusammenhang in den Erkennungsaufgaben zwischen der Anzahl der Punkte im Läufer und der Punkte im Rauschen lässt sich durch Verarbeitung von Bewegungssignalen erklären.

- Der nicht-lineare Zusammenhang, der in den Unterscheidungsaufgaben beobachtet wird, kann durch das Form analysierende Modell erklärt werden.

7.4   Schlussbemerkungen

In dieser Arbeit habe ich ein Modell entwickelt, das seine Informationen nur aus den verfügbaren Forminformationen zieht. Dieses Modell erlaubte es, den quantitativen Einfluss von Forminformationen auf die Wahrnehmung biologischer Bewegung zu untersuchen. Des Weiteren konnte dieser Modellansatz zu einem neuronal plausiblen Modell weiterentwickelt werden.

Die allgemein akzeptierte Theorie zur Wahrnehmung biologischer Bewegung besagt, dass die Forminformation in Lichtpunkt-Läufern nicht ausreicht, um auf diese Weise den Wahrnehmungsprozess zu erklären. Aus diesem Grund wird angenommen, dass

Wahrnehmung biologischer Bewegung durch Analyse der vorhandenen Bewegungssignale erfolgt. Durch meine Arbeit konnte ich jedoch zeigen, dass die Forminformation entgegen der oft vertretenen Meinung doch ausreichend ist. Durch zeitliche Integration dieser Information können sowohl psychophysische als auch neurophysiologische Ergebnisse und Resultate aus bildgebenden Studien erklärt werden.

In dieser Arbeit habe ich mich in erster Linie mit Lichtpunkt-Läufern beschäftigt, die einen von der Seite aus betrachteten Läufer darstellten. Darüber hinaus habe ich mich mit Unterscheidungsaufgaben beschäftigt. Die Resultate können somit streng genommen nur auf diese Aufgaben und Reize angewendet werden. Ob Bewegungssignale für die Wahrnehmung biologischer Bewegung im allgemeinen deshalb unwichtig sind, lässt sich nur schwer sagen. Allerdings habe ich die Resultate in mehreren Aufgaben getestet. Da ich immer dieselben Ergebnisse erzielte, selbst wenn der Läufer im Rauschen präsentiert wurde, gibt es jedoch Anzeichen, dass sich die Ergebnisse verallgemeinern lassen und nicht nur für einzelne Experimente gelten. Ein Einfluss von Bewegungssignalen in anderen, eventuell schwierigeren Aufgaben lässt sich allerdings nicht ganz ausschließen.

Ein allgemeines Problem von Schablonen-Modellen wie dem in dieser Arbeit präsentierten, liegt darin, dass die erzielten Ergebnisse streng genommen nur für die benutzten Schablonen gelten. Aus diesem Grund waren die Aufgaben, die in dieser Arbeit benutzt wurden, auf Unterscheidungsaufgaben beschränkt. Denn nur in diesen Aufgaben ist ein beschränkter und definierter Satz von Schablonen vorgegeben. Deshalb sind Verallgemeinerungen auf die allgemeine Erkennung von biologischer Bewegung schwierig. Allerdings habe ich in meiner Arbeit auch Aufgaben wie die Vorwärts/Rückwärts-Unterscheidung benutzt. Im Gegensatz zu den anderen Unterscheidungsaufgaben, setzt diese Aufgabe implizit eine Erkennung der globalen Bewegung und damit die globale Erkennung des Läufers voraus. Da auch in diesen Aufgaben der Modellansatz eine Erklärung für die beobachteten Ergebnisse liefert, können die Schlussfolgerungen mit realistischem Grund auf allgemeine Erkennungsmechanismen übertragen werden.

Die Verallgemeinerung der Modellsimulationen auf andere Handlungen als Laufen und auf andere Aufgaben stellt einen wichtigen Punkt für zukünftige Modellierungs- und Experimentalaufgaben dar.

Bibliography

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.*, 4(3):165–78.

Ahlstrom, V., Blake, R., and Ahlstrom, U. (1997). Perception of biological motion. *Perception*, 26:1539–1548.

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the sts region. *Trends Cogn. Sci.*, 4(7):267–78.

Anderson, K. C. and Siegel, R. M. (2005). Three-dimensional structure-from-motion selectivity in the anterior superior temporal polysensory area, stpa, of the behaving monkey. *Cereb. Cortex.*

Astafiev, S. V., Stanley, C. M., Shulman, G. L., and Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature Neurosci.*, 7:542–548.

Astafiev, S. V., Stanley, C. M., Shulman, G. L., and Corbetta, M. (2005). Is the extrastriate body area involved in motor actions? reply. *Nature Neurosci.*, 8:125–126.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.*, 154:600–9.

Barlow, H. B. (1997). The knowledge used in vision and where it comes from. *Philos.Trans.R.Soc.Lond. B*, 352:1141–1148.

Batelli, L., Cavanagh, P., and Thornton, I. M. (2003). Perception of biological motion in parietal patients. *Neuropsychologia*, 41:1808–1816.

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., and Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat. Neurosci.*, 7(11):1190–2.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., and Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, 34:149–159.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., and Martin, A. (2003). FMRI responses to video and point-light displays of moving humans and manipulable objects. *J. Cogn. Neurosci.*, 15(7):991–1001.

Beintema, J. A., Georg, K., and Lappe, M. (2005). Perception of biological motion from limited lifetime stimuli. *Percept. Psychophys.* (in press).

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

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

Bidet-Caulet, A., Voisin, J., Bertrand, O., and Fonlupt, P. (2005). Listening to a walking human activates the temporal biological motion area. *Neuroimage*.

Blake, R. (1993). Cats perceive biological motion. *Psychol.Sci.*, 4:54–57.

Blakemore, S. J. and Decety, J. (2001). From the perception of action to the understanding of intention. *Nat. Rev. Neurosci.*, 2(8):561–7.

Bobick, A. F. and Davis, J. W. (2001). The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine*, 23(3):257–67.

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

Boussaoud, D., Ungerleider, L. G., and Desimone, R. (1990). Pathways for motion analysis: Cortical connections of the medial superior temporal visual areas in the macaque. *J.Comp.Neurol.*, 296:462–495.

Braddick, O. (1974). Short-range process in apparent motion. *Vision Res.*, 14(7):519–527.

Bradley, D. C., Chang, G. C., and Andersen, R. A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature*, 392:714–717.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., and Freund, H.-J. (2001). Short commnunication - action observation activates premotor and parietal areas in a somatotopic manner. An fMRI study. *Eur.J.Neurosci.*, 13:400–404.

Buccino, G., Lui, F., N.Canessa, Patteri, I., Lagravinese, G., Benuzzi, F., Porro, C., and Rizzolatti, G. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: an fmri study. *J. Cogn. Neurosci.*, 16(1):114–26.

Bülthoff, H. and Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc.Nat.Acad.Sci.USA*, 89:60–64.

Bülthoff, I., Bülthoff, H., and Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nat. Neurosci.*, 1(3):254–7.

Campbell, R., Zihl, J., Massaro, D., Munhall, K., and Cohen, M. M. (1997). Speechreading in the akinetopsic patient, l.m. *Brain*, 120:1793–1803.

Casile, A. and Giese, M. A. (2005). Critical features for the recognition of biological motion. *J. Vis.*, 5(4):348–60.

Cavanagh, P. (1992). Attention-based motion perception. *Science*, 257(5076):1563–65.

Chan, A. W. Y., Peelen, M. V., and Downing, P. E. (2004). The effect of viewpoint on body representation in the extrastriate body area. *Neuroreport*, 15(15):2407–10.

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

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

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

Cowey, A. and Vaina, L. M. (2000). Blindness to form from motion despite intact static form perception and motion detection. *Neuropsychologia*, 38:566–578.

Cutting, J. E. (1978). A program to generate synthetic walkers as dynamic point-light displays. *Behav.Res.Meth.Instrumentation*, 10:91–94.

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

Cutting, J. E. and 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., and Morrison, R. (1988). Masking the motions of human gait. *Percep.Psychophys.*, 44:339–347.

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

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

Fox, R. and McDaniel, C. (1982). The perception of biological motion by human infants. *Science*, 218 (4571):486–487.

Georg, K. (2002). Die Bedeutung von Forminformation bei der Wahrnehmung biologischer Bewegung. Diplomarbeit in Biologie, Ruhr-Universität Bochum.

Giese, M. and Lappe, M. (2002). Measurement of generalization fields for the recognition of biological motion. *Vision Res.*, 42:1847–1858.

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

Giese, M. A. and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Neurosci.*, 4:179–192.

Grossman, E., Batelli, L., and Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vis. Res.*

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

Grossman, E. D. and Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron*, 35:1167–1175.

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

Hamker, F. H. (2004). Predictions of a model of spatial attention using sum- and max-pooling functions. *Neurocomputing*, 56:329–343.

Heberlein, A. S., Adolphs, R., Tranel, D., and Damasio, H. (2004). Cortical regions for judgments of emotions and personality traits from point-light walkers. *J. Cogn. Neurosci.*, 16(7):1143–58.

Hirai, M. and Hiraki, K. (2005). The relative importance of spatial versus temporal structure in the perception of biological motion: An event-related potential study. *Cognition.* in press.

Hoffman, D. D. and Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biol. Cybern.*, 42(3):195–204.

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

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M.-C., Mazziotta, J. C., and Rizzolatti, G. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proc.Nat.Acad.Sci.USA*, 98:13995–13999.

Ivry, R. B. and Spencer, R. M. (2004). The neural representation of time. *Curr. Opin. Neurobiol.*, 14(2):225–32.

Jacobs, A. and Shiffrar, M. (2005). Walking perception by walking observers. *J. Exp. Psychol. Hum. Percept. Perform.*, 31(1):157–69.

Jellema, T., Baker, C. I., Wicker, B., and Perrett, D. I. (2000). Neural representation for the perception of the intentionality of actions. *Brain Cogn.*, 44(2):280–302.

Jellema, T. and Perrett, D. I. (2003). Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? *Neuropsychologia*, 41:1728–1737.

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

Jokisch, D., Troje, N. F., Koch, B., Schwarz, M., and Daum, I. (2005). Differential involvement of the cerebellum in biological and coherent motion perception. *Eur J Neurosci*, 21(12):3439–46.

Kobatake, E. and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.*, 71(3):856–67.

Kourtzi, Z. and Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *J.Cog.Neurosci.*, 12:48–55.

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

Krekelberg, B., Kubischik, M., Hoffmann, K.-P., and Bremmer, F. (2003). Neural correlates of visual localization and perisaccadic mislocalization. *Nature*, 37:1–20.

Lee, J. and Wong, W. (2004). A stochastic model for the detection of coherent motion. *Biol Cybern.*, 91(5):306–14.

Logothetis, N. K. and Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex*, 5:270–288.

Logothetis, N. K., Pauls, J., Bülthoff, H. H., and Poggio, T. (1994). View-dependent object recognition by monkeys. *Curr.Biol.*, 4:401–414.

Logothetis, N. K., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Curr.Biol.*, 5:552–563.

Loula, F., Prasad, S., Harber, K., and Shiffrar, M. (2005). Recognizing people from their movement. *J. Exp. Psychol. Hum. Percept. Perform.*, 31(1):21–20.

Lu, X. and Ashe, J. (2005). Anticipatory activity in primary motor cortex codes memorized movement sequences. *Neuron*, 5(6):967–73.

Lu, Z. L. and Sperling, G. (1995). Attention-generated apparent motion. *Nature*, 377:237–239.

Marr, D. and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proc.Royal.Soc.London B*, 200:269–294.

Marr, D. and Vaina, L. (1982). Representation and recognition of the movements of shapes. *Proc.Royal.Soc.London B*, 214:501–524.

Mather, G. and Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proc.Royal.Soc.London B*, 258:273–279.

Mather, G., Radford, K., and West, S. (1992). Low-level visual processing of biological motion. *Proc.Royal.Soc.London B*, 249:149–155.

McLeod, P., Dittrich, W., Driver, J., Perrett, D., and Zihl, J. (1996). Preserved and impaired detection of structure from motion by a 'motion-blind' patient. *Vis.Cognition*, 3:363–391.

Michels, L., Lappe, M., and Vaina, L. (2005). Visual areas involved in the perception of human movement from dynamic form analysis. *NeuroReport*, 16(10):1037–41.

Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends.Neurosci.*, 6:414–417.

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

Oram, M. W. and Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to 'biological motion' stimuli. *J.Cog.Neurosci.*, 6:99–116.

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

Parish, D. H., Sperling, G., and Landy, S. L. (1990). Intelligent temporal subsampling of american sign language using event boundaries. *J. Exp. Psychol. Hum. Percept. Perform.*, 16(2):282–94.

Pavlova, M., Krägeloh-Mann, I., Sokolov, A., and Birbaumer, N. (2001). Recognition of point-light biological motion displays by young children. *Perception*, 30:925–933.

Pavlova, M. and Sokolov, A. (2000). Orientation specificity in biological motion perception. *Percep.Psychophys.*, 62:889–899.

Peelen, M. V. and Downing, P. E. (2005a). Is the extrastriate body area involved in motor actions? *Nature Neurosci.*, 8:125.

Peelen, M. V. and Downing, P. E. (2005b). Selectivity for the human body in the fusiform gyrus. *J. Neurophysiol.*, 93(1):603–8.

Perrett, D., Smith, P., Potter, D., Mistlin, A., Head, A., Milner, A., and Jeeves, M. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc.Royal.Soc.London B*, 223(1232):293–317.

Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., Chitty, A. J., Hietanen, J. K., and Ortega, J. E. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *J. Exp. Biol.*, 146:87–113.

Perrett, D. I., Harries, M. H., Mistlin, A. J., and Hietanen, J. K. (1990). Social signals analysed at the cell level: someone is looking at me, something touched me, something moved. *Int. J. Comp. Psychol.*, 4:25–54.

Peuskens, H., Vanrie, J., Verfaillie, K., and Orban, G. A. (2005). Specificity of regions processing biological motion. *Eur. J. Neurosci.*, 21(10):2864–75.

Pinto, J. and 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., and Braden, V. (2001). Recognising the style of spatially exaggerated tennis serves. *Perception*, 30(3):323–38.

Pollick, F. E., Lestou, V., Ryu, J., and Cho, S. (2002). Estimating the effiency of recognizing gender and affect from biological motion. *Vision Research*, 42(20):2345–55.

Ptito, M., Faubert, J., Gjedde, A., and Kupers, R. (2003). Separate pathways for contour and biological-motion cues in motion-defined animal shapes. *NeuroImage*, 19:246–252.

Puce, A., Allison, T., Bentin, S., Gore, J. C., and McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *J. Neurosci.*, 18(6):2188–99.

Reed, C. L., Stone, V. E., Bozova, S., and Tanaka, J. (2003). The body-inversion effect. *Psychol Sci.*, 14(4):302–8.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature*, 2:1019–1025.

Rizzolatti, G. and Arbib, M. A. (1998). Language within our grasp. *Trends.Neurosci.*, 21:188–194.

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

Rizzolatti, G., Fogassi, L., and Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Curr. Opin. Neurobiol.*, 12(2):149–54.

Rohr, K. (1994). Towards model-based recognition of human movements in image sequences. *Comp.Vis.Graph.Image Proc.: Image Understanding*, 59:94–115.

Rolls, E. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical areas. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 335:11–21.

Sakai, K., Ramnani, N., and Passingham, R. E. (2002). Learning of sequences of finger movements and timing: frontal lobe and action-oriented representation. *J. Neurophysiol.*, 88(4):2035–46.

Sakreida, K., Schubotz, R. I., Wolfensteller, U., and von Cramon, D. Y. (2005). Motion class dependency in observers' motor areas revealed by functional magnetic resonance imaging. *J. Neurosci.*, 25(6):1335–42.

Santi, A., Servos, P., Vatikiotis-Bateson, E., Kuratate, T., and Munhall, K. (2003). Perceiving biological motion: Dissociating visible speech from walking. *J. Cogn. Neurosci.*, 15(6):800–9.

Saygin, A., Wilson, S., Hagler, D. J., Bates, E., and Sereno, M. (2004). Point-light biological motion perception activates human premotor cortex. *J. Neurosci.*, 24(27):6181–8.

Schenk, T. and Zihl, J. (1997). Visual motion perception after brain damage: II. deficits in form-from-motion perception. *Neuropsychologia*, 35:1299–1310.

Schubotz, R. I. and von Cramon, D. Y. (2004). Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observation and imagery. *J. Neurosci.*, 24(24):5467–74.

Schultz, J., Friston, K. J., O'doherty, J., Wolpert, D. M., and Frith, C. D. (2005). Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. *Neuron*, 45(4):625–35.

Sekiyama, K., Kanno, I., Miura, S., and Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neurosci Res.*, 47(3):277–87.

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

Shiffrar, M., Lichtey, L., and Heptulla Chatterjee, S. (1997). The perception of biological motion across apertures. *Percep.Psychophys.*, 59:51–59.

Shipley, T. (2003). The effect of object and event orientation on perception of biological motion. *Psychol. Sci.*, 14(4):377–80.

Sinha, P. and Poggio, T. (1996). Role of learning in three-dimensional form perception. *Nature*, 384:460–463.

Sumi, S. (1984). Upside-down presentation of the Johansson moving light pattern. *Perception*, 13:283–286.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu.Rev.Neurosci.*, 19:109–139.

Thompson, J. C., Clarke, M., Stewart, T., and Puce, A. (2005). Configural processing of biological motion in human superior temporal sulcus. *J Neurosci.*, 25(39):9059–66.

Thornton, I., Rensink, R., and Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception*, 31(7):837–53.

Thornton, I. A. and Vuong, Q. C. (2004). Incidental processing of biological motion. *Curr. Biol.*, 14(12):1084–89.

Thornton, I. A., Vuong, Q. C., and Bülthoff, H. H. (2003). Last but not least. *Perception*, 32:377–383.

Thornton, I. M., Pinto, J., and Shiffrar, M. (1998). The visual perception of human locomotion. *Cogn. Neuropsychol.*, 15:535–552.

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

Troje, N. F. (2003). Reference frames for orientation anisotropies in face recognition and biological-motion perception. *Perception*, 32(2):201–10.

Troje, N. F., Westhoff, C., and Lavrov, M. (2005). Person identification from biological motion: effects of structural and kinematic cues. *Percept Psychophys.*, 67(4):667–75.

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

Ungerleider, L. (1995). Functional brain imaging studies of cortical mechanisms for memory. *Science*, 270:769–775.

Ungerleider, L. G. and Desimone, R. (1986). Cortical connections of visual area MT in the macaque. *J Comp Neurol*, 248(2):164–89.

Urgesi, C., Berlucchi, G., and Aglioti, S. M. (2004). Magnetic stimulation of extrastriate body area impairs visual processing of nonfacial body parts. *Curr. Biol.*, 14:2130–4.

Vaina, L. and Bennour, Y. (1985). A computational approach to visual recognition of arm movements. *Percept.Mot.Skills*, 60:203–228.

Vaina, L., Lemay, M., Bienfang, D. C., Choi, A. Y., and 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., Cowey, A., LeMay, M., Bienfang, D. C., and Kikinis, R. (2002). Visual deficits in a patient with kaleidoscopic disintegration of the visual world. *European Journal of Neurology*, 9:463–77.

Vaina, L. M. and Gross, C. G. (2004). Perceptual deficits in patients with impaired recognition of biological motion after temporal lobe lesions. *Proc.Nat.Acad.Sci.USA*, 101(48):16947–51.

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

Verfaillie, K. (2000). Perceiving human locomotion: priming effects in direction discrimination. *Brain and Cognition*, 44:192–213.

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

Wolpert, D. M., Miall, R. C., and Kawato, M. (1998). Internal models in the cerebellum. *Trends.Cogn.Sci.*, 2:338–347.

Wright, T., Pelphrey, K., Allison, T., McKeown, M., and McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex*, 13(10):1034–43.

Zihl, S., von Cramon, D., and Mai, N. (1983). Selective disturbance of vision after bilateral brain damage. *Brain*, 106:313–40.

Wissenschaftlicher Werdegang und Publikationen

2000-2006   Promotion

Thema: Wahrnehmung Biologischer Bewegung durch Formanalyse

2005-2006: Philipps-Universität Marburg

2002-2005: Westf. Wilhelms-Universität Münster

2000-2002: Ruhr-Universität Bochum

2000   Diplom in Physik an der Westf. Wilhelms-Universität Münster

Thema: Charakterisierung des Muskelstoffwechsels bei Patienten mit Poly- und Dermatomyositis mittels ^{31}P Magnet Resonanz Spektroskopie

1994-2000   Studium der Physik an der Westf. Wilhelms-Universität Münster

Article

Lange J. and Lappe M. (2005). A model of biological motion perception from configural form cues *submitted*

Lange J. and Lappe M. (2005). Visual perception of biological motion by form: a template-matching analysis *submitted*

Pfleiderer B., Lange J., Loske K.-D., Sunderkötter C. (2004). Metabolic disturbances during short exercises in dermatomyositis revealed by real-time functional P-31 magnetic resonance spectroscopy *Rheumatology* 43 (6): 696-703

Proceedings

Lange J. and Lappe M. (2002). Ideal-observer-model and psychophysical studies on the role of form information in biological motion perception. *In: R.Würtz & M. Lappe, editors, Dynamic Perception, Infix Verlag*, 109-115

Abstracts

Lange J. & Lappe M. (2005) Discrimination of biological motion in noise *Perception 34 (ECVP Abstracts)* 2005

Lange J. & Lappe M. (2004) The role of form analysis in the perception of biological motion *Society for Neuroscience Abstracts* 2004

Lange J. & Lappe M. (2004) Dynamic model of form-based biological motion recognition *Perception 33 (ECVP Abstracts)* 2004

Lange J. & Lappe M. (2003) Biological motion discrimination by form analysis *Perception 32: 34-34 Suppl. S* 2003

Lange J. & Lappe M. (2003) Form-based ideal-observer model for the recognition of biological motion *European Science Conference on Three-Dimensional Sensory and Motor Space (Abstracts)* 2003

Lange J. & Lappe M. (2003) Formbasierte Erkennung von Bewegungsabläufen Biologischer Bewegung *TWK (Abstracts)* 2003

Lange J. & Lappe M. (2002) Ideal-Observer-Modell zur Erkennung Biologischer Bewegung *TWK (Abstracts)* 2002

Danksagung

Auch wenn die Titelseite dieser Dissertation mich als einzigen Autor ausgibt, so bin ich mir doch im Klaren, dass das Gelingen der Arbeit auch von vielen anderen Leuten abhing. Während des Schreibens habe ich versucht, diese Tatsache durch den Plural „we" zum Ausdruck zu bringen. In dieser abschließenden Kapitel möchte ich nun den Personen danken, die direkt oder indirekt hinter diesem „we" stehen.

Mein größter Dank gilt den beiden Personen, die im wissenschaftlichen Sinne direkt an dieser Arbeit beteiligt waren: Prof. Markus Lappe für die Betreuung der Arbeit. Ich danke ihm für seine stets hilfreichen und fruchtbaren Diskussionen, seinen wertvollen Rat und für die Freiräume, eigene Ideen und Ansätze zu verwirklichen. Und ich danke Prof. Frank Bremmer insbesondre für seine Hilfe, Mühen und sein Engagement, damit der Versuch, den Dr. rer. nat. zu erlangen, nicht schon an der bürokratischen Hürde scheitert.

Danken möchte ich an dieser Stelle auch Ursula Husemann. Sie hat immer für eine positive Atmosphäre innerhalb der gesamten Arbeitsgruppe gesorgt und mir und allen anderen immer den Rücken frei von allen finanziellen und bürokratischen Formalitäten gehalten.

Ganz herzlich danke ich Dr. Jaap Beintema. Gerade die nicht einfache Startphase in die Neurowissenschaften hat er mir durch seine große Hilfsbereitschaft und seine vielen kritischen Diskussionen sehr vereinfacht.

Dr. Marc de Lussanet danke ich dafür, dass er mir bei der Programmierung der Stimuli immer mit Rat und Tat zur Seite gestanden hat. Und wenn die Statistik drohte, mich in ihre Untiefen zu ziehen, konnte ich mich auf die Hilfe von Maximilian Schmidt, Dr. Farid Kandil und Marc Zirnsak verlassen. Mit letzterem hatte ich auch das Vergnügen, ein Büro zu teilen, was nicht das schlechteste Los war. Auch dem Rest der Arbeitsgruppe möchte ich für die nette Atmosphäre und die vielen anregenden Diskussionen danken. Besonders bei meinen jetzigen und ehemaligen Mitdoktoranden Dr. Holger Awater, Dr. Harald Frenz, Karsten Georg, Katharina Georg, Dr.