Neural Correlates of Learning of Complex Movement Recognition

Dissertation

zur Erlangung des Grades eines Doktors
der Naturwissenschaften

der Fakultät für Biologie
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von

Jan Jastorff
aus Bremen

2006

Tag der mündlichen Prüfung:	27.10.2006
Dekan der Fakultät für Biologie:	Prof. Dr. F. Schöffl
Dekan der Medizinischen Fakultät:	Prof. Dr. I. B. Autenrieth
1. Berichterstatter:	Prof. Dr. H. P. Thier
2. Berichterstatter:	Prof. Dr. H. H. Bülthoff
Prüfungskommission	Prof. Dr. H. H. Bülthoff Prof. Dr. H. P. Thier Prof. Dr. W. Grodd Prof. Dr. W. Lutzenberger PD Dr. M. A. Giese

Für Muddel und Berni

Acknowledgments

First and foremost, I would like to convey my special gratitude to Dr. Martin Giese, my Ph.D. advisor, for providing me with the opportunity to work in his group on this topic. His door was always open and despite severe work load and time constraints, he took the time to help his student. Through very healthy criticisms and suggestions (although sometimes difficult), he helped me improve my work substantially. Thank you very much Martin for introducing and inspiring me to the depth of cognitive research.

Moreover, I would like to express my sincere thanks to Prof. Dr. Zoe Kourtzi, for sparking my interest in functional imaging and for guiding my work not only during the fMRI studies but also the psychophysical part. I am very grateful for the many hours I spend together with Martin and Zoe discussing the results of the experiments and interpreting the data. These times helped me broaden my understanding of how different parts finally come together to form a wider picture.

Additionally, I am indebted to Prof. Dr. Heinrich H. Bülthoff, who gave me the possibility to work in the very inspiring and challenging atmosphere of the Max Planck Institute for Biological Cybernetics in Tübingen and allowed me to make use of all the sophisticated equipment he possesses in his department.

Special acknowledgments I owe Prof. Dr. Peter Thier, who very much supported my fMRI work at the University Clinic Tübingen and to Prof. Dr. Wolfgang Grodd, Fotini Scherer and Franziska Hösl for their guidance and assistance when acquiring the functional data presented in Chapter 4.

My work here in Tübingen would not have started without the Graduate School of Neural and Behavioural Sciences. I am very thankful to the program coordinator Prof. Dr. Horst Herbert, who always had time to listen to any problem, small or big, that came in the way and was never short of tips how to deal with the tricky administrative work.

Special thanks go to Dr. Christian F. Altmann, from whom I learned how to run functional imaging experiments and data analysis. Moreover, I would like to thank Elisabeth Huberle, Pegah Sarkheil, Dr. James Scott McDonald, and Theresa Cooke for spending their weekends together with me at the University Clinic in front of the fMRI scanner and for cheering me up when things went wrong. Additionally, my work would have been much more tedious without the exquisite programming skills of Mario Kleiner and Chandrasekaran Chandramouli. I am very thankful to Dr. Isabelle Bülthoff for her continuous help

and suggestions during the psychophysical studies of my work, Dr. Winfried Ilg for support with the motion capturing, Dr. Antonino Casile for his instructions in the theoretical part presented in Chapter 5 and Dr. Kamil Uludag for introducing me to the physics of fMRI.

Thanks also to my lab-mates Claire Roether and Lars Omlor, who were a big help especially at the end of my thesis, trying to cheer me up with chocolates, tea and cake.

The everyday life at the institute would not have been so much fun without the coffee breaks and pizza parties I enjoyed with Karin S. Pilz, Dr. Anne-Marie Brouwer, Andries Hof, Dr. Quoc C. Vuong, Lewis Chuang and Joost Maier. Thanks!

Finally, I am very grateful to my parents, my sisters and Dr. Hans Hettler. Without their loving support this work would have not been possible. I am thinking of my late mother who missed my graduation by only a few months. Danke für alles, Muddel!

Very special thanks go to my wife Archana, who knows the best, how I felt during the last years, tolerated without complaining, when our short weekends became even shorter due to work and encouraged me when times where tough.

This work was financially supported by the German Research Foundation through the SfB 550.

Contents

1 Introduction **1**

1.1 The Human Visual Cortex 2

1.2 Perceptual Learning 7

1.3 Biological Motion Perception 9

1.4 Aim of Thesis and Experimental Questions 14

2 Methods **17**

2.1 Psychophysics 17

2.2 Magnetic Resonance Imaging 20

2.2.1 Principles of fMRI 21

2.2.2 Experimental Design 26

2.2.3 Data Analysis 31

2.3 Motion Morphing 36

3 Psychophysical Investigations **39**

3.1 Introduction 40

3.2 Experimental Specifications 42

3.2.1 Participants 42

3.2.2 Stimuli 42

3.2.3 Procedure 46

3.2.4 Approximation of Motion Morphs 46

3.3 Results 48

3.3.1 Experiment 1 48

3.3.2 Experiment 2 51
3.3.3 Experiment 3 53
3.3.4 Experiment 4 54
3.4 Discussion 56

4 Functional Imaging Studies **61**
4.1 Introduction 61
4.2 Experimental Specifications 65
4.2.1 Participants 65
4.2.2 Stimuli 65
4.2.3 Design and Procedure 67
4.2.4 Imaging 70
4.2.5 fMRI Data Analysis 70
4.3 Results 73
4.3.1 Experiment 5 73
4.3.2 Experiment 6 77
4.3.3 Experiment 7 79
4.3.4 Learning in Retinotopic Areas 81
4.4 Discussion 83

5 Theoretical Modeling **91**
5.1 Introduction 91
5.2 Learning of the Feedback Connectivity 95
5.2.1 Recurrent Neural Network Model 96
5.2.2 Learning Rules: 98
5.2.3 Stability Analysis 99
5.2.4 Simulation Results 101
5.2.5 Conclusion 102
5.3 Learning of Feedforward Connectivity 103
5.4 Modeling of BOLD Activity Changes 106

6 General Discussion **111**

6.1 Summary 111

6.2 Outlook 116

7 References **119**

A Curriculum Vitae **139**

B Bibliography **143**

Chapter 1

Introduction

Interaction between individuals is a constantly encountered situation in the animal kingdom. The ability to recognize the movements of the counterparts is an inevitable part of their communication. The recognition of complex movement patterns is important for example, to identify possible threats by predators, or to select a weaker animal as prey. Additionally, intra-species communication, especially of higher animals relies to a large degree on the interpretation of facial or body movements. While ecological studies have provided evidence for innate mechanisms of movement recognition in lower vertebrates and also in higher species (Lorenz, 1965), most likely, also learning mechanisms influence the processing of complex movements.

Learning mechanisms have so far been studied intensively in the field of object recognition. Poggio and Edelman (1990) have shown that the representation of three-dimensional objects can be learned exclusively from the presentation of two-dimensional views of the objects. Moreover, Bülthoff and colleagues created a set of arbitrary artificial complex shapes and demonstrated that humans are able to learn to recognize those shapes (Bülthoff and Edelman, 1992; Tarr and Bülthoff, 1998). Neurophysiological recordings in monkeys identified neural populations in the inferotemporal cortex, which changed their tuning properties to become selectively activated by particular views of previously learned complex objects (Logothetis et al., 1995; Logothetis and Sheinberg, 1996).

Interestingly, the learned representation was view-dependent. This means that once the object is learned from a particular view, the recognition performance drops significantly, if the object is presented from a different viewpoint. Only within very similar viewing angles, does the performance stay high. Similarly,

this effect was also observed in the neurophysiological recordings. Many of the recorded neurons significantly reduced their firing rate, when the object was presented from a viewpoint that differed significantly from the training view.

Orientation dependence has also been reported in experiments investigating the recognition of complex movements (e.g Bertenthal and Pinto, 1994; Pavlova and Sokolov, 2000; Shiffrar et al., 1997; Sumi, 1984). This motivated the hypothesis that objects and movements might be encoded in similar ways (Verfaillie et al., 1994). That in turn suggests the possibility that object and movement recognition might also share similar learning mechanisms. Additionally, a theoretical model developed recently by Giese and Poggio (2003) also supports the possible involvement of learning mechanisms in the recognition of complex movements. Their model, which represents complex movements on the basis of learned prototypical example movements, replicates many of the experimental results obtained in the field of movement recognition.

The similarities between movement and object recognition, together with the theoretical work provided the motivation for my study to investigate, how learning shapes the processing of complex movements. In the following sections, I will briefly review the concepts of visual processing, learning and movement recognition, which serve as a foundation for the experimental questions. The different methodological approaches that were used to study the neural correlates of learning are described in Chapter 2. The psychophysical investigations in Chapter 3 examine the influence of learning on the recognition of complex movements and the invariant properties of the learned representation. The functional imaging experiments reported in Chapter 4 investigate the contribution of different visual areas to the learning process by comparing BOLD activity changes before and after learning. Finally, the theoretical work based on the already existing neural model for movement recognition (Giese and Poggio, 2003) intends to evaluate different possible plasticity mechanisms, by which the learning of complex movements could be accomplished (Chapter 5).

1.1 The Human Visual Cortex

Most of the visual information from the retina gets transferred via the Lateral Geniculate Nucleus to the primary visual cortex (V1), located at the posterior pole of the brain. This transfer takes place in a topographically organized manner, that is, neighboring positions on the retina project to neighboring sites in area V1. The cortical mapping of the visual space is not a one to one representation but occurs in a logarithmic fashion, with much larger representations

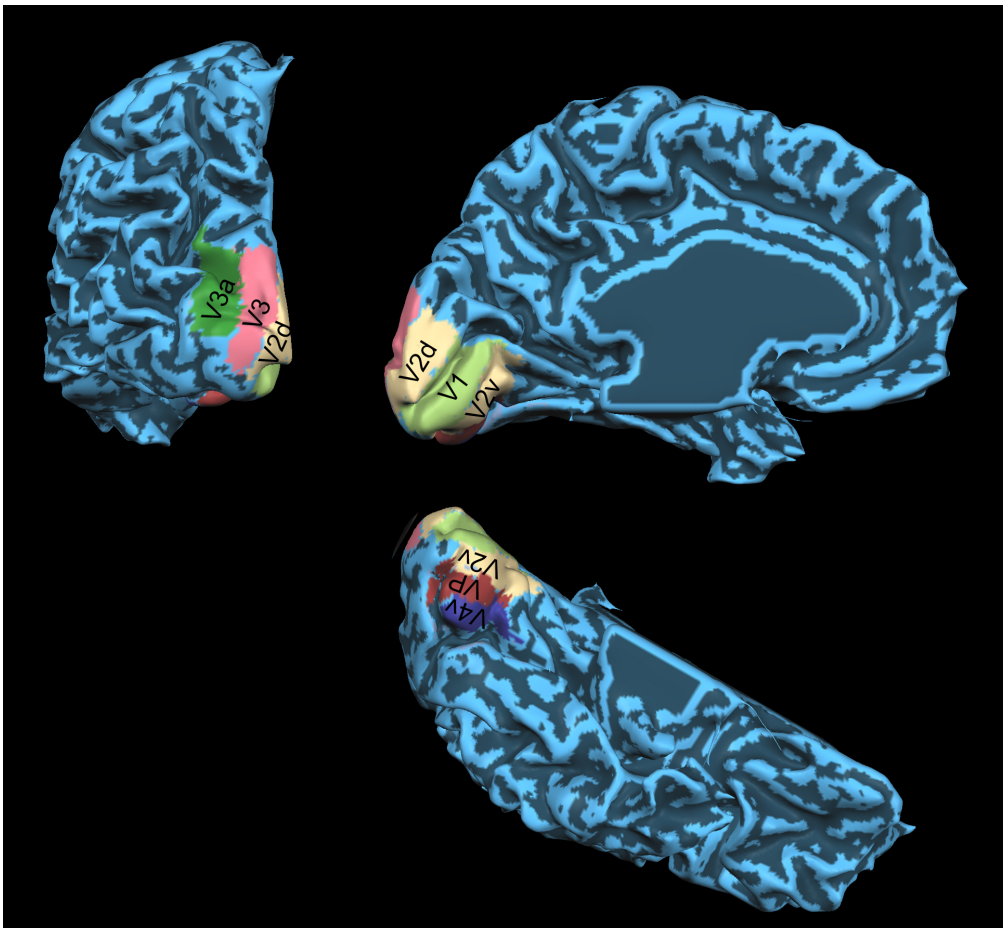

Figure 1.1: Left hemisphere of a human brain seen from different viewpoints. Early retinotopic areas are mapped onto the cortical surface at the occipital pole. Primary visual cortex (V1) is bordered by two cortical areas that form area V2. Dorsal V2 represents the lower quadrant and ventral V2 represents the upper quadrant of the visual field. More dorsal primarily motion processing areas are V3 followed by V3a and more ventral areas primarily involved in form processing are VP and V4v.

for the foveal region of the retina and gradually smaller representations depending on the distance of the stimulus from the fovea (Duncan and Boynton, 2003; Schwartz et al., 1985). While the fovea is represented at the posterior pole, more peripheral regions are represented anteriorly. This so called retinotopic organization is observed not only in V1 but also throughout lower- and mid level visual areas (Levy et al., 2001; DeYoe et al., 1996) (see Figure 1.1).

The description of visual cortical areas has mainly been focussed on two concepts (see Grill-Spector and Malach, 2004, for review). On the one hand, pro-

cessing along the visual areas can be described in a hierarchical fashion. Visual analyses at lower levels of this hierarchy are thought to occur within brief temporal intervals and small spatial neighborhoods. The results of this low-level or local analysis are then passed onto higher levels in the hierarchy, which process information across larger spatio-temporal extents, resulting in a so-called global analysis of the scene (DeYoe and Essen, 1988). Alternatively, visual processing can be described based on functional specialization. A very famous example of functional specialization divides the visual system into two parallel processing streams specialized for specific aspects of the visual input. Mishkin and colleagues (Mishkin et al., 1983) proposed the existence of a dorsal stream also called *"where"* stream, involved in spatial localization and a ventral *"what"* stream, involved in object and form recognition. Even though both streams are not completely separated anatomically or functionally, this concept has been followed up by many other researchers (e.g. Goodale et al., 1991; Goodale and Milner, 1992) and proved to be a valid principle in understanding the visual cortex. In the following sections I will focus on the processing of two aspects of the visual input, namely the processing of motion and the processing of form information within the visual cortex.

Motion Processing

Most of the motion information from the retina reaches primary visual cortex via the magnocellular pathway terminating in layer 4B. From layer 4B, the information gets transferred to areas V2, V3 and V3a. In addition, V1 also has direct connections to the middle temporal area (MT) (Maunsell and Essen, 1983; Felleman and Essen, 1991). The anatomical location of these areas suggests that the perception of motion is mainly accomplished by a network of brain regions in the dorsal pathway of the human visual system (see Figures 1.1 and 1.2).

In human imaging studies a region called human MT+ (hMT+) complex has been found to play a central role in the analysis of motion information (Watson et al., 1993; Dupont et al., 1994; Tootell et al., 1995). This region seems to be a homologue of the extrastriate motion selective area MT/V5 in the macaque (Rees et al., 2000). hMT+ is selectively activated by moving versus static stimuli. It is also involved in the processing of apparent motion (e.g. Goebel et al., 1998), illusory motion (e.g. Tootell et al., 1995), implied motion (e.g. Kourtzi and Kanwisher, 2000a; Senior et al., 2000), and imagined motion (Goebel et al., 1998). Furthermore, hMT+ has been implicated in the analysis of shape properties like the perception of object structure from motion (e.g. Orban et al., 1999), and the processing of static objects that are associated with motion, such as tools (Chao et al., 1999).

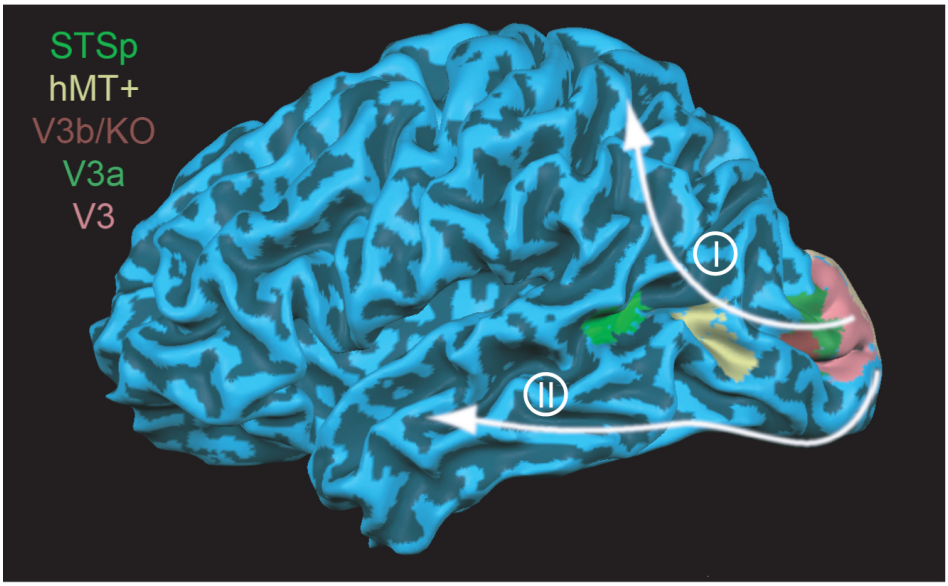

Figure 1.2: Left hemisphere of a human brain with some areas of the dorsal stream highlighted. I) symbolizes the dorsal *"where"* or *"perception for action"* pathway and II) symbolizes the ventral *"what"* or *"perception for identificaion"* pathway (Mishkin et al., 1983; Goodale and Milner, 1992).

However, presentation of stimuli containing motion defined contours, like two patches of dots moving in opposite directions creating the impression of a boundary at the position where the dot motion differs, leads to selective activation in a different region of the human brain. An area, responding more strongly to such motion-defined contour stimuli compared to uniform motion stimuli has been termed kinetic occipital area (KO) (Oostende et al., 1997; Dupont et al., 1997). In contrast, Zeki and colleagues reported that the time course of activity in KO correlates better with activity in area V3 than with activity in the two adjacent areas, hMT+ and LO. Moreover, when humans perceived shapes generated from kinetic boundaries or from equiluminant colors, area KO was activated, independent of how the shapes were derived. They concluded therefore that area KO is not specialized for the processing of kinetic contours (Zeki et al., 2003). Another study reported an area that they termed V3b, which gets activated by second order motion (Smith et al., 1998). Because their area V3b shows close proximity with area KO, it is most likely that both names refer to the same cortical area. Therefore this region is also called V3b/KO.

Functional imaging experiments investigating responses for more complex movement stimuli show activation in areas further up the dorsal stream. One

area that seems to be particularly involved in the processing of biological movements like full body human or animal movements but also lip, mouth and hand movements is located in the posterior part of the superior temporal sulcus (STSp) (e.g. Bonda et al., 1996; Allison et al., 2000; Grossman et al., 2000; Grezes et al., 2001; Puce and Perrett, 2003). The same area becomes activated by stationary images that are relevant for actions, such as body postures (Allison et al., 2000). Recently, another extrastriate area that responds selectively to movies and pictures of human body parts has been described (Downing et al., 2001).

Depending on the type of complex movements, also higher brain areas can be activated, like the amygdala or the orbitofrontal cortex (Allison et al., 2000). Additionally, the 'mirror neuron system' has been proposed to play a role in movement analysis. The term 'mirror neuron' refers to the finding of neurons in area F5 located in the premotor cortex of the monkey that discharge when the monkey performs a movement as well as when it observes the very same movement. Functional imaging studies in humans have identified a possible homologue of area F5 in broca areas 6 and 44. Possible interpretations of the functional role of the mirror neurons include action understanding, imitation, understanding of intentions and empathy (see Gallese et al., 2004; Rizzolatti et al., 2001; Rizzolatti and Craighero, 2004, for review). While this system promises to become a very interesting research field with respect to movement recognition, the focus of my thesis concentrates on visual areas involved in the processing of complex movements.

Form Processing

The processing of form information begins in primary visual cortex, which contains neurons analyzing edges, color and brightness of the visual input. This information is passed on to certain subdivisions of V2 and the processing continues via VP and V4 mainly along ventral brain regions into the temporal cortex. However, also presumably dorsal areas in the parietal cortex have been shown to be selectively activated by the presentation of objects (see Grill-Spector, 2003, for review).

When talking about form processing, two areas seem primarily to be involved in the task. The lateral occipital complex (LOC) and a ventral occipito-temporal region (VOT), which is comprised of a mainly face-selective area (fusiform face area, FFA) and a house- or scene-selective area (parahippocampal place area, PPA) (Grill-Spector and Malach, 2004). Interestingly, all these areas seem to respond independent of whether the object is defined by luminace, texture, motion or stereo cues (e.g. Grill-Spector et al., 1998; Kastner et al., 2000; Kriegeskorte

et al., 2003; Gilaie-Dotan et al., 2002; Kourtzi and Kanwisher, 2001). While activity in early visual areas represents the presence of a physical stimulus, activity in LOC and VOT is correlated with the actual recognition of the object, deduced from higher activity in trials, where the object was recognized compared to trials where observers failed to identify the object (Grill-Spector et al., 2004; Bar et al., 2001).

The exact type of representation of categories of objects, faces and scenes within the occipital temporal cortex is still very much under debate. A number of functional imaging experiments have shown selective activation for specific categories in certain regions along the temporal cortex. For example, different areas have been described for the processing of faces, animals, body parts, tools, places and letter strings (e.g. Kanwisher et al., 1997; Martin et al., 1996; Downing et al., 2001; Beauchamp et al., 2002; Ishai et al., 1999; Cohen et al., 2000). Based on these findings, Kanwisher and collaborators proposed the hypothesis that object recognition is accomplished by a set of areas, all specific to a certain category (Spiridon and Kanwisher, 2002). On the other hand, Haxby and colleagues follow the idea that the whole temporal cortex serves as a distributed system for object recognition with overlapping representations for individual categories (Haxby et al., 2001). A third hypothesis has been put forward by Tarr and Gauthier, suggesting that objects are not clustered according to their visual properties but to the type of processing required for recognition. In this view, the FFA would not be an area specialized for face processing but for the processing of fine details between members of a category, the observer has a lot of experience with (Tarr and Gauthier, 2000).

Having introduced the different visual areas that might play a role in the recognition of complex movements, the different concepts by which learning might influence the processing within these areas will be reviewed in the coming section.

1.2 Perceptual Learning

The term perceptual learning refers to changes in perception caused by experience. In principle, perceptual learning takes place in any perceptual modality and throughout adulthood. Improvements can be observed for auditory tasks like pitch discrimination, somatosensory tasks like spatial resolution and of course visual tasks like the discrimination of hue (see Gibson, 1953, for review). With regard to vision, perceptual learning can take place at basically any step of the visual hierarchy. Early stages of visual analysis are involved in performance

improvements in orientation discrimination (e.g. Vogels and Orban, 1985) or vernier acuity, meaning small differences in the offset of two lines (e.g. Poggio et al., 1992). Intermediate stages are involved in discriminating motion direction (e.g. Ball and Sekuler, 1982) or in the perception of depth from random-dot stereograms (e.g. Ramachandran and Braddick, 1973) and high level areas facilitate the recognition of individual faces or objects (see Tarr and Bülthoff, 1998; Tarr and Cheng, 2003, for review).

Several theories have been introduced as a general framework of perceptual learning. One of which, the reverse hierarchy theory first introduced by Ahissar and Hochstein (1997), argues, that learning starts at high levels in the cortex and progresses backwards towards lower levels if those are needed for fine grained discriminations. A similar theory, called late selection theory, proposes that higher cortical areas are predominantly involved in perceptual learning. Irrelevant information gets filtered out at later stages leaving all information available for processing. On the other hand, an early selection theory argues for cortical changes due to perceptual learning as early as possible in the visual hierarchy to eliminate irrelevant signals already at the beginning (see Fahle, 2004, for comparison of early and late selection theory). However, up to now the consensus seems to be that different processing levels are involved in perceptual learning and that it is modulated by top-down influences (see Fahle, 2005, for review).

In most of the experiments, perceptual improvements are shown to be specific with respect to the trained stimulus and the training location. This means that improvements of observers to discriminate between a pair of motion directions at a specific location in the visual field do not transfer to a pair of different directions at the same spatial location or to the same two directions at a different location (Ball and Sekuler, 1982). This phenomenon has been reported for many other perceptual learning tasks (see Gilbert, 1994; Gilbert et al., 2001, for review). However, Liu and Weinshall repeated the motion discrimination experiment by Ball and Sekuler (1982) and analyzed the observers' performance after changing the training direction. Even though the initial performance for the different training direction dropped to baseline, the perceptual learning for the new direction became much faster. This indicates some transfer to the new direction suggesting generalization (Liu and Weinshall, 2000). In another series of experiments the same researchers (Liu, 1995) showed that the lack of transfer or generalization occurs only if the task is 'difficult'. Similarly, Ahissar and Hochstein (1997) found transfer in a visual search task with line elements, only when the difference between target and distractor was large enough to allow for easy detection of the target. They have termed this effect 'Eureka' effect and explained the phe-

nomenon with the possibility, that the easy task allows the observer to allocate attention to the relevant dimensions.

The underlying plasticity mechanisms enabling perceptual improvements are still very much under debate. One possible hypothesis would explain improvements in performance with changes in the tuning curves of neurons relevant for the task. Depending on the task, tuning curves could be shifting, sharpening or broadening with practice. A similar explanation follows the idea that the weights with which each neuron contributes to the response are adapted in a way that the best-tuned neuron for the discrimination would be weighted more strongly (see e.g. Fine and Jacobs, 2002, for review). Experimental evidence from primate studies investigating changes in tuning curves of individual neurons has been reported at different levels in the visual hierarchy. Schoups and colleagues (Schoups et al., 2001) have found changes in the slope of neural tuning within V1 as a function of training using an orientation discrimination task. Neurons in area V4 with receptive fields in the trained region of the visual field, narrow their orientation tuning and increase their responses as a result of training (Yang and Maunsell, 2004). Also high level areas have been shown to be modulated by experience. Neurons in inferotemporal cortex (IT) respond to particular objects, faces and shapes (e.g. Desimone et al., 1984) and the firing of individual neurons reflects even emerging sensitivity for trained complex shapes the monkey had no previous experience with (Logothetis et al., 1995).

Literature review on perceptual learning shows that this topic has been investigated extensively over the last decades. While most of these studies have used stationary stimuli in order to examine possible learning processes, my work focusses on influences of learning on the recognition of complex movements. The research that has been carried out so far using complex human like movements will be discussed in the following section.

1.3 Biological Motion Perception

Psychophysical Investigations

One of the most striking examples of the ability of the visual system to recover object information from sparse input is provided by the phenomenon known as biological motion. This term has been introduced by Gunnar Johansson in 1973 (Johansson, 1973). In his experiments, he attached small light bulbs at the major joints of a human actor. Following, the actor was filmed walking in the dark, leaving only the light points visible but not the silhouette of the actor

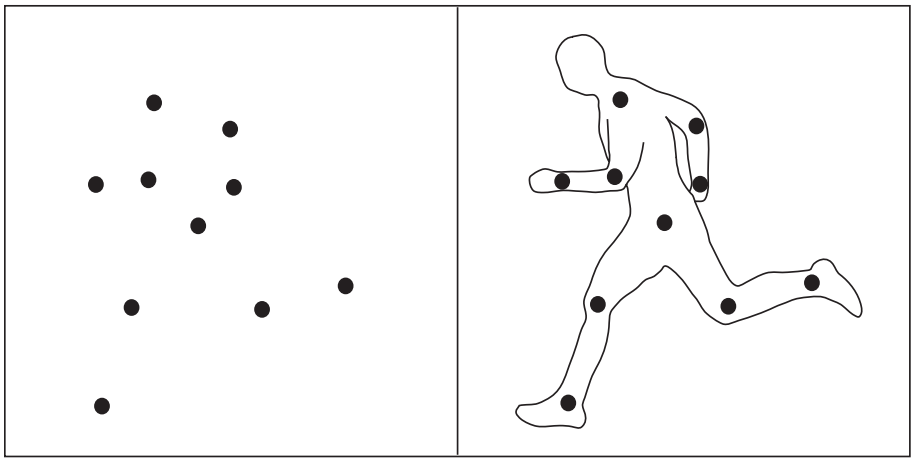

Figure 1.3: Positions of the point-lights on the major joints of the actor. Without the silhouette, a static frame does not contain much information about the movement but a sequence of a few frames is enough for human observers to identify the portrayed action.

(see Figure 1.3). Even though each individual light point undergoes only translational, elliptical and/or pendular motions, human observers were readily able to perceive a human person walking. This experiment showed that the human visual system is able to integrate the local motion signals that are conveyed by the individual light points over time and to organize them into a global percept. Even though for each instantaneous "snapshot" of the movement sequence, a very large number of different groupings and connections are possible, the human visual system organizes and interprets these kind of stimuli in less than 200 ms (Johansson, 1976). Following the first discovery, these stimuli have been used in numerous psychophysical and functional imaging experiments involving humans and animals.

The sensitivity of the human visual system extends beyond simple detection of human locomotion in biological motion displays. In 1977 Cutting and Kozlowski were the first ones to demonstrate that humans are able to judge the gender of the portrayed person and even identify familiar individuals based on their point-light animations (Kozlowski and Cutting, 1977). These studies have been replicated recently showing that humans are able to learn to discriminate between different point-light individuals (Troje et al., 2005). However, biological motion displays do not only convey information about human locomotion. Dittrich (Dittrich, 1993) compared the observers ability to identify different forms of locomotion (like walking, jumping and leaping), instrumental behavior (like hammering, bouncing or stirring) and social actions involving two individuals (like dancing, boxing and greeting). Even though, recognition performance was

best for locomotion, observers also performed way above chance level for the instrumental and the social movements. Moreover, point-light movements even convey information about the emotional state of the actor. Observers could readily recognize different emotions like fear, anger or joy portrayed in point-light animations of professional dancers (Dittrich et al., 1996) and even from animations showing only one arm while drinking or knocking (Pollick et al., 2001).

Evidence from psychophysical studies indicates that the perception of biological motion relies on more than just low-level motion detectors specialized for human movement. Adults easily tolerate changes in temporal delay between the frames (Thornton et al., 1998), or in dot contrast and spatial frequency (Ahlstrom et al., 1997). Moreover, the point-lights do not have to be positioned directly at the joint. Dittrich found only a small decrease in recognition performance when placing the points between two joints instead directly on the joints (Dittrich, 1993). Even distortions caused by embedding an upright point light figure in simultaneous moving dot masks do not reduce the perceptual salience of point light displays substantially. Only very complex masks are able to completely camouflage the figure (Bertenthal and Pinto, 1994; Cutting et al., 1988; Pinto and Shiffrar, 1999; Thornton et al., 1998).

Besides masking studies, also investigations of the effect of display inversion support the hypothesis of global processing of biological motion. Upside-down presentation of biological motion patterns impairs the ability of infants to discriminate between a point light walker and similar displays (Bertenthal et al., 1984). The same holds true for adult observers. When the motion patterns were presented at orientations between 0 and 180 degrees, only an upright oriented walker was reliably identified (Bertenthal and Pinto, 1994; Pavlova and Sokolov, 2000; Shiffrar et al., 1997). Even when observers were aware of the manipulation of the display, or when they had previous experience with upside down displays, their performance was much lower than for upright presentations (Sumi, 1984). Troje suggested that the system for biological motion processing operates in egocentric coordinates. Recognition performance was essentially identical for conditions in which the observer and the stimuli were oriented upright compared to a condition where observer and stimuli were rotated by 90 degrees. If however, the stimuli were presented upright and the observer was lying on the side, or if the stimuli were rotated 90 degrees and the observer was standing upright, performance was significantly reduced (Troje, 2003).

Fox and McDaniel (1982) were the first ones to systematically investigate the perception of biological motion by human infants. Testing three groups of infants at 2, 4 and 6 months they showed that a preference for biological motion

displays appears by 4 months of age. Their finding led to the hypothesis that the sensitivity of the human visual system to biological motion displays might be an innate capacity of the visual system rather than one acquired through experience. A number of psychophysical studies however, support the involvement of learning in the recognition of biological motion. It has been shown, that 3-month-old infants respond to the absolute and relative motions within a single limb, whereas infants by the age of 5 months primarily respond to the relation between the limbs and to their bilateral symmetry (Booth et al., 2002). These findings suggest that the older infants developed perceptual skills that bias them to organize biological motion displays at the level of the human form. Additionally, biological motion recognition seems to improve with experience. Pavlova and colleagues found that recognition of point-light displays representing human and non-human movements by 5-year-old children was substantially improved compared to children at the age of three (Pavlova et al., 2001).

Animals have also been tested with point-light animations. Using this technique, it was shown that cats are able to learn to discriminate point-light animation sequences depicting a cat walking from animations showing the same dot motion but with randomized starting positions (Blake, 1993). Furthermore, pigeons trained to discriminate between video sequences of moving pigeons showed transfer of this learning to point-light displays showing the same movements (Dittrich et al., 1998). Another study showed that pigeons could learn to discriminate between the movement of a pigeon compared to the movement of a toy dog both depicted as point-light animations and some could even transfer from the point-light movement to the real movement (Omori and Watanabe, 1996). Recently Vallortigara and colleagues have used an imprinting procedure to investigate any possible natural predisposition to attend preferentially to biological motion stimuli. In a series of experiments they showed that newly hatched chicks without any previous visual experience show a preference for point-light displays of a hen compared to displays showing random or rigid motion. Interestingly, no preference was observed when the chicks had to choose between the point-light hen and a spatially scrambled version of the same stimulus, leaving the local motion vectors of the individual points unchanged (Vallortigara et al., 2005). These results indicate the existence of a predisposition that helps the young chicks to orient their attention towards a class of stimuli which in a natural environment would most likely correspond to their mother. However, they also show that the predisposition only exists for a general class of stimuli, namely semi-rigid objects, independent of the underlying shape of the stimulus. In Chapter 3, I will use the same stimulus manipulations to investigate, whether the perceived shape actually influences the learning process.

Neurophysiology and Brain Imaging

Functional imaging experiments indicate that the perception of point-light displays leads to activation patterns that are distinct from the activations elicited by rigid motion (e.g. Grossman et al., 2000; Grezes et al., 2001). Several imaging studies using biological motion stimuli have found specific activation of the posterior part of the superior temporal sulcus (Beauchamp et al., 2003; Bonda et al., 1996; Grezes et al., 2001; Grossman et al., 2000; Howard et al., 1996; Puce and Perrett, 2003; Saygin et al., 2004; Vaina et al., 2001). Interestingly, this region shows similar activation for full body movements compared to point-light displays (Grossman and Blake, 2002) and its activation seems to be orientation-dependent with lower activation for inverted- compared to upright point-light stimuli (Grossman and Blake, 2001). Even imagination of biological motion stimuli results in significantly higher activation compared to baseline (Grossman and Blake, 2001).

Furthermore, the presentation of point-light displays results in strong activation in a number of other brain regions. Similar to the results obtained from the superior temporal sulcus, also the fusiform face area (FFA) (Kanwisher et al., 1997) shows higher activation for intact point-light stimuli compared to their scrambled versions. This result has been interpreted in several different ways. Grossman and colleagues (2002) proposed that the FFA might receive direct feedback projections from the STSp. Another possible explanation is provided by Peuskens and colleagues (2005) proposing a model where the complex motion information processed at the level of hMT+ is not only transferred to STSp for the extraction of the action but also to the ventral areas (e.g. the FFA) for the analysis of the figure. A third, completely different explanation for the consistent activation of the FFA when contrasting point-light stimuli with their scrambled versions could result from a theory originally proposed by Gauthier and colleagues in the context of face recognition (Gauthier et al., 1999; Tarr and Gauthier, 2000). According to their theory, the FFA is involved in all types of subordinate level discrimination, if sufficient expertise is present for the given stimulus class, which is certainly the case for biological motion stimuli. However, up to now, no consensus has been reached on this topic.

Even though, the motion related areas hMT+ and the kinetic occipital area (KO) do not show significantly higher activation for point-light compared to scrambled stimuli, they are believed to provide strong afferent connections to the STSp (Grossman et al., 2000; Grezes et al., 2001; Vaina et al., 2001; Howard et al., 1996). During the presentation of biological motion, frequently activation in the cerebellum has also been observed (e.g. Grossman et al., 2000; Vaina et al., 2001). The activity patterns induced by complex movement stimuli are task-

dependent and can change fundamentally if subjects perform a low-level motion vision task, rather than biological motion recognition (Vaina et al., 2001). An imaging study specifically investigating the learning of direction discrimination has demonstrated learning-dependent changes of the activity in different cortical areas, including area MT and in the cerebellum (Vaina et al., 1998).

Recently, also selective activation of premotor areas during the presentation of point-light stimuli has been observed (Saygin et al., 2004). Neurons in these areas seem to be activated during the observation of an action as well as during the execution of the the same action. Although first observed in monkeys, (e.g. Gallese et al., 1996; Rizzolatti et al., 1996b, 2001), studies in humans have also demonstrated the involvement of motor and premotor areas in action observation indicating that humans may use their own motor representations in order to understand the actions of others (e.g. Grafton et al., 1996; Rizzolatti et al., 1996a; Decety et al., 1997).

Neurophysiological studies investigating point-light animations have so far focussed on the superior temporal polysensory area (STPa) in monkeys, a possible homolog of the STSp area in the human. Oram and Perrett (1994) found neurons responding selectively to particular whole body movements (e.g. a walking human). The majority of these cells were also tuned to the direction of walking and responded similarly when the movements were presented as point-light animations. They interpreted their findings in a way that cells in the STPa are able to compute form from motion inputs alone, without relying on the presence of additional form inputs.

1.4 Aim of Thesis and Experimental Questions

As pointed out in the previous sections, the recognition of complex movements is a fundamental function of our everyday life. Results from object recognition, biological motion processing and theoretical modeling indicate that learning could play an important role in shaping the processing of complex movements in the human brain. The aim of this thesis is to systematically investigate visual learning mechanisms, focussing on three main questions:

1. Are humans able to learn to discriminate between very similar complex movement patterns and what are the possible constraints that influence the learning process?

2. What are the neural correlates of the learning process in the different areas of the visual cortex and is there a difference between the learning of human like movements compared to artificial articulated movement patterns at the level of the BOLD response?
3. Is it be possible to implement biologically plausible learning mechanisms into an already existing model for biological movement recognition and can the model be used to simulate the BOLD activity changes obtained from the functional imaging experiments in order to test different hypothesis about the underlying plasticity mechanisms?

According to the three main research questions, the thesis is divided into three experimental chapters.

Chapter 3 reports a series of psychophysical experiments that investigated whether humans are able to learn to differentiate between complex movement patterns. These patterns belonged to three different groups of movements and were all presented as point-light animations. The first group was composed of natural human movements, the second one consists of movements of artificial skeleton models containing nine segments moving in an articulated fashion and the third one consisted of the same human movements as the first one, but this time the spatial positions of the individual points were scrambled. The main focus of the psychophysical investigations was to identify possible differences in the learning process between the three different groups with respect to the time scale of the learning and the invariance properties of the learned representation. To control for the learning history of the subjects, all stimuli were generated by motion morphing. In this way it was possible to create stimuli that were completely novel to the observer. Additionally, motion morphing by linear combination of prototypical movements allowed to precisely control the spatio-temporal similarity between the individual stimuli.

Chapter 4 presents the results of a series of functional imaging experiments that were carried out to identify possible neural correlates of the learning process. Because the stimuli across the conditions that had to be compared were very similar, a special fMRI adaptation paradigm was used to acquire the images. This technique allows to identify possible sub-populations of neurons within the same voxel that contribute to the encoding of different movement stimuli. In addition to the experimental runs, several localizer runs were acquired for every observer to reliably detect the visual areas involved in low-level, mid-level and high-level motion and form processing (namely V1, V2, V3, V3a, VP, V4v, V3b/KO, hMT+/V5, FFA and STSp). By analyzing the fMRI signal separately for each of these areas, it became possible to identify learning processes at all

stages of the visual hierarchy. The main focus of the imaging experiments was to pinpoint learning induced neural plasticity mechanisms in the visual cortex and to identify possible differences between the learning of natural human-like movements compared to artificial articulated movement patterns.

The final experimental chapter of this thesis deals with the theoretical implementation of the experimental results. The theoretical part was based on an already existing neural model for biological motion recognition, which was extended by the implementation of biologically plausible learning rules. Specific detectors of complex form and optic flow fields were learned automatically along with the temporal order with which these features arise during the movement sequence. Additionally, a neural adaptation mechanism was implemented at the highest level of the model. The goal of the theoretical part was to be able to simulate a whole run of a real fMRI experiment and to determine whether the simulated BOLD responses are in accordance with measured BOLD responses. In the future, this model could then be used to test different hypothesis about how learning shapes the processing of complex movements in the human brain.

Chapter 2

Methods

2.1 Psychophysics

The term psychophysics entered the scientific world in 1860, when Gustav Fechner (a German physicist) published his work entitled "Elemente der Psychophysik". In this book, he introduced the concept of a psychological world, that exists next to the physical world. Like in the physical world, where objects can be measured in physical units, corresponding sensations can be measured in psychological units. In the so called Weber's law

$$\Delta\Phi = k\Phi \tag{2.1}$$

he tried to formulate a mathematical relationship between the two, where an increase in a stimulus that is just noticeably different ($\Delta\Phi$) is a constant proportion k of the stimulus. Thereby, measuring a single discrimination threshold for one stimulus intensity predicts the discrimination threshold for many other stimulus intensities. Weber's Law can be applied to many different sensory modalities like brightness, sound, mass, pressure or temperature. While the size of the Weber fraction k varies across modalities, it tends to be a constant within a specific modality. Even though Fechner's idea that sensations might become representable by numbers and therefore, psychology might become an exact science did not hold true, he still was the founder of a "new" psychology, investigating human performance with the help of exact scientific apparatus.

Today, the term psychophysics stands for the study of the relationship between the physical world and the psychological world. By studying questions regarding detection or discrimination thresholds, or reaction times, psychologists try to infer underlying neural processing steps purely from the "outside". The fact that

it is non invasive is one of the main advantages of psychophysics. This makes it possible to study even complex experimental questions using human participants as subjects. However, the drawback is that it is very difficult to to design experiments that are suitable to exactly pinpoint the underlying neuronal mechanisms involved in the processing. Therefore, a combination of psychophysical investigations with other experimental procedures like neurophysiological recordings, functional imaging or theoretical modeling, may advance the understanding the processing mechanisms that form the basis of the observed behavior.

In detection threshold experiments, the stimulus is presented in several intensities and the task for the observer is to report, at which intensity he is able to perceive it. Discrimination threshold experiments measure how different two stimuli have to be, so that the observer can reliably tell the difference. Because the stimuli will be presented in a variety of intensities, psychophysical experiments usually result in the measure of a so called psychometric function. A psychometric function usually resembles a sigmoid function with the observer's response to the stimuli, displayed on the ordinate and the variable stimulus parameter on the abscissa. For example, the measure of detectability used in a psychometric function is the percentage of detections at a given intensity. However, there most certainly will be a range of intensities, for which the observer will be uncertain about the detection of the stimulus. An additional complication when measuring the percentage of detection is that the observer might show a response bias. This bias would shift the whole psychometric function towards the right or left and might mislead the interpretations. A more detailed analysis of the performance of the observer that is independent of a possible response bias will be discussed in the next section.

Signal Detection Theory

A measure of accuracy independent of a possible bias of the observer can be obtained using signal detection theory Swets et al. (1961). In the psychophysical experiments reported in Chapter 3, observers view two complex movement stimuli presented in succession. The task for the observer is to decide whether the two stimuli were identical or whether they were different. The trials, in which the stimuli were actually identical correspond to the signal present condition, whereas the trials, where the two movements were different belong to the signal absent or noise condition. The ability of the observer to discriminate between signal present and signal absent (or noise) trials is represented by the sensitivity index or d-prime (d').

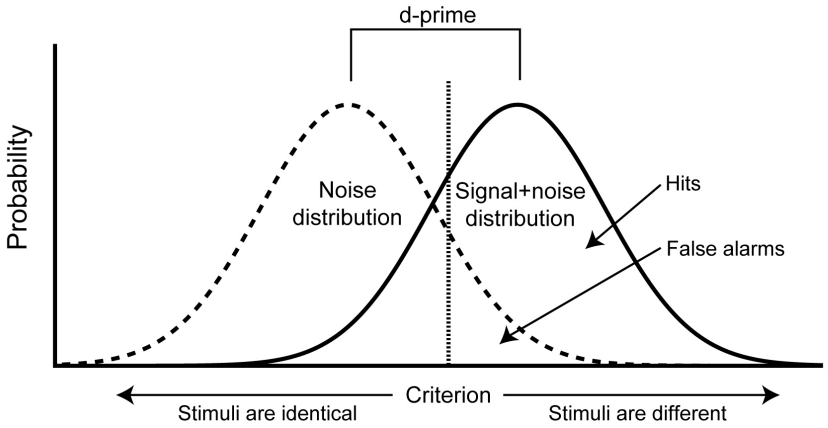

Figure 2.1: Concept of signal detection theory. The dotted line represents the criterion of the observer. Depending on the shift of the criterion towards right or left, the hits and the false alarms increase and decrease, respectively. The measure of d' depends only on the differences of the means of the two distributions and is therefore independent of the criterion and therewith a possible bias of the observer.

The design of the reported experiments is a so called two alternative forced choice (2AFC) paradigm. In terms of signal detection theory the responses of the observer can be classified into four different categories. He could correctly identify the two stimuli as being different (hit, *H*), or as the same (correct rejection, *CR*). Yet, he could also falsely decide that the two stimuli were different even though they were identical (false alarm, *FA*) or judge the two stimuli as being identical, even though they were different (miss, *M*). The measure of d' is independent of a possible response bias of the observer, because it takes into account not only the hits, but also the false alarms. If for example the observer would show a bias for the response "different", we would obtain a high *hit rate* ($\frac{H}{H+M}$), but also a high *false alarm rate* ($\frac{FA}{FA+CR}$), because he would frequently answer "different" even though the stimuli were identical. Exactly the opposite result would be found if the observer shows a bias towards answering "identical".

Because the task is initially very difficult, there will be always some uncertainty for the observer as to whether the stimuli were different (signal present) or identical (signal absent or noise). Whether the observer is sensitive to the difference depends on the stimulus strength but also on the amount of noise present. This noise could come from external sources like the monitor, the stimuli are presented on, but also from internal sources like the visual system of the observer. If the noise is random, sampling it over time will produce a normal distribution, with its mean at the average noise level (the dotted curve in Figure 2.1). If there is a recurring signal present, its energy will be added to the background noise.

Therefore, sampling the signal present trials over time would also result in a normal distribution, but with a higher mean (the solid curve in Figure 2.1). The spread of the two curves depends on the amount of noise in the system. The less the noise present, the more separated the two curves become.

The decision of the observer is mainly influenced by two factors. One of it is the *signal strength* or in our case the difference between the two successive stimuli. If the difference is increased, the task becomes easier, which results in less overlap of the two curves. The second one is called *criterion*. Because the observer has to take a binary decision as to whether the stimuli were identical or different, he will set a level of "difference" based on which he determines if the stimuli were different. The criterion is displayed in Figure 2.1 by a vertical dotted line which divides the graph into four parts corresponding to hits, misses, false alarms, and correct rejections. Even though, the criterion might be influenced by certain expectations or a bias of the observer towards one of the two possible answers, a shift of the vertical line will alter both the hit rate and the false alarm rate together.

The sensitivity index (d') is a measure of the difference between the means of the signal present and the signal absent distributions. It can be calculated solely from the observers hit and false alarm rates. This is accomplished by converting the proportions of hits and false alarms to z-scores (a measure that normalizes for the mean and the variance of a distribution).

$$d' = z(H) - z(FA) \tag{2.2}$$

In our case, larger absolute values of d' would indicate that the observer is more sensitive to the differences between the two presented stimuli. The formula also shows that the value of d' is independent of the criterion the observer has adopted.

2.2 Magnetic Resonance Imaging

In Chapter 4, the results of the functional imaging experiments investigating the neural correlates of the learning of complex movements will be discussed. The present section intends to provide an introduction to the method of fMRI and to highlight certain concepts and processing steps involved in the analysis of functional imaging data.

2.2.1 Principles of Functional Magnetic Resonance Imaging

Basic Physics of Magnetic Resonance Imaging

All protons, neutrons and electrons forming an atom possess an intrinsic angular momentum called spin. While the magnitude of this spin is the same for all of them, they can vary in their main axis of spin. Important for the image acquisition is the spin of the protons. When protons and neutrons combine to form a nucleus, they combine in pairs of oppositely oriented spins leaving atoms with even numbers of protons with no net spin. Hydrogen however, with only one proton at its nucleus has a net spin. This spin leads to the formation of a magnetic dipole, providing the Hydrogen atom with the properties of a small magnet.

The main components of a nuclear magnetic resonance (NMR) experiment are a large magnetic field (B_0), Hydrogen atoms and a coil. The coil has to act as a transmitter as well as a receiver. A single Hydrogen atom can be present in two different spins (+1/2 and -1/2). However, because of the huge number of Hydrogen atoms contained in a small brain volume, in principle, all spin axes are present. If the Hydrogen atoms are placed inside B_0, their dipoles tends to align with the magnetic field. Because of the spin of the proton, the alignment does not happen at once but the spin axis of the proton precesses around the field axis. The frequency of this precession is the resonant frequency of nuclear magnetic resonance and is proportional to the magnetic field. Over time, the dipole tends to gradually align with the magnetic field. The time constant of this relaxation is called T_1. A typical value for T_1 in the human body is around 1 second. In reality, only a small number of dipoles gets oriented along the main field because other forces like thermal motions prevent the dipoles from settling. Nevertheless, this small amount of correctly oriented dipoles creates a weak equilibrium magnetization (M_0) or longitudinal magnetization aligned with the field. M_0 is however very small compared to the strong magnetic field B_0.

In the transmitting state, a brief oscillating current (for a few milliseconds) is applied to the coil which induces an oscillating magnetic field (B_1) in the sample that is perpendicular to B_0. Even though B_1 is much smaller than B_0, applied in the correct frequency (radio frequency, RF) it is sufficient to tip M_0 away from B_0, resulting in transverse magnetization M_T. After the transmission, the coil switches into the receiving state. M_T is now precessing around B_0 and induces a current in the coil. This current decays exponentially over time with a time constant called T_2 or T_2^* because the dipoles start to get out of phase. After some time, the transverse magnetization is no longer measurable, whereas M_0 builds up again along the main magnetic field B_0.

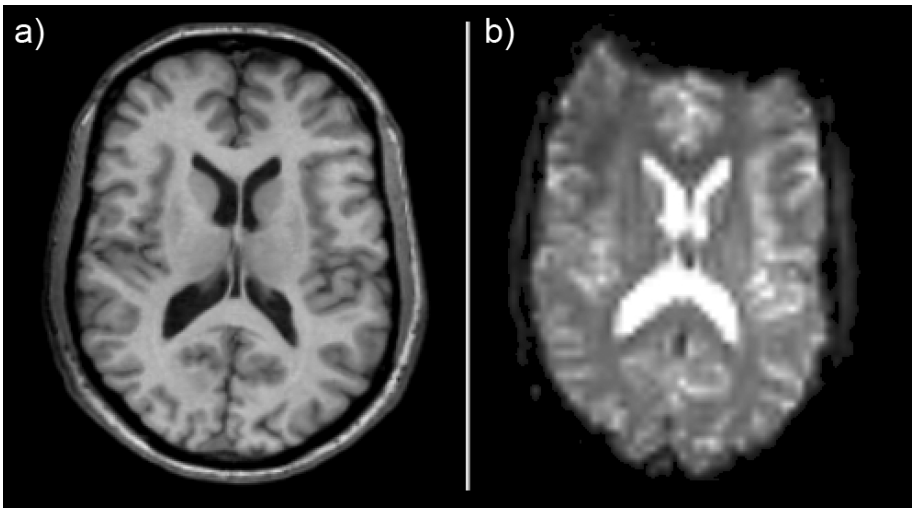

Figure 2.2: MR images of the same section with different tissue contrasts. a) shows a T_1-weighted high resolution volume scan and b) a low resolution T_2^*-weighted image acquired with echo planar imaging (EPI). Depending on the TR chosen, it is possible to highlight different aspects in the image like for example the gray-white matter boundary in a).

Because most of the brain volume is composed of water, the hydrogen atoms in each of the water molecules will be the main source of the signal. However, different parts of the brain contain slightly different amount of water. Nerve cells, for example, are relatively rich in water, whereas the myelin around the long nerve fibers contain less of it. This can be used to generate contrast between the gray matter, the underlying white matter and the cerebrospinal fluid (CSF) of the brain in a proton density weighted image.

Another way to create contrast in the image is by using the two time constants T_1 and T_2^*. At a magnetic field of 1.5 Tesla (which is about 30000 times stronger than the natural magnetic field at the surface of the earth), T_1 corresponds to about 900 ms for gray matter, 700 ms for white matter and 4000 ms for CSF. If the repetition time *TR* (the time between two radio frequency pulses) is short, the signal for white matter will recover more strongly than the signal for CSF, resulting in bright white matter and dark CSF in the image. In a similar way, also T_2 can be used to produce different contrasts. T_2^* for white matter is about 70 ms for gray matter 90 ms and 400 ms for cerebrospinal fluid. When the signal acquisition time is delayed, gray and white matter will appear dark and CSF white (Figure 2.2).

To regain the initial longitudinal magnetization, the next RF pulse can only be applied at a time several times larger than T_1 (about 20 seconds). Because it is crucial to acquire as many images as possible to increase the statistical power of

the results, it is desirable to reduce the time between two RF pulses (repetition time or TR). If TR is small, the recovery of the longitudinal magnetization is incomplete and therefore smaller than the initial M_0. This in turn results in a decreased transverse magnetization and thus the measured MR signal will also be smaller. However, if the repetition time is kept constant, the subsequent MR signals will be smaller than the first one but, nevertheless, constant and are therefore also called steady-state signals.

The location in the brain where the measured MR signal is obtained from, is encoded in three steps for the three main axes x, y and z. Usually encoding the position along the x, y and z axis are called *frequency encoding*, *phase encoding* and *slice encoding*, respectively. Because the resolution with which each encoding step takes place is limited, the final image of the whole brain is described in terms of voxels (rectangular blocks) containing a certain volume of brain tissue.

In principle, each position encoding requires a new RF pulse, leading to a very high number of pulses, each separated by the fixed TR. This leads to very long scanning times for the acquisition of a single image. Since such long acquisition times are not feasible for functional imaging, one possibility to reduce image acquisition time is echo planar imaging (EPI). With this technique, a single RF pulse can be used to generate the full data set for a low resolution image (see Figure 2.2). This means that, depending on the hardware, the image of a single "brain slice" can be acquired within 30 - 100ms giving the opportunity to acquire many slices to cover the whole brain within about 2 seconds.

While EPI imaging is suitable for functional scans, the resolution of the images is low. To relate individual functional differences to specific brain structures, at least one high resolution 3D scan of the whole brain is needed (see also section 2.2.3). A commonly used volume-imaging sequence is called MP-RAGE (Magnetization Prepared Rapid Gradient Echo Image), which acquires one volume in about eight minutes (see Figure 2.2).

Brain Metabolism

Neurons fire action potentials to communicate with each other. Action potentials are based on rapid influx of Sodium (Na^+) ions into the cell membrane downstream a concentration gradient. To maintain the concentration gradient Na^+ ions have to be transported out of the cell and Potassium ions (K^+) into the cell. This transport requires energy, which is delivered mainly in the form of adenosine triphosphate (ATP). The transport of Na^+ and K^+ ions against their

concentration gradient is coupled to the hydrolysis of ATP to adenosine diphosphate (ADP). The Na-K-ATPase (also known as Na/K pump) transports three sodium ions and two potassium ions with the breakdown of each ATP.

Another important mechanism that consumes energy is the transport of calcium ions. The fusion of the synaptic vesicles with the synaptic membrane and thereby the release of neurotransmitters at the synapse is induced by massive influx of calcium ions into the synapse along a concentration gradient. Like for the sodium ions, the calcium ions have to be transported outside the synapse against the concentration gradient, which is again achieved by the consumption of ATP. Additional energy in form of ATP is needed for re-uptake of the neurotransmitters.

To maintain the concentration gradients and thereby the functioning of the neurons, the reservoir of ATP has to be constantly refilled. To restore the supply of ATP, ADP has to be converted back into ATP. This is done in two steps namely, *glycolysis* and the *trans-carboxylic acid cycle* involving glucose and oxygen. The supply of these two molecules is provided by the capillary system.

Taken together, the activity of neurons results in an increase of energy consumption. The restoration of this energy depends on the availability of glucose and oxygen delivered by the blood stream. This results in an increase of cerebral blood flow (CBF) around the activated region. (The exact spatial distribution of this increase is still under debate). For reasons not yet fully understood, the increase in CBF is much higher than the actual cerebral metabolic rate of oxygen consumption ($CMRO_2$) (Fox and Raichle, 1986; Fox et al., 1988). This fact leads to a substantial increase of oxygenated haemoglobin relative to deoxyhaemoglobin in the venous blood at the activated region. The next section will show how this change can be used to measure the blood oxygenation level dependent (BOLD) change for functional magnetic resonance imaging.

Blood Oxygenation Level Dependent fMRI

The term magnetic susceptibility describes the fact that all samples placed in a magnetic field become partly magnetized. The strength of this magnetization is called susceptibility and it depends on the composition of the sample. When materials with different susceptibilities are close together, it leads to distortions of the magnetic field. While this can be a problem for imaging areas where air and bone interface, this effect is essential for functional imaging.

Fully oxygenated blood has about the same susceptibility as other brain tissues but deoxyhaemoglobin is paramagnetic and thus alters the susceptibility

of the blood. An increase in deoxyhaemoglobin therefore leads to local distortions of the static magnetic field B_0 around the blood vessels. Spins in this non-uniform magnetic field now precess at different frequencies causing more rapid phase dispersal and decay of the NMR signal. Therefore, changes in blood oxygenation can cause changes in the MR decay parameter, T_2^*, leading to changes in image intensity in T_2^*-weighted images. This process is the basis of the BOLD contrast on which most fMRI studies are built up.

Although the fact that deoxyhaemoglobin is paramagnetic and leads to local field distortions around red blood cells was known already since 1982 (Thulborn et al., 1982), Ogawa and colleagues provided the first evidence that changes in brain oxygenation could be followed with MR imaging (Ogawa et al., 1990). Shortly afterwards, several studies reported increase in MR signal specific to the stimulated brain area (Bandettini et al., 1992; Frahm et al., 1992; Ogawa et al., 1992). This was the start of functional magnetic resonance imaging based on the BOLD effect.

At a first glance it seems counterintuitive that the local MR signal increases with brain activation based on an increased blood oxygenation. Whereas at rest, about 40% of the blood oxygen gets metabolized, the previous section reviewed that an increase of brain activity leads to a strong increase of CBF, although $CMRO_2$ only increases moderately. By this mechanism, the venous blood at an active state is more oxygenated than at rest.

The measured BOLD response to a brief stimulation is called the impulse hemodynamic response. Measuring the time course of the hemodynamic response typically results in a gradually increasing signal starting about 2 seconds after stimulation onset. This delay has been attributed to the fact that the blood needs some time to flow from arteries to capillaries and draining veins (Kwong et al., 1992). After 6 to 8 seconds, the hemodynamic response reaches a plateau phase and slowly decreases again after stimulation offset. Often the signal does not return straight to baseline but decreases even further resulting in a post-stimulus undershoot for tens of seconds (Buxton et al., 1998; Frahm et al., 1996; Logothetis et al., 1999). The exact shape of the hemodynamic response function differs between observers but also between stimulated regions of the cortex (Glover, 1999).

Neural Correlates of the BOLD Signal

Recently, much research has been dedicated to investigate the neuronal events underlying the BOLD signal changes (Logothetis et al., 2001). Electrophysiological studies in awake behaving animals usually measure extracellular field

potentials (EFPs). The release of an action potential by a neuron is characterized by a local influx of positive ions (e.g. Na^+) into the axon. This so called inward current traveling down the core of the axon or dendrite has to be compensated by an outward current at an inactive site, flowing back through the extracellular medium to the site where the inward current took place. The current flowing within this closed loop is called extracellular field potential and can be measured with an electrode placed nearby in the tissue. A single electrode measures the spatio-temporal mean of all EFPs in its surrounding. One way of extracting the spiking activity of individual neurons from the mean EFP is by placing the electrode very close to the spiking neuron. In this way, the mean EFP is dominated by the individual firing rate of the neuron.

While the main output of a neuron is determined by the action potential, the occurrence and the frequency of an action potential is determined by the integration of sub-threshold local communication between the neuron and its surrounding excitatory and inhibitory neurons, via its dendrites. To separate the local integration processes from the spiking activity, the mean EFP can be divided into two components by applying specific filtering techniques. High-pass filtering (cut-off 300-400 Hz) measures the multiunit spiking activity (MUA), while low-pass filtering (cut-off $<$ 200 Hz) reflects the local synaptic voltages, also called local field potentials (LFP) (Logothetis, 2002, 2003).

To investigate the neural origin of the BOLD response, Logothetis and colleagues measured local field potentials, multi-unit spiking activity and BOLD responses simultaneously (Logothetis et al., 2001). In general, they found a very good correlation of both physiological measurements with the BOLD response. However, investigating the effect of stimulus duration provided a tool to dissociate between LFPs and MUA. While visual stimulation of 12 seconds and 24 seconds resulted in very similar multi-unit activity measurements, the local field potentials as well as the BOLD response showed a prolonged increase for the 24 second stimulation. This was the first evidence that the BOLD response primarily reflects the input and the local processing of neuronal information rather then the output measured in terms of action potentials. However, despite much research in this area, the exact correlation between the measured BOLD signal and the underlying neuronal activity still remains largely unclear.

2.2.2 Experimental Design

A suitable experimental design is critical for any functional imaging study. Usually, only about a few percent of the overall BOLD signal measured by fMRI is related to the experimental question. In this section, the two most commonly

used experimental designs will be described and their strengths and limitations to answer certain questions will be discussed.

Block Design

The most commonly used experimental design in functional imaging studies is called the Block- or Boxcar Design. In this design, multiple trials of the same condition are grouped into blocks, and blocks belonging to different conditions are presented alternatively (see Figure 2.3). Each block is usually presented for 16 to 20 seconds. That is because the hemodynamic response to the stimulus is additive, which means the integral of the hemodynamic response function (HRF) increases, as more stimuli are presented in succession. However, presenting the same condition for a long time, might lead to the result that low frequency noise and linear trends present in the signal, become correlated with one of the conditions.

Ideally, the different conditions presented as alternating blocks differ only with respect to one modality. The logic behind this paradigm is therefore to compare the signal for the individual conditions and those areas showing significant higher activation for one condition compared to the other, should be involved in the processing of this modality. This interpretation is based on the concept of pure insertion, where the change in one modality does not affect the processing of the contents already present in both conditions. Yet, if the assumption fails, it might actually be the case that the subtraction of the conditions highlights the interaction of the modality of interest with all the other processing steps already involved in both conditions, which would mislead the subsequent interpretations (Aguirre and D'Esposito, 2000).

The main disadvantage of the block design is that stimuli from the same condition are presented consecutively for several trials, which makes the stimulus itself very predictable for the observer. Since it is well established that the BOLD signal is strongly modulated by attention (e.g. O'Craven et al., 1997; Ress et al., 2000), this could lead to additional confounds in interpreting the results.

Event Related Design

An alternative approach is to present individual single trails of different conditions, which are spaced in time so that each trial elicits its own hemodynamic response function. Subsequently, all HRFs belonging to the same condition are averaged. In this way it is possible to compare the size and shape of the HRFs across conditions. Savoy and colleagues have shown that stimuli as short

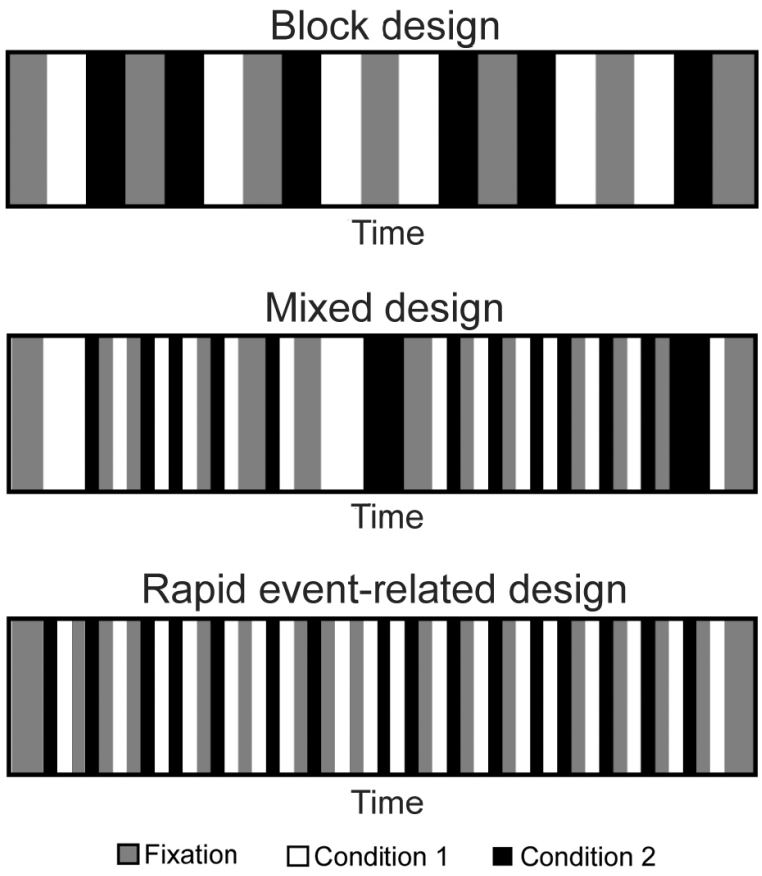

Figure 2.3: Illustration of the different experimental designs in functional imaging experiments.

as 34 ms are capable of eliciting a measurable change of the BOLD response, which means that even stimuli of very short duration can be used in event related designs (Savoy et al., 1995).

The main advantage of this design is that individual trials can be presented in random order because trials of the same condition do not have to be presented in blocks anymore. This allows for a much better control in terms of potential confounds like habituation, anticipation or strategy effects. Another strength of event related designs is that it allows post hoc analyses. For example, trials can be grouped afterwards based on correct or incorrect answers. While it is easier to control this design in terms of attention, the main downside is that only very few HRFs can be measured per condition. Because the change of the BOLD signal in response to a stimulus is very low and noisy, this design has a much lower statistical power compared to a block design.

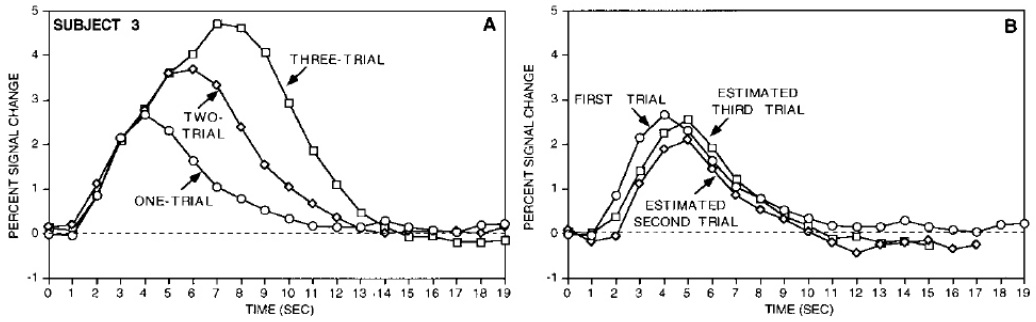

Figure 2.4: A: Raw signal data from one-, two- and three-trial clusters are superimposed onto one graph. It shows a larger and prolonged hemodynamic response as the number of trials increase. B: The estimated contribution of each trial event is shown. The hemodynamic response adds in a roughly linear fashion. (Image from Dale and Buckner, 1997)

A way of increasing the statistical power in event related designs is to decrease the time interval between the individual trials (Figure 2.3). Dale and Buckner (1997) reported that the BOLD signal increases roughly linear, if two or more stimuli are presented in rapid succession (Figure 2.4). Based on this finding, it became possible to decrease the inter-stimulus interval (ISI) because the HRF for an individual stimulus could be estimated despite the overlap in HRF's for successive stimuli. For this design, also called rapid event related design, it is necessary however, to precisely control the order of the individual trials. To avoid possible confounds by preceding trials, the order of the trials has to be counterbalanced to ensure that trials of each condition are preceded and followed by trials of the other conditions equally often.

fMRI Adaptation

Because the resolution of functional imaging scans is limited, the BOLD activity changes reported are actually mean activities averaged over many thousands of neurons. These neurons can be grouped into separate populations, all of which having different response characteristics. One problem with classical functional imaging designs is that due to the averaging it becomes impossible to visualize these differences. One possibility to circumvent this problem is a design called fMRI adaptation. This method is based on the finding that the repetition of the same stimulus over time results in a decrease of neural activity, which is evident at the level of single neurons as well as the hemodynamic response (e.g. Li et al., 1993; Henson et al., 2000). Applied in the correct way, this repetition suppres-

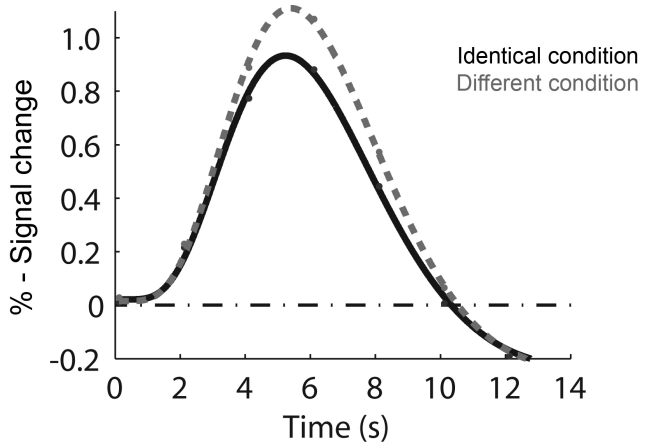

Figure 2.5: Fitted time courses of the hemodynamic response to the presentation of two identical point-light stimuli and two different point-light stimuli.

sion or adaptation effect can be utilized to characterize neuronal populations at a spatial resolution beyond the single voxel (e.g. Grill-Spector and Malach, 2001).

The basic paradigm contains two conditions. In the first condition, identical stimuli are presented, leading to an immediate decrease of the fMRI signal. The second condition holds stimuli that are varied along one dimension. If the neural populations within one voxel are insensitive to the stimulus change, the measured fMRI signal would be similar to the one obtained from the first condition. If however, the neurons are sensitive to the transformation, the neural response would not adapt and the measured signal would be higher compared to the first condition (Figure 2.5).

fMRI adaptation has been widely used by researchers investigating the encoding of various stimulus properties in the visual cortex. Recent human fMRI studies have tested specifically, whether the obtained adaptation effects for different stimulus features are consistent with results from electrophysiological experiments with monkeys, e.g. to the direction of motion (e.g. Huk and Heeger, 2000). Other human fMRI studies used adaptation paradigms to test for selectivity to shape in higher visual areas, e.g. in the lateral occipital complex (LOC) (e.g. Malach et al., 1995). The same technique was used to test the effect of different stimulus transformations, namely position, size, orientation, and illumination change, on the BOLD signal in the LOC (e.g. Grill-Spector et al., 1999). Adaptation effects have also been observed in higher cognitive tasks, such as semantic classification of objects (Buckner and Koutstaal, 1998) and procedural motor learning (Karni et al., 1995). The fMRI adaptation paradigm is also particularly useful in learning studies. With this technique it would become possible to investigate whether sensitivity to trained stimulus changes would be accompanied by an increase of the fMRI signal after learning.

Several models have been proposed to explain how and why adaptation effects occur (see Grill-Spector et al., 2006, for review). One possibility would be that all neurons within a population show a similar degree of reduction in their firing rate in response to repeated presentations of the same stimulus. This so called *"fatigue model"*, would result in a general decrease of activity that would affect all neurons in the same way (Grill-Spector and Malach, 2001). A different explanation for reduced activity has been proposed by Desimone and colleagues. In their *"sharpening model"* repetition effects are explained in terms of a learning process in which tuning curves are sharpened resulting in a sparser representation of the stimulus (Desimone, 1996; Wiggs and Martin, 1998). A third model called *"facilitation model"* illustrates adaptation as a result of faster processing and shorter latencies or durations of neural firing (James and Gauthier, 2006). However, additional research in neurophysiology as well as functional imaging is needed to verify these theories.

2.2.3 Data Analysis

Prior to every fMRI experiment, one has to decide, on the areas of brain that are of interest. During the scan, a predefined number of slices will be scanned covering that area (Figure 2.6). Depending on the size of the area, the distance between each of the slices and the speed of the scanner, the whole set of slices, called one volume, will be acquired within 1 to 4 seconds. In the end, the whole fMRI experiment consists of hundreds of such volumes. This creates several complications when analyzing the data. Since the slices are acquired individually, there will be a time difference of several seconds between the most ventral and the most dorsal slices. Another problem arises from head movement since the subjects might slightly change position over the course of the whole experiment. Furthermore, linear or non-linear trends might be evident during the experiment because of changes in the biorhythm of the subject or changes in physical parameters like temperature etc.. Finally, in order to circumvent individual variations of the brain anatomy, the data has to be spatially normalized

Slice Timing Correction

The fact that the different slices are collected at different points in time means that the corresponding hemodynamic response functions are acquired at different points with respect to time. Hence the HRF appears to rise faster for a voxel that is collected later in time compared to the voxel in an previous slice. With the help of slice timing correction, the data can be modified in a way that the HRF is

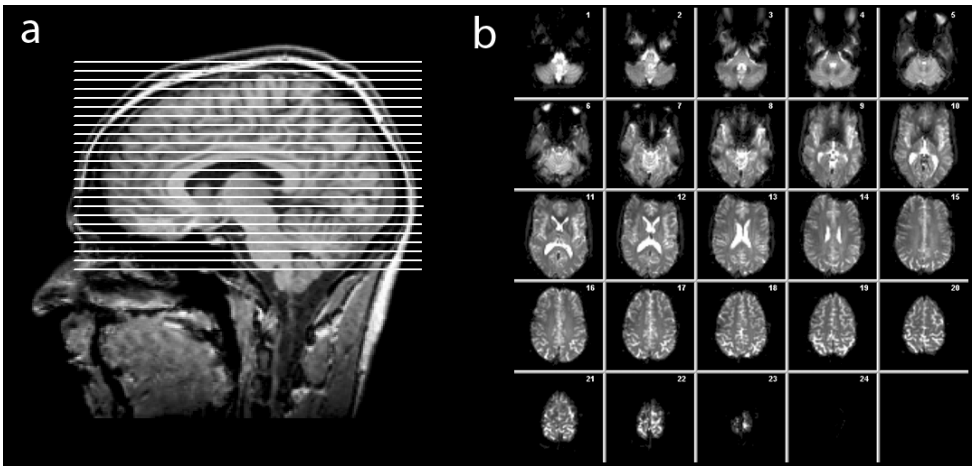

Figure 2.6: a: Sagital section of the brain. The white lines show the positions of the 24 slices per volume during the fMRI experiment. b: Presented are the 24 individual slices acquired per volume.

sampled for all slices at the same time. This is necessary to be performed before the statistical analysis, where all slices are assumed to be acquired at once.

For this purpose, one slice of the volume has to be chosen as a reference slice. Usually this will be either the first or the middle slice of each volume. Following, all other slices of the volume will be shifted in time to match with the reference slice by interpolation. All slices acquired earlier in time than the reference slice will be weighted with their corresponding slice in the subsequent volume. Yet, all slices measured after the reference slice will be weighted with their corresponding slice in the previous volume. The only slice that is not shifted is the reference slice. In this way, the time series of each voxel will be shifted in time, as if the whole volume was acquired at the same point in time. Alternatively, instead of only using only one model for the hemodynamic response function, one can additionally include its temporal derivative to allow for deviations in the timing (Smith, 2001).

Certainly, the interpolation will lead to the introduction of artifacts into the data. These errors will be most prominent for slices that have to be shifted the most, that is the ones lying most distal from the reference slice. Therefore, it is important to maximally reduce the acquisition time per volume and to choose the reference slice in a way that it overlaps with the areas where the effects are expected to occur.

Motion Correction

Because the whole fMRI scan lasts for several minutes, head movement has to be accounted for. The result of the head movements is that individual voxels might not correspond to the same location in the brain throughout the scan. This however, will make the data very difficult to interpret because every comparison across conditions is based on the assumption that the area covered by each voxel does not change. In general, head movement will introduce an additional source of noise, which reduces statistical power. In the worst case, the head movement could be correlated with one of the conditions, which might lead to significant activation differences not due to the stimuli or task (Brammer, 2001).

To minimize the influence of head motion, all volumes are realigned to a reference volume. The realignment process works as a rigid transformation not affecting the size or shape of the brain, but only allows transformation and rotation of the whole volume in x, y and z direction. The new value of the fMRI signal after realignment for each voxel is calculated by interpolating the signal of neighboring voxels.

The motion correction cannot be accomplished completely, if the head movement is too large. In this case the algorithm which tries to minimize the least square difference between the individual volume and the reference volume, might stay in a local minimum. Another problem is that the brain is not completely rigid. Due to heart beat and respiration, the shape and size of the brain might vary slightly. Also the imaging process itself might interfere with the motion correction because head movements change the overall magnetic field leading to inhomogeneities causing reduced signal in the affected areas. Finally, the interpolation method to estimate the fMRI signal in the voxels leads to interpolation errors, further reducing the quality of the data (Ashburner and Friston, 2000; Frackowiak et al., 1997).

Temporal Filtering

Changes in global blood flow, changes of physical parameters or changes in the scanner hardware can alter the mean intensity of the image independent of functional activity (see Figure 2.7). One possibility to reduce this noise is by high pass filtering the signal. This high-pass filtering will remove low frequency variations in the data, without significantly altering signals correlated to the stimulus. For each pixel in the image, the time course is extracted, Fourier transformed, multiplied by a Gaussian filter, and inverse Fourier transformed. A possible confound might arise, if only very few blocks per condition are acquired and if the

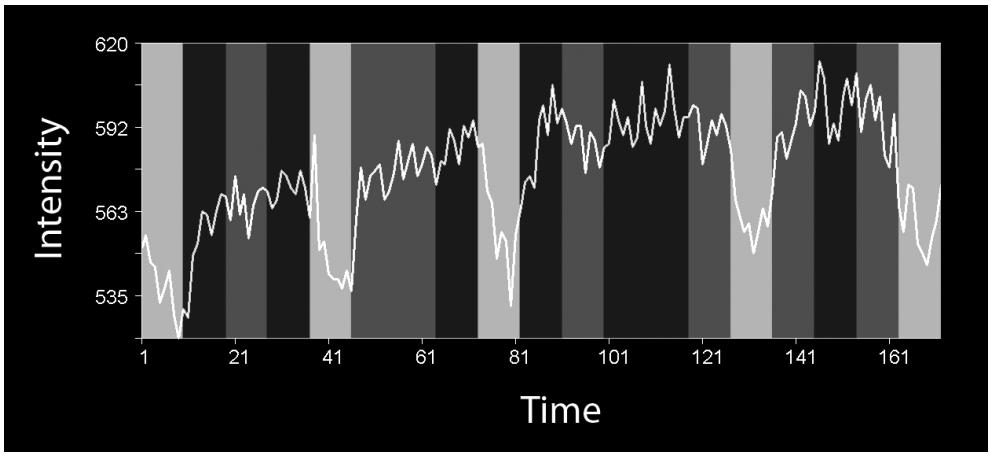

Figure 2.7: The figure shows the changes in BOLD intensity during the course of the experiment. While the signal varies across conditions as expected, one can also observe a linear trend with generally lower intensities at the beginning of the experiment and higher ones in the end. This trend can be removed using temporal filtering.

conditions are additionally separated in time. In this case, temporal filtering might reduce the differences between the conditions because they would appear in the low frequency domain.

Spatial Smoothing

To use a multi-subject analysis it might be useful to spatially smooth the data. In this way, the fMRI data will be blurred prior to statistical analysis. This might initially reduce spatial resolution, but it eventually results in an increased signal-to-noise ratio by reducing high frequency noise. Also small differences in frequencies will be reduced making the comparison between several subjects easier.

To smoothen the data, the volume data is convolved with a 3D Gaussian kernel, the width of which has to be adjusted to the spatial resolution of the fMRI data. The smoothing kernel is defined by the Full Width Half Maximum (FWHM), which is usually two to three times larger than the voxel size. If the size of the area of interest is known, one can adjust the FWHM accordingly since the size of the FWHM determines the spatial clustering that is emphasized.

Spatial Normalization

An fMRI experiment requires data acquisition from several subjects to increase statistical power. However, individual variations with respect to the size and

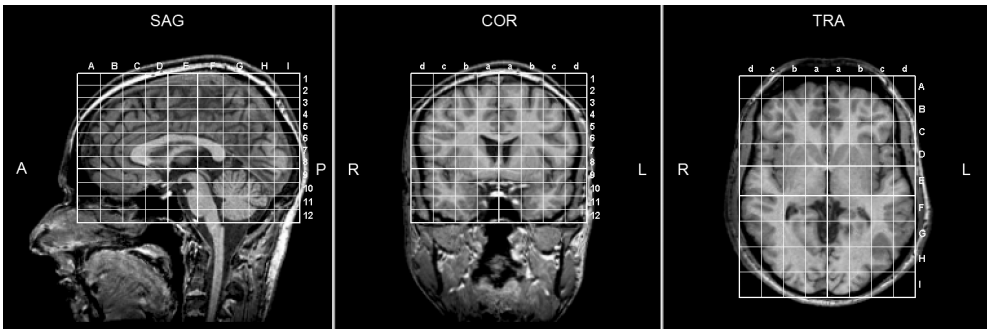

Figure 2.8: The figure displays the talairach coordinate system superimosed on a high resolution 3D scan. Assigning locations of activation in terms of Talairach space allows to compare results between different research groups.

organization of the brain requires the alignment of all individual brains to one standard brain. In this way, activations can be compared across subjects and inferences about underlying brain structures can be made.

Spatial normalization between the individual brain and a template brain to match overall size, shape and orientation is done using 12 linear parameters (translation in x, y, and z direction, roll, pitch, yaw and resizing in all directions and shearing in all directions). Each linear transformation affects the entire 3D volume in the same way. This results in a good fit of the overall structure. To correct for individual parts of the brain that might be shaped differently between the subjects, it is possible to apply a set of nonlinear transformations after the initial normalization step (Jenkinson, 2001).

A template brain that allows to compare specific activation sites across research groups is the Talairach brain or the Talairach space (Figure 2.8). In this space, the anterior commissure serves as reference point and every voxel in the brain can be referred to as the distance in x, y and z direction from this point. The problem with this method and spatial normalization in general is that individual variability across different brains is very large and therefore a precise matching of gyri and sulci would create severe local distortions. To reduce this distortions, most normalization software smoothes the images before normalization. Therefore, even though the brains are normalized, individual gyri and sulci do not match exactly. As a consequence, stereotaxic coordinates do not necessarily refer to a specific location along a particular sulcus. Because of this, stereotaxic coordinates have also been referred to as probabilistic (Mazziotta et al., 1995). An additional problem in inter-individual comparisons is that brain functions may not always be found at a specific location of the sulcus. Zilles and colleagues noted that ”sulci are not generally valid landmarks of the microstructural

organization of the cortex" (Zilles et al., 1997). However, up to now, describing brain function in terms of stereotaxic coordinates and cortical cytoarchitecture is the best way to compare between individual subjects.

2.3 Motion Morphing

All the stimuli used in the experiments reported in the following chapters of the thesis were derived from motion morphing. Motion morphing algorithms interpolate between prototypical trajectories, resulting in new trajectories that blend between the properties of the prototypes (e.g. Bruderlin and Williams, 1995; Wiley and Hahn, 1997). The method used to generate the morphed stimuli for the experiments is called *spatio-temporal morphable models* and was developed by Giese and Poggio in 2000. This method allows the generation of new trajectories by linear combination of different prototypical trajectories in space-time.

A complex movement presented as a point-light display is composed of individual movement of 10 to 12 feature points. Each of these points moves along a certain trajectory over time. When two different movements are combined, the spatio-temporal correspondence problem between these points has to be solved. That is, corresponding points in space-time of the trajectories of the dots between the two movement patterns have to be defined. These points differ with respect to space and time. For example, the maximum extension of the ankle during a walking movement and a running movement are corresponding points. If the ankle point of the walking movement serves as a reference, the corresponding ankle point of the running movement will be characterized by a shift in time, because the ankle will reach its maximum extension earlier in the gait cycle. It is also characterized by a shift in space because the ankle might be extended farther.

Figure 2.9 depicts the concept of spatio-temporal correspondence. The trajectories of the prototypical movement patterns n can be described by the time-dependent vector $x_n(t)$. The different trajectories x_1 and x_2 differ from each other by the spatial shifts $\xi(t)$ and the temporal shifts $\tau(t)$. The transformation, by which x_2 can be transformed into x_1 can be described as

$$x_2(t) = x_1(t + \tau(t)) + \xi(t) \tag{2.3}$$

The spatio-temporal correspondences are calculated by minimizing the sum of the spatial and the temporal shifts under the constraint, that the temporal shifts define a new time variable that is always monotonically increasing. Further details about the underlying algorithm are described in Giese and Poggio (2000).

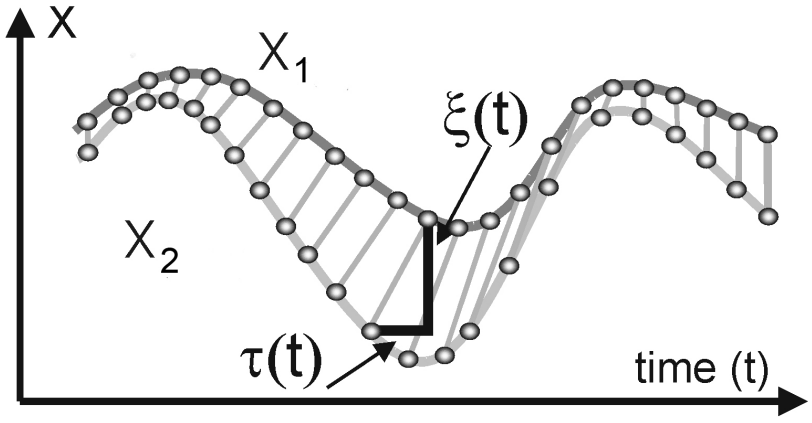

Figure 2.9: The figure shows the concept of spatio-temporal correspondence. Spatio-temporal correspondence between two trajectories $x_1(t)$ and $x_2(t)$ is defined by the spatial shifts $\xi(t)$ and the temporal shifts $\tau(t)$ that map the two trajectories onto each other. These shifts are indicated by the gray lines. (Figure modified from Giese and Poggio (2000).

To generate smooth interpolations between several movement patterns, one of the movements has to be defined as the reference trajectory and the spatio-temporal shifts for the other movements will be computed with respect to that reference movement. In a next step, linearly combined spatial and temporal shifts between the individual movement n and the reference movement are computed as:

$$\boldsymbol{\xi}(t) = \sum_{n=1}^{N} w_n \boldsymbol{\xi}_n(t) \tag{2.4}$$

$$\tau(t) = \sum_{n=1}^{N} w_n \tau_n(t) \tag{2.5}$$

The weigths w_n define the contribution of the individual movement to the linear combination. For all the experiments reported in this thesis, the sum of the weights was constrained to 1. The new trajectory of the morphed movement pattern can be computed with the help of Equation 2.3 , where x_1 refers to the reference movement, and by applying the spatial and temporal shifts $\boldsymbol{\xi}(t)$ and $\tau(t)$ to the reference movement.

The technique of spatio-temporal morphable models has already been used in a number of psychophysical experiments (e.g. Giese and Lappe, 2002; Giese et al., 2003). In these experiments it has been verified that this technique results in smooth interpolations between different movement patterns like walking and running. Moreover, when observers were asked to rate the naturalness of the

presented movements, their rating for the individual prototypical movements was very similar to their rating for the morphed stimuli. Additionally, the results obtained from observers judging the perceptual similarity between several morphed stimuli matched quite well with the actual distance between these movements in the morphing space.

These results motivated the use of spatio-temporal morphs to investigate learning processes in the recognition of complex movements. The following chapter reports a series of psychophysical experiments, where these stimuli have been applied.

Chapter 3

Psychophysical Investigations

This chapter reports the results of a series of psychophysical experiments to investigate the involvement of learning mechanisms for the recognition of complex movements. The main research questions that will be discussed are:

1. Are humans able to learn to discriminate between very similar complex movement patterns?
2. Does the learning process depend on the perceived naturalness of the stimuli or on the compatibility of the movement with human kinematics?
3. Do the learned representations for the different stimuli share the same invariant properties, known from natural movement recognition?

To investigate these questions, we trained human observers to discriminate novel motion patterns that were generated by motion morphing. We tested the learning of different classes of novel movement stimuli. One group of stimuli was fully consistent with human movements. A second class of stimuli was based on artificial skeleton models that were inconsistent with human and animal bodies. A third group of stimuli specified the same local motion information as human movements, but was inconsistent with an underlying articulated shape. Participants learned both classes of articulated movements very fast in an orientation-dependent manner. Learning speed and accuracy were strikingly similar, and independent of the similarity of the stimuli with biologically relevant body shapes. For the class of stimuli without underlying articulated shape, however, we did not observe significant improvements of the recognition performance after training. Our results indicate the existence of a fast visual learning process for complex articulated movement patterns, which likely forms the basis of biological motion perception. This process seems to operate independently of

the consistency of the patterns with biologically relevant body shapes. However, it seems to require the compatibility of the learned movements with a global underlying shape[1].

3.1 Introduction

The ability to recognize complex movements and actions is critical for the survival of many species. Consequently, the human visual system is very skilled in the extraction of information from movements, even for strongly impoverished stimuli like point-light displays (Johansson, 1973). Neonates are able to imitate facial and manual gestures (Meltzoff and Moore, 1977), suggesting that the recognition of complex movements might, at least partially, depend on innate mechanisms for the processing of biologically important human movements (Fox and McDaniel, 1982).

However, motion recognition might also critically depend on visual learning. A dependence on learning has been demonstrated for a broad range of visual tasks ranging from lower-level vision tasks like orientation discrimination, hyperacuity or direction discrimination (e.g. Ball and Sekuler, 1987; Mayer, 1983; Poggio et al., 1992) to higher level tasks such as face- or object recognition (see Fine and Jacobs, 2002; Goldstone, 1998; Tarr and Bülthoff, 1998, for review). Studies on object recognition indicate that observers are able to learn novel complex shapes, and that the resulting representation shows view-dependence (Edelman and Bülthoff, 1992). Moreover, neurophysiological experiments in monkeys support the existence of neurons in the inferotemporal cortex that learn to respond selectively to novel complex 3D shapes (Logothetis et al., 1995). Many of these neurons show view-dependent tuning.

The central role of learning in the visual recognition of complex shapes motivates the hypothesis that the recognition of complex movement patterns might also be based on learning. Evidence supporting this hypothesis was provided by studies showing that human observers learn to recognize individuals based on their facial or full-body movements (e.g. Hill and Pollick, 2000; Kozlowski and Cutting, 1977; O'Toole et al., 2002; Troje et al., 2005). Moreover, the detection of point-light walkers in dynamic noise can be improved through visual learning (Grossman et al., 2004). Furthermore, the recognition of biological motion is dependent on stimulus orientation, like the recognition of stationary objects

[1]The paper reporting the work presented in this chapter is currently under review at the *Journal of Vision*. (Jastorff J, Kourtzi Z, Giese MA: Learning to recognize complex movements: Biological vs. artificial trajectories)

(Bertenthal et al., 1987; Pavlova and Sokolov, 2000; Sumi, 1984). Consistent with these psychophysical findings, biological-motion sensitive neurons in the superior temporal sulcus (STS) of monkeys show view-dependent modulation of their firing rate (Perrett et al., 1985), and imaging studies indicate reduced fMRI activity in human STS for the presentation of inverted point-light walkers (Grossman and Blake, 2001). This suggests that complex movements and static shapes might be encoded by similar view-dependent mechanisms (c.f. Verfaillie et al., 1994). That such learning mechanisms provide a computationally powerful explanation of biological motion recognition has been shown by theoretical models that account for many experimental results (Giese, 2000; Giese and Poggio, 2003).

Our study investigated the learning of complex motion patterns that were either biologically relevant or artificial. One class of non-biological stimuli was articulated, but based on skeleton models without biological relevance. The other class of artificial stimuli had the same local motion information as human motion, but no underlying global shape. All three stimulus classes were carefully balanced in terms of low-level stimulus properties. In particular, we tried to address the question which biological properties of complex motion stimuli might be critical for the learning process. For the generation of stimuli with highly controlled spatio-temporal properties we exploited new techniques for trajectory manipulation by motion morphing, and for the approximation of such morphs by real human movements.

We conducted four main experiments and two control experiments. Experiment 1 shows that humans can learn novel articulated movements very quickly. In addition, it demonstrates that human-like and artificial articulated movements are learned equally fast and accurate. The similarity of the learning of the two stimulus classes is confirmed by Experiment 2, which demonstrates that the learned representation for both stimulus types is orientation-dependent, like normal biological motion perception. Experiment 3 rules out the possibility that the observed similarity between the learning of human-like and artificial articulated movement patterns is due to the fact that the both stimulus classes were generated by motion morphing. Consistent with Experiment 1, we obtained striking similarities between the learning of (non-morphed) real human movements and artificial patterns. Experiment 4 demonstrates that the presence of an underlying global shape is crucial for the fast learning of complex movement patterns.

3.2 Experimental Specifications

3.2.1 Participants

34 individuals (15 male, 19 female, mean age: 27.6 years) participated in this study (11 in Experiment 1, 9 in Experiment 2, 7 in Experiment 3 and 7 in Experiment 4). All participants had normal or corrected-to-normal vision. Many of them had participated in psychophysical experiments before, but no one had ever been exposed to the same or similar morphing stimuli. Participants were tested individually and gave written informed consent to participate in the study and were paid for participation.

3.2.2 Stimuli

Visual Stimulus Presentation

The stimuli were presented as point-light walkers consisting of 10 dots. The dot trajectories were generated by motion morphing (see below) between three prototype trajectories (natural human or artificial movements). The stimuli were displayed using an Apple Macintosh G4 computer and a Sony color monitor (75 Hz frame rate; 1024x768 pixels resolution) that was viewed binocularly from a distance of 40 cm. Stimulus presentation and recording of the participants' responses was accomplished with the MATLAB Psychophysics Toolbox (Brainard, 1997). The stimuli were shown as black dots on a gray background, and each dot had a diameter of 0.5 degrees visual angle (Figure 3.1). In order to prevent participants from using low-level strategies for accomplishing the task, the stimulus dots were not presented on the major joints. Instead, for every frame the dot positions were chosen randomly and uniformly distributed along the bone segments that were immediately adjacent to the relevant joint (c.f. Beintema and Lappe, 2002, for a similar stimulus manipulation). The maximum displacements were 30 % of the bone length away from the joints. This manipulation does not disrupt the perception of biological motion. However, it efficiently prohibits the use of local strategies, like comparing the relative positions of individual dot pairs (Jastorff et al., 2003). The lifetime of the dots was 1 frame. The size of the stimuli was 5 x 10 degrees, and their position was randomized within an area of ± 2 degrees horizontally and vertically. During each trial, the stimuli were presented for four movement cycles (90 frames per cycle).

Tracking of Human Movements

Prototypical human movement trajectories were obtained by tracking the two-dimensional joint positions in videos showing an actor performing different movements. The actor moved along a line that was orthogonal to the view direction of the camera. All movements were executed periodically for multiple cycles, but only a single cycle of each movement was used for motion morphing. First the translation of the body center was subtracted by fitting the translation of the hip with a linear function of time. The resulting movement looks like a person performing the movement on a treadmill. The body points that were tracked manually were the head, shoulders, elbows, wrists, hip, knees and ankles. For the generation of the point-light stimuli the positions of the shoulder and the head markers were averaged, resulting in a stimulus with 10 dots. The tracked trajectories were time-normalized and smoothed by fitting them with a second order Fourier series. The resulting periodic trajectories served as prototypes for the generation of human-like novel movements. We used four groups of human movements, each containing three prototypical movements.

Generation of Artificial Articulated Movements

Prototypical trajectories for the artificial stimuli were generated by animation of multiple artificial skeleton models with 9 bones that were linked in the same way as a human skeleton (Figure 3.1b). The shapes of the skeletons were chosen to be highly dissimilar from biologically relevant body structures. The trajectories of the joint angles $\alpha_n(t)$ of the 8 joints of the skeletons were given by sinusoidal functions of time:

$$\alpha_n(t) = \alpha_n + \beta_n sin(\omega t + \phi_n) \tag{3.1}$$

The frequency ω, and amplitudes β_n were matched with typical values of joint trajectories of human actors during natural movements. In this way, low-level properties for human-like and artificial stimuli were coarsely balanced. Twelve different prototypical artificial movements were generated and divided into four groups, each containing three prototypes.

Motion Morphing

Motion morphing algorithms interpolate between prototypical trajectories, resulting in new trajectories that blend between the style properties of the prototypes (e.g. Bruderlin and Williams, 1995; Wiley and Hahn, 1997). Recently,

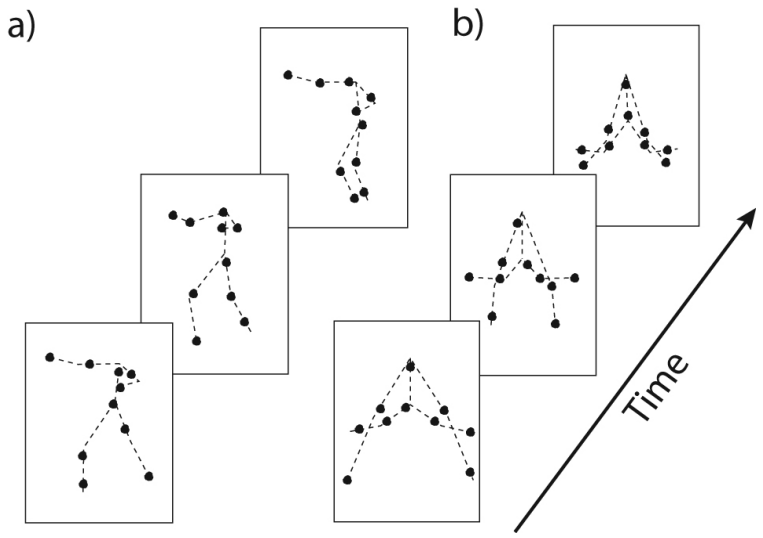

Figure 3.1: Human-like and artificial articulated stimuli. Snapshots from the sequence of a human-like (a) and an artificial articulated (b) point-light stimulus. Point positions were randomly displaced along the bones of the skeleton in each frame to minimize the influence of low-level strategies (dashed lines not shown during the experiment).

such methods have been exploited to generate stimuli for psychophysical experiments (Giese and Lappe, 2002; Troje, 2002). For our experiment we applied a method already introduced in Section 2.3 that creates new trajectories by linear combination of three prototypical trajectories in space-time (Giese and Poggio, 2000). It has been shown that this method results in natural-looking morphs for interpolations between different natural gaits (see Giese and Lappe, 2002, for details). Formally, the morphs were given by the equation

$$\text{New motion pattern} = c_1\ (\text{Prototype 1}) + c_2\ (\text{Prototype 2}) + c_3\ (\text{Prototype 3})$$

where the weights c_i determine how much the individual prototypes contribute to the morph. When the weight of a prototype is high, the linear combination strongly resembles this prototype. (Weight combinations always fulfilled $c_1 + c_2 + c_3 = 1$). The weight vectors (c_1, c_2, c_3) define a metric space, and the distances between these vectors provide a measure for the spatio-temporal similarity of the corresponding trajectories. By varying the distance between the weight vectors, we were able to control the difficulty of the stimulus discrimination precisely.

Using this algorithm, we generated three different stimulus groups for each type of movement (human-like and artificial patterns). The first group, called

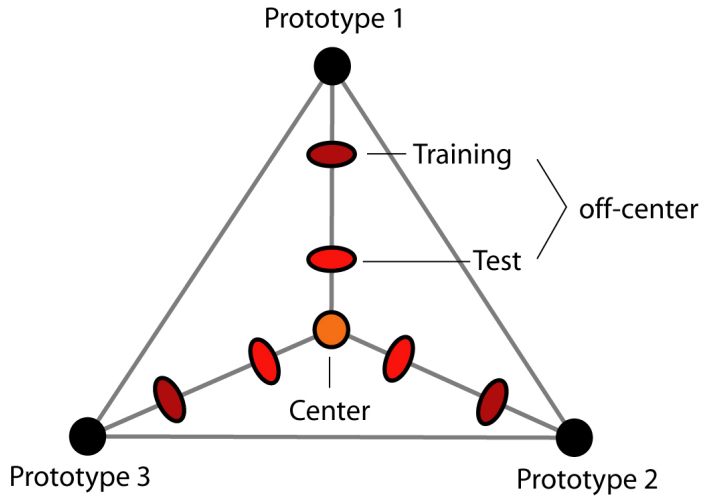

Figure 3.2: Pattern space defined by motion morphing. Morphs were generated by linear combination of the joint trajectories of three prototypical patterns (Prototypes 1 to 3). Three groups of stimuli were generated by choosing different combinations of linear weights: Center Stimuli with equal weights of all three prototypes; Test Off-Center Stimuli with one prototype weighted slightly higher than the others; and Training Off-Center Stimuli for which the weight of one prototype significantly exceeded the other two.

Center stimuli in the following, is characterized by equal weights of all prototypes in the linear combination ($c_1 = c_2 = c_3$). The other two groups were characterized by non-equal weights of the prototypes in the motion morph. For the group of *Test Off-Center* stimuli, the weights for one prototype only slightly exceeded the weights of the other two ($c_i > c_j = c_k; i, j, k$ defining the three prototypes). For the third group, called *Training Off-Center* stimuli in the following, one prototype was weighted significantly higher than the others ($c_i >> c_j = c_k$) (Figure 3.2).

We confirmed that the physical distances between the trajectories of the Center and Test Off-Center stimuli for human-like and artificial stimuli were comparable by computing the mean Euclidean distance between the dot trajectories (distance 0.073 for human-like and scrambled human-like stimuli vs. 0.085 for the artificial patterns). This makes it very unlikely that our results can be explained by simple low-level motion or spatial differences between the dot trajectories.

The triples of human movements serving as prototypes for the generation of the human-like stimuli were carefully chosen to guarantee that the resulting interpolations looked like a human actor could execute them (e.g. by morphing between three different types of boxing or locomotion). For the generation of

the artificial movements, we chose triples of prototypes leading to smooth interpolated movements without particularly salient features. Example stimuli can be downloaded from: *http://www.uni-tuebingen.de/uni/knv/arl/demo/index.htm*

3.2.3 Procedure

For each observer, two of the four human-like stimulus groups and two out of the four artificial stimulus groups were chosen randomly. The experiment started with a brief practice session of four trials, showing four example stimuli without feedback (2 human-like and 2 artificial stimuli). Participants were instructed to respond immediately after making their decision. However, no explicit time constraint was imposed.

In all trials participants had to compare *Center* stimuli with *Off-Center* stimuli in a pair comparison paradigm. Each trial started with the presentation of a Center Stimulus for four gait cycles, followed either by the same Center Stimulus or by an Off-Center Stimulus (generated from the same triple of prototypes, four movement cycles). The prototype contributing with the highest weight to the Off-Center Stimulus was chosen randomly. In a two-alternative forced choice paradigm, participants had to report whether the second stimulus matched the first one.

The experiment consisted of three test blocks that were interleaved by two training blocks. In the test blocks, Center stimuli had to be discriminated from Test Off-Center stimuli. Each stimulus group was presented 3 times in random order, resulting in 12 trials overall. During test trials no feedback about correct discrimination was provided. Based on a pilot experiment with a different set of observers, we adjusted the similarity of the Test Off-Center to the Center Stimuli in order to achieve an average performance level of about 60 % before training. The training blocks consisted of 32 trials (8 repetitions per stimulus group). During training participants had to discriminate between Center and Training Off-Center stimuli and received feedback about their performance.

3.2.4 Approximation of Motion Morphs by Real Human Movements

We have developed a special method to approximate trajectories generated by motion morphing by real human movements. Tests with such approximating stimuli are important because even human-like motion morphs, generated by space-time interpolation, might violate important constraints of real human

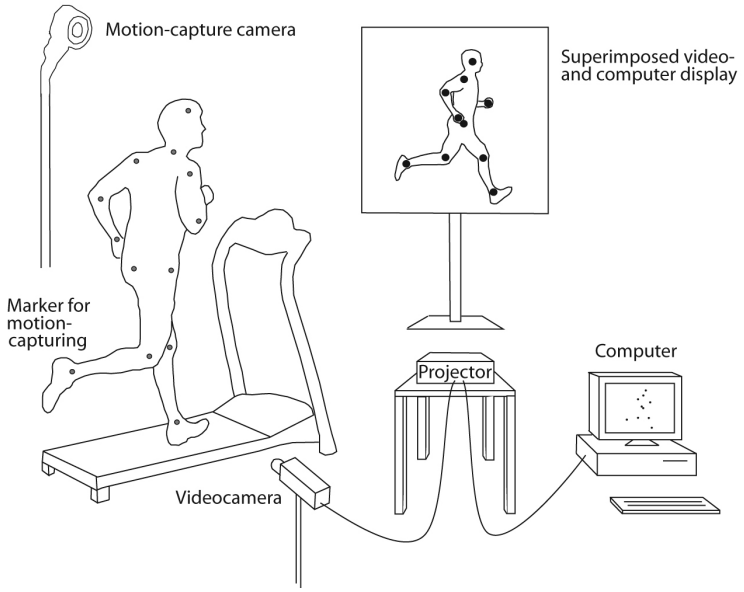

Figure 3.3: Setup for approximating human-like motion morphs by real human movements. A point light walker, animated with the trajectories generated by motion morphing, is superposed to an on-line video of the actor walking on a treadmill. The actors movements are motion captured (using reflecting markers on shoulders, elbows, wrists, knees and ankles, head and hip). The recorded trajectories closely approximate the trajectories of the human-like motion morph, but fulfill exactly the laws of motion of real human movements. Based on the processed trajectories, point light stimuli with ten dots were generated by averaging the positions of the head and the shoulder markers as well as the hip markers.

movements. In this case, a comparison between artificial and human-like morphs would not permit conclusions about differences between biological and artificial patterns, since similarities might just reflect the fact that all morphs violate these constraints.

We used a special setup that is illustrated in Figure 3.3. A human actor on a treadmill tried to imitate the movements of (human-like) morphs. The actor viewed a superposition of an online video of his own movement and a point-light stimulus, whose movements were defined by the motion morph. The actor tried to align the joints of his body, as accurately as possible, with the positions of the dots of the point-light walker, monitoring himself on the video screen. After several minutes of training, the actor was able to accomplish a relatively accurate reproduction of the movements of the point-light walker (reproducing 88 % of the variance of the two-dimensional joint trajectories).

The movements of the actor were recorded using a VICON 612 motion capture system with 6 cameras. The 3D positions of 22 reflecting markers were

recorded with a sampling frequency of 120 Hz, and an accuracy below 1 mm. The resulting trajectories were processed using commercial software by VICON. From each movement several movement cycles were recorded, and the cycle that was most similar (after projection on the two-dimensional plane) to the imitated point-light stimulus was selected for stimulus generation. Based on the processed marker trajectories, point-light stimuli with ten dots were generated and presented with the same frame rate as the original stimuli.

3.3 Results

3.3.1 Experiment 1: Learning of Human-like vs. Artificial Articulated Movements

The first experiment compared the discrimination learning between two types of articulated motion stimuli: Movements that closely approximated human movements, and movements based on artificial skeleton models that were quite dissimilar from biologically relevant movements of humans and animals. With this experiment we tried to clarify two questions: (1) Are humans able to learn the discrimination between artificial articulated motion patterns without biological relevance, and how fast is this learning? (2) Is there a difference between the learning of biologically relevant and artificial articulated movements?

Participants were trained with two types of point-light stimuli: human-like stimuli, generated by morphing between real human movements, and artificial articulated movements, generated by morphing between trajectories that were generated with artificial skeleton models (see Methods). If the human visual system contained special mechanisms for the learning of biologically important movements, one would expect that learning of human-like patterns should be faster and potentially more accurate than the learning of completely artificial patterns. If however, the visual system was disposing of a general mechanism for the learning of movements that is independent of biological relevance, no difference would be expected.

Results and Discussion

Participants perceived the human-like stimuli as human movements, whereas the artificial articulated patterns typically resulted in very inconsistent interpretations between subjects (e.g. ”mechanical device” or ”weird spider”). Figure 3.4a

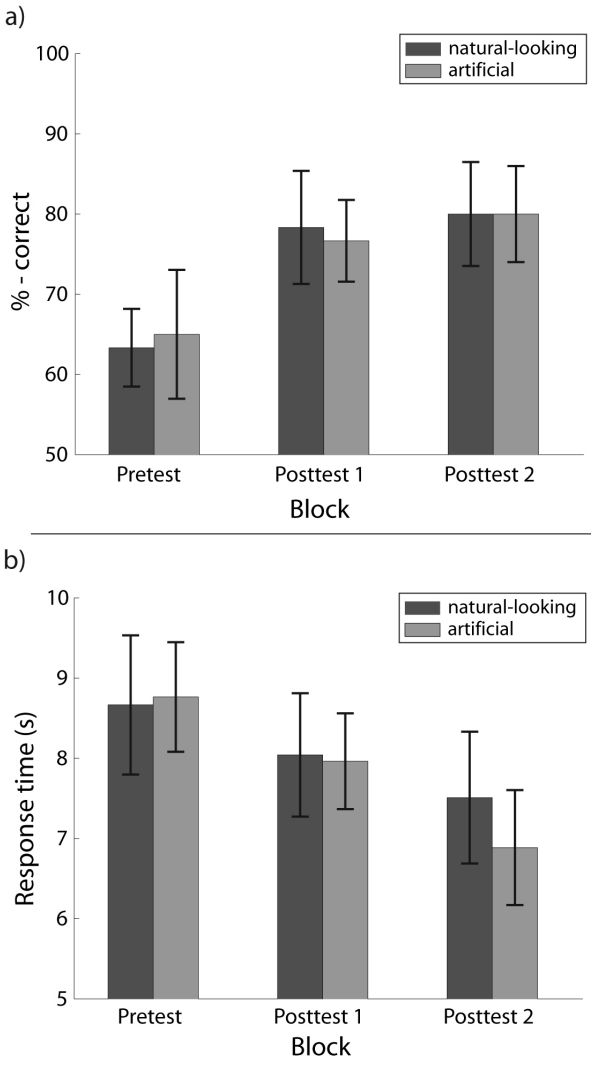

Figure 3.4: Discrimination learning for human-like vs. artificial ar-ticulated movements. Panel a shows the mean percentages of correct responses for Experiment 1 (+ s.e.m.) in the three test blocks for the human-like stimuli (black), and for artificial articu-lated movements (gray). Panel b shows the mean response times (+ s.e.m.) in the three test blocks separately for human-like and the artificial articulated movements (N=11).

shows the recognition performance (percent correct) for the human-like stimuli (black bars) and the artificial articulated patterns (gray bars) for the three test blocks. Starting close to chance level, participants show very similar improvements of discrimination performance for both stimulus types. Two training blocks with only 16 repetitions of each center stimulus were sufficient to improve performance to a level above 80 % correct responses for both stimulus types. A two-way repeated measures ANOVA reveals a significant main effect of the number of the test block ($F_{(2,20)}$ = 7.9, p $<$ .01). Neither the main effect of stimulus type ($F_{(1,10)} < .1$, p = .98), nor the interaction were significant ($F_{(2,20)} <1$, p = .61).

Figure 3.4b shows the response times for the two stimulus types. Consistent with the increase in performance, the response times decrease after training for both stimulus types in a very similar way. This observation was confirmed by an analysis of variance showing a significant main effect of the number of test blocks ($F_{(2,20)}$ = 6.8, p $<$ .01), but no significant influence of stimulus type ($F_{(1,10)} < 1$, p = .54) and interaction ($F_{(2,20)}$ = 1.7, p = .32).

These results indicate that the human visual system disposes of a learning mechanism for articulated movement patterns that works equally well for biologically relevant and artificial patterns. Furthermore, less than 20 stimulus repetitions during training were sufficient for participants to increase their performance significantly. This indicates relatively fast visual learning for such articulated movement patterns.

Since the discrimination during the test blocks of our experiments (discrimination between Center and Test off-Center stimuli) was more difficult than the discrimination during the training blocks (discrimination between Center and Training off-Center stimuli (see Methods), our result implies also that the successful learning of a simpler discrimination facilitated the more difficult discrimination during the test blocks. The same phenomenon has been reported in other perceptual learning experiments before (e.g. Ahissar and Hochstein, 1997; Liu and Weinshall, 2000; Mackintosh, 1974).

To further investigate the selectivity of the learning process, we conducted two additional control experiments. The first control experiment tested the transfer of the learning to novel untrained stimuli of the same type. Two human-like and two artificial articulated stimuli were randomly chosen for training. In the test blocks, however, all four stimuli from each stimulus type (human-like and artificial movements) were presented. In this way, we could test whether the improvements due to training would transfer to novel similar stimuli that had not been part of the training set.

Our results show that improvements during training did not transfer to novel, untrained stimuli. The observers' performance for the novel stimuli did not change significantly, while we obtained significant improvements for the trained stimuli. This was confirmed by a three-way ANOVA that revealed a significant main effect of the number of the test block ($F_{(2,16)}$ = 4.7, p $<$ .05), and a significant interaction ($F_{(3,24)}$ = 3.5, p $<$ .05) between the number of test block and the familiarity (Trained vs. Novel stimuli). An additional contrast analysis based on the significant interaction revealed significant improvements for the trained groups (human-like: $F_{(1,8)}$ = 8.6, p $<$ .01 ; artificial articulated: $F_{(1,8)}$ = 13.8, p $<$ .01) but not for the untrained groups (human-like: $F_{(1,8)} < 1$, p = .43 ; artificial articulated: $F_{(1,8)} < 1$, p = .98). This result implies that the learning was highly specific for the trained patterns and did not transfer to untrained patterns generated from different prototypical movements. Moreover, it implies that the observed effects cannot be explained by general factors, like increasing familiarity with the task or increasing efficiency of the processing of biological motion.

The second control experiment tested the necessity of the training blocks with feedback for the learning process. The three test blocks were presented without intermediate training. In this case, discrimination performance did not increase significantly (no significant main effect of the number of the test block; $F_{(2,18)} < 1$, p = .91). Like for Experiment 1, we did not observe significant differences between human-like and artificial articulated stimuli in this control experiment (no main effect of stimulus type $F_{(1,9)} < 1$, p = .93, and no interaction $F_{(2,18)} < 1$, p = .46). This implies that training together with feedback is essential for the improvement during learning. This experiment provides also additional evidence against the explanation of the observed changes by unspecific familiarity or practice effects.

3.3.2 Experiment 2: Orientation Dependence of the Learned Representation

A characteristic property of the recognition of biological motion is its strong view and orientation dependence. Rotation of point-light walkers in the image plane against the familiar upright orientation substantially degrades recognition performance (e.g. Bertenthal et al., 1987; Bülthoff et al., 1998; Pavlova and Sokolov, 2000; Sumi, 1984). Experiment 2 tested whether such orientation dependence also applies to newly learned representations of human-like and artificial articulated patterns.

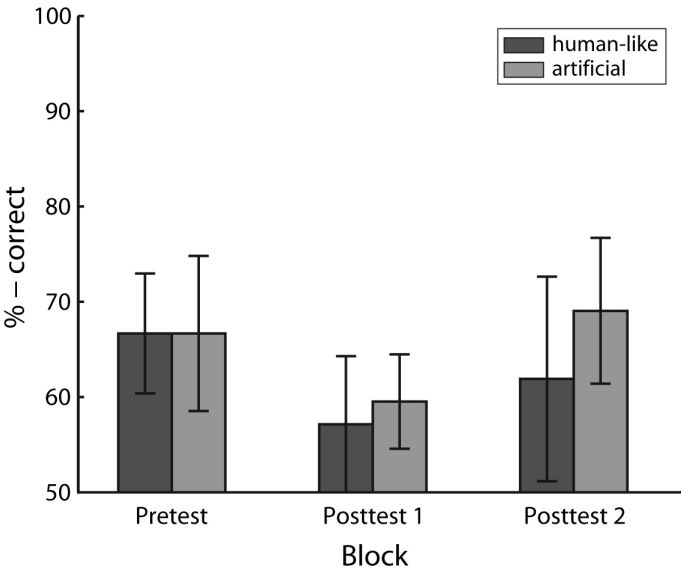

Figure 3.5: Orientation dependence of the learned representation. Mean percentages of correct responses (+ s.e.m.) for three test blocks. Training stimuli were rotated against the test stimuli in the image plane (by 90 deg). Recognition rates are shown for the human-like movements (black), and for the artificial articu-lated movements (gray), (N=9).

To test orientation dependence, we modified Experiment 1 by training the participants with stimuli that were rotated by 90 degrees in the image plane against the test stimuli, which were presented upright.

Results and Discussion

In accordance with the orientation dependence of normal biological motion recognition, we did not find any significant improvement of recognition performance in the test blocks for both stimulus types when the training stimuli were rotated against the test stimuli (Figure 3.5). This was confirmed by a repeated-measures ANOVA, which did not show a significant influence of the number of test blocks ($F_{(2,16)}$ = 1.6, p = .22). In accordance with the previous experiments, we did not obtain significant differences between the two stimulus types ($F_{(1,8)} < 1$, p = .89) and no significant interaction ($F_{(2,16)} < 1$, p = .76).

This result implies that, like for normal biological motion patterns, the representations for the novel learned patterns are strongly orientation dependent. This result seems consistent with the hypothesis that biological motion recognition is based on the learning of orientation or view-dependent templates (Giese and Poggio, 2003; Verfaillie, 2000). This suggests that biological motion recognition might be based on similar principles as the view-based recognition of complex static shapes (see Tarr and Bülthoff, 1998, for review).

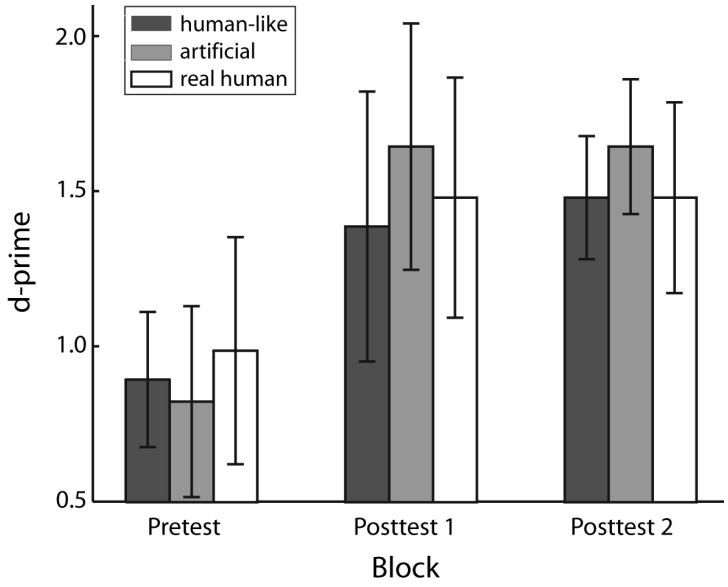

Figure 3.6: Real human movements compared with artificial movements. Shown are the d-prime values for the human-like stimuli (black), the artificial articulated stimuli (gray), and real human movements approximating human-like morphs (white) for the three test blocks (+ s.e.m.) (N = 7).

3.3.3 Experiment 3: Comparison with Real Human Movements

A possible criticism of Experiment 1 is that, since all stimuli were generated by motion morphing, none of them might provide an appropriate approximation of real human movements. This would explain why we did not find differences between human-like and artificial stimuli. In order to exclude this possibility, we have developed a method for approximating human-like motion morphs by movement trajectories of a real human actor (see Methods).

In Experiment 3 we compared the learning of three stimulus classes: artificial articulated movements, human-like morphs and real human movements. Otherwise, the design of this experiment was identical to Experiment 1.

Results and Discussion

The results of Experiment 3 are summarized in Figure 3.6. Based on a different set of participants, this experiment replicates the results of Experiment 1. For a more detailed analysis of differences between the different stimulus types we applied a signal detection analysis and computed d-prime values. In accordance with the improvement in performance in Experiment 1, the d' values increase significantly with the number of test blocks ($F_{(2,12)}$ = 3.8, p ¡ .05), whereas there is no significant influence of the stimulus type ($F_{(2,12)} < 1$, p = .48) and no interaction ($F_{(4,24)} < 1$, p = .86). In particular, there is no significant difference

between real human movements and the artificial articulated patterns. These results confirm that the similarities between the learning of human-like and artificial articulated patterns in Experiment 1 are not an artifact that was induced by the fact that the stimuli were generated by motion morphing. The same similarity is found for the comparison between real human movements and artificial articulated patterns.

3.3.4 Experiment 4: Learning of Stimuli With and Without Global Underlying Shape

Experiments 1-3 have demonstrated strong similarities for the learning of human-like and artificial articulated movement patterns. This raises the question if any movement pattern of similar complexity can be learned, even if a global underlying shape or skeleton cannot be perceived. To test this question we compared the learning of human-like movements with the learning of movement patterns without underlying global shape.

To generate stimuli without global underlying shape we spatially scrambled the human-like movement patterns, i.e. we added temporally constant random position offsets to the individual dot positions. This operation destroys the consistency of the movements with an underlying articulated shape. However, since the offsets are temporally constant, it does not affect the local motion information present in the stimuli. The offsets were constrained to ensure that the scrambled stimuli covered the same spatial area as the original human-like stimuli. None of the tested participants was able to recognize an articulated shape in any of these scrambled stimuli.

The experimental design was identical to Experiment 1. For each subject, two human-like morphs were randomly chosen and presented intact, while the remaining two human-like morphs were presented as scrambled stimuli.

Results and Discussion

The results of Experiment 4 from the three test blocks are presented in Figure 3.7. In this case, the learning process seems to be different for the two stimulus classes. While the initial d-prime values for both stimulus types are close to 0, only the performance for the natural-looking morphs is increasing significantly (like in the previous experiments). The performance for the scrambled stimuli in the test blocks did not increase over the training. This observation is confirmed by a repeated measures ANOVA with the factors stimulus type (intact / scrambled) and number of test block (Pretest, Post-test 1 and

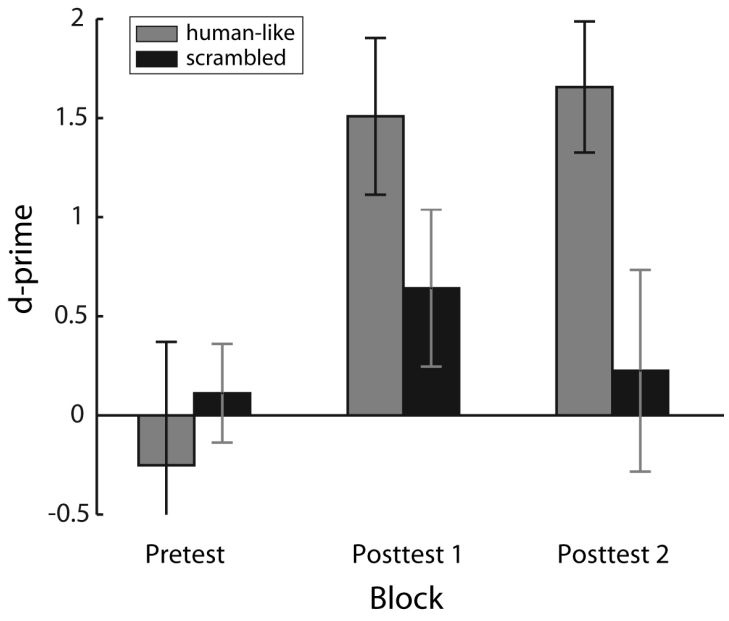

Figure 3.7: Human-like movement stimuli compared with scrambled human-like stimuli. Shown are the d-prime values for the intact human-like stimuli (gray) and the scrambled stimuli (black) in the individual test blocks (+ s.e.m.), (N = 7).

Post-test 2). We obtained a significant main effect of stimulus type ($F_{(1,6)}$ = 11.1, p $<$ .05) and of the number of the test block ($F_{(2,12)}$ = 8.3 , p $<$ .01). The interaction was marginally significant ($F_{(2,12)}$ = 3.8 , p = .05).

Interestingly, a separate analysis of the trials in the training blocks suggests no differences in performance during training between the intact and the scrambled stimuli. A repeated measures ANOVA reveals a significant improvement from the first training block to the second training block ($F_{(1,6)}$ = 10.6 , p $<$ .05) for both stimulus types (no significant effect of stimulus type ($F_{(1,6)} < 1$, p = .41) and interaction ($F_{(1,6)} < 1$, p = .79). This result suggests that subjects learned to discriminate between the training stimuli. However, they were not able to generalize to the test stimuli with smaller distances between center and Off-center stimuli in the morphing space.

An additional analysis of the response times shows that the observed difference between intact and scrambled movements in the test blocks cannot be attributed to a speed-accuracy trade off. Comparing the response times between the two stimulus types (in the test blocks) we found no significant differences ($F_{(1,6)} < 1$, p = .72). This clearly indicates that subjects were trying to succeed for both stimulus types.

Experiments 1-3 focused on the learning of articulated movements, i.e. movements with an underlying global shape. Experiment 4 suggests that the presence of such shapes seems to be critical for the learning process, or at least for the

generalization to more difficult discriminations of very similar stimuli. Learning seems thus to be substantially facilitated in presence of a skeleton, or potentially also other global shapes.

3.4 Discussion

Our study investigated the visual learning of complex movement patterns that shared specific properties with natural human movements. We studied discrimination learning for stimulus pairs that were generated by motion morphing, allowing for a precise control of the spatio-temporal similarities between the stimuli and their low-level properties. Our experiments show that humans can learn to recognize novel articulated movement patterns very quickly, after less than 20 stimulus repetitions. Like for normal biological motion, the learned visual representation seems to be strongly orientation-dependent, i.e. stimuli were recognized only if they were presented with the same orientation as the training patterns. Interestingly, learning speed and accuracy for human movements and completely artificial articulated patterns were very similar. Familiarity or biological relevance of the underlying kinematics or skeleton seems thus not to be critical for the visual learning process. Contrasting with this result, motion stimuli without underlying global shape could not be learned equally fast, even when their local motion properties were identical to human motion. This suggests that the binding or grouping of the individual stimulus elements into a global percept might strongly facilitate the learning. The similarity of our stimuli with normal biological motion stimuli, and the fact that the learning was strongly orientation-dependent makes it likely that the investigated learning process plays also a central role in normal biological motion recognition.

The observed strong similarity between learning of human-like and artificial articulated movements was highly reproducible. This similarity was observed in Experiment 1-3 and in the two additional control experiments (in total for 46 participants and 8 different movement stimuli). In addition, we found reproducible improvements of performance with training, and strongly selective differences between trained and untrained stimuli. These observations, and the fact that we found a difference between the learning of normal and scrambled human stimuli rule out the possibility that the observed similarity between the two classes of articulated stimuli just reflects a lack of sensitivity of our paradigm or behavioral measures.

Our study provides some insights that contribute to the question of what is 'special' about biological motion perception, or at least for the visual learning

of such stimuli. Biological movements are characterized by several properties that might be critical for their perception: (1) general smoothness properties of the trajectories, (2) the consistency with an underlying global shape or skeleton, and (3) the familiarity or biological relevance of these shapes and the associated motion patterns. Our experiment shows that the last property is not critical for fast learning of complex movements. However, the second property seems to be critical.

By construction, all our stimuli shared the first property, similar smoothness of the trajectories. Smoothness is closely related to the consistency of the movements with 'laws of motion' that are typical for motor behavior (Viviani and Flash, 1995). An example is the 'two-thirds power law' (Lacquaniti et al., 1983) signifying that curvature and speed of planar human movements are linked by a power law. Psychophysical experiments have shown that simple motion stimuli fulfilling this law appear smoother (Viviani and Stucchi, 1992). The joint trajectories used for the generation of our stimuli were in agreement with this motor invariant. (The exponents of the power law were determined by linear regression, applied to the logarithms of velocities and curvatures, yielding exponents ranging between .31 and .36 for human-like and artificial articulated movements.) It remains to be clarified in future experiments whether a violation of general smoothness properties impairs fast visual learning. Well-controlled experiments of this type might be very difficult to realize, since it would have to be excluded that differences between stimuli with different smoothness are not just reflecting differences in low-level motion processing, induced by the different motion energy distributions of such stimuli.

There are several possible explanations why learning of human-like and artificial articulated patterns are similar, whereas learning of scrambled patterns is much more difficult. First, there might be a mechanism that recognizes complex movements by matching the underlying articulated shape (e.g. Beintema and Lappe, 2002; Giese, 2000; Giese and Poggio, 2003; Marr and Vaina, 1982; Vaina and Bennour, 1985; Webb and Aggarwal, 1982). This mechanism might operate independent of the biological relevance of such shapes. Second, the recognition of motion patterns might be based on features of intermediate complexity which only arise for motion that is derived from smoothly deforming or articulated shapes. And third, there could be top-down influences of shape recognition that facilitates the learning of motion patterns. The existence of such-top down influences is suggested by a number of psychophysical and imaging studies showing that local motion perception and the activity in motion-related brain areas are modulated by the recognition of shapes, if they are typically associ-

ated with body movements (e.g. Chatterjee et al., 1996; Kourtzi and Kanwisher, 2000a; Peuskens et al., 2005; Senior et al., 2000).

The importance of top-down mechanisms seems also consistent with our observation that more difficult discriminations between articulated movements were facilitated by previously learned simpler discriminations of similar patterns. The same phenomenon has been observed in other visual learning experiments (e.g. Liu and Weinshall, 2000; Mackintosh, 1974), and has also been termed perceptual *"Eureka"* (Ahissar and Hochstein, 1997). It has been explained by the learning of a more effective allocation of attention to features or stimulus dimensions that are relevant for the discrimination. A high importance of top-down influences in motion recognition is suggested by several psychophysical experiments (Cavanagh et al., 2001; Thornton et al., 2002). The results of our experiments suggest that an underlying global shape, and potentially form recognition, might strongly facilitate such top-down processes.

Another implication of our study is that consistency with a preexisting internal (dynamic) body model seems not to be required for fast visual learning of complex movements. Such internal models likely contribute to the perception of imitable body movements (e.g. Fadiga et al., 1995; Grafton et al., 1997; Prinz, 1997; Wilson and Knoblich, 2005; Wolpert et al., 2003). Recent imaging experiments suggest that neural structures that are involved in the representation of such internal models are also activated by point-light walkers (Saygin et al., 2004). However, we think that it is unlikely that the learning of the artificial articulated patterns was based on such internal body models, since their kinematics differed strongly from human bodies and specified non-imitable movements. The contribution of internal models to the recognition of artificial stimuli could potentially be clarified in brain imaging studies that compare activity distributions for the two articulated stimulus types.

Several studies have suggested that infants have an innate preference for the processing of biological motion (Fox and McDaniel, 1982; Grezes et al., 2001; Johansson et al., 1980; Meltzoff and Moore, 1977; Pavlova et al., 2003). A preference for biological motion stimuli has also been observed in animals (e.g. Blake, 1993). For example, unexperienced newly-hatched chicks demonstrate an innate predisposition to approach motion stimuli that share specific low-level properties with biological movements (Vallortigara et al., 2005). However, this preference seems not to be selective for the global stimulus structure, since the animals were equally attracted by scrambled and intact point-light displays of hens. Also, this innate preference seemed not to be selective for movements of different species, e.g. own species vs. predators. Yet, an innate predisposition to attend to stimuli with low-level properties that are typical for biological move-

ments might be very helpful to support the learning of more subtle biologically important differences between complex motion stimuli. Our results complement these studies about potential unspecific innate factors by providing a detailed investigation of stimulus properties that seem to be critical for the learning of detailed distinctions between different biological movements.

Summarizing, we have demonstrated the existence of a fast visual learning process for the holistic structure of complex motion patterns. This process shares important properties with normal biological motion perception and seems not to differentiate between biologically relevant and non-biological articulated movements, as long as they convey the percept of a global form. Learning might thus play a central role for understanding the perception of biological motion and actions.

Chapter 4

Functional Imaging Studies

The psychophysical results from the previous chapter showed that the discrimination between very similar complex movements can be learnt. In a series of functional imaging studies, we wanted to verify which visual areas might be involved in this learning process and if we would obtain differences at the BOLD activity level between the human-like movements and the artificial movement patterns. Our results indicate that learning novel human-like actions shaped higher-level processing of known action categories by enhancing motion analysis in hMT+/V5, V3B/KO and generalization to novel movements similar to prototypical actions in STSp, FFA. However, learning artificial movements bolstered the formation of representations for novel movement patterns by enhancing the processing in these areas as well as retinotopic regions. These findings support a central role of experience-based plasticity in the recognition of complex movements and action understanding[1].

4.1 Introduction

The recognition and understanding of complex movements and actions is critical for survival and social interactions in the dynamic environments we inhabit. Thus, it is no surprise that the human visual system is highly skilled in action recognition even from highly impoverished articulated point light displays (e.g. Beintema and Lappe, 2002; Bertenthal et al., 1984; Fox and McDaniel, 1982;

[1]The paper reporting the work presented in this chapter is currently under review at *Neuron*. (Jastorff J, Giese MA, Kourtzi Z: Visual learning shapes the processing of actions in the human visual cortex).

Hill and Pollick, 2000; Johansson, 1973; Thornton et al., 1998). Understanding the neural mechanisms that mediate this skill has been a topic of increasing interest in cognitive neuroscience. Several recent studies have investigated the link between action recognition and production (see Hommel et al., 2001; Wilson and Knoblich, 2005, for review) and its neural substrates in the monkey (Rizzolatti and Craighero, 2004) and the human (e.g. Decety et al., 1997; Iacoboni et al., 1999) brain. While some studies propose innate mechanisms for the recognition of body movements (e.g. Fox and McDaniel, 1982), others provide evidence for the importance of learning. In particular, action understanding has been proposed to be accomplished by internal simulation of motor behavior (Haruno et al., 2001) that forms the basis for understanding other people's mental states and emotions (Blakemore and Decety, 2001; Frith and Frith, 1999; Gallese et al., 2004). On the other hand, studies showing that visual learning improves observers' performance in detecting human actions in noise (Grossman et al., 2004) and recognizing individuals from their facial or body movements (Hill and Pollick, 2000; Knappmeyer et al., 2003; O'Toole et al., 2002; Troje et al., 2005) suggest that action understanding is enhanced and refined by experience. Despite the importance of learning in the recognition of biological movements, several questions remain open. Does learning of novel complex movements dependent on their similarity to motor behavior and how can artificial movement patterns that do not appear to match existing motor programs be leant? What are the neural substrates of this learning in the human visual cortex? Does learning result in localized or distributed experience-based plasticity across stages of visual analysis?

We addressed these questions using concurrent psychophysical and fMRI measurements. To understand the role of experience-based plasticity in the processing of complex movements, we compared discrimination learning of human-like movements that are known to the observers (e.g. walking, running), and artificial complex movements that had biological movement properties but were unknown to the observers and were interpreted as dissimilar from typical human or animal movements (Figure 4.1a,b). To study the learning of novel exemplars for each of these stimulus classes we generated movement sequences with well-defined spatio-temporal properties using a motion morphing technique that models new movements as linear combinations of sets of three prototypical example trajectories (Giese and Lappe, 2002; Giese and Poggio, 2000). By appropriate choice of the weights of the prototypes in the linear combinations, we were able to vary parametrically and smoothly the similarity between the generated movements. That is, similar stimuli corresponded to movements that were close in the continuous morphing space, while dissimilar stimuli corresponded to more distant movements (Figure 4.1c). For each stimulus class, observers were pre-

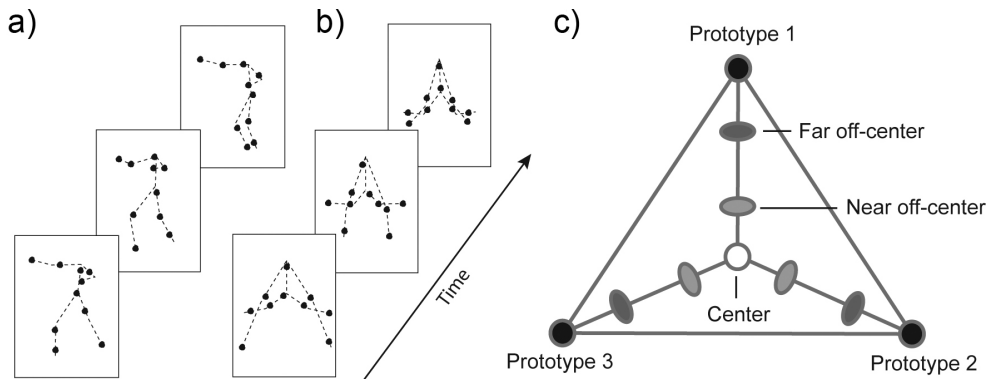

Figure 4.1: Individual frames of point light stimuli representing a) a human like stimulus and b) an artificial pattern (the dashed lines connecting the joins were not shown during the experiment). c) Metric space defined by motion morphing. Morphs were generated by linear combination of the joint trajectories of three prototypical patterns (Prototypes 1-3). Three groups of stimuli were generated by choosing different combinations of linear weights: 1) Center Stimuli, for which each of the three prototypes contributed to the resulting morph with equal weights (33.33%), 2) near Off-Center Stimuli, for which the weight for one of the prototypes was on average 60% and for each of the other two prototypes 20%, and 3) far Off-Center Stimuli, for which the weight for one of the prototypes was on average 75% and for each of the other two prototypes 12.5%.

sented with two movements sequentially and judged whether they were the same or different. Initially, observers performed this task in the scanner. Observers were then trained in the laboratory with feedback for three consecutive days, and then tested again in the scanner.

The fMRI measurements employed an event-related adaptation paradigm (Grill-Spector et al., 2001; Kourtzi and Kanwisher, 2000b). This technique exploits the phenomenon of adaptation or repetition suppression that results in decreased fMRI responses for repeated presentation of the same stimulus compared to presentation of different stimuli. Recent electrophysiological evidence suggests that this phenomenon may reflect neural adaptation in stimulus-selective ensembles of neurons (Sawamura et al., 2006). We compared the adapted response evoked by the same movement shown twice with responses evoked by movements that differed in their properties (very similar, less similar, dissimilar) depending on their distance from each other in the morphing space. fMRI selective adaptation indicated by higher fMRI responses for different that identical movements would suggest sensitivity in the measured neural populations to the differences between movements critical for the discrimination task. We tested fMRI selective adaptation in areas implicated in the processing of biological motion (STSp, FFA) (Beauchamp et al., 2003; Bonda et al., 1996; Grezes et al., 2001; Grossman and Blake, 2002; Peuskens et al., 2005; Puce and Perrett,

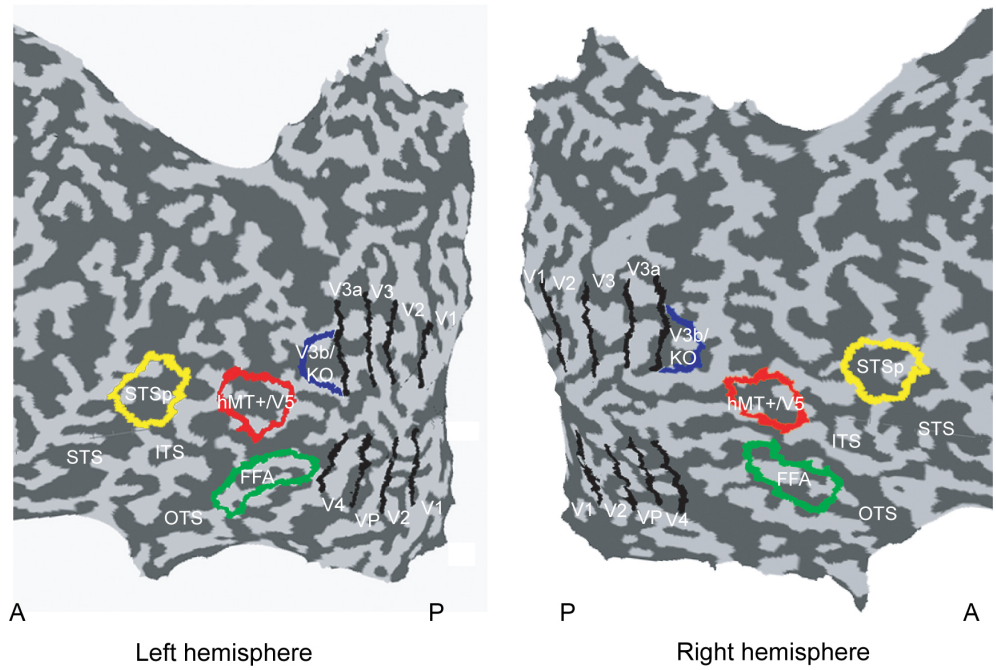

Figure 4.2: Regions of interest visualized on the flattened cortical surface of the left and the right hemisphere of one subject. The functional activation maps represent the early retinotopic regions, V3b/KO, hMT+/V5, the fusiform face area (FFA) and the posterior part of the superior temporal sulcus (STSp). Dark gray: sulci; light gray: gyri. STS: Superior temporal sulcus, ITS: Inferior temporal sulcus, OTS: Occipital temporal sulcus. A, anterior; P, posterior.

2003; Vaina et al., 2001), physical motion (hMT+/V5) (Tootell et al., 1995), relative motion (V3b/KO) (Dupont et al., 1997), and retinotopic visual areas that were independently localized in each observer (Figure 4.2; Table S1). We reasoned that improvement of the observers' performance in discriminating between movements and increased fMRI selective adaptation after training would provide evidence for experience-based plasticity in the measured cortical areas that mediate behavioral learning.

Our findings provide novel evidence that the human brain learns to recognize novel biological actions by reorganizing the processing of complex movements across stages of visual analysis. In particular, improvement in observers performance for the discrimination of movements after training was coupled with increased fMRI selective adaptation across multiple visual areas, suggesting enhanced learning-dependent sensitivity to the critical features for the discrimination of movements in these areas. Learning of artificial movements dissimilar from familiar action patterns bolstered the processes critical for building representations for unknown motion patterns; that is, integration of novel dynamic

configurations in early visual areas (V3a, Vp, V4), analysis of global movement patterns in higher motion-related areas (hMT+/V5, V3b/KO), and processing of the biological movement characteristics in biological motion-related areas (STSp and FFA). However, learning of human-like movements enhanced primarily higher-level motion processes; that is, global motion processing in motion-related areas, and generalization (Giese and Poggio, 2003; Poggio and Edelman, 1990) to novel exemplars based on their similarity to prototypical members of known movement categories (human actions) in biological motion-related areas.

4.2 Experimental Specifications

4.2.1 Participants

Thirty four students from the University Tübingen participated in this study: twelve in Experiment 5, eleven in Experiment 6 and eleven in Experiment 3. The data from two subjects in Experiment 1 and two subjects in Experiment 7 were excluded either due to excessive head movement or poor psychophysical performance in the training sessions. Informed consent was obtained from all subjects and the experiments were approved by the local ethics commission of the University Tübingen.

4.2.2 Stimuli

All stimuli (5 x 10 degrees of visual angle) were presented as 10 white dots (0.5 degrees of visual angle) on a black background. In order to minimize the effect of low-level position cues, the points were not presented exactly at the joint position, but were randomly jittered along the segments of the skeleton. The displacement varied randomly between 0% and 20% of the segment length in every frame of the animation. A similar manipulation had been applied in previous studies of biological motion (Beintema and Lappe, 2002).

The stimuli used in Experiments 5, 6 and 7 were identical to the stimuli used in Chapter 3. The human-like stimuli used for Experiments 5 and 6 (Figure 4.1a) were obtained by tracking the two-dimensional joint positions in video movies of a human actor facing orthogonally to the view axis of the camera while performing different movements. 18 movements were recorded including locomotion, dancing, aerobics and martial arts sequences. Twelve points were tracked manually (head, shoulders, elbows, wrist, hip, knees and ankles) but the positions of the shoulder and the head markers were averaged for stimulus presentation.

The artificial movement stimuli used in Experiment 7 (Figure 4.1b) were generated by animation of eighteen different artificial skeleton models with 9 segments. The skeletons were chosen to be highly dissimilar from naturally occurring body structures. Debriefing of the observers showed that they did not provide any consistent interpretations of these stimuli in contrast with the human like movement stimuli that were recognized accurately by all observers. None of the observers interpreted any artificial pattern as a human action. Some observers interpreted these artificial movements as peculiar animal movements (38% of observers), dynamic mechanical devices (21% of observers) or perceived moving disconnected dots without any particular global interpretation (35% of observers). The joint angle trajectories $\alpha_n(t)$ of each skeleton were defined in the same way like in the psychophysical experiments in Chapter 3:

$$\alpha_n(t) = \alpha_n + \beta_n sin(\omega t + \phi_n) \tag{4.1}$$

Their frequency ω and amplitudes β_n were matched with that of typical joint trajectories of human actors during natural movements. This procedure resulted in 18 artificial movement prototypes. In addition, the segment length and the area covered by the artificial stimuli were matched with the human-like movements in order to control for differences in the low-level properties of the two stimulus classes.

All stimuli were generated using motion morphing. We applied an algorithm known as spatio-temporal morphable models that was introduced already in Section 2.3. Each stimulus was defined as linear combinations of three prototypical movements:

New motion pattern = c_1 (Prototype 1) +
c_2 (Prototype 2) + c_3 (Prototype 3)

The weights c_i determined how much the individual prototypes contributed to the morph. When the weight of one prototype was very high, the linear combination strongly resembled this prototype. (Weight combinations always fulfilled $c_1 + c_2 + c_3 = 1$). The weight vectors (c_1, c_2, c_3) defined a Euclidian space of movement patterns that provided a metric of the spatio-temporal similarity between the patterns. This has been verified in additional studies applying multidimensional scaling to similarity judgments between stimuli generated by this method (Giese et al., 2003). This metric space allowed us to precisely manipulate the difficulty of the discrimination task by varying the distance between the movements in morphing space. This morphing technique was used to generate three different classes of stimuli: a) Center Stimuli, for which each of the three

prototypes contributed to the resulting morph with equal weights (33.33%), b) near Off-Center Stimuli, for which the weight for one of the prototypes was 60% and for each of the other two prototypes 20%, and c) far Off-Center Stimuli, for which the weight for one of the prototypes was 75% and for each of the other two prototypes 12.5% (Figure 4.1c). These weights resulted in a gradual change in the physical similarity between morphs as indicated by measurements of the Euclidean distances between the stimulus trajectories. For human-like movements, the mean distance between the trajectories of the Center and the near Off-Center stimuli was 0.073, the Center and the far Off-Center stimuli 0.117, and the Center and the Prototype stimuli 0.172. For the artificial movements, the Euclidean distance between the artificial Center and near Off-Center Stimuli was 0.085, indicating that the physical differences in the stimulus space generated for the human-like movements and the artificial patterns were very similar.

To ensure that the human-like morphed movements appeared natural, we morphed between prototypes from the same movement category (e.g. running, limping and marching or three different types of boxing movements). Previous studies have shown that the technique of spatio-temporal morphing interpolates smoothly between quite dissimilar gait patterns (e.g. walking and running), resulting in motion morphs that look highly natural (Giese and Lappe, 2002). In addition, in a pilot experiment we collected naturalness ratings (scale 1: unnatural, 5: natural) for each of the morphed stimuli. Only human-like stimuli with high naturalness ratings (4 or 5) were used in Experiments 5 and 6.

Finally, the scrambled point light stimuli used for the localizer of biological motion-related areas were generated by randomizing the initial starting position of every point in the intact point-light displays of these human actions, while preserving the original motion vector of each individual point. The scrambled and intact point-light displays were matched for area and dot density.

4.2.3 Design and Procedure

Each subject participated in one of three experiments. In addition, for each subject we mapped the early retinotopic areas (1 run), the kinetic occipital area, V3b/KO (1 run), the middle temporal area: hMT+/V5 (1 run) and biological motion-related areas, namely the posterior superior temporal sulcus: STSp and the fusiform face area: FFA (2 runs).

Experiment 5: Learning Novel Biological Movements

Pre-Training scanning session (day 1): The observers were presented with six different groups of human-like biological movements. Each group consisted of movement morphs between three prototypical natural human actions from the same category (e.g. three different locomotion patterns or three kinds of boxing movements). Four different conditions were tested: a) Identical: the same center stimulus was presented twice, b) Very Similar: a center stimulus was presented, followed by a near off-center stimulus, c) Less Similar: a center stimulus was followed by a far off-center stimulus, and d) Dissimilar: a center stimulus was followed by a prototype. The scanning session consisted of four event-related runs without feedback. The observers were instructed to judge whether two successively presented stimuli were the same or different. Each run started and ended with an eight second fixation epoch. Every run consisted of 20 experimental trials for each of the four conditions and 20 fixation trials that were interleaved with the experimental trials. Each trial lasted 4 seconds and started with the first stimulus (one movement cycle) presented for 1300ms, followed by a 100ms blank, a second stimulus for 1300ms followed by a blank interval of 1300ms. The history of the conditions was matched so that each condition, including the fixation condition, was preceded equally often by trials from each of the other conditions (Kourtzi and Kanwisher, 2000b).

Training sessions (day 2-4): Each subject participated in three training sessions in the lab on consecutive days. Each session consisted of a total of 282 trials, 47 trials per movement group. Each trial consisted of two sequentially presented movements. The first movement was a center stimulus, while the second one was either the same center (Identical), a near off-center (Very Similar) or a far off-center (Less Similar). Observers were not trained in the Dissimilar condition, as their accuracy performance was high before training. The Less Similar condition was included during training, as previous studies have shown that an easier discrimination facilitates learning in a difficult task (Ahissar and Hochstein, 1997; Liu, 1999). Each movement was presented for 5200 ms (four cycles) with an ISI of 500 ms between movements followed by a blank interval of 1100 ms. In a two alternative forced choice task, subjects were instructed to report, whether the two stimuli in a trial were the same or different. Feedback was provided throughout training. After each training session, the observers' performance was tested in one experimental run without feedback, similar to those during scanning (Figure 4.3).

Post-Training scanning session (day 5): After completion of the training sessions, the subjects were tested on the following day in the scanner on the same stimuli and conditions as in the pre-training scanning session.

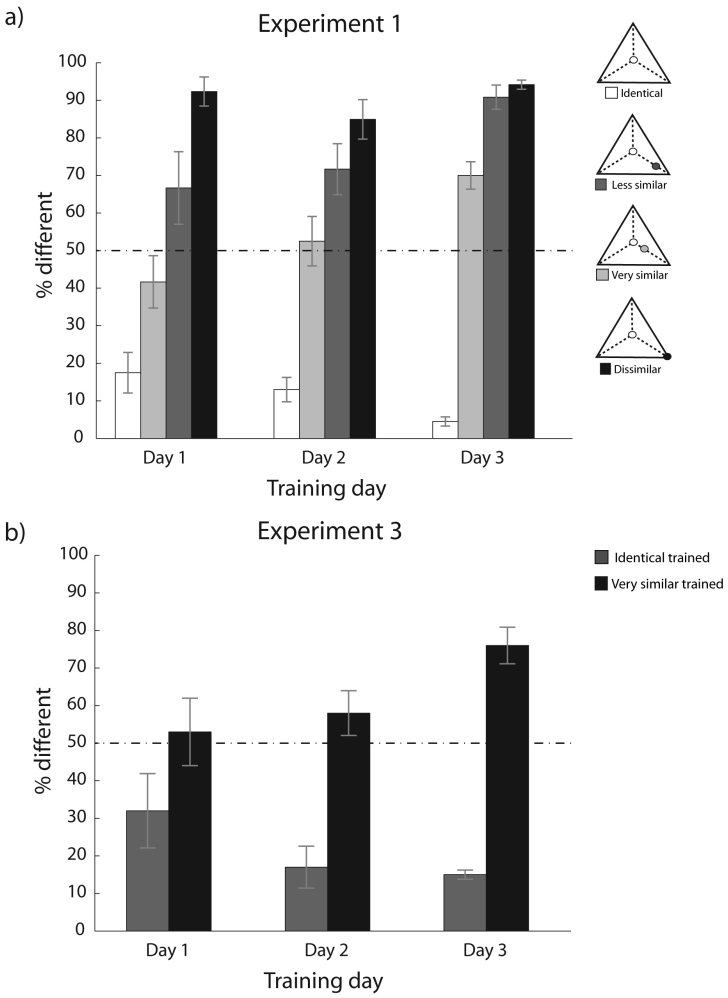

Figure 4.3: Mean behavioral performance (% different) during the training sessions in the laboratory for Experiments 5 (panel a) and 7 (panel b).

Experiment 6: Trained vs. Untrained Human-like Movements

The experiment consisted of three days of training, followed by one scanning session (day 4). The training procedure was similar to that of Experiment 5 with the exception that three out of the six different movement groups were selected for training (trained stimuli) while the other three groups were only presented during the scanning session without any training (untrained stimuli). The stimuli used for training and the untrained stimuli were counterbalanced across subjects. During each training session, observers were trained with feedback for 141 trials; that is 47 trials for each of the three movement groups.

Post-Training scanning session: Four conditions were tested: a) Identical Trained: the same trained center stimulus was presented twice in a trial, b) Very Similar Trained: a trained center stimulus was followed by a trained near off-

center stimulus, c) Identical Untrained: the same untrained center stimulus was presented twice in a trial, d) Very Similar Untrained: an untrained center stimulus was followed by an untrained near off-center stimulus. As in Experiment 5, the scanning session consisted of four event-related runs without feedback. The observers were instructed to judge whether two successively presented stimuli were the same or different. Each run consisted of 20 experimental trials for each of the four conditions and 20 fixation trials that were balanced for their history. Each trial lasted 4 seconds and started with the first stimulus (one movement cycle) for 1300ms, followed by a 100ms blank, the second stimulus for 1300ms followed by a blank interval of 1300ms.

Experiment 7: Learning Novel Artificial Complex Movements

The experiment consisted of a Pre-Training scanning session (day 1) followed by three consecutive days of training (day 2-4), and a Post-Training scanning session (day 5). The procedure and design for the training and scanning sessions was the same as in Experiment 6 with the exception that artificial movement stimuli were used.

4.2.4 Imaging

Data were collected in a 3T Siemens scanner (University Clinic, Tübingen) with gradient echo pulse sequence (TR=2s, TE=30ms, FA=80) for 24 axial slices (3x3x5 resolution) using a standard head coil.

4.2.5 fMRI Data Analysis

The fMRI data were processed using Brain Voyager 2000. After removing linear trends, temporal filtering as well as correcting for head movements, the 2D functional data were aligned to the 3D anatomical data and co-registered across scanning sessions.

For each individual subject, the retinotopic visual areas, hMT+/V5 and V3B/KO were localized based on standard mapping procedures (Dupont et al., 1997; Engel et al., 1994; Tootell et al., 1995). Area STSp was defined as the set of contiguous voxels along the superior temporal sulcus that showed significantly stronger activation ($p < 10^{-4}$) for intact than scrambled point-light walkers, consistent with previous studies (Bonda et al., 1996; Grossman and Blake, 2002;

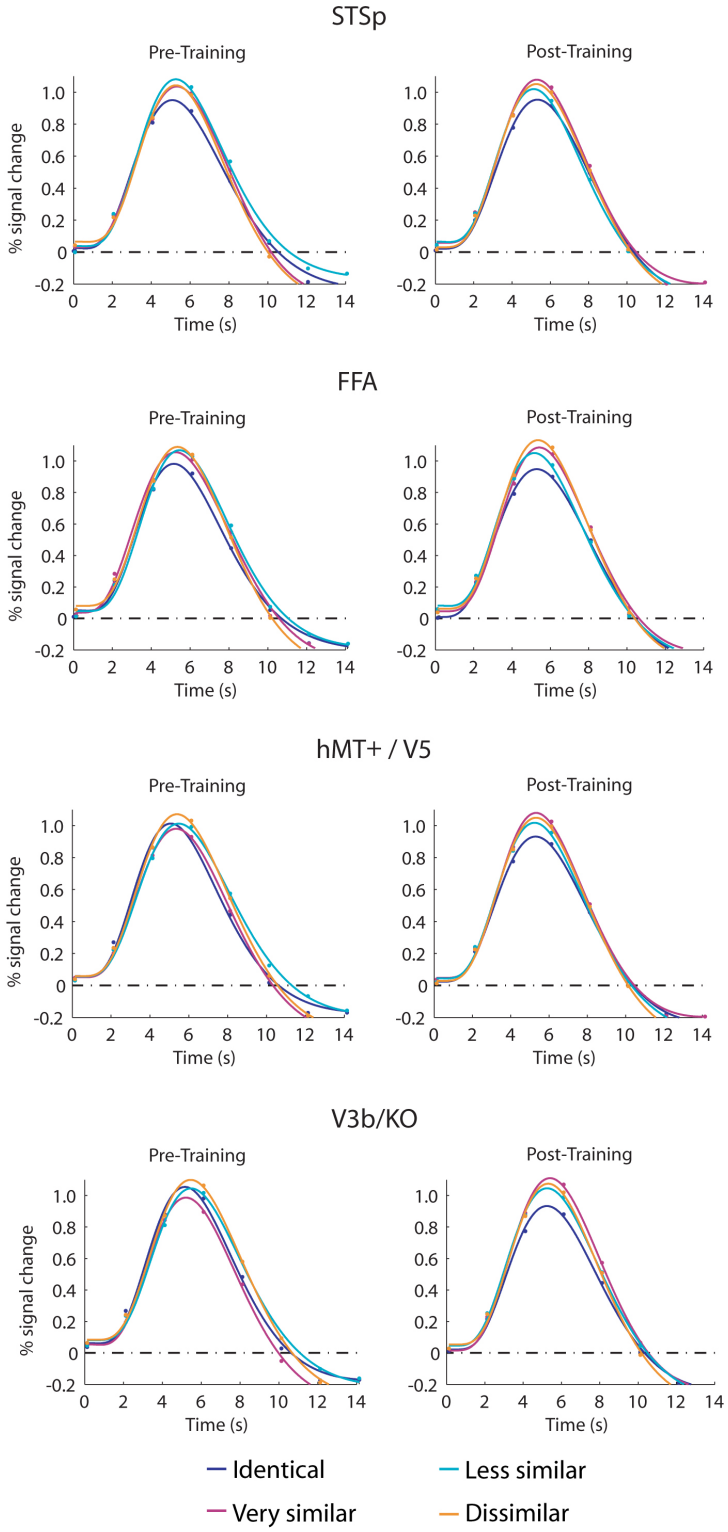

Figure 4.4: Time courses of the fMRI responses and fits across conditions for the Pre-Training and the Post-Training scanning sessions in Experiment 5. The peak points of the fMRI time courses were obtained by fitting the percent signal change in each ROI with a hemodynamic response model based on the difference of two gamma functions, as described previously (Boynton and Finney, 2003). Fits and peak time points were obtained for all experiments.

Vaina et al., 2001) (Table 4.1). Similarly, stronger activation ($p < 10^{-4}$) for intact than scrambled point-light walkers was observed in a region at the fusiform gyrus that overlaps with the face selective area FFA, consistent with previous studies (Grossman and Blake, 2002). This overlap was verified by significantly stronger responses to intact than scrambled faces tested in five of the observers that participated in Experiment 5 (Table 4.1). A recent fMRI study of high spatial resolution (1.4x1.4x2.0) showed that only a subregion of the FFA is involved in the processing of body parts and therefore possibly of biological movements (Schwarzlose et al., 2005). Identifying this subregion could have strengthened our effects in the FFA, as it would minimize partial volume effects from neural populations in this region that are non-selective to biological motion. Unfortunately, the resolution used in our study (3x3x5) did not allow us to discern this subregion.

For each ROI, we calculated the fMRI response by extracting the signal intensity for every trial from trial onset (0 - 14 sec) and averaging across trials for each condition. The resulting time courses were converted to percent signal change relative to the fixation condition and averaged across runs and subjects, as described previously. Because of the hemodynamic response properties, the peak of the BOLD signal is expected to occur with a delay of several seconds after trial onset. Fitting the time course data with the hemodynamic response function and ANOVA analysis across time points for each ROI indicated that peak time fMRI responses occurred between four and six seconds after trial onset, consistent with the hemodynamic response properties (Figure 4.4, 4.5) Thus, we used the average fMRI response at these time points for further statistical analysis of the differences between conditions. Repeated-measures ANOVAs and contrast analyses on significant interactions (Greenhouse-Geisser) for Training

	LH			RH		
Area	X	Y	Z	X	Y	Z
SPSp	-48 (±4)	-49 (±2)	10 (±3)	48 (±3)	-47 (±4)	13 (±3)
FFA	-38 (±3)	-46 (±4)	-13 (±4)	37 (±4)	-43 (±5)	-13 (±4)
FFA (face)	-36 (±2)	-47 (±4)	-16 (±3)	35 (±3)	-41 (±4)	-14 (±4)
hMT+ / V5	-38 (±3)	-63 (±3)	3 (±4)	42 (±3)	-60 (±2)	4 (±4)
V3b / KO	-33 (±3)	-83 (±3)	2 (±3)	33 (±4)	-80 (±3)	4 (±3)

Table 4.1: Mean Talairach coordinates and standard deviations (in parentheses) for biological motion-related areas: STSp: posterior superior temporal sulcus; FFA: fusiform face area localized with intact and scrambled biological motion stimuli; FFA (face): fusiform face area localized with intact and scrambled faces and motion-related areas; hMT+ / V5: middle temporal area; V3b / KO: kinetic occipital area.

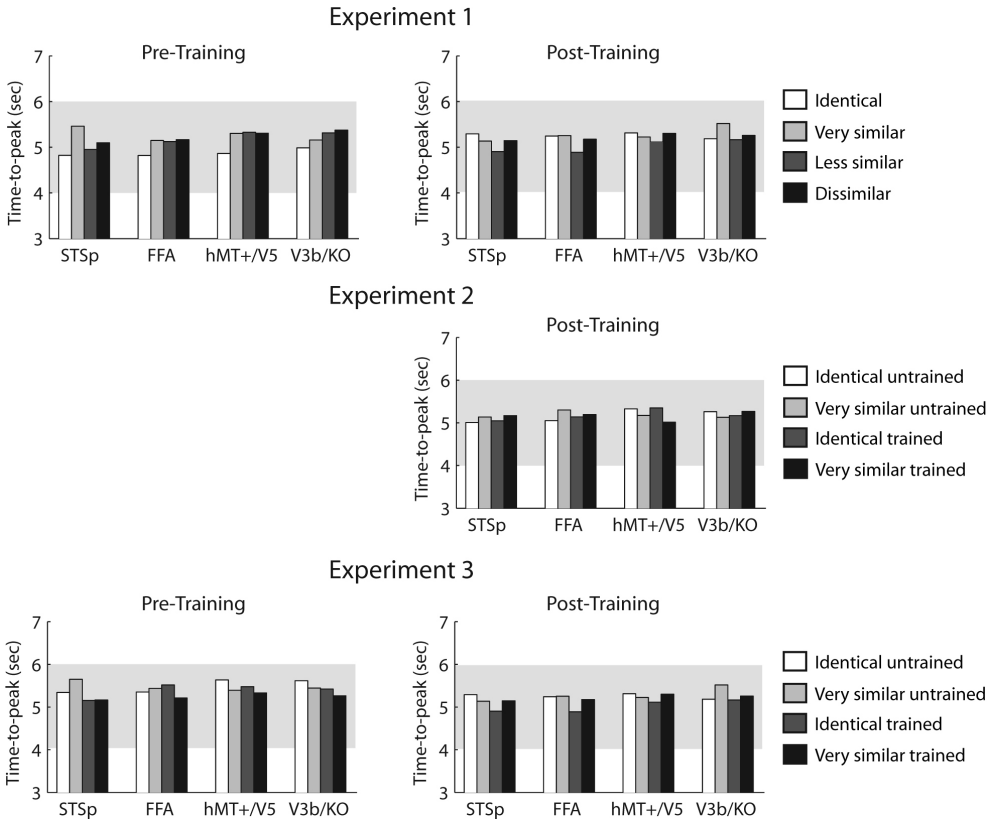

Figure 4.5: Time-to-peak for each ROI and experiment based on the fits of the fMRI time courses with the hemodynamic response model (Figure 4.4). All time courses across conditions, ROIs and experiments peaked between four and six seconds, consistent with the hemodynamic response properties. The shaded gray regions represent these time points that were averaged to quantify the fMRI responses for each individual subject in each ROI.

(Pre-Training, Post-Training scanning session), Condition (Identical, Very Similar, Less Similar, Dissimilar), and Stimulus (Trained, Untrained) were used for the analysis of the psychophysical and fMRI time course data.

4.3 Results

4.3.1 Experiment 5: Learning Novel Human-like Movements

Behavioral performance: Figure 4.6a, shows the observers' performance (percentage of different responses) in discriminating novel human-like movements during scanning before (Pre-Training) and after (Post-Training) training. Ob-

servers were highly accurate in discriminating identical and dissimilar stimuli before training. However, the observers' performance was significantly improved for very similar and less similar stimuli after training ($F_{(3,27)}$ = 26.69, p $<$ 0.001).

fMRI data: biological motion-related areas: Figure 4.6b shows the fMRI responses in areas involved in biological motion processing (STSp and FFA) before and after training in discriminating novel human-like movements. We observed stronger fMRI responses when different movements were presented in a trial (Very Similar, Less Similar, Dissimilar) than when the same movements were presented twice, consistent with previous adaptation studies (Grill-Spector et al., 2001). Interestingly, fMRI selective adaptation was observed both before (STSp: $F_{(3,27)}$ = 6.38, p $<$ 0.01, FFA: $F_{(3,27)}$ = 3.25, p $<$ 0.05) and after training (STSp: $F_{(3,27)}$ = 6.09, p $<$ 0.01, FFA: $F_{(3,27)}$ = 10.06, p $<$ 0.001) in these areas.

These fMRI selective adaptation effects before and after training are illustrated more clearly by means of a rebound index (Figure 4.6c). This index was calculated by dividing fMRI responses in each condition by responses to the identical condition. An index of one indicates adaptation due to repetition of the same stimulus, while an index higher than one indicates recovery from adaptation and thus neural sensitivity to differences between the stimuli presented in a trial. Figure 4.6c shows neural sensitivity in areas STSp and FFA (rebound index higher than 1) for differences between human-like movements both before and after training. Comparison of the rebound index before and after training showed enhanced fMRI selective adaptation after training (STSp: $F_{(2,18)}$ = 4.11, p $<$ 0.05, FFA: $F_{(2,18)}$ = 2.89, p $<$ 0.05). Thus, these results suggest that learning facilitates the discrimination of biological movements by enhancing the sensitivity of neural populations in biological motion-related areas (STSp, FFA) to the differences between novel human-like movements.

fMRI data: motion-related areas: In contrast to the activations in biological motion-related areas, fMRI selective adaptation was observed in motion-related areas (hMT+/V5, V3b/KO) after, but not before training (Figure 4.6d). That is, significantly stronger fMRI responses were observed when different movements were presented in a trial than when the same movements were presented twice after (hMT+/V5: $F_{(3,27)}$ = 5.54, p $<$ 0.01, V3b/KO:$F_{(3,27)}$ = 4.02, p $<$ 0.05) but not before training (hMT+/V5: $F_{(3,27)}$ = 1.11, p = 0.36, V3b/KO: $F_{(3,27)}$ = 1.05, p = 0.39).

Figure 4.6e summarizes these learning effects showing that rebound indices were not significantly different from one before training indicating adaptation, but were significantly higher than one after training indicating recovery from adaptation and emerging sensitivity to stimulus differences. Comparison of the rebound indices before and after training showed significantly higher fMRI se-

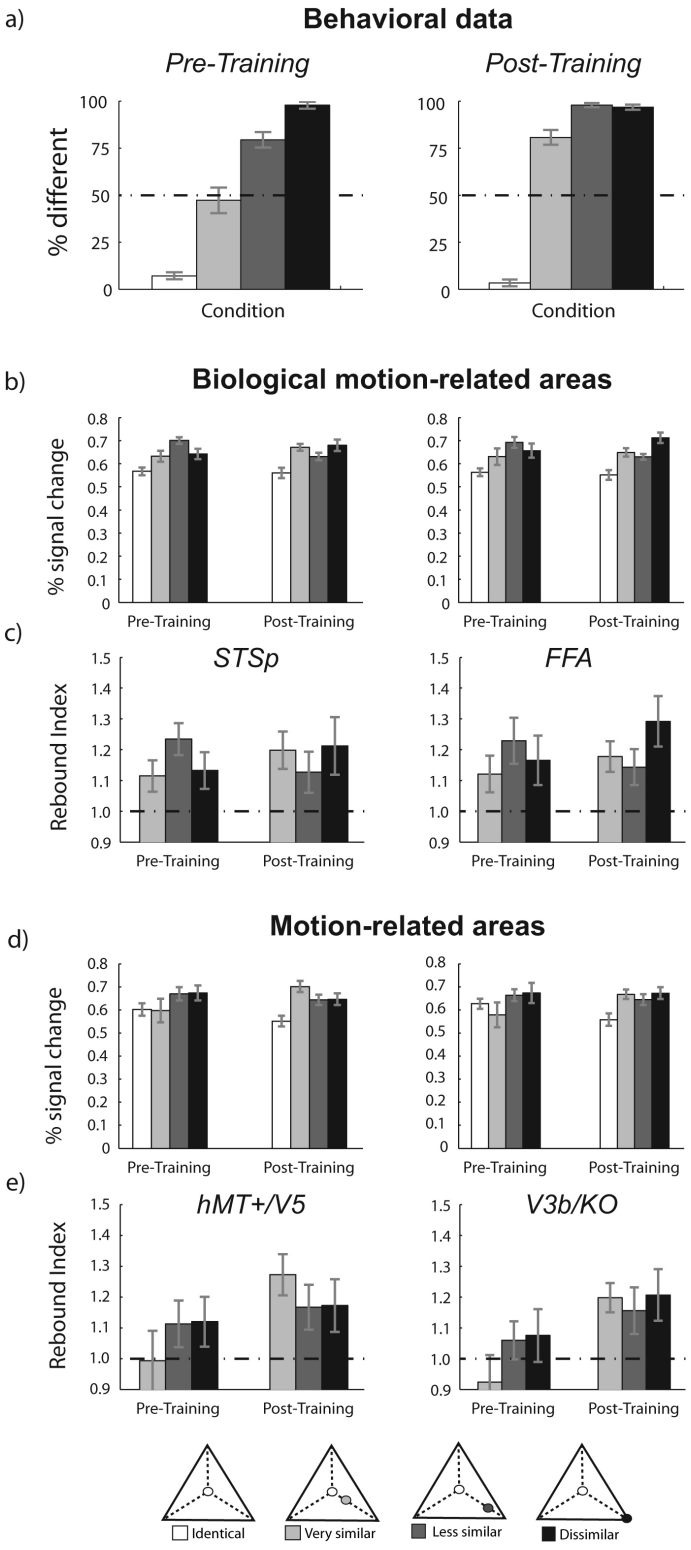

Figure 4.6: a) Psychophysical data obtained during the scanning sessions before training (Pre-Training) and after training (Post-Training). The data are expressed as percent of different judgments to the successive presentation of two movement stimuli per trial (% different). b) Average peak fMRI response across subjects in biological motion-related areas (STSp, FFA) before and after training. c) Rebound indices for the fMRI responses in biological motion-related areas (STSp, FFA) before and after training. d) Average peak time points of the fMRI time course across subjects in motion-related areas (hMT+/V5, V3B/KO) before and after training. e) Rebound indices for the fMRI responses in motion-related areas (hMT+/V5, V3B/KO) before and after training. Error bars indicate standard error of the mean across subjects (SEM)

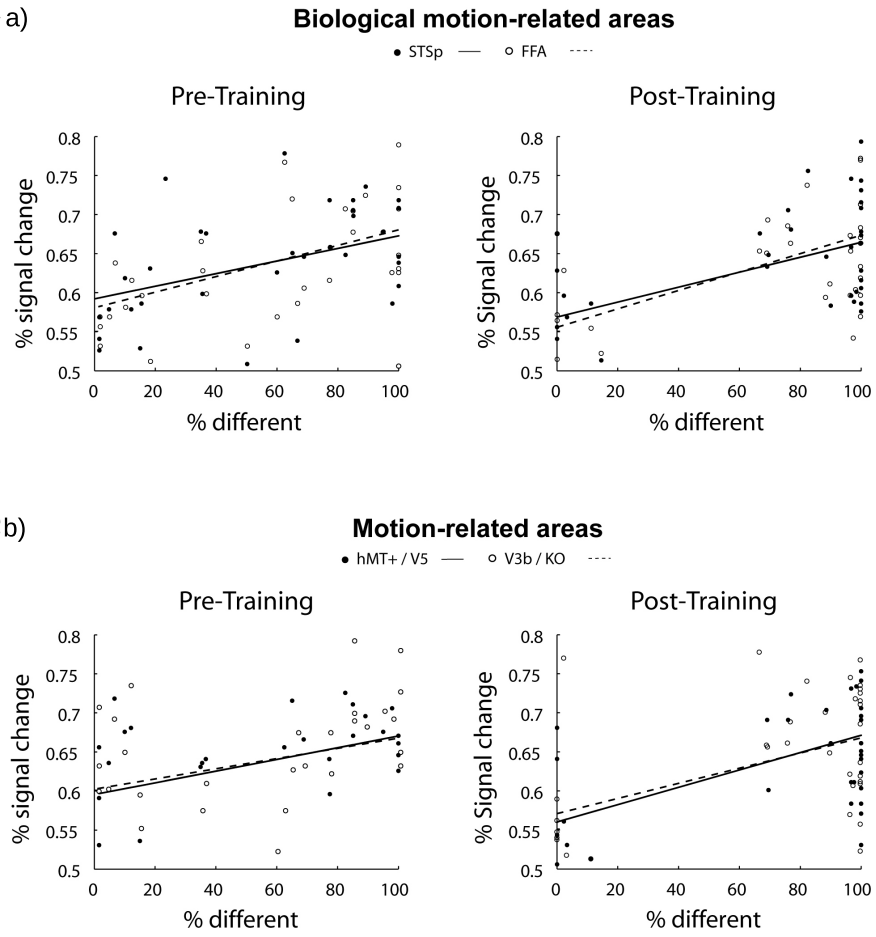

Figure 4.7: Normalized fMRI data of each subject are plotted against the corresponding psychophysical response in the different conditions for Pre-Training and Post-Training scanning sessions separately for a) biological motion-related areas and b) motion-related areas. Regressions for STSp and hMT+/V5 are shown by solid lines, whereas for FFA and V3B/KO by dashed lines. See Table 4.2 for regression statistics.

lective adaptation after training (hMT+/V5: $F_{(1,18)}$ = 9.02, p $<$ 0.01, V3b/KO: $F_{(1,18)}$ = 8.95, p $<$ 0.01). These results suggest that neural sensitivity to differences between human-like movements, as measured by fMRI, emerges in motion-related areas only after training, in contrast to biological motion-related areas that exhibit some sensitivity to these differences before training.

Comparison of behavioral performance and fMRI responses To further quantify the relationship between the behavioral and fMRI learning effects, we conducted a regression analysis on the individual subjects psychophysical and fMRI data across areas (Figure 4.7). Before training, a significant correlation was observed

between behavioral performance and fMRI responses in biological motion but not motion-related areas. However, after training these correlations were significant across all areas (Table 4.2). This analysis provides additional evidence for a link between behavioral improvement and experience-dependent neuronal changes. In particular, improvement in the discrimination of novel human-like movements due to training is reflected by enhancement of the neural sensitivity to the critical differences between movements for this discrimination in biological motion-related areas and emerging sensitivity in motion-related areas.

4.3.2 Experiment 6: Learning Specificity vs. Generalization

We investigated whether the learning effects observed for novel human-like movements were specific to the trained movements or generalized to untrained stimuli. We focused on the Identical and Very Similar conditions as the data from Experiment 5 showed the most prominent learning effects for very similar movements. We followed the same training procedure as in Experiment 5, but conducted only one scanning session after training, as these conditions had been already tested before training in Experiment 5. Behavioral improvement and enhancement of fMRI selective adaptation for trained stimuli would indicate learning. Similar learning effects for trained and untrained stimuli would indicate generalization to novel stimuli, whereas stronger effects for trained than for untrained stimuli would suggest learning that is specific for the trained movement patterns.

Behavioral performance: Figure 4.8a shows that the observers' accuracy was significantly higher ($F_{(1,10)}$ = 8.03, p < 0.05) for trained compared to untrained stimuli in discriminating very similar movements but did not differ for identical trained and untrained movements. Thus, the learning effects observed in Experiment 5 were specific to the trained movements as indicated by a) the similar

	Pre-Training				Post-Training			
Area	STSp	FFA	hMT+	KO	STSp	FFA	hMT+	KO
R	0.41	0.38	0.24	0.20	0.51	0.58	0.51	0.45
p	0.01	0.01	0.16	0.24	< 0.01	< 0.01	< 0.01	< 0.01
$F_{(1,34)}$	6.69	5.87	2.10	1.43	12.01	17.34	12.20	8.73

Table 4.2: Regression analysis of fMRI and behavioral responses before (Pre-Training) and after (Post-Training) training across regions of interest for Experiment 5.

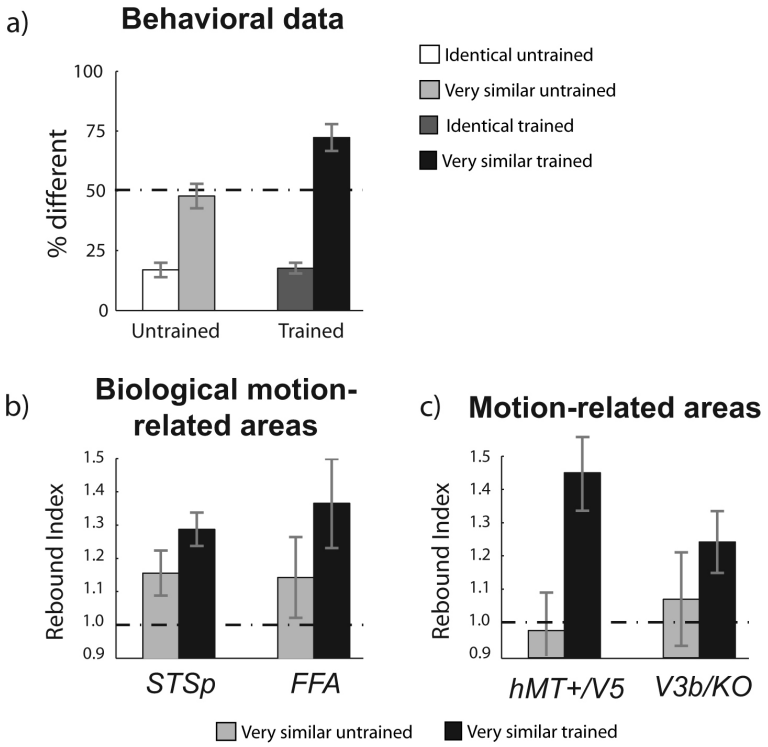

Figure 4.8: a) Psychophysical data (% different) during scanning after training. Rebound indices for b) biological motion-related areas (STSp, FFA) and c) motion-related areas (hMT+/V5, V3B/KO). Error bars represent SEM.

performance for untrained very similar movements in this experiment and the same movements before training in Experiment 5, and b) the similar behavioral improvement for these stimuli after training in the two experiments.

fMRI data: biological motion-related areas: Figure 4.8b shows fMRI selective adaptation effects (rebound indices > 1) for trained and untrained very similar movements in areas involved in biological motion processing (STSp, FFA); that is, fMRI responses were significantly stronger for very similar than identical movements (STSp: $F_{(1,10)} = 32.33$, $p < 0.001$, FFA: $F_{(1,10)} = 10.56$, $p < 0.01$). This fMRI selective adaptation was enhanced after training, as indicated by the higher rebound indices for trained than untrained movements (STSp: $F_{(1,10)} = 5.41$, $p < 0.05$, FFA: $F_{(1,10)} = 6.31$, $p < 0.05$). These results are consistent with the findings from Experiment 5 showing that significant fMRI selective adaptation for very similar movements before training in biological motion-related areas was enhanced after training. Importantly, these findings suggest that experience-based plasticity is specific to the trained movements and does not generalize to untrained stimuli.

fMRI data: motion-related areas: fMRI selective adaptation (rebound indices) for very similar movements was observed for trained but not untrained stimuli (Figure 4.8c) in motion-related areas (hMT+/V5, V3b/KO). That is, fMRI

responses for very similar stimuli were significantly higher than responses for identical stimuli for the trained movements (hMT+/V5: $F_{(1,10)}$ = 8.59, p $<$ 0.05, V3B/KO: $F_{(1,10)}$ = 6.42, p $<$ 0.05), but not for the untrained movements (hMT+/V5: $F_{(1,10)} < 1$, p = 0.82, V3b/KO: $F_{(1,10)} < 1$, p = 0.38). Thus, consistent with the results from Experiment 5, these results suggest that neural populations in motion-related areas develop sensitivity for very similar novel human-like movements after training that is specific to the trained stimuli.

4.3.3 Experiment 7: Learning Novel Artificial Complex Movements

It is possible that the learning effects observed for novel human-like movements were due to the fact that observers were familiar with the prototypical actions used to generate the novel spatio-temporal movement morphs. As a result, training could have enhanced the observers' ability to generalize and discriminate between novel exemplars from the same class of movements. To control for the possible effect of prior knowledge about the stimulus class, we used novel artificial complex movements. These patterns shared similar biological motion characteristics with human actions (i.e. articulation, approximately sinusoidal joint motion, consistency with the 2/3 power law), but were constructed based on artificial skeletons that were chosen to be highly dissimilar from naturally occurring body structures and thus were not interpretable as typical human or animal actions (see section 4.2). As for the human-like stimuli, we tested the effect of training on behavioral performance and fMRI responses when observers discriminated novel artificial movement patterns that were generated based on spatiotemporal linear interpolation between prototypical patterns. Observers were tested with identical and very similar movements in two scanning sessions, one before and one after training. For each scanning session, half of the stimuli were the same as those presented during the training session (trained), while the rest of the stimuli were presented only during the scanning sessions (untrained).

Behavioral performance: Figure 4.9a shows that observers exhibited significantly higher accuracy for trained than for untrained movements after ($F_{(1,8)}$ = 15.06, p $<$ 0.01) but not before ($F_{(1,8)} < 1$, p = 0.37) training. As in Experiment 5, before training the observers' performance for very similar stimuli was less accurate than for identical stimuli. However, after training the observers' performance in discriminating very similar movements improved significantly while their performance with identical movements did not change ($F_{(1,8)}$ = 29.41, p $<$ 0.01).

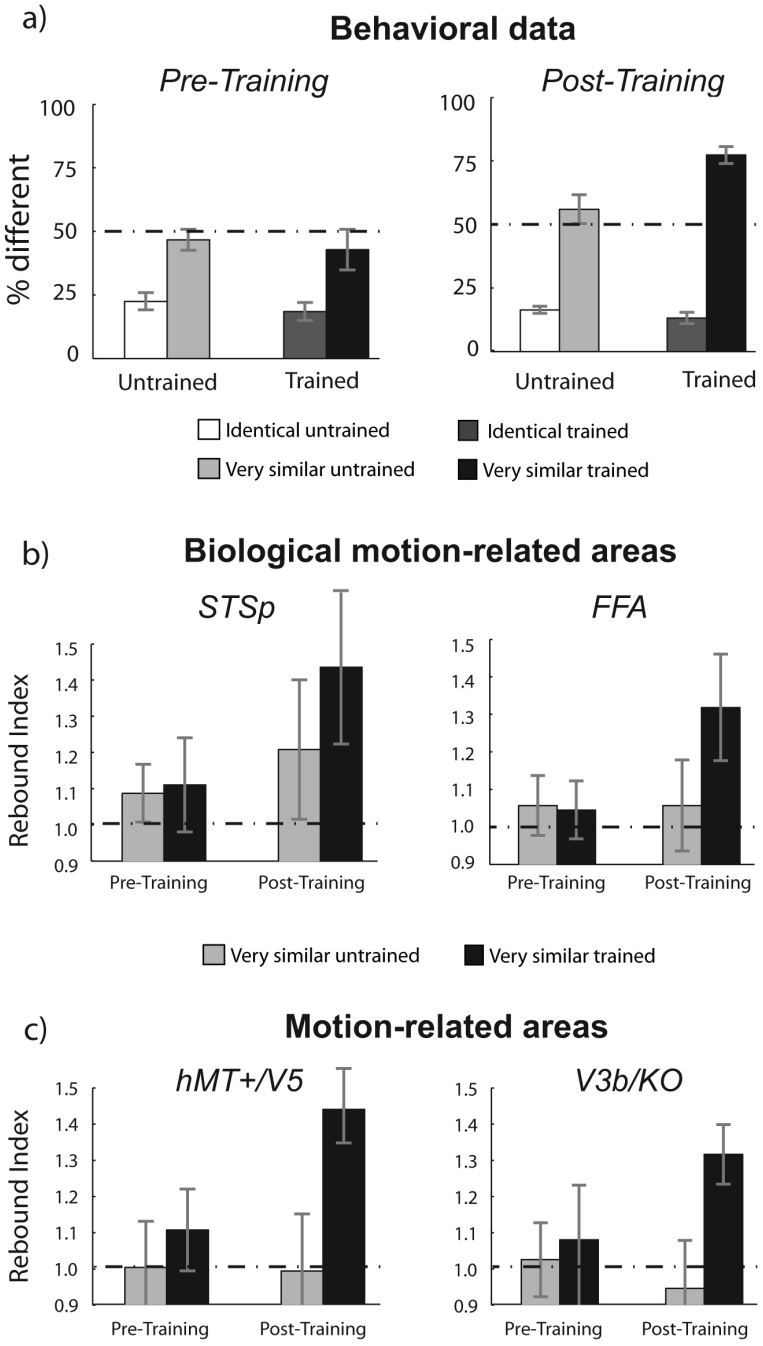

Figure 4.9: a) Psychophysical data (% different) during scanning before (Pre-Training scanning session) and after training (Post-Training scanning session). Rebound indices for b) biological motion-related areas (STSp, FFA) and c) motion-related areas (hMT+/V5, V3B/KO) before and after training. Error bars represent SEM.

fMRI data: biological motion-related areas: Figure 4.9b shows fMRI selective adaptation effects (rebound indices) for very similar movements specific to the trained stimuli after but not before training in biological motion-related areas (STSp, FFA). That is, we observed significantly stronger fMRI responses for very similar artificial movements compared to identical movements when the observers were tested with the trained stimuli (STSp: $F_{(1,8)}$ = 7.06, p $<$ 0.05, FFA: $F_{(1,8)}$ = 6.65, p $<$ 0.05), but no significant differences when the observers were tested with the untrained stimuli (STSp: $F_{(1,8)}$ = 2.71, p = 0.14, FFA: $F_{(1,8)} < 1$, p = 0.58). No significant fMRI selective adaptation was observed before training (STSp: $F_{(1,8)}$ = 1.84, p = 0.21, FFA: $F_{(1,8)} < 1$, p = 0.39). In contrast to fMRI selective adaptation for human-like movements before training in biological motion-related areas (Experiment 5), neural sensitivity for artificial movements emerged only after training and was specific to the trained stimuli.

fMRI data: motion-related areas: Similar to the activations in biological motion-related areas, fMRI selective adaptation (Figure 4.9c) specific to the trained movements was observed after but not before training in motion-related areas (hMT+/V5, V3b/KO). That is, no significant differences were observed between very similar and identical artificial movements before training (hMT+/V5; $F_{(1,8)} < 1$, p = 0.36, V3b/KO: $F_{(1,8)}$ = 1.31, p = 0.28). However, after training, we observed significantly stronger fMRI responses for very similar compared to identical movements when the observers were tested with the trained stimuli (hMT+/V5: $F_{(1,8)}$ = 5.40, p $<$ 0.05, V3b/KO: $F_{(1,8)}$ = 5.25, p $<$ 0.05), but no significant differences when the observers were tested with the untrained stimuli (hMT+/V5: $F_{(1,8)} < 1$, p = 0.93, V3b/KO: $F_{(1,8)} < 1$, p = 0.52). Thus, in motion-related areas neural sensitivity to the differences between very similar artificial movements emerged only after training, and this learning was specific to the trained stimuli.

4.3.4 Learning of Novel Complex Movements in Retinotopic Visual Cortex

We further examined whether retinotopic visual areas are engaged in the learning of biological movements. We tested fMRI responses for human-like and artificial movements in Experiments 6 and 7 that tested comparable conditions (trained, untrained stimuli).

Figure 4.10a shows fMRI selective adaptation (rebound indices) for trained and untrained human-like movements across retinotopic visual areas (Experiment 6). No significant differences were observed in the fMRI responses for very similar and identical movements when the observers were presented with

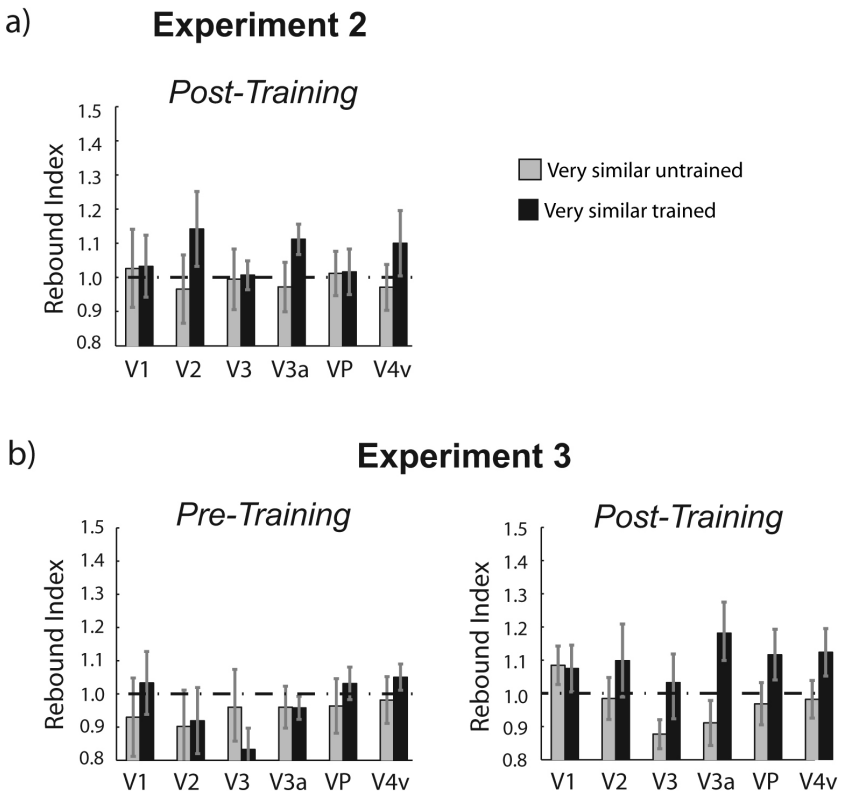

Figure 4.10: Rebound indices based on the activations in the retinotopic visual areas for a) Experiment 6 (Post-Training scanning session) and b) Experiment 7 (Pre- and the Post-Training scanning sessions). Error bars represent SEM.

trained or untrained stimuli. Areas V3a and V4 showed a trend for stronger fMRI selective adaptation for trained compared to untrained stimuli, but these effects did not reach significance (Table 4.3). Figure 4.10b shows fMRI selective adaptation effects (rebound indices) for the artificial movement stimuli (Experiment 7). Before training, no significant fMRI selective adaptation was observed for the stimuli used in the training sessions or the untrained stimuli. However, after training we observed stronger fMRI selective adaptation for trained than untrained movements in V3a, Vp and V4 but not in V1, V2, or V3 (Table 4.3).

The lack of fMRI selective adaptation for very similar biological and artificial movements in V1 and V2 suggests that our findings in higher visual areas were due to learning differences in the global structure rather than the local features of the movements. Interestingly, ventral (VP, V4) and dorsal (V3a) retinotopic areas, downstream from primary visual cortex showed enhanced fMRI selective adaptation specific to trained movements. These effects were more prominent for artificial than human-like movements suggesting that learning novel artificial patterns may recruit earlier (middle level) stages of visual analysis, whereas learning novel human-like movements may shape higher-level motion processes.

4.4 Discussion

Our results provide novel evidence for experience-based neural plasticity in the human visual cortex that mediates the learning of novel human-like and artificial complex movements. In particular, our study reveals the following main findings: First, by combining psychophysical and fMRI selective adaptation measurements, we demonstrate that training to discriminate highly similar complex movements results in behavioral improvement and increased neural sensitivity to differences between these movements that may represent the critical features for their discrimination (Figure 4.11). These learning effects were specific to the trained movements and did not generalize to untrained stimuli, suggesting movement-specific learning rather than a general improvement on the discrimination task that generalizes across different stimuli. Second, our results reveal that this experience-based plasticity is distributed not only across areas that are known to be involved in the processing of biological motion (STSp, FFA) but also across areas implicated in the analysis of physical motion (hMT+/V5, V3b/KO). Interestingly, we observed differences in the learning for different classes of complex movements across these areas. That is, biological motion-related areas exhibited a small but significant sensitivity to differences between very similar human-like movements before training that was enhanced after training. This finding is consistent with the specificity of these areas in the analysis of biological motion (Beauchamp et al., 2003; Bonda et al., 1996; Grezes et al., 2001; Grossman and Blake, 2002; Peuskens et al., 2005; Vaina et al., 2001) and their role in action understanding (Blakemore and Decety, 2001; Rizzolatti and Craighero, 2004). In contrast, such sensitivity emerged only after training in motion-related areas, consistent with their role in learning of global motion configurations (Vaina et al., 1998; Zohary et al., 1994). Third, behavioral improve-

	Experiment 6		Experiment 7			
	Post-Training		Pre-Training		Post-Training	
ROI	$F_{(1,10)}$	p	$F_{(1,8)}$	p	$F_{(1,8)}$	p
V1	< 1.0	0.45	1.21	0.15	1.09	0.34
V2	1.87	0.10	< 1.0	0.47	2.85	0.07
V3	< 1.0	0.45	< 1.0	0.19	2.05	0.09
V3a	2.33	0.08	< 1.0	0.49	6.19	< 0.01
VP	< 1.0	0.47	2.42	0.08	5.53	< 0.05
V4v	1.76	0.11	1.82	0.11	3.35	< 0.05

Table 4.3: Statistical analysis (repeated measures ANOVA) on fMRI responses across conditions in retinotopic areas for Experiments 6 and 7.

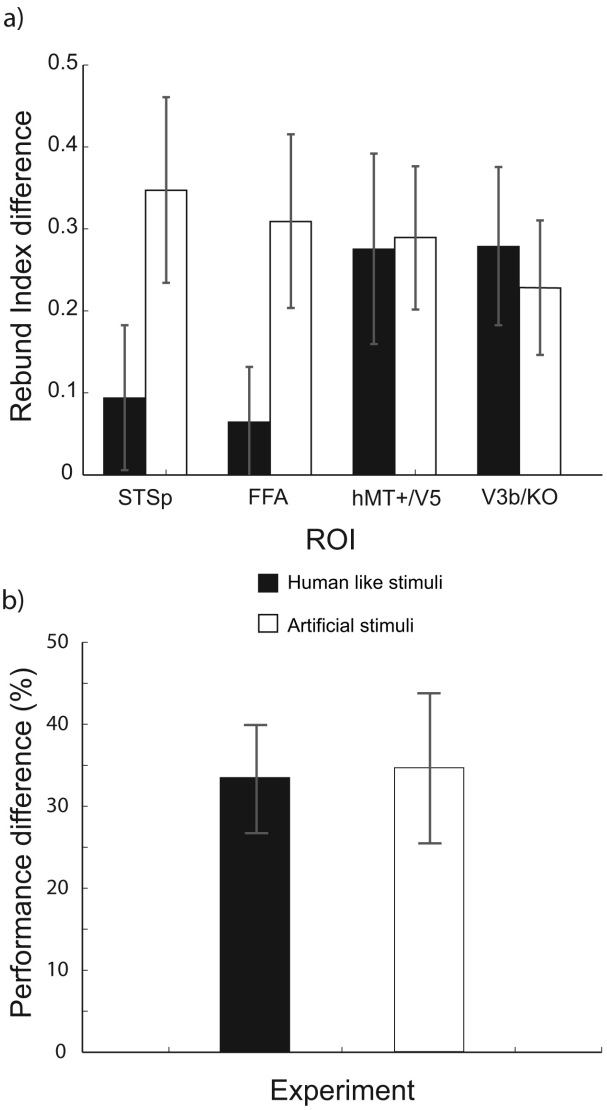

Figure 4.11: a) Differences in the rebound indices before and after training (Post-Training rebound index minus Pre-Training rebound index) for the very similar human-like (Experiment 5) and artificial movements (Experiment 7; trained stimuli) in biological motion- and motion-related areas. These learning effects were similar for human-like movements and artificial patterns in motion-related areas. However, in biological motion-related areas differences in the fMRI selective adaptation before and after training were stronger for artificial patterns than human-like movements, consistent with fMRI selective adaptation for human-like movements before training in these areas. b) Differences in behavioral performance (% different responses) between the Pre-Training and Post-Training sessions for the same experiments. Error bars represent SEM.

ment and emerging neural sensitivity after but not before training were observed in motion and biological motion-related areas for artificial biological movements with which the observers had no previous experience. Further, learning of artificial rather than human-like movements engaged retinotopic visual areas (V3a, Vp, V4).

Could these differences in the learning of human-like and artificial movements be due to low level differences in the stimuli or differences in the observers' performance? We think that this is unlikely, as the two stimulus classes were matched for their low-level properties (area, number of dots, number of segments), and the distances between stimuli in the morphing space (see Sec-

tion 4.2). Further, both stimulus types fulfilled the 'two-thirds power law' that links the curvature and the speed of approximately planar human movements. Jittering of the dots on the segments ensured that observers learned the global movement configurations rather than simply the local positions of the dots[2] (see Experimental Specifications, 4.2). Finally, the psychophysical performance of the observers and the behavioral learning effects were similar for human-like and artificial movements (Figure 4.11b), suggesting that the differences in the fMRI activation patters for these stimulus classes could not be due to differences in the observer's performance.

It is also unlikely that differential attentional allocation could account for the observed pattern of fMRI responses. In contrast to our fMRI results, an attentional load explanation would predict higher fMRI responses when the discrimination task was hardest as difficult conditions require prolonged, focused attention resulting in higher fMRI responses (Ress et al., 2000). We observed the opposite effect: fMRI responses were highest for conditions in which the discrimination was easiest and subjects responded fastest (Figure 4.12). Further, the quick succession of randomly interleaved trials ensured that the observers could not attend selectively to particular conditions. Moreover, it is unlikely that the fMRI learning effects observed were due to the fact that the observers were more attentive after than before training for the trained compared to the untrained movements, as the task was more demanding before training and for the untrained stimuli. Reaction times in these difficult discrimination conditions were the slowest (Figure 4.12) rather than very fast, as it would be expected if the observers had given up and were simply guessing in these conditions. These psychophysical data indicate that observers were engaged in the task and not responding randomly both before and after training. Further, any increases in general alertness or arousal would result in increases in fMRI response across the visual areas (Ress et al., 2000). This is not consistent with the lack of fMRI learning effects that we observed in the primary visual cortex that is known to be modulated both by learning (Furmanski et al., 2004; Kourtzi et al., 2005) and attention (Ress et al., 2000). Finally, it is not likely that our learning results could be significantly confounded by eye movements. Eye movement recordings showed that the subjects were able to fixate for long periods of time and any saccades did not differ systematically in their number, amplitude or duration before and after training.

[2]This observation has been verified in preceding piloting experiment that are not reported in this thesis.

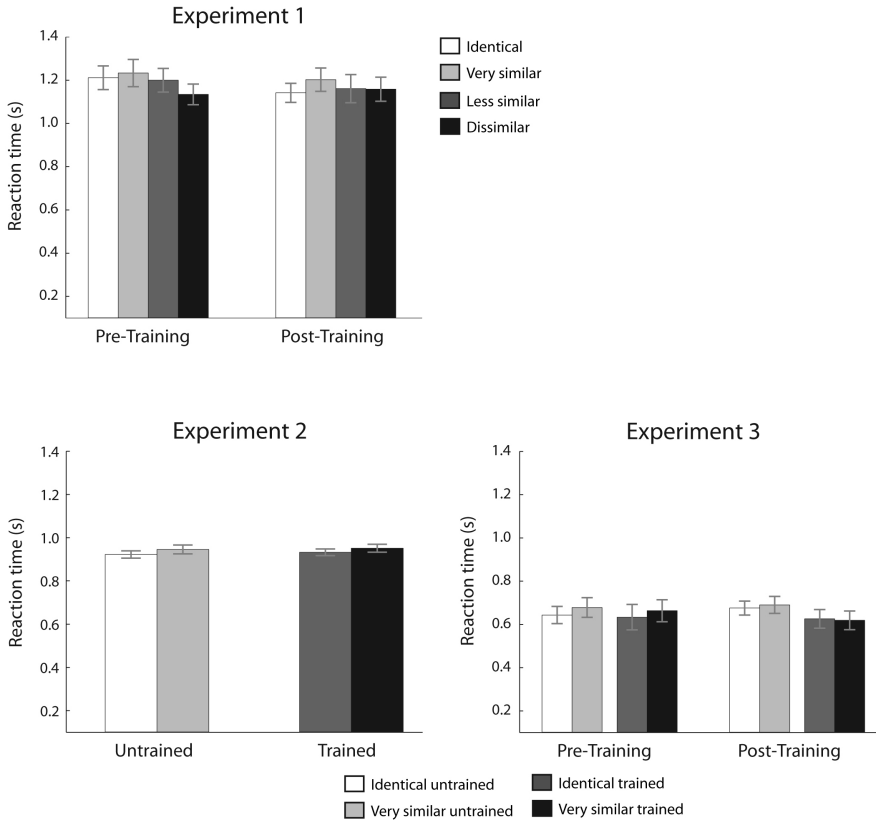

Figure 4.12: Reaction times (seconds) across conditions for all three Experiments. No significant differences were observed in the reaction times across conditions in Experiment 5 (Pre-Training: $F_{(3,27)}$ = 2.56; p = 0.07; Post-Training: $F_{(3,27)}$ = 1.42; p = 0.26), Experiment 6 (Post-Training: $F_{(1,10)}$ = 3.34; p = 0.12) and Experiment 7 (Pre-Training: $F_{(1,8)}$ = 3.22; p = 0.11; Post-Training: $F_{(1,8)} < 1$; p = 0.82). In Experiment 7, significantly faster reaction times were observed for the trained than untrained stimuli in the Post-Training scan ($F_{(1,8)}$ = 9.07; p = 0.02) but not in the Pre-Training one ($F_{(1,8)} < 1$; p = 0.43).

Distributed Experienced-based Plasticity for Visual Learning

Visual learning has been demonstrated in a variety of tasks (see Fine and Jacobs, 2002; Goldstone, 1998, for review) from the detection and discrimination of form and motion properties to face and object recognition. Electrophysiological (Gilbert et al., 2001; Zohary et al., 1994) and imaging (Dolan et al., 1997; Grezes et al., 1999; Kourtzi et al., 2005; Tarr and Cheng, 2003; Vaina et al., 1998) studies have investigated the neural correlates of these learning effects across visual areas. Learning of biological motion has been studied psychophysically in the context of facial movements and recognition of individual actors (Hill and Pollick, 2000; Knappmeyer et al., 2003; Troje et al., 2005). A recent

fMRI study (Grossman et al., 2004) has shown that learning improves observers' performance in detecting biological movements in noise and enhances fMRI responses in STSp and FFA. Unlike in our experiments, the learning effects in this study transferred to novel untrained stimuli, suggesting neural changes related to improved segmentation of the target movement from the motion of the background noise rather than specific encoding of novel complex motion patterns.

Our study advances our understanding of the experience-based plasticity mechanisms involved in the learning of novel complex biological movements in several respects. First, our experiments investigated the role of learning in the discrimination of novel biological movements that mediate recognition of individual actions rather than detection from their background that engages scene segmentation processes. Such learning entails analysis of the distinctive features of movements that are critical for the discrimination task, and thus results in representations specific to the trained stimuli. Second, the use of complex artificial movements that were matched for low level and biological movement features with human-like movements, but lack consistent interpretations related to familiar human actions allowed us to investigate in a controlled manner the learning of biological movements independent of the observers previous knowledge about specific movement patterns. Third, the combination of advanced spatiotemporal morphing techniques for the stimulus generation and fMRI selective adaptation paradigms provided us with sensitive tools for studying parametrically the link between changes in behavioral performance and the sensitivity of neural populations for the critical features involved in the discrimination of biological movements after training.

Our findings provide novel insights in understanding this link between experience-based behavioral improvement and neural plasticity that mediates learning of biological movements across visual areas in the human brain. Specifically, similarities in the learning substrates for human-like and artificial movements suggest similar mechanisms for the learning of these classes of complex movements. In particular, increased sensitivity for these movements was observed after training in motion and biological motion-related areas. It is likely that training results in processing of the common features shared by these two movement classes (i.e. local motion configurations and biological properties) in these areas. Interestingly, some differences in the learning substrates for human-like and artificial movements were also observed. Learning to discriminate between members of novel categories of artificial movements resulted in emerging sensitivity only after training in these higher-motion analysis areas and retinotopic areas that are involved in the integration of motion (V3a) and form (Vp, V4). However, learning new exemplars from known human movement cate-

gories appeared to shape existing representations of human actions in biological motion-related areas. This interpretation is consistent with the observation that neural sensitivity to movement differences that was evident in these areas before training was further enhanced after training.

These findings for learning of biological movements extend previous neurophysiological (Logothetis et al., 1995) and imaging (Gauthier et al., 1999) studies showing enhanced responses in higher occipito-temporal areas for learning novel object classes and generalizing across novel exemplars presented at orientations close to those of known class members. An area of special interest for the study of experienced-based plasticity in the temporal cortex is area FFA, as it has been implicated in the learning of novel object classes (Gauthier et al., 1999) as well as the selective processing of faces (Kanwisher et al., 1997) and biological motion (Grossman and Blake, 2002; Peuskens et al., 2005). A recent fMRI study suggests that high spatial resolution imaging may facilitate the identification of separate subregions within the fusiform cortex specialized for these functions (Schwarzlose et al., 2005). The resolution used in our experiments did not allow us to localize such subregions in FFA. As a result, the learning-induced changes we observed in the FFA are consistent with specialized mechanisms for the processing of biological features and learning of novel stimulus classes.

Measuring Experience-based Neural Plasticity with fMRI

One of the advantages of using fMRI for the study of learning is that it allows us to investigate experience-based plasticity changes in multiple human brain areas simultaneously. However, the spatio-temporal resolution of this method is suitable for studying neural plasticity changes at the level of large-scale neural populations rather than the single neuron. As a result, our study can not discern whether the increased fMRI selective adaptation for very similar movements after training is due to changes in the tuning of neurons selective for these movements or recruitment of larger numbers of neurons that become sensitive to these differences after training (Gilbert et al., 2001).

To study learning-dependent changes in subpopulations of neurons at a resolution beyond that of the typical human fMRI voxel, we used an fMRI selective adaptation paradigm (Grill-Spector et al., 2001; Kourtzi and Kanwisher, 2000b). This paradigm allow us to study neural sensitivity to feature differences that mediate discrimination between stimuli, as neurons responding to these features are intermingled within each voxel and thus could not be measured at the standard fMRI resolution. Recent electrophysiological studies provide evidence that fMRI adaptation effects are related to neural adaptation at the level of single

neurons (Sawamura et al., 2006; Tolias et al., 2005). fMRI selective adaptation has been used extensively for studying the processing of motion patterns (Huettel et al., 2004; Huk et al., 2002; Tolias et al., 2001). Our study is the first to introduce this paradigm as a sensitive tool for studying the processing and learning of complex biological movement patterns. In particular, we reasoned that enhanced fMRI selective adaptation for differences between stimuli (movements) after training provides a measure of changes in neural sensitivity due to learning. Given the complexity of the BOLD signal, fMRI selective adaptation studies can not discern whether the neural sensitivity observed at the scale of large neural populations reflects selectivity as measured by the spike output of individual neurons or adapted input from other neural populations within or across cortical areas (Sawamura et al., 2006; Tolias et al., 2005). These limitations notwithstanding, fMRI selective adaptation for motion has been verified in several single cell recording studies in area MT (e.g. Petersen et al., 1985). In our study, the lack of fMRI selective adaptation after training in V1 and V2 suggests that learning-dependent changes in neural sensitivity in higher visual areas could not be simply attributed to adaptation of input responses from the primary visual cortex.

Finally, studies using methods with higher temporal resolution than fMRI would be helpful in discerning feedforward from feedback processes in the learning of biological movements. Specifically, it is possible that learning of artificial movements could be implemented in a bottom-up manner from retinotopic to higher visual areas by enhancing processes of increasing complexity ranging from the integration of local configurations to the analysis of global motion features and biological motion properties. Alternatively, learning could begin at higher visual areas by enhancing the processing of the global biological characteristics of these movements and proceed to early retinotopic areas that have higher resolution necessary for the finer discrimination of differences between movements at the level of local configurations (Ahissar and Hochstein, 1997). Similarly, learning novel human-like movements could implicate top-down processes that support generalization to novel movements from known templates of human actions via feedback from motor areas that are thought to represent models for these actions (Saygin et al., 2004; Wolpert et al., 2003). The fMRI selective adaptation observed for differences between human-like movements before training in biological motion-related areas (STSp / FFA) could be attributed to such feedback processes. It is possible that the lack of fMRI selective adaptation before training in other areas or for artificial movements was due to limited fMRI sensitivity in detecting small differences in the activity of neural populations. An alternative explanation is that feedback processes from motor areas may modulate more strongly processing of movements in biological motion-related areas

that are involved in the action understanding network (Blakemore and Decety, 2001; Rizzolatti and Craighero, 2004), rather than motion-related or retinotopic areas. Similarly, internal motor models are more likely to exist before training for human-like than for the artificial movements. However, it is possible that the similarity of these stimulus classes in their stimulus characteristics (e.g. compliance with the 2/3 power law) may encourage interpretations of novel artificial movements using internal motor models which could be further enhanced by learning. Further work is necessary to investigate whether learning to discriminate artificial movements results in shaping not only visual representations but also motor models in parieto-frontal networks.

In conclusion, despite the spatio-temporal limitations of the fMRI resolution, our study advances our understanding of experience-based plasticity mechanisms for the learning of complex movements in the human visual cortex at the level of large-scale neural populations. Learning a novel class of movements is implemented by enhancing the sensitivity of local, global pattern motion and biological movement detectors across multiple stages of visual analysis, whereas learning novel exemplars of a known class (human-like actions) shapes the class representations by enhancing the sensitivity of global motion detectors and the generalization of biological movement detectors to novel class members. These findings provide novel insights into the mechanisms of human action understanding and the basis for future combined fMRI and neurophysiological studies that will shed light to neural plasticity mechanisms at the level of single neurons and their interactions within and across cortical areas involved in the visual analysis of movements and the planning of actions.

Chapter 5

Theoretical Modeling

5.1 Introduction

Although the functional imaging experiments have identified changes in BOLD activity in various areas of the visual cortex, these findings do not conclusively explain how learning shapes the processing of complex movements. While we were successfully utilizing an fMRI adaptation paradigm to increase the spatial resolution, we are limited in explaining how the properties of single neurons or at least small populations of neurons change in the context of learning. One possible way to test different predictions about the underlying plasticity mechanisms is the use of theoretical models.

In the previous years, many different research groups have started to investigate theoretical mechanisms by which the recognition of complex movements can be accomplished (see Aggarwal and Cai, 1999; Gavrila, 1999; Giese, 2006; Moeslund and Granum, 2001, for review). However, only a small number of these groups have tried to constrain their models in a way that the utilized mechanisms could actually mimic biologically plausible processes.

The model that will be used in this chapter has been recently developed by Giese and Poggio (2003) (Figure 5.1). Their model accomplishes movement recognition on the basis of learned prototypical movement patterns, which are encoded by neural feature detectors specialized for the detection of complex shapes and optic flow patterns. The receptive field sizes and the tuning properties of the model neurons are closely matched with those of real neurons in the visual cortex. Additionally, several simulations have shown that the model reproduces many results from psychophysical experiments and physiology that have been discovered in the context of biological motion recognition.

The model consists of two parallel processing streams that are analogues to the dorsal and ventral pathway proposed for the human visual cortex (Mishkin et al., 1983; Goodale and Milner, 1992). These streams analyze form and optic flow information separately but converge at the level that corresponds to the superior temporal sulcus in humans and monkeys. Both pathways consist of four hierarchy levels which extract either form or optic flow features with increasing complexity as well as position and size invariance along the different levels. In the following section, I will briefly review the composition of the individual layers of the form and the motion pathway (see also Table 5.1). The exact details of how the properties of the neurons in the individual layers are modeled are described in Giese and Poggio (2002).

Form Pathway

Analog to the simple cells in primary visual cortex, the neurons of the first hierarchy level are modeled as Gabor filters. These local orientation detectors exist in two different spatial scales and eight orientations.

The neurons at the next hierarchy level show a certain degree of position and scale invariance, which is achieved by pooling the responses of neurons with the same orientation preference but different receptive field positions and spatial scales. The pooling is accomplished by a maximum-like operation to guarantee a high degree of feature selectivity and invariance (e.g. Riesenhuber and Poggio, 2002). Such properties have been reported for *complex cells* in V1 and neurons in areas V2 and V4.

In the next hierarchy level, single neurons represent the shape of whole body configurations. An example could be the different body postures during a walking cycle (see also Figure 5.3a). Since these neurons are selective for specific snapshots of the movement cycle, they are termed *'snapshot neurons'*. Such snapshot neurons could be analogous to the view-tuned neurons discovered in the infero-temporal cortex of the monkey (Logothetis et al., 1995). The snapshot neurons are modeled by radial basis functions that are trained with the output vectors of the complex cells in the previous hierarchy level. Additionally, the snapshot neurons are embedded in a recurrent neural network that assures that their responses are sequence selective. That is, all snapshot neurons possess lateral connections to all the other snapshot neurons in this layer. These connections are however asymmetric in a way, that they have excitatory connections to neurons that code subsequent body configurations and inhibitory connections to neurons coding preceding body configurations.

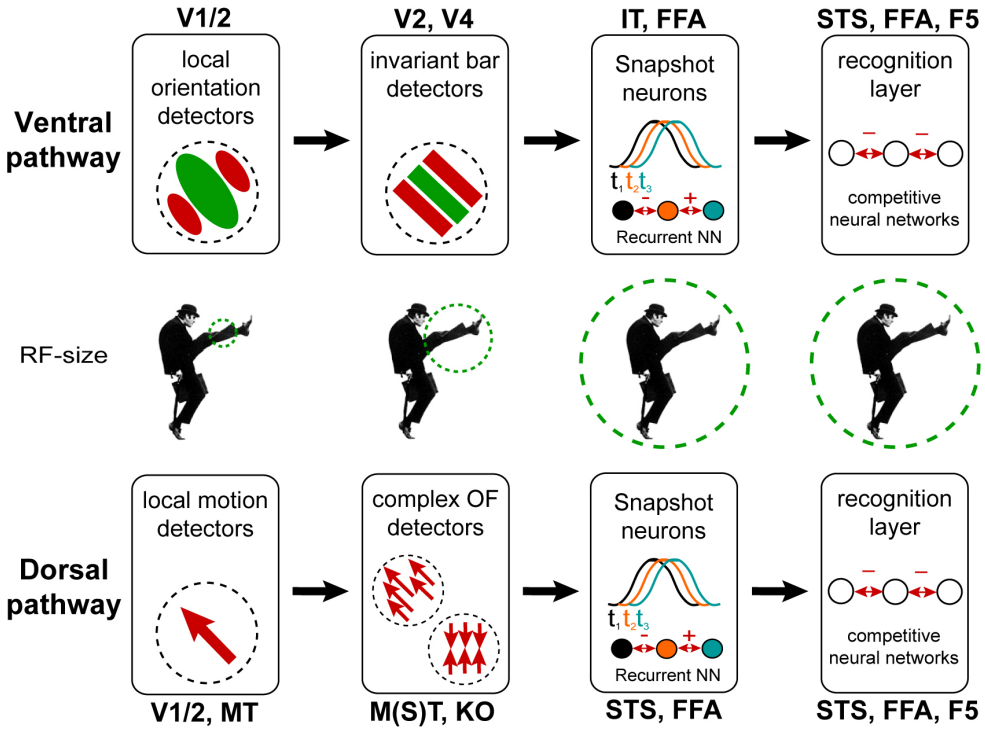

Figure 5.1: Schematic representation of the neural model for biological motion recognition modified from Giese and Poggio (2003). The model consists of two separate pathways for the analysis of from and optic flow information. Both pathways are organized in a hierarchical fashion starting with local analysis (small receptive fields (RF)) at lower levels of the hierarchy.

The highest hierarchy level of the model contains neurons that temporally smooth and summate the activity of the individual snapshot neurons for a certain movement. In the end, each of these *'motion pattern neurons'* would correspond to a whole movement sequence like walking, limping or running. However, this is an oversimplification of the physiological processes it is more likely that biological movements are encoded on the basis of population responses in the superior temporal sulcus.

Motion Pathway

Like the form pathway, also the motion pathway consists of four hierarchy levels that analyze optic flow information with increasing complexity. The first level consists of local motion energy detectors that correspond to motion direction selective neurons in primary visual cortex and component motion selective detectors in MT, with their corresponding receptive field size.

Model neurons	**Area**	**Number of neurons**	**Receptive field size**
	Form pathway		
Simple cells	V1, V2	1010	0.6° , 1.2°
Complex cells	(V1, V2) V4	128	4°
(View-tuned) snapshot neurons	IT, EBA, STS, FFA	63-840	$> 8°$
Motion pattern neurons	STS, FAA, F5	3-40	$> 8°$
	Motion pathway		
Local motion detectors	V1, V2, MT	1147	0.9°
Local OF pattern detectors	MT, MST *transl.* MST KO *expan./contr.*	72 2 x 50	3.5°
Complex OF pattern detectors	STS, FFA	63-840	$> 8°$
Motion pattern neurons	STS, FAA, F5	3-40	$> 8°$

Table 5.1: The table represents some of the details of the model. OF = optic flow, transl. = translation, expan. = expansion, contr. = contraction.

The second level of the motion pathway consist of two types of local optic flow detectors. The first type is selective for translation flow at four different orientations and two different speeds. Their responses are computed by summation of the activity of local motion detectors coding similar motion directions and speeds. The second type is selective for opponent motion. The response of these neurons is computed by multiplying the responses of adjacent local motion detectors with opposite direction preference. Whereas the first type of optic flow detectors mimics the properties of neurons in area MT, the second type models the features of neural populations observed in area MST and KO.

The neurons in the next layer of the model are motion pattern neurons that encode, like the snapshot neurons of the form pathway, the complex optic flow patterns that occur during the movement. One example could be the opponent motion information coming from the arms and legs that move in opposite directions during the walking cycle. Similar to the form pathway, the motion pattern neurons are modeled using radial basis functions that receive their inputs form the previous layer of the model. In the same way, they also have asymmetric lateral connections to achieve sequence selectivity.

The final layer of the motion pathway consists of the same motion pattern neurons already described in the form pathway. If form and optic flow information has to be analyzed separately, it is possible to introduce individual sets of motion pattern neurons for the form and the motion pathway.

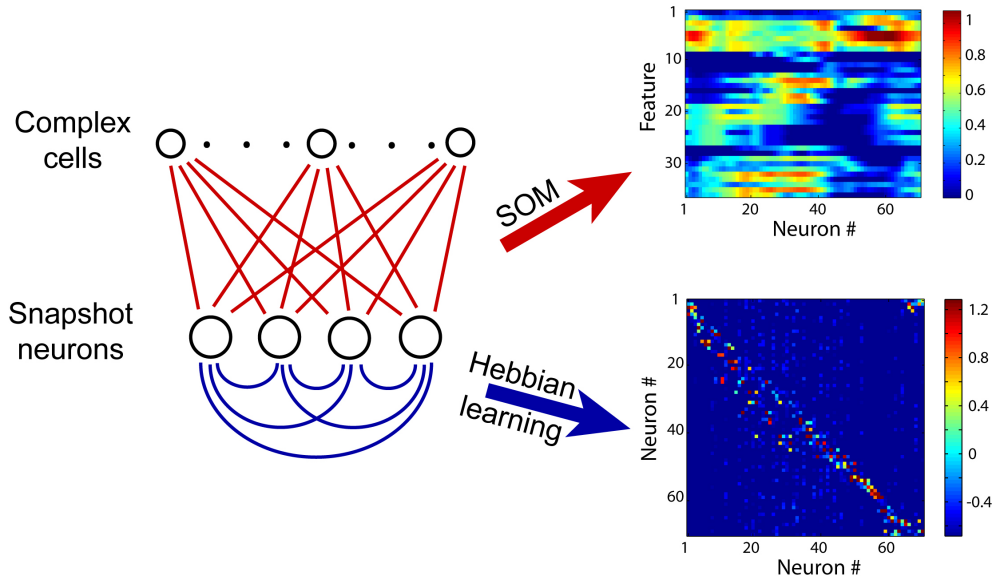

Figure 5.2: Illustration of the different learning steps implemented in the model. The neuronal representation (”snapshot neurons”) for the features of the individual frames was learned using a Kohonen self organizing map (SOM) (see equation 5.11) Every feature vector corresponds to the activity of 35 complex cells, which showed the highest variance over the course of the animation. The correct asymmetric lateral connectivity between the individual snapshot neurons was learned using a time dependent hebbian learning rule (see equation 5.4)

Taken together, the properties of the neurons in the individual layers of the model are designed to mimic the properties of real neurons in the visual processing streams. One of the goals of the modeling part of my thesis was to introduce biologically plausible learning mechanisms into the model, so that the feedforward connections of the *complex cells* (layer 2) of the model to the *snapshot neurons* in layer 3 can be learned automatically. Additionally, we wanted to explore different mechanisms, by which the correct asymmetric lateral connectivity between the individual snapshot neurons can be learned (feedback connectivity)(Figure 5.2). For the presented modeling work, we concentrated on the form pathway, but the mechanisms discussed will be equally applicable to the motion pathway.

5.2 Learning of the Feedback Connectivity

The recognition of biological movements and actions is an important visual function. The perception of body as well as facial movements can be explained by the recognition of temporal sequences of form and optic flow patterns (God-

dard, 1992; Giese and Poggio, 2003). A neural encoding of such temporal sequences can be realized in a physiologically plausible way by recurrent neural networks with asymmetric lateral connections (Giese and Poggio, 2003; Mineiro and Zipser, 1998; Xie and Giese, 2002). One possible mechanism for the learning of such connections is time-dependent hebbian plasticity. Recent experiments show that spikes initiated at the axon hillock can back-propagate into the dendrites, due to their active properties (Stuart and Sakmann, 1994). These backpropagating signals provide information about previous activity states of the neuron. It also has been shown that the occurrence of LTP vs. LTD depends critically on the timing between pre- and postsynaptic spikes (Markram et al., 1997; Bi and Poo, 1998). Time-dependent synaptic plasticity is suitable for the realization of temporal sequence learning. In particular, it has been shown that spike-timing-dependent plasticity in cortical neurons can be related to the temporal difference learning rule (Rao and Sejnowski, 2001), where an activated synapse is strengthened or weakened depending on the sign of the difference between two temporally subsequent output signals (Sutton, 1988; Montague and Sejnowski, 1994)[1].

After a short description of the network *(section 5.2)*, we propose different Hebbian learning rules that can be used to establish the required form of lateral connectivity *(section 5.3)*. After a mathematical analysis of the stability properties of these learning rules *(section 5.4)*, we will present a number of simulations that compare the learning rules with respect to their efficiency and robustness *(section 5.5)*.

5.2.1 Recurrent Neural Network Model

The recurrent neural network that we used for the encoding of temporal sequences of body shapes or optic flow patterns has been originally proposed by Amari (1972). Sequence selectivity arises in this Hopfield-like network, if a suitable asymmetric form is chosen for the lateral connections. Signifying by $\mathbf{u}(t)$, the activity vector of all N_s neurons in the network, the dynamics is given by the differential equation:

$$\tau \left(\frac{d\mathbf{u}(t)}{dt} \right) + \mathbf{u}(t) = \tilde{\mathbf{W}}(t) f(\mathbf{u}(t)) + \mathbf{s}(t) - h \tag{5.1}$$

[1]The Learning of the feedback connectivity has been published as: Jastorff J, Giese MA (2004): Time-dependent hebbian rules for the learning of templates for visual motion recognition. In Ilg U, Bülthoff HH, Mallot H (eds): *Dynamic Perception*; Infix, Berlin, 151-156.

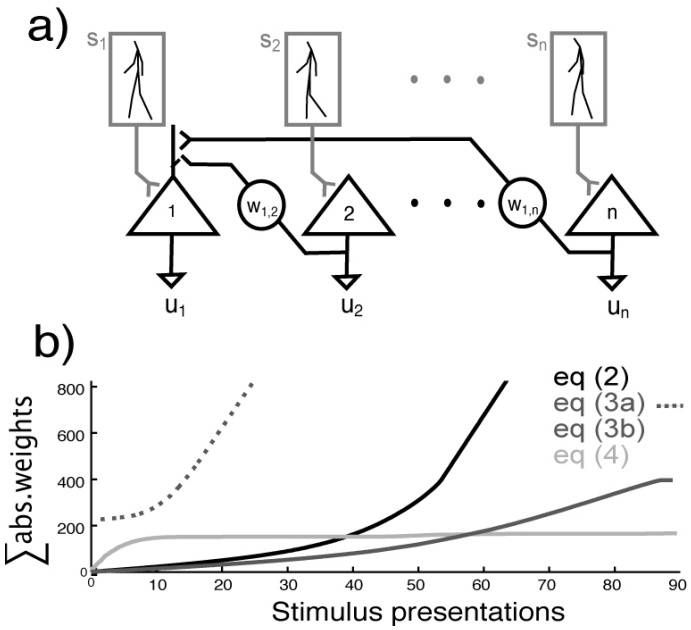

Figure 5.3: (a) Illustration of the laterally connected network with units encoding body shapes arising during a walking cycle. (b) Sum of the absolute values of the elements of the weight matrix as a function of time for the learning rules 5.2, 5.3 (a, weights having positive and negative values, and b, weights being restricted to non-negative values), and for learning rule 5.4).

The matrix $\tilde{\mathbf{W}}(t) = -w_I\mathbf{M} + \mathbf{W}^T(t)$ defines the synaptic strength of the lateral connections. It consists of a constant inhibitory part that ensures a sufficient level of lateral inhibition, and a second term that changes during learning. ($\mathbf{M}$ is a matrix with only one elements, and w_I is a positive constant.) We used a step activation function with $f(z) = 1$ for $z > 0$ and $f(z) = 0$ otherwise. The time-dependent feed-forward inputs are given by the signal vector $\mathbf{s}(t)$. The time constant τ was chosen to be 200ms and the positive parameter h defines the resting activity level.

For the simulations, we derived the input signal vector from the complete model for biological motion recognition (Giese and Poggio, 2003) showing a walker as stimulus. The input signal can be approximated by a localized positive activity pulse that moves along the neural network over time. In our implementation the recurrent network contained 20 neurons (Figure 5.3). After the training of the lateral connections the network was tested with respect to its stability, sequence selectivity, and tuning with respect to stimulus speed and direction.

5.2.2 Learning Rules:

The most simple form of a time-dependent Hebbian rule can be written in matrix form by the differential equation:

$$\tau_w \frac{d\mathbf{W}}{dt} = \mathbf{s}(t) f\left(\mathbf{u}^T(t-\Delta t)\right) \tag{5.2}$$

Consistent with the classical Hebbian postulate, the synaptic weight matrix $\mathbf{W}$ changes with the product of the presynaptic activity vector $\mathbf{s}(t)$ and the postsynaptic activity vector $f(\mathbf{u}(t))$, where the postsynaptic activity level enters the equation with the positive time delay Δt. It is crucial that the time constant τ_w of the learning rule is much larger than the time constant τ of the network dynamics and the duration of the encoded actions.

It is well-known that the simple Hebbian rule is unstable. This implies that the elements of the weight matrix tend to become unbounded after several stimulus presentations (Dayan and Abbott, 2001). This can be prevented by introduction of competition between the different synapses by weight normalization. One possibility to realize such competition is to modify learning rule 5.2 by subtraction of a term that depends on the sum of the input signal vector $\mathbf{s}(t)$. Let $\mathbf{m}$ be an N_s-dimensional vector whose components are all one, then a modified learning rule is given by:

$$\tau_w \frac{d\mathbf{W}}{dt} = \left[\mathbf{s}(t) - \frac{\mathbf{m}^T\mathbf{s}(t)\mathbf{m}}{N_s}\right] f\left(\mathbf{u}^T(t-\Delta t)\right) \tag{5.3}$$

It is shown in section 5.2.3 that this form of normalization imposes a rigid constraint on the sum of the weights in each column of the matrix $\mathbf{W}$. This implies that increases of some weights lead to decreases of others, stabilizing the behavior of the weights during learning. We tested another version of this learning rule where we restricted the weights to non-negative values.

Another way to stabilize the learning process is to impose a constraint on the sum of synaptic weights that is supported by each neuron. A limitation of the maximum number of synapses that is supported by individual neurons seems to be suggested by electrophysiological data (Miller, 1996). A learning rule that implements this constraint can be written in matrix form:

$$\tau_w \frac{d\mathbf{W}}{dt} = \left[\mathbf{s}(t)\mathbf{m}^T - \mathbf{m}\mathbf{m}^T\mathbf{W}(t) + \alpha\mathbf{M}\right]\mathbf{D}_\mathbf{u} \tag{5.4}$$

where $\mathbf{D}_\mathbf{u}$ is a diagonal matrix with the elements $\mathbf{D}_{\mathbf{u},kk} = f(u_k(t-\Delta t))$. The positive constant α determines the sum of the weights in each column of the weight matrix $\mathbf{W}$ (see section 5.2.3). In the final implementation the weights were additionally constrained to be non-negative.

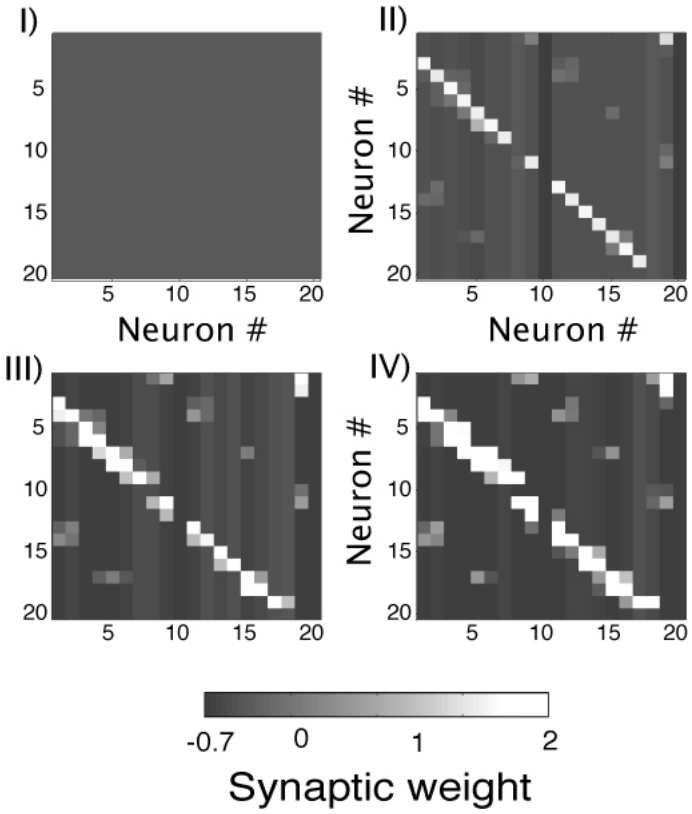

Figure 5.4: Plot of the weight matrix learned with rule 5.4. (I) Initial weight matrix with all elements set to zero. (II), (III) and (IV) show the weight matrix after 15, 30 and 45 stimulus repetitions.

5.2.3 Stability Analysis

For the chosen simple learning rules the effect of the normalization term can be analyzed mathematically if the positivity constraint for the weights is dropped. We assume in the following an idealized input signal vector $\mathbf{s}(t)$ that is given by a positively activated region with constant shape that moves along the neural network .

Stability Analysis for Learning Rule (5.3):

By multiplying both sides of equation 5.3 with $\mathbf{m}$ we can derive a differential equation for the vector $\nu^T(t) = \mathbf{m}^T\mathbf{W}(t)$ that defines the sums of the weights in the columns of the weight matrix. The sum $\sigma = \mathbf{m}^T\mathbf{s}(t)$ does not depend on time because of the constant shape of the moving input peak. The resulting differential equation

$$\tau_w \frac{d\mathbf{m}^T\mathbf{W}}{dt} = \tau_w \dot{\mathbf{v}}^T(t) = \left[\sigma - \frac{\mathbf{N}_s}{\mathbf{N}_s}\sigma\right] f\left(\mathbf{u}^T(t-\Delta t)\right) \equiv \mathbf{0}^T \qquad (5.5)$$

implies $\mathbf{v}^T = \mathbf{m}^T\mathbf{W} = const.$ This implies that the column vectors of the weight matrix remain within hyperplanes with normal vector $\mathbf{m}$. Still the weights can diverge tangentially to these hyperplanes. If the weights are additionally constrained to positive values they remain bounded within the interval $0 \leq W_{mn} \leq \max(\mathbf{m}^T\mathbf{W}(0))$.

Stability Analysis of Learning Rule (5.4):

Multiplication of equation 5.4 with $\mathbf{m}$ yields:

$$\tau_w \frac{d\mathbf{m}^T\mathbf{W}}{dt} = \tau_w \dot{\mathbf{v}}^T(t) = \left[(\sigma + \alpha N_s)\mathbf{m}^T - N_s \mathbf{v}^T(t)\right] \mathbf{D_u} \tag{5.6}$$

Let T signify the cycle time of the walking stimulus. The temporal dynamics of the learning process is much slower than the activation dynamics of the network ($\tau_w \gg T, \tau$), so that we can average the dynamical equation 5.6 over time (Guckenheimer and Holmes, 1983). If the activation of the network is close to a form-stable solution that moves with the stimulus signal along the network we obtain with the time-averaged signals

$$\mathbf{y}(t) = \frac{1}{T}\int_t^{t+T} \mathbf{v}(t)dt \tag{5.7}$$

$$\frac{1}{T}\int_t^{t+T} f\left(\mathbf{u}(t - \Delta t)\right) dt \approx f_0 \mathbf{m} \tag{5.8}$$

where the constant f_0 fulfills $0 < f_0 < 1$. We obtain the following approximative dynamical equation for the averaged quantities:

$$\tau_w \frac{d\mathbf{y}}{dt} = \tau_w \dot{\mathbf{y}}(t) = \left[(\sigma + \alpha N_s)\mathbf{m}^T - N_s \mathbf{y}(t)\right] f_0 \tag{5.9}$$

The last dynamics is stable because N_s and f_0 are always positive. The stationary solution can be easily computed to be

$$\mathbf{y}^* = \mathbf{m}\frac{(\sigma + \alpha N_s)}{N_s} \tag{5.10}$$

The stable sum of weights of the columns of $\mathbf{W}$ depends on the constant α, and on the strength of the input signal $\mathbf{s}(t)$. Again the weight vector can change orthogonal to the vector $\mathbf{m}$ without affecting the previous dynamical equations. If the weights are restricted to positive values the learning rules produce stable results.

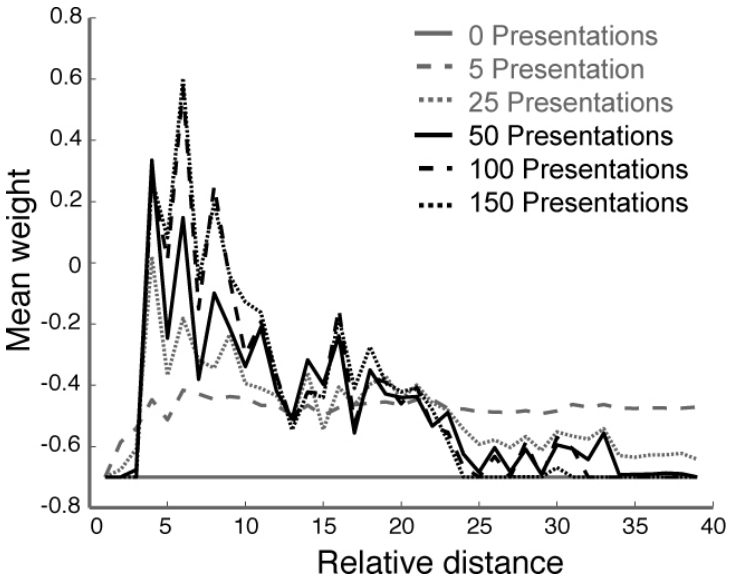

Figure 5.5:
Development of the interaction kernel over time for learning rule 5.4. The weight kernel develops quickly with only marginal changes after 25 stimulus repetitions.

5.2.4 Simulation Results

The sum of the absolute values of the weight matrix learned with different learning rules is shown in Figure 5.4. As predicted, the simple Hebbian learning rule 5.2 leads to unbounded growth of the weight vector. No direction selectivity arises in the network because all units become activated after some time. Also learning rule 5.3 is suboptimal. If the weights were not constrained to positive values, consistent with the mathematical analysis, the sum of the weights remains constant, but the absolute values of the weights grow and become unbounded, since positive and negative weights compensate each other in the sum (curve 3a). Only if the weights were restricted to non-negative values, the learning rule converges to a stable weight kernel, after a large number of gait cycles (curve 3b). For learning rule 5.4 we found a much faster stabilization of the learned weight distribution. After only eight gait cycles the sum of the weights reaches its plateau.

Figure 5.5 presents the learned weight distributions using rule 5.4 with additional positivity constraint for the W_{ij}. The learned weight distribution is periodic with respect to the neuron number, reflecting the periodic nature of the stimulus. The figure shows the learned lateral weight kernel as function of the index difference between the connected neurons. This kernel can be obtained by averaging the weights that connect neurons with the same relative distance in the network over all neurons.

After 25 stimulus presentations the form of the kernel remains quite stable. This fast learning is compatible with the psychophysical data obtained in Chap-

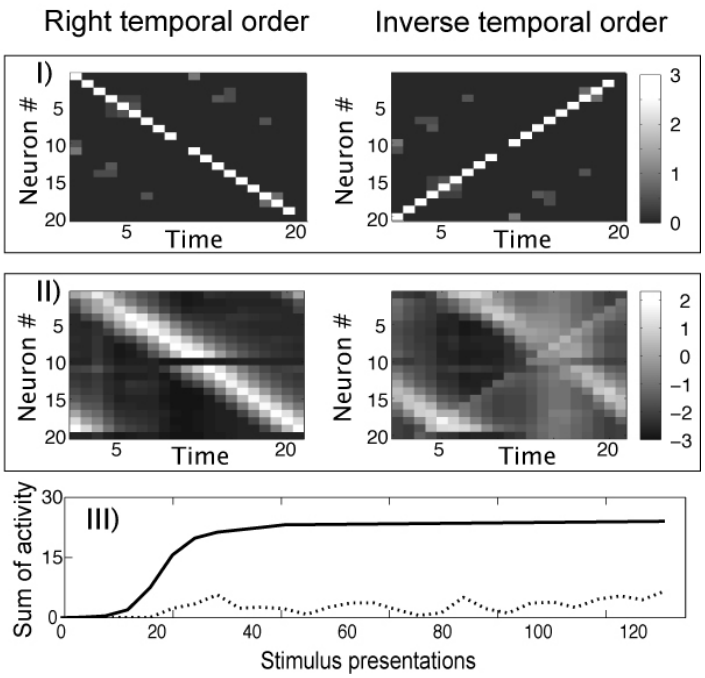

Figure 5.6: Development of direction selectivity for learning rule 5.4: (I) shows the stimulus signal $\mathbf{s}(t)$ for the stimulus in normal and inverse temporal order. (II) illustrates the output activity after 25 stimulus presentations, and (III) the sum of the activity of all neurons for the correct (solid line) and the inverse temporal order (dashed line).

ter 3, demonstrating substantial improvements during the learning of novel biological motion patterns after 20 stimulus presentations.

Figure 5.6 illustrates that during the learning process substantial sequence selectivity arises. After 25 presentations we find strong dependence on temporal order. If the stimulus (panel I) is presented in reverse order much weaker activity arises in the network than for the correct temporal order of the stimulus frames (panel II). The proposed learning rule is thus suitable for establishing sequence selectivity with a small number of training trials. Panel (III) shows that the activity level for correct temporal order substantially increases already after 10 stimulus presentations and reaches a plateau after about 40 presentations. The activity for the reverse temporal order stimulus remains at low levels throughout the whole training.

5.2.5 Conclusion

We have discussed different time-dependent Hebbian learning rules to establish sequence selectivity in a recurrent neural network for biological motion recognition. Most efficient were learning rules that enforce a constant sum of synaptic weights that are supported by each neuron, combined with a lower bound for the connection strength. Both assumptions seem physiologically plausible, because

usually excitatory synapses cannot change into inhibitory ones, and because neurons can only supply transmitter for a limited number of synapses (Miller, 1996). Given the high number of recurrent connections in the visual cortex, it seems plausible that lateral connections play an important role, potentially also for the realization of sequence selectivity.

5.3 Learning of Feedforward Connectivity

In many areas of the cortex, groups of neighboring neurons are engaged in the processing of similar features. In the visual cortex, these neuronal populations are also called microcolumns, because they are usually organized along a vertical cylindrical volume of cortex. Such a Population could serve for example, to analyze a particular stimulus feature like the orientation of an edge present in the stimulus. At the next higher level of organization, microcolumns engaged in the processing of similar features are arranged in specialized areas. These areas could be involved for example, in the analysis of edge orientation, colors or optic flow fields.

In terms of theoretical modeling, the described organization of the cortex can be interpreted as a continuous, two-dimensional feature map. In the literature, multiple approaches have described, how such feature maps can be learned without supervision. The one I am going to apply to map the responses of the different complex cells onto individual snapshot neurons was originally developed by Kohonen (1982). The benefit of this algorithm is that the similarity in the signals will be automatically converted into proximity of neurons in the network so that neurons representing similar features will in this way be grouped together, much like in the cortex. Kohonen maps, also called self organizing feature maps (SOM) have already been applied for example, for the modeling of orientation selective neurons in primary visual cortex (e.g. Obermayer et al., 1990).

To visualize the learning process we will for now concentrate only on the form pathway, since the mechanisms with which the connectivity for the motion pathway can be learned are identical. Layer two of the model consists of 128 complex cells (see Table 5.1). For each frame of the animation, these cells are more or less activated, depending on whether their preferred orientation was present in their receptive field. Out of these 128 neurons, the 35 with the highest variance over all frames of the animation are chosen as it is assumed that they provide the most useful information to identify the movement. Because the whole movement was represented by a movie containing 72 frames, we chose a chain of 72 neurons for the Kohonen map. After the training process, these 72

neurons would correspond to the different snapshot neurons of the layer three of the model. To map the 35-dimensional feature vector for each frame of the animation onto a certain neuron we applied the Kohonen algorithm.

The initial values for the features were chosen randomly for each snapshot neuron. The training procedure consists of, finding the neuron with its feature vector having the minimum Euclidean distance compared to the feature vector of the input frame. This neuron would correspond to the one that is best tuned to represent the input frame. The next step is called the learning step. Here, the feature vector of the chosen neuron will be updated in a way that after learning, this neuron is tuned sharper to the features of the input frame. Mathematically, this is done by:

$$\mathbf{w}_r^{new} = \mathbf{w}_r^{old} + \epsilon h_{rr'} \left(\mathbf{v} - \mathbf{w}_r^{old} \right) \tag{5.11}$$

where the vector $\mathbf{w}_r$ refers to the feature vector of the winning neuron and the vector $\mathbf{v}$ to the responses of the different complex cells of the input frame.

There are two additional parameters influencing the learning process. ϵ determines the size of each learning step ($0 < \epsilon < 1$). We chose ϵ to be a function $\epsilon(t)$, which decreases exponentially with the number t of the learning steps from large initial values to small final values. This results in rapid learning at the beginning of the training process which allows to coarsely correct the feature vectors of the different snapshot neurons. For large ϵ, the fluctuation of the map caused by each learning step is very large. Therefore, ϵ has to decrease over time to finally stabilize the map in an equilibrium state.

The second parameter h is called the neighboring function. The neighboring function refers to the fact that not only the feature vector with the smallest euclidean distance gets updated but also the neighboring neurons are affected by an amount that decays with their distance from the best tuned neuron. This distance dependence is modeled by the following Gaussian function:

$$h_{rr'} = e^{\left(\frac{-(r-r')^2}{2\sigma_E^2} \right)} \tag{5.12}$$

where r' is the position of the neuron with the smallest euclidean distance and r is the the position of the neuron that gets updated. The feature vector of the best tuned neuron will hence be updated the most since $r' = r$, in that case.

The radius σ_E of the Gaussian determines the distance at which the input stimulus frame causes corrections to the map ($0 < \sigma < 1$). Similar to the learning parameter ϵ, also σ is chosen as a function $\sigma(t)$, starting at a large initial

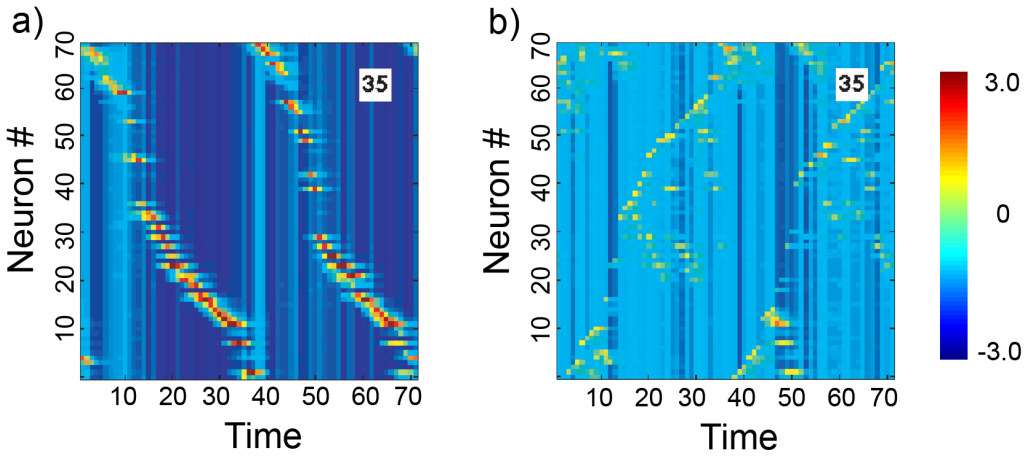

Figure 5.7: Activity of the individual snapshot neurons after training the model for 35 repetitions of the same stimulus. Panel a) shows the activity if the stimulus is presented in the correct temporal order (the stimulus, the model has been trained with) and panel b) represents the activity of the same neurons if the temporal order of the training stimulus is reversed.

value $\sigma(0)$ and decreasing exponentially with the number of learning steps, towards a small final value. This can be interpreted as gradually increasing the *"selectivity"* of the individual neurons in the course of the learning process.

Because the feature vectors vary smoothly over time, this learning process will lead to a neuronal chain of snapshot neurons, in which neighboring frames of the animation will be represented by neighboring snapshot neurons. To prevent that a single snapshot neuron is continuously selected, we introduced an additional adaptation term, which increases the Euclidean distance if the same neuron wins continuously and thereby reduces the likelihood that it gets selected for the next input frame.

Figure 5.7 displays the activity in the network that combines the unsupervised learning of the feedforward and the feedback connections. The model was trained with 35 presentations of a human person walking towards the left, viewed from the side. If the stimulus is presented in the correct temporal order as depicted in panel a), a wave of activity is traveling across the network representing the activity of the individual snapshot neurons for the different frames of the animation. Panel b) shows the activity of the snapshot neurons when the temporal order of the stimulus is reversed. These results indicate that we were successfully implementing the simultaneous learning of a neural representation for the individual frames occurring during the gait cycle (snapshot neuron) as well as the correct lateral connectivity between them to guarantee sequence selectivity.

In the next section, we will further extend the model to simulate BOLD activity changes observed during our functional imaging experiments (see Chapter 4).

5.4 Modeling of BOLD Activity Changes

One possible way of testing the different hypotheses about the underlying plasticity mechanisms of the reported BOLD activity changes in Chapter 4, is by applying theoretical models. As described previously, the model we are employing in this chapter consists of two parallel streams that are specialized for the processing of form and optic flow information. The advantage of using this model is that the different visual areas we were analyzing in Chapter 4, are also represented in the different layers of the model. This would allow us to directly test possible explanations of how learning influences the processing of complex movements, by changing the properties of the model neurons at the different layers or alter the connections between the layers. However, to identify, how changes in the internal structure of the model relate to BOLD signal activity, we have to extend the model in a way that the activity of the model neurons gets translated into BOLD responses. Hence later on it will be possible to compare the simulated BOLD responses with the measured BOLD responses in Chapter 4 and to see how changes at the level of the model neurons affect the simulated BOLD activity.

As an equivalent of real neural activity, we used the activity of the different motion pattern neurons in the highest hierarchy layer of the model (see Figure 5.1). Like in the previous sections, we concentrated on the activity of the form pathway, but the reported mechanisms are equally applicable to the motion pathway. As a starting point, we trained the model with five different movement stimuli (walking, runnning, marching, jumping and boxing), which resulted in the formation of 5 different motion pattern neurons, each of which specifically tuned to the features of one of the stimuli.

To match the functional imaging experiments as closely as possible, we stimulated the model after training not only with a single stimulus but with a whole sequence of stimuli. To this end, we created a continuous series of trials belonging to three different conditions. The first condition, which corresponded to the *"identical" condition* in Chapter 4, contained the same movement stimulus presented twice. The second condition was analogous to the *"dissimilar" condition* and consisted of two different movement stimuli presented in succession. The third condition contained no stimulus and corresponded to the *"fixation" condition*.

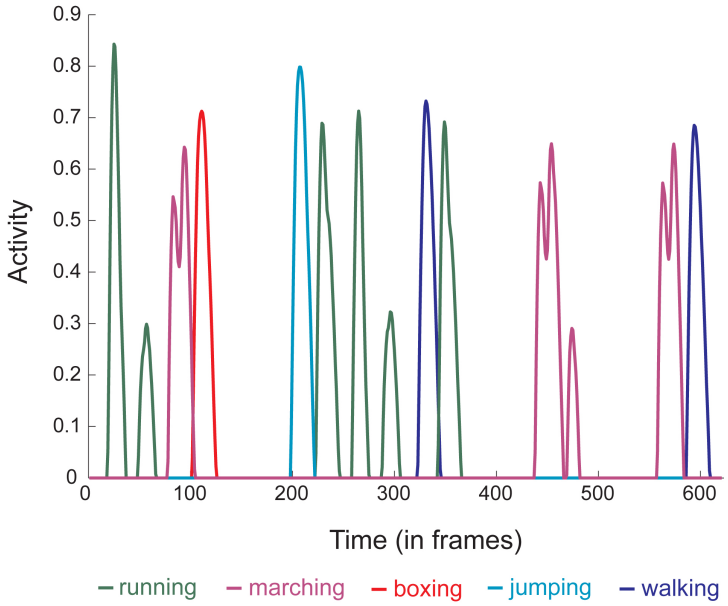

Figure 5.8: Activity of the different motion pattern neurons for the presentation of a sequence of movement stimuli. We observed reduced activity (adaptation), if the same movement was presented within a very short time interval.

The timing of the trials was identical to the timing used in Chapter 4, where a single trial lasted for four seconds. Each movement stimulus used for the simulations consisted of 21 frames. To match the timing of the fMRI experiments, every trial started with the presentation of the first movement for 21 frames, followed by a blank for 2 frames. Afterwards the second movement was presented for 21 frames and the trial ended with another 21 frames containing no stimulus. Taken together, every trial of the simulation contained 65 frames, which corresponds to four seconds of the real fMRI experiment.

The whole simulation consisted of 60 trials, with 20 trials per condition. The order of the trials was counterbalanced in exactly the same way as in the functional imaging experiments (e.g. Kourtzi and Kanwisher, 2000b).

One possible explanation for the reduced BOLD activity we observed for the presentation of two identical stimuli compared to the presentation of two different stimuli during our functional imaging experiments could be a repetition suppression mechanism at the neural level (see section 2.2.2 for discussion of different adaptation mechanisms). To model this adaptation mechanism, we had to extend the model by the implementation of an adaptation term, which reduces the activity of the motion pattern neuron gradually, if the same neuron is active over a longer period of time. Before the implementation of the adaptation term, the dynamics of the motion pattern neuron was described by:

$$\tau_u \dot{u}(t) + u(t) = \sum_n f(z_n(t)) \tag{5.13}$$

where u represents the activity of the motion pattern neuron, f is a sigmoidal nonlinear threshold and z_n describes the activity of the individual snapshot neurons that are involved in the representation of this movement. τ_u is the time constant for the temporal smoothing of the motion pattern neuron (150ms).

The adaptation term is modeled analogous to a leaky integrator given by the following equation:

$$\tau_a \dot{a}(t) + a(t) = u(t) \tag{5.14}$$

with $\tau_a >> \tau_u$ (1 second) and $a(t) \geq 0$, because the adaptation term is restricted to contain positive values.

Consequently, the simulated activity of the motion pattern neurons was derived by combining the two equations in the following way:

$$\tau_u \dot{u}(t) + u(t) = \sum_n f(z_n(t)) - \gamma a(t) \tag{5.15}$$

where γ was set to 0.015 to scale the adaptation term accordingly.

The parameter τ_a was adjusted in a way that the activity of the motion pattern neuron is significantly reduced if the same movement is presented twice in a single trial, but recovers from adaptation for subsequent trials. The effect of the adaptation mechanism on the activity of the motion pattern neurons is displayed in Figure 5.8. For trials in which the same stimulus is presented twice, the same motion pattern neuron is active over the whole trial. This results in an increase of the adaptation term, which in turn reduces the activity of the neuron. It can clearly be seen that if the same neuron is activated within a very short interval, the activity of the second activation is reduced compared to the first one. If however, two different stimuli are presented during one trial, different motion pattern neurons get activated, which results in no adaptation.

To simulate BOLD activity changes, we convolved the train of activity of the motion pattern neurons with a hemodynamic response function, which was obtained in a real fMRI experiment performed with exactly the same conditions [2]. Because the fMRI experiment was scanned with a TR of two seconds, the time course of the hemodynamic response function was sampled only every two seconds. To obtain the values for a continuous BOLD response that are needed for the convolution, we fitted the percent signal change within the superior temporal sulcus for the *"different" condition* with a hemodynamic response model based on the difference of two gamma functions. The difference between two

[2]The experiment served as a pilot experiment for the functional imaging experiments in Chapter 4 and is not reported in this thesis.

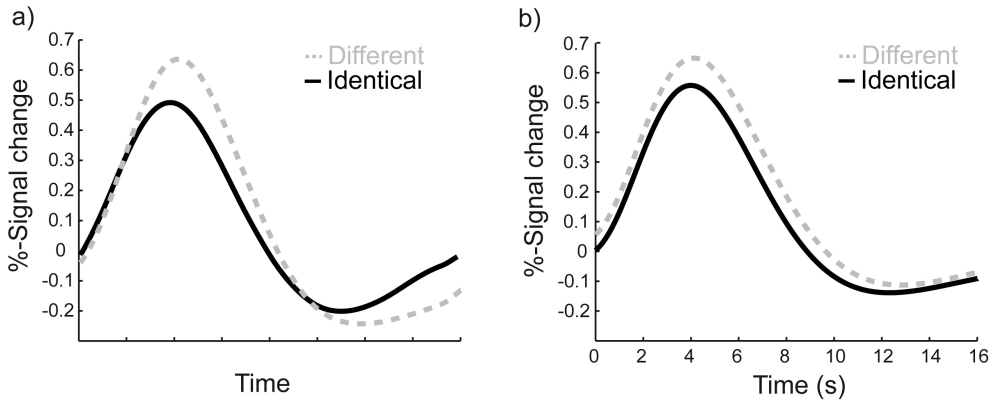

Figure 5.9: Simulated and real BOLD activity changes. Panel a) shows the simulated BOLD activity changes for the presentation of either two identical movement stimuli or two different stimuli per trial. Panel b) shows the measured BOLD activity changes in the superior temporal sulcus for the presentation of either twice the same point-light stimulus or two different point-light stimuli.

gamma functions predicts a biphasic response when the second function, $h_2(t)$, has a slower time course than the first. This smooth parametric function provides a continuous description of the BOLD response and provides an estimate of parameters such as maximum height and delay (Boynton and Finney, 2003).

$$h(t) = h_1(t) - h_2(t) + k \tag{5.16}$$

where $h_i(t)$ is defined as

$$h_i(t) = \frac{((t-\delta_i)/\tau_i)^{(n_i-1)} e^{-(t-\delta_i)/\tau_i}}{\tau_i (n_i - 1)!} \tag{5.17}$$

The phase delay parameters n_1 and n_2 were set to 5.5 and 11.5, respectively, and the remaining parameters were allowed to vary freely. The baseline parameter k was set to 0. The best fit of the time course was obtained for δ_1 = 2.46, $\tau_1 = 0.66$, δ_2 = 6.27 and $\tau_2 = 0.18$.

Following the convolution, we performed exactly the same data analysis that was used in Chapter 4 to extract the time courses for the individual conditions. That is, we extracted the simulated hemodynamic response function separately for each trial and converted it to percent signal change by subtracting the averaged activity of the preceding two trials. Afterwards, we averaged the percent signal change for all trials belonging to the same condition. That way, the simulations would also test whether the assumption of linearity holds true, when the order of the conditions is precisely counterbalanced. This assumption is the foundation of rapid event-related fMRI designs.

Figure 5.9 displays our simulated BOLD response obtained from the model (panel a) and the BOLD activity changes obtained in the real functional imaging experiment (panel b). Even though they do not match exactly, we believe that our simulation provides a good approximation of the real fMRI data.

Based on this result, we can conclude that an adaptation mechanisms that is based on repetition suppression provides a possible explanation for the effect that the fMRI BOLD signal decreases when the same stimulus is presented repeatedly. Furthermore, the assumption of linearity of the BOLD signal, which is the foundation of rapid event-related fMRI designs, seems to hold true, because the very same processing steps applied for the analysis of real and simulated fMRI data, lead to very similar hemodynamic response functions. In the future, the extended model could be used to test different predictions about the underlying activity changes observed in our functional imaging experiments.

Chapter 6

General Discussion

6.1 Summary

Experimental evidence in the field of object recognition suggests that learning plays a key role in our ability to identify novel objects. Three dimensional objects seem to be represented in the cortex on the basis of learned two dimensional views of the objects (e.g. Bülthoff and Edelman, 1992; Logothetis et al., 1995; Logothetis and Sheinberg, 1996; Poggio and Edelman, 1990; Tarr and Bülthoff, 1998). In a theoretical model developed recently by Giese and Poggio (2003), the concept of learning has been transfered to the of domain of movement recognition. The simulation results of this model, which represents complex movements on the basis of learned prototypical patterns, show a high degree of similarity with the experimental findings in the field of movement recognition. The aim of this thesis was to investigate learning processes and their underlying neural correlates in the recognition of complex movements.

In a series of psychophysical experiments (Chapter 3), we have shown that humans are able to learn to discriminate between very similar exemplars of complex movements presented as point-light stimuli very quickly after less than 20 stimulus repetitions. This learning process applies not only for movements that are similar to natural human movements, but is equally efficient for completely novel artificial skeletons. These artificial stimuli were composed of nine segments and shared certain low level properties with human movements (e.g. smoothness of the trajectories, frequency and amplitude of the movement) but differed with respect to their kinematics and biological relevance. Like for normal biological motion recognition, the learned representation was orientation dependent for both stimulus groups (e.g Bertenthal and Pinto, 1994; Pavlova and Sokolov,

2000; Shiffrar et al., 1997; Sumi, 1984), which indicates that the same neural mechanisms might be involved in their encoding. However, motion stimuli that consisted of the same local motion trajectories as the human like movements, but with perturbed spatial relations between the dots so that the stimulus appeared no longer articulated, could not be learned equally fast.

These results indicate that learning of complex movements is strongly facilitated by the binding or grouping of individual dot or joint movements on the level of limbs and possibly even at the entire body level. The organization of the individual dot movements into a global percept could be achieved via a combination of bottom-up and top-down processes, where mid level feature detectors could signal the relationship between neighboring dots and top down processes help to organize these configurations at the level of a global shape. This information could finally converge in higher level motion and form areas.

The involvement of top-down processing in biological motion recognition is supported by several psychophysical investigations (e.g Cavanagh et al., 2001; Thornton et al., 2002). Moreover, the observation that learning generalizes from an easier discrimination to a more difficult discrimination has been obtained in other visual learning experiments (e.g Liu and Weinshall, 2000; Mackintosh, 1974) and provided the basis for the formulation of the reverse hierarchy theory, which postulates that learning starts at higher areas in the cortex and progresses backwards towards lower levels (Ahissar and Hochstein, 1997).

Another important observation deduced from the psychophysical experiments was that there seems to be no advantage for the learning of stimuli that are in accordance with the human kinematics. Recently, several studies have investigated the possible link between action recognition and action production (see Hommel et al., 2001; Wilson and Knoblich, 2005, for review). It has been proposed that movement recognition can be accomplished by the internal simulation of the motor programs that lead to that movement (Haruno et al., 2001) and that this internal simulation provides the basis for understanding other peoples emotions and mental states (e.g Blakemore and Decety, 2001; Frith and Frith, 1999; Gallese et al., 2004). Since the kinematics of the artificial articulated movements differed significantly from the one of normal human movements, it seems unlikely, that their recognition was based on such internal simulations. However, it cannot be completely ruled out, that learning also involves the formation of motor models for the artificial stimuli, even though in this case, one might expect a slight disadvantage for the learning of these stimuli, since motor models of human-like movements should already exist prior training.

Taken together, since our human-like stimuli were very similar to natural biological movements, and the learned representation showed orientation-

dependency, it seems plausible that the investigated learning mechanisms play a central role in normal biological motion recognition. To investigate possible neural correlates of the learning process, we conducted a series of functional imaging studies that were described in Chapter 4.

Concentrating on the visual areas involved in the processing of biological motion, we could demonstrate that learning to discriminate between very similar complex movements results in increased neural sensitivity to the differences between these movements. These learning induced changes in BOLD activity did not generalize to novel untrained movement patterns, suggesting that the investigated learning mechanism is stimulus specific rather than an effect of familiarity with the task.

Even though the psychophysical investigations in Chapter 3 showed a high degree of similarity between the learning of human-like and artificial articulated movements, our fMRI experiments highlighted small but significant differences in the corresponding BOLD activity changes. Using an fMRI adaptation paradigm, we found that the presentation of very similar human like movements led to significant recovery from adaptation in the classical biological motion areas (posterior superior temporal sulcus and fusiform face area) already before training. This result indicates the existence of partially distinct neural populations for the different human like movements already prior to training, which is consistent with the findings showing the specificity of these areas in the analysis of biological motion (Beauchamp et al., 2003; Bonda et al., 1996; Grezes et al., 2001; Grossman and Blake, 2002; Peuskens et al., 2005; Vaina et al., 2001). However, after training selective adaptation effects not only increased in the STSp and the FFA, but also became evident in the primarily mid to high level motion processing middle temporal (hMT+/V5) and kinetic occipital (V3b/KO) areas, consistent with their role in learning of global motion configurations (Vaina et al., 1998; Zohary et al., 1994).

In contrast to the human like movements, we observed no selective adaptation effects for the artificial articulated movements, with which the subjects had no prior experience, in any of the investigated areas before the training. Yet after training, we found very similar adaptation effects, like in the case of human like movements. That is, selective recovery from adaptation in the STSp, FFA, hMT+/V5 and V3b/KO was observed for the consecutive presentation of two similar movement patterns in contrast to the presentation of two identical patterns. Additionally, we also obtained this effect in mid level motion and form processing areas, namely V3a, VP and V4v. The involvement of retinotopic areas in the learning of artificial articulated patterns could signal the fact that these stimuli were completely novel to the observers. The recruitment of mid level fea-

ture areas could thereby help the observer to interpret the internal link structure of these novel skeleton models.

The main findings of the imaging experiments can be summarized as follows: I) Discrimination learning between similar complex movement stimuli involves distributed plasticity processes occurring across a whole range of visual areas sparing only low level areas V1 and V2. II) While partially distinct neural populations exist prior training in classical biological motion areas for human-like movements, which get refined due to the training process, distinct neural representations for artificial articulated movements emerge in the very same areas only after training. III) Additional mid level motion and form areas are recruited for the learning of artificial articulated patterns, possibly to support the grouping of the individual dots for previously unknown skeleton structures.

Although, our results imply plasticity processes in a wide range of visual areas, the technique of functional imaging is not suitable to determine the exact neural mechanisms that underlie this plasticity. As a result, we cannot discern whether the increased fMRI selective adaptation for very similar movements after training is due to changes in the tuning of neurons selective for these movements or due to the recruitment of larger numbers of neurons that become sensitive to these differences after training (Gilbert et al., 2001).

Additionally, we cannot discern bottom-up from top-down processes in the learning of biological movements. It is possible, that the learning of artificial articulated patterns could be implemented in a bottom-up manner from retinotopic to higher visual areas by enhancing processes of increasing complexity ranging from the integration of local configurations to the analysis of global motion features and biological motion properties. Alternatively, learning could begin at higher visual areas by enhancing the processing of the global biological characteristics of these movements and proceed to early retinotopic areas that have higher resolution necessary for the finer discrimination of the differences between the movements.

Similarly, learning of novel human-like movements could implicate top-down processes that support generalization to novel movements from known templates of human actions. These templates could be stored in the biological motion-related areas (STSp/FFA). Alternatively, motor areas that are thought to represent models for human actions (Saygin et al., 2004; Wolpert et al., 2003), could influence the processing in the biological motion-related areas via feedback processes. The selective adaptation effects prior training in these areas could be attributed to either of the two explanations.

To investigate the different explanations in terms of bottom-up or top-down processing, we proposed to use a theoretical model, which would allow to simu-

late possible learning mechanisms and connectivity changes due to the learning process. However, to utilize this model for testing the different hypotheses, we had to extend it by the implementation of physiologically plausible learning rules as well as a mechanism, by which BOLD signal changes observed in the functional imaging experiments could be simulated.

The simulations were based on an already existing neural model for biological motion recognition (Giese and Poggio, 2003). Their model consists of two parallel pathways and accomplishes movement recognition on the basis of learned prototypical movement patterns, which are encoded by neural feature detectors specialized for the detection of complex shapes and optic flow patterns. Both pathways consist of four hierarchy levels which extract either form or optic flow features with increasing complexity as well as position and size invariance along the different levels. Taken together, the properties of the neurons in the individual layers of the model are designed to mimic the properties of real neurons in the visual processing streams.

To learn the feedforward connections from the complex cells (layer 2) of the model to the snapshot neurons in layer 3, we applied an algorithm that was originally developed by Kohonen (1982). The outcome of this algorithm was that each snapshot neuron learned to represent a single frame of the animation in a fully unsupervised way.

To guarantee the sequence selectivity of the recognition, we additionally had to implement a mechanism by which the correct asymmetric lateral connectivity between the individual snapshot neurons could be learned (feedback connectivity). After testing several time dependent hebbian learning rules, we developed a physiologically plausible learning rule, which achieves stable learning of lateral connectivity by imposing only two constraints. These were that the maximum synaptic weight each neuron can support is constrained and that the neurons are not allowed to change from an excitatory into and inhibitory one and vice versa. Following the extension of the model, our simulations showed that the implemented learning rules led to specific recognition of individual movement patterns in a fully unsupervised way. Moreover, stable recognition could be achieved after only 30 stimulus repetitions, which is consistent with the learning speed of human observers obtained in the psychophysical experiments in Chapter 3.

To simulate real BOLD activity changes, we additionally implemented a neural adaptation mechanism at the highest hierarchy level of the model. Following simulations verified that consecutive activation of the same motion pattern neuron resulted in a decrease of activity of this neuron. This effect is analogous to a repetition suppression effect currently discussed

as one possibility for observed adaptation effects in fMRI experiments (Sawamura et al., 2006; Tolias et al., 2005).

Finally, we used the model to test, if the implemented adaptation mechanism would explain the reduced BOLD activity observed in response to the consecutive presentation of two identical stimuli. After simulating a whole fMRI experiment and performing exactly the same data analysis used for real fMRI data, we obtained modeled hemodynamic response functions that show great similarity with actually measured hemodynamic responses in the superior temporal sulcus. This shows that the implemented adaptation mechanism based on repetition suppression provides a suitable explanation for the reduced BOLD activity observed for the consecutive presentation of identical stimuli. In addition, the simulation verified that the assumption of linearity of the BOLD signal seems to hold true and can be exploited in rapid event-related designs, once the stimulus sequence is sufficiently counterbalanced.

We propose that after the extensions, this model could provide the foundation to test the different hypotheses about plasticity mechanisms at the neuronal level that underlie the observed BOLD activity changes during our functional imaging experiments.

6.2 Outlook

Certainly, the presented results lead to further questions with regard to the learning of complex movements. While our experiments identify articulation as one decisive factor in the learning process, other stimulus manipulations could highlight different processing constraints. A possible stimulus manipulations could involve for example, a violation of the 2/3 power law, which is believed to be an inherent principle of biological movements (Lacquaniti et al., 1983; Viviani and Flash, 1995). However, those manipulations might lead to different low level properties in the stimuli, which would be difficult to match.

By design, our functional imaging experiments focussed on visual areas involved in the processing of form and motion. It is however undoubted that also other cortical areas are involved in the recognition of complex movements. One possibility is the so called action-perception circuit, linking the superior temporal sulcus with parietal and prefrontal areas (see Hommel et al., 2001; Rizzolatti and Craighero, 2004; Wilson and Knoblich, 2005, for review). A very interesting question would be to investigate the contribution of the premotor cortex in this respect. Recently, Saygin and colleagues (2004) reported selective activation

of premotor areas for biological motion stimuli. Investigating selective adaptation processes in premotor cortex during discrimination learning could identify possible sites of top-down modulations in terms of internal motor models (e.g. Wolpert et al., 2003). Additionally, it would be very interesting to explore possible differences between human-like stimuli, which could be internally simulated and artificial articulated movements, for which, at least prior learning, no such models should exist.

The modeling work completed so far provides a theoretical framework to investigate the underlying plasticity mechanisms observed in our experiments. Although so far the model has only been tested in two experimental conditions serving as a 'proof of principle', further modeling should involve the exact simulation of our experiments to compare the simulated data with the real BOLD activity. By training the model with the prototypical movement stimuli used for the human like movements, it would become possible to examine the hypothesis that representations for similar movements already present in the biological motion-related areas provide the basis of the selective adaptation effects observed, already prior to training. By changing connections between and within the individual layers of the model, one could test the modifications that would lead to the best approximation of the data in order to replicate the adaptation effects at different levels in the hierarchy. Furthermore, modeling of parietal or prefrontal areas would enable us to investigate possible top-down influences on the activity changes.

In addition, the model could be used to investigate, if the proposed learning mechanisms would also be suitable to learn more realistic stimuli, like natural movements extracted from video sequences. Furthermore, the model could be used to derive quantitative predictions that could in turn be tested psychophysically or in additional functional imaging studies.

Taken together, additional experiments possibly together with neurophysiological recordings are necessary to completely understand the neural correlates of movement recognition.

Chapter 7

References

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

Aguirre, G. K. and D'Esposito, M. (2000). *Experimental design for brain fMRI*, pages 369 – 380. Functional MRI. Springer-Verlag Berlin, Heidelberg.

Ahissar, M. and Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, 387(6631):401 – 406.

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*, 4:267 – 278.

Amari, S. I. (1972). Learning patterns and pattern sequences by selforganizing nets of threshold elements. *IEEE Trans. on Computers*, 21:1197 – 1206.

Ashburner, J. and Friston, K. J. (2000). Image registration. In Moonen, C. T. W. and Bandettini, P. A., editors, *Functional MRI*, pages 285 – 299. Springer-Verlag Berlin, Heidelberg.

Ball, K. and Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science*, 218:697 – 698.

Ball, K. and Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Res*, 27(6):953 – 965.

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., and Hyde, J. S. (1992). Time course epi of human brain function during task activation. *Magn Reson Med.*, 25(2):390 – 397.

Bar, M., Tootell, R. B., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., Rosen, B. R., and Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron*, 29(2):529 – 535.

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*, 15:991 – 1001.

Beintema, J. A. and Lappe, M. (2002). Perception of biological motion without local image motion. *Proc. Natl. Acad. Sci. USA*, 99(8):5661 – 5663.

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

Bertenthal, B. I., Proffitt, D. R., and Cutting, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions. *J Exp Child Psychol*, 37:213 – 230.

Bertenthal, B. I., Proffitt, D. R., and Kramer, S. J. (1987). Perception of biomechanical motions by infants: implementation of various processing constraints. *J Exp Psychol Hum Percept Perform*, 13(4):577 – 585.

Bi, G. Q. and Poo, M. M. (1998). Synaptic modification in cultured hippocampal neurons: Dependence on spike timing, synaptic strength and postsynaptic cell type. *J. Neurosci.*, 18:10464 – 10472.

Blake, R. (1993). Cats perceive biological motion. *Psychological Science*, 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 – 567.

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.

Booth, A. E., Bertenthal, B. I., and Pinto, J. (2002). Perception of the symmetrical patterning of human gait by infants. *Dev Psychol*, 38:554 – 563.

Boynton, G. M. and Finney, E. M. (2003). Orientation-specific adaptation in human visual cortex. *J Neurosci*, 23:8781 – 8787.

Brainard, D. H. (1997). The psychophysics toolbox. *Spat Vis*, 10(4):433 – 436.

Brammer, M. J. (2001). Head motion and its correction. In Jezzard, P., Matthews, M, P., Smith, and M, S., editors, *Functional MRI: an introduction to methods*, pages 243 – 250. Oxford University Press Inc., New York.

Bruderlin, A. and Williams, L. (1995). Motion signal processing. *Computer Graphics*, 29:97 – 104.

Buckner, R. L. and Koutstaal, W. (1998). Functional neuroimaging studies of encoding, priming, and explicit memory retrieval. *Proc. Natl. Acad. Sci. USA*, 95:891 – 898.

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

Bülthoff, I., Bülthoff, H., and Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nat Neurosci.*, 1:254 – 257.

Buxton, R. B., Wong, E. C., and Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med.*, 39:855 – 864.

Cavanagh, P., Labianca, A. T., and Thornton, I. M. (2001). Attention-based visual routines: sprites. *Cognition*, 80:47 – 60.

Chao, L. L., Haxby, J. V., and Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci*, 2:913 – 919.

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(4):916 – 929.

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., and Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123:291 – 307.

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

Dale, A. M. and Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human brain mapping*, 5:329 – 340.

Dayan, P. and Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and mathematical modeling of neural systems*. MIT Press, Cambridge, MA.

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

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. USA*, 93:13494 – 13499.

Desimone, R., Albright, T. D., Gross, C. G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.*, 4:2051 – 2062.

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., and Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc. Natl. Acad. Sci. USA*, 93:2382 – 2386.

DeYoe, E. A. and Essen, D. C. V. (1988). Concurrent processing streams in monkey visual cortex. *Trends Neurosci*, 11:219 – 226.

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

Dittrich, W. H., Lea, S. E. G., Barrett, J., and Gurr, P. R. (1998). Categorization of natural movements by pigeons: Visual concept discrimination and biological motion. *Journal of the Experimental Analysis of Behavior*, 70:281 – 299.

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

Dolan, R. J., Fink, G. R., Rolls, E., Booth, M., Holmes, A., Frackowiak, R. S., and Friston, K. J. (1997). How the brain learns to see objects and faces in an impoverished context. *Nature*, 389:596 – 599.

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.

Duncan, R. O. and Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, 38:659 – 671.

Dupont, P., Bruyn, B. D., Vandenberghe, R., Rosier, A. M., Michiels, J., Marchal, G., Mortelmans, L., and Orban, G. A. (1997). The kinetic occipital region in human visual cortex. *Cereb Cortex*, 7:283 – 292.

Dupont, P., Orban, G. A., Bruyn, B. D., Verbruggen, A., and Mortelmans, L. (1994). Many areas in the human brain respond to visual motion. *J Neurophysiol*, 72:1420 – 1424.

Edelman, S. and Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Res*, 32(12):2385 – 2400.

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., and Shadlen, M. N. (1994). fMRI of human visual cortex. *Nature*, 369:525.

Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G. (1995). Motor facilitation during action observa-tion: a magnetic stimulation study. *J Neurophysiol*, 73(6):2608 – 2611.

Fahle, M. (2004). Perceptual learning: a case for early selection. *J Vis*, 4:879 – 890.

Fahle, M. (2005). Perceptual learning: specificity versus generalization. *Curr Opin Neurobiol*, 15:154 – 160.

Felleman, D. J. and Essen, D. C. V. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex*, 1(1):1 – 47.

Fine, I. and Jacobs, R. A. (2002). Comparing perceptual learning tasks: a review. *J Vis*, 2(2):190 – 203.

Fox, P. T. and Raichle, M. E. (1986). Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proc. Natl. Acad. Sci. USA*, 83(4):1140 – 1144.

Fox, P. T., Raichle, M. E., Mintun, M. A., and Dence, C. (1988). Nonoxidative glucose consumption during focal physiologic neural activity. *Science*, 241:462 – 464.

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

Frackowiak, R. S. J., Friston, J, K., Frith, D, C., Dolan, J, R., Mazziotta, and C, J. (1997). *Human brain function*. Academic Press, San Diego.

Frahm, J., Bruhn, H., Merboldt, K. D., and Hanicke, W. (1992). Dynamic mr imaging of human brain oxygenation during rest and photic stimulation. *J Magn Reson Imaging*, 2(5):501 – 505.

Frahm, J., Kruger, G., Merboldt, K. D., and Kleinschmidt, A. (1996). Dynamic uncoupling and recoupling of perfusion and oxidative metabolism during focal brain activation in man. *Magn. Reson. Med.*, 35:143 – 148.

Frith, C. D. and Frith, U. (1999). Interacting minds - a biological basis. *Science*, 286(5445):1692 – 1695.

Furmanski, C. S., Schluppeck, D., and Engel, S. A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Curr Biol*, 14:573 – 578.

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

Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn Sci*, 8:396 – 403.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci*, 2:568 – 573.

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

Gibson, E. J. (1953). Improvement of perceptual judgments as a function of controlled practice or training. *Psycho bull*, 50:401 – 431.

Giese, M. A. (2000). Neural model for the recognition of biological motion. In *Dynamic Perception*, pages 105–110. Infix Verlag, Berlin.

Giese, M. A. (2006). Computational principles for the recognition of biological movements: Model versus feature-based approaches. In Knoblich, G., Thornton, I. M., Grosjean, M., and Shiffrar, M., editors, *Perception of the Human Body from the Inside Out*, pages 323–359. Oxford University Press, New York.

Giese, M. A. and Lappe, M. (2002). Measurement of gen-eralization fields for the recognition of biological motion. *Vision Res*, 42(15):1847 – 1858.

Giese, M. A. and Poggio, T. (2000). Morphable models for the analysis and synthesis of complex motion patterns. *International Journal of Computer Vision*, 38(1):59 – 73.

Giese, M. A. and Poggio, T. (2002). Neural mechanisms for the recognition of biological movements and actions. *AI Memo*, 219:012/CBCL.

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

Giese, M. A., Thornton, I. M., and Edelman, S. (2003). Metric category spaces of biological motion. *Journal of Vision*, 3:83a.

Gilaie-Dotan, S., Ullman, S., Kushnir, T., and Malach, R. (2002). Shape-selective stereo processing in human object-related visual areas. *Hum. Brain Mapp*, 15:67 – 79.

Gilbert, C. D. (1994). Early perceptual learning. *Proc. Natl. Acad. Sci. USA*, 91:1195 – 1197.

Gilbert, C. D., Sigman, M., and Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, 31:681 – 697.

Glover, G. H. (1999). Deconvolution of impulse response in event-related bold fMRI. *Neuroimage*, 9:416 – 429.

Goddard, N. H. (1992). The perception of articulated motion: Recognizing moving light displays. Technical report, Dept. of Computer Science, Univ. of Rochester.

Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H., and Singer, W. (1998). The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *Eur J Neurosci*, 10:1563 – 1573.

Goldstone, R. L. (1998). Perceptual learning. *Annu Rev Psychol*, 49:585 – 612.

Goodale, M. A., Millner, A. D., Jakobson, L. S., and Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, 349:154 – 156.

Goodale, M. A. and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.*, 15:20 – 25.

Grafton, S. T., Fagg, A. H., Woods, R. P., and Arbib, M. A. (1996). Functional anatomy of pointing and grasping in humans. *Cereb Cortex*, 6:226 – 237.

Grafton, S. T., L.Fadiga, Arbib, M. A., and Rizzolatti, G. (1997). Premotor cortex activation during obser-vation and naming of familiar tools. *Neuroimage*, 6(4):231 – 236.

Grezes, J., Costes, N., and Decety, J. (1999). The effects of learning and intention on the neural network involved in the perception of meaningless actions. *Brain*, 122:1875 – 1887.

Grezes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C., and Decety, J. (2001). Does perception of biological motion rely on specific brain regions? *Neuroimage*, 13(5):775 – 785.

Grill-Spector, K. (2003). The neural basis of object perception. *Curr. Opin. Neurobiol*, 13:159 – 166.

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci.*, 10(1):14 – 23.

Grill-Spector, K., Knouf, N., and Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci.*, 7:555 – 562.

Grill-Spector, K., Kourtzi, Z., and Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Res.*, 41:1409 – 1422.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, 24(1):187 – 203.

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., and Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*, 21:191 – 202.

Grill-Spector, K. and Malach, R. (2001). fmr-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol. (Amst.)*, 107:293 – 321.

Grill-Spector, K. and Malach, R. (2004). The human visual cortex. *Annu Rev Neurosci.*, 27:649 – 677.

Grossman, E. D. and Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Res*, 41(10-11):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., Blake, R., and Kim, C. Y. (2004). Learning to see biological motion: brain activity parallels behavior. *J Cogn Neurosci*, 16(9):1669 – 1167.

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

Guckenheimer, J. and Holmes, P. (1983). *Nonlinear Oscillations, Dynamic Systems and Bifurcations of Vector Fields*. Springer-Verlag, New York.

Haruno, M., Wolpert, D. M., and Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Comput*, 13:2201 – 2220.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293:2425 – 2430.

Henson, R., Shallice, T., and Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science*, 287:1269 – 1272.

Hill, H. and Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychol Sci*, 11(3):223 – 228.

Hommel, B., Musseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (tec): a framework for perception and action planning. *Behav Brain Sci*, 24:849 – 878.

Howard, R. J., Brammer, M., Wright, I., Woodruff, P. W., Bullmore, E. T., and Zeki, S. (1996). A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. *Curr Biol.*, 6:1015 – 1019.

Huettel, S. A., Obembe, O. O., Song, A. W., and Woldorff, M. G. (2004). The bold fMRI refractory effect is specific to stimulus attributes: evidence from a visual motion paradigm. *Neuroimage*, 23:402 – 408.

Huk, A. C., Dougherty, R. F., and Heeger, D. J. (2002). Retinotopy and functional subdivision of human areas MT and MST. *J. Neurosci.*, 22:7195 – 7205.

Huk, A. C. and Heeger, D. J. (2000). Neuronal basis of the motion aftereffect reconsidered. *Neuron*, 32:6 – 8.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., and Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, 286:2526 – 2528.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., and Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proc. Natl. Acad. Sci. USA*, 96:9379 – 9384.

James, T. W. and Gauthier, I. (2006). Repetition-induced changes in bold response reflect accumulation of neural activity. *Hum Brain Mapp.*, 27:37 – 46.

Jastorff, J., Kourtzi, Z., and Giese, M. A. (2003). Role of learning in biological motion recognition. *Journal of Vision*, 3(9):84a.

Jenkinson, M. (2001). Registration, atlases and cortical flattening. In Jezzard, P., Matthews, M, P., Smith, and M, S., editors, *Functional MRI: an introduction to methods*, pages 271 – 293. Oxford University Press, New York.

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

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. an experimental and theoretical analysis of calculus-like functions in visual data processing. *Psychol Res*, 38:379 – 393.

Johansson, G., von Hofsten, C., and Jansson, G. (1980). Event perception. *Annu Rev Psychol*, 31:27 – 63.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.*, 17:4302 – 4311.

Karni, A., Meyer, G., Jezzard, P., Adams, M. M., Turner, R., and Ungerleider, L. G. (1995). Functional mri evidence for adult motor cortex plasticity during motor skill learning. *Nature*, 377:155 – 158.

Kastner, S., Weerd, P. D., and Ungerleider, L. G. (2000). Texture segregation in the human visual cortex: a functional mri study. *J. Neurophysiol.*, 83:2453 – 2457.

Knappmeyer, B., Thornton, I. M., and Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Res*, 43:1921 – 1936.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 24:59–69.

Kourtzi, Z., Betts, L. R., Sarkheil, P., and Welchman, A. E. (2005). Distributed neural plasticity for shape learning in the human visual cortex. *PLoS Biol*, 3:e204.

Kourtzi, Z. and Kanwisher, N. (2000a). Activation in human MT/MST by static images with implied motion. *J Cogn Neurosci*, 12(1):48 – 55.

Kourtzi, Z. and Kanwisher, N. (2000b). Cortical regions involved in perceiving object shape. *J. Neurosci.*, 20:3310 – 3318.

Kourtzi, Z. and Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, 293:1506 – 1509.

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

Kriegeskorte, N., Sorger, B., Naumer, M., Schwarzbach, J., van den Boogert, E., Hussy, W., and Goebel, R. (2003). Human cortical object recognition from a visual motion flowfield. *J. Neurosci.*, 23:1451 – 1463.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., Cheng, H. M., Brady, T. J., and Rosen, B. R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci. USA*, 89(12):5675 – 5679.

Lacquaniti, F., Terzuolo, C., and Viviani, P. (1983). The law relating the kinematic and figural aspects of drawing movements. *Acta Psychol (Amst)*, 54(1-3):115 – 130.

Levy, I., Hasson, U., Avidan, G., Hendler, T., and Malach, R. (2001). Center-periphery organization of human object areas. *Nat. Neurosci.*, 4:533 – 539.

Li, L., Miller, E. K., and Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *J Neurophysiol.*, 69:1918 – 1929.

Liu, Z. (1995). Learning a visual skill that generalizes. Technical report, NEC Research Institute.

Liu, Z. (1999). Perceptual learning in motion discrimination that generalizes across motion directions. *Proc. Natl. Acad. Sci. USA*, 96:14085 – 14087.

Liu, Z. and Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vision Res*, 40(1):97 – 109.

Logothetis, N. K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos. Trans. R. Soc. London, Ser. B*, 357:1003 – 1037.

Logothetis, N. K. (2003). The underpinnings of the bold functional magnetic resonance imaging signal. *J. Neurosci.*, 23:3963 – 3971.

Logothetis, N. K., Guggenberger, H., Peled, S., and Pauls, J. (1999). Functional imaging of the monkey brain. *Nat. Neurosci.*, 2:555 – 562.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412:150 – 157.

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

Logothetis, N. K. and Sheinberg, D. L. (1996). Visual object recognition. *Annu Rev Neurosci*, 19:577 – 621.

Lorenz, K. (1965). *Evolution and Modification of Behaviour*. University of Chicago Press, Chicago.

Mackintosh, N. J. (1974). *The Psychology of Animal Learning*. London: Academic.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. I., Brady, T. J., Rosen, B. R., and Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl. Acad. Sci. USA*, 92:8135 – 8139.

Markram, H., Lübke, J., Frotscher, M., and Sackmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275:213 – 215.

Marr, D. and Vaina, L. (1982). Representation and recogni-tion of the movements of shapes. *Proc R Soc Lond B Biol Sci*, 214(1197):501 – 524.

Martin, A., Wiggs, C. L., Ungerleider, L. G., and Haxby, J. V. (1996). Neural correlates of categoryspecific knowledge. *Nature*, 379:649 – 652.

Maunsell, J. H. and Essen, D. C. V. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J. Neurosci.*, 3(12):2563 – 2586.

Mayer, M. J. (1983). Practice improves adults' sensitivity to diagonals. *Vision Res*, 23(5):547 – 550.

Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., and Lancaster, J. (1995). A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage*, 2:89 – 101.

Meltzoff, A. N. and Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198(4312):74 – 78.

Miller, K. D. (1996). Synaptic economics: Competition and cooperation in synaptic plasticity. *Neuron*, 17:371 – 374.

Mineiro, P. and Zipser, D. (1998). Analysis of direction selectivity arising from recurrent cortical interactions. *Neural Computation*, 10:353 – 371.

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

Moeslund, T. B. and Granum, G. (2001). A survey of computer vision-based human motion capture. *Comp. Vis. Image Underst.*, 81:231 – 268.

Montague, P. R. and Sejnowski, T. J. (1994). The predictive brain: Temporal coincidence and temporal order in synaptic learning mechanisms. *Learning and Memory*, 1:1 – 33.

Obermayer, K., Ritter, H., and Schulten, K. (1990). A principle for the formation of the spatial structure of cortical feature maps. *Proc. Natl. Acad. Sci. USA*, 87:8345 – 8349.

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., and Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron*, 18(4):591 – 598.

Ogawa, S., Lee, T. M., Nayak, A. S., and Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn Reson Med.*, 14(1):68 – 78.

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., and Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA*, 89(13):5951 – 5955.

Omori, E. and Watanabe, S. (1996). Discrimination of johansson's stimuli in pigeons. *Int J Comp Psychol*, 9:92.

Oostende, S. V., Sunaert, S., Hecke, P. V., Marchal, G., and Orban, G. A. (1997). The kinetic occipital (KO) region in man: an fMRI study. *Cereb Cortex*, 7:690 – 701.

Orban, G. A., Sunaert, S., Todd, J. T., Hecke, P. V., and Marchal, G. (1999). Human cortical regions involved in extracting depth from motion. *Neuron*, 24:929 – 940.

O'Toole, A. J., Roark, D. A., and Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn Sci*, 6(6):261 – 266.

Pavlova, M., Krageloh-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. *Percept Psychophys*, 62(5):889 – 899.

Pavlova, M., Staudt, M., Sokolov, A., Birbaumer, N., and Krageloh-Mann, I. (2003). Perception and production of biological movement in patients with early periventricular brain lesions. *Brain*, 126:692 – 701.

Perrett, D. I., Smith, P. A., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., Broennimannand, R., Milner, A. D., and Jeeves, M. A. (1985). Visual analysis of body movements by neurones in the temporal cortex of the macaque monkey: a preliminary report. *Behav Brain Res*, 16(2-3):153 – 170.

Petersen, S. E., Baker, J. F., and Allman, J. M. (1985). Direction-specific adaptation in area MT of the owl monkey. *Brain Res*, 346:146 – 150.

Peuskens, H., Vanrie, J., Verfaillie, K., and Orban, G. A. (2005). Specificity of regions processing biological motion. *Eur J Neurosci*, 21(10):2864 – 2875.

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

Poggio, T. and Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343:263 – 266.

Poggio, T., Fahle, M., and Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, 256(5059):1018 – 1021.

Pollick, F. E., Paterson, H. M., Bruderlin, A., and Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82:B51 – 61.

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

Puce, A. and Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philos. Trans. R. Soc. London Ser. B Biol. Sci.*, 358:435 – 445.

Ramachandran, V. S. and Braddick, O. (1973). Orientation-specific learning in stereopsis. *Perception*, 2:371 – 376.

Rao, R. P. N. and Sejnowski, T. J. (2001). Spike-timing-dependent hebbian plasticity as temporal difference learning. *Neural Computation*, 13:2221 – 2237.

Rees, G., Friston, K., and Koch, C. (2000). Adirect quantitative relationship between the functional properties of human and macaque v5. *Nat. Neurosci.*, 3:716 – 723.

Ress, D., Backus, B. T., and Heeger, D. J. (2000). Activity in primary visual cortex predicts performance in a visual detection task. *Nat Neurosci.*, 3(9):940 – 945.

Riesenhuber, M. and Poggio, T. (2002). Neural mechanisms of object recognition. *Curr Opin Neurobiol.*, 12:162 – 168.

Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system. *Annu Rev Neurosci*, 27:169 – 192.

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

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., and Fazio, F. (1996b). Localization of grasp representations in humans by pet: 1. observation versus execution. *Exp Brain Res.*, 111:246 – 252.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci*, 2:661 – 670.

Savoy, R. L., Bandettini, P. A., O'Craven, K. M., Kwong, K. K., Davis, T. L., Baker, J. R., Weisskoff, R. M., and Rosen, B. R. (1995). Pushing the temporal resolution of fMRI: Studies of very brief visual stimuli, onset variability and asynchrony and stimulus-correlated changes in noise. *Proc Soc Magn Reson Med 3rd Annual Meeting*, 2:450.

Sawamura, H., Orban, G. A., and Vogels, R. (2006). Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the fMRI adaptation paradigm. *Neuron*, 49:307 – 318.

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

Schoups, A. A., Vogels, R., Qian, N., and Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, 412:549 – 553.

Schwartz, E., Tootell, R. B., Silverman, M. S., Switkes, E., and Valois, R. L. D. (1985). On the mathematical structure of the visuotopic mapping of macaque striate cortex. *Science*, 227:1065 – 1066.

Schwarzlose, R. F., Baker, C. I., and Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *J. Neurosci.*, 25:11055 – 11059.

Senior, C., Barnes, J., Giampietro, V., Simmons, A., Bullmore, E. T., Brammer, M., and David, A. S. (2000). The functional neuroanatomy of implicit-motion perception or representational momentum. *Curr Biol*, 10(1):16 – 22.

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

Smith, A. T., Greenlee, M. W., Singh, K. D., Kraemer, F. M., and Hennig, J. (1998). The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging (fMRI). *J. Neurosci.*, 18(10):3816 – 3830.

Smith, S. M. (2001). Preparing fMRI data for statistical analysis. In Jezzard, P., Matthews, M, P., Smith, and M, S., editors, *Functional MRI: an introduction to methods*, pages 229 – 241. Oxford University Press Inc., New York.

Spiridon, M. and Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? an fMRI study. *Neuron*, 35:1157 – 1165.

Stuart, G. J. and Sakmann, B. (1994). Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature*, 367:69 – 72.

Sumi, S. (1984). Upside-down presentation of the johansson moving light-spot pattern. *Perception*, 13(3):283 – 286.

Sutton, R. S. (1988). Learn to predict by the method of temporal differences. *Machine Learning*, 3:9–44.

Swets, J. A., Tanner, W. P., and Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, 68(5):301 – 340.

Tarr, M. J. and Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, 67(1-2):1 – 20.

Tarr, M. J. and Cheng, Y. D. (2003). Learning to see faces and objects. *Trends Cogn Sci*, 7:23 – 30.

Tarr, M. J. and Gauthier, I. (2000). Ffa: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat. Neurosci.*, 3:764 – 769.

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

Thornton, I. M., Rensink, R. A., and Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception*, 31(7):837 – 853.

Thulborn, K. R., Waterton, J. C., Matthews, P. M., and Radda, G. K. (1982). Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochim. Biophys. Acta*, 714(2):265 – 270.

Tolias, A. S., Keliris, G. A., Smirnakis, S. M., and Logothetis, N. K. (2005). Neurons in macaque area V4 acquire directional tuning after adaptation to motion stimuli. *Nat Neurosci*, 8:591 – 593.

Tolias, A. S., Smirnakis, S. M., Augath, M. A., Trinath, T., and Logothetis, N. K. (2001). Motion processing in the macaque: revisited with functional magnetic resonance imaging. *J Neurosci*, 21:8594 – 8601.

Tootell, R. B., Reppas, J. B., Dale, A. M., Look, R. B., Sereno, M. I., Malach, R., Brady, T. J., and Rosen, B. R. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.*, 15:3215 – 3230.

Troje, N. F. (2002). Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. *J Vis*, 2(5):371 – 387.

Troje, N. F. (2003). Reference frames for orientation anisotropies in face recognition and biological-motion perception. *Perception*, 32:201 – 210.

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 – 675.

Vaina, L. and Bennour, Y. (1985). A computational ap-proach to visual recognition of arm movements. *Percept Mot Skills*, 60(1):203 – 228.

Vaina, L. M., Belliveau, J. W., des, R. E., and Zeffiro, T. A. (1998). Neural systems underlying learning and representation of global motion. *Proc. Natl. Acad. Sci. USA*, 95:12657 – 12662.

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

Vallortigara, G., Regolin, L., and Marconato, F. (2005). Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biol*, 3(7):e208.

Verfaillie, K. (2000). Perceiving human locomotion: priming effects in direction discrimination. *Brain Cogn*, 44(2):192 – 213.

Verfaillie, K., Troy, A. D., and Rensbergen, J. V. (1994). Transsaccadic integration of biological motion. *J Exp Psychol Learn Mem Cogn*, 20(3):649 – 670.

Viviani, P. and Flash, T. (1995). Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. *J Exp Psychol Hum Percept Perform*, 21(1):32 – 53.

Viviani, P. and Stucchi, N. (1992). Biological movements look uniform: evidence of motor-perceptual interactions. *J Exp Psychol Hum Percept Perform*, 18(3):603 – 623.

Vogels, R. and Orban, G. A. (1985). The effect of practice on the oblique effect in line orientation judgments. *Vision Res*, 25:1679 – 1687.

Watson, J. D., Myers, R., Frackowiak, R. S., Hajnal, J. V., Woods, R. P., Mazziotta, J. C., Shipp, S., and Zeki, S. (1993). Area v5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cereb Cortex.*, 3:79 – 94.

Webb, J. A. and Aggarwal, S. J. (1982). Structure from motion of rigid and jointed objects. *Artif. Intell*, 19:107 – 130.

Wiggs, C. L. and Martin, A. (1998). Properties and mechanisms of perceptual priming. *Curr. Opin. Neurobiol.*, 8:227 – 233.

Wiley, D. and Hahn, J. (1997). Interpolation synthesis of articulated figure motion. *IEEE Computer Graphic and Application*, 17:39 – 45.

Wilson, M. and Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychol Bull*, 131(3):460 – 473.

Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philos Trans R Soc Lond B Biol Sci*, 358(1431):593 – 602.

Xie, X. and Giese, M. A. (2002). Nonlinear dynamics of direction-selective nonlinear neural media. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 65:051904.

Yang, T. and Maunsell, T. H. (2004). The effect of perceptual learning on neuronal responses in monkey visual area. *J. Neurosci.*, 24:1617 – 1626.

Zeki, S., Perry, R. J., and Bartels, A. (2003). The processing of kinetic contours in the brain. *Cerebral Cortex*, 13(2):189 – 202.

Zilles, K., Schleicher, A., Langemann, C., Amunts, K., Morosan, P., Palomero-Gallagher, N., Schormann, T., Mohlberg, H., Bürgel, U., Steinmetz, H., Schlaug, G., and Roland, P. E. (1997). Quantitative analysis of sulci in the

human cerebral cortex: Development, regional heterogeneity, gender difference, asymmetry, intersubject variability and cortical architecture. *Human Brain Mapping*, 5:218 – 221.

Zohary, E., Celebrini, S., Britten, K. H., and Newsome, W. T. (1994). Neuronal plasticity that underlies improvement in perceptual performance. *Science*, 263:1289 – 1292.

Appendix A

Curriculum Vitae

Jan Jastorff

Date / place of birth:	1/18/1977, Bremen (Germany)
Citizenship:	German
Marital status:	Married

Education:

2/2006	Dissertation (Ph.D.) to be completed at the Graduate School of Neural and Behavioural Sciences, Tübingen, Germany.
since 1/2003	Graduate Student, Laboratory for Action Representation and Learning, Dept. for Cognitive Neurology, University Clinic Tübingen, Germany.
10/2002	Masters Degree in Neural and Behavioural Sciences, Graduate School of Neural and Behavioural Sciences, Tübingen, Germany.
2002	Masters Thesis at the Laboratory for Action Representation and Learning, Dept. for Cognitive Neurology, University Clinic Tübingen Germany.
2000-2002	Studies of Neural and Behavioural Sciences, Graduate School of Neural and Behavioural Sciences, Tübingen, Germany.
1999-2000	Studies of Neurobiology, Swiss Federal Institute of Technology Zürich, Switzerland.
1999	Vordiplom (B.Sc.), Dept. of Biology, University of Rostock, Germany.
1997-1999	Studies of Biology, University of Rostock, Germany.

Research Experience:

since 2003	PhD fellow, Laboratory for Action Representation and Learning, Dept. for Cognitive Neurology, University Clinic Tübingen , Germany.
1/2002-4/2002	Laboratory rotation, Brain Research Institute, Dept. of Neuropharmacology, Bremen, Germany.
2000-2001	Studentische Hilfskraft (research student), Institute of Medical Psychology and Behavioral Neurobiology. Tübingen, Germany.
1999	Studentische Hilfskraft (research student), Dept. of Biochemistry, University of Zürich, Switzerland.

Honors, awards:

2002 Best Poster Prize, 5. Interdisciplinary College, Günne, Germany.

Appendix B

Bibliography

Journal Papers and Peer-Reviewed Conference Papers:

- Jastorff J, Kourtzi Z, Giese MA (2002): Learning of the discrimination of artificial complex biological motion. In Würz RP, Lappe M (eds): Dynamic Perception; Infix, Berlin, 133-138.
- Jastorff J, Giese MA (2004): Time-dependent hebbian rules for the learning of templates for visual motion recognition. In Ilg U, Bülthoff HH, Mallot H (eds): Dynamic Perception; Infix, Berlin, 151-156.
- Jastorff J, Kourtzi Z, Giese MA: Learning to recognize complex movements: Biological vs. artificial trajectories. (under review: Journal of Vision)
- Jastorff J, Giese MA, Kourtzi Z: Visual learning shapes the processing of actions in the human visual cortex. (under review: Neuron)
- Jastorff J, Kourtzi Z, Giese MA: Learning in biological motion: Local vs. global processing. (in preparation)
- Sarkheil P, Jastorff J, Giese MA, Kourtzi Z: Categorization of complex dynamic patterns in the human brain. (in preparation)

Conference Publications:

Oral Presentations:

- Jastorff J, Kourtzi Z, Giese MA (2003) Role of learning in biological motion recognition. Journal of Vision; 3, 84a.
- Jastorff J, Kourtzi Z, Giese MA (2004) fMRI adaptation dissociates multiple areas involved in biological motion recognition. In Kerzel D, Franz V, Gegenfurtner K (eds): BeitrŁge zur 46. Tagung experimentell arbeitender Psychologen; Papst Science Publishers, Lengerich 126.
- Jastorff J, Giese MA, Kourtzi Z (2004) Perceptual learning of novel biological movements in the human visual brain. 34th Annual Meeting of the Society for Neuroscience, San Diego, United States.
- Jastorff J, Kourtzi Z, Giese MA (2005) Visual Learning of Complex Movements: Investigation of Neural Plasticity Mechanisms. 28th European Conference on Visual Perception.

Poster Presentations:

- Jastorff J, Kourtzi Z, Giese MA (2002): Recognition of artificial complex biological motion generated by motion morphing. 5. Inderdisciplinary College 2002, Gnne, Germany.
- Jastorff J, Giese MA (2002): Neural model for the learning of biological motion. Perception; 31 Suppl.119.
- Giese MA, Jastorff J, Kourtzi Z (2002): Learning of the discrimination of artificial complex biological motion. Perception; 31 Suppl. 117.
- Giese MA, Jastorff J, Kourtzi Z (2002): Learning of artificial biological motion patterns. 32nd Annual Meeting of the Society for Neuroscience, Orlando, United States.
- Jastorff J, Kourtzi Z, Giese MA (2002): Role of learning in biological motion recognition. 2nd Workshop on Biologically Motivated Computer Vision, Tübingen, Germany.
- Jastorff J, Kourtzi Z, Giese MA (2003): Learning of artificial biological motion: Comparison between natural and synthetic trajectories. In: Bülthoff HH, Gegenfurtner KR, Mallot HA, Ulrich R, Wichmann FA (eds): Beiträge zur 6. Tübinger Wahrnehmungskonferenz; Knirsch, Kirchentellinsfurt 152.

- Jastorff J, Kourtzi Z, Giese MA (2003) Learning of natural and synthetic biological motion. In: N. Elsner and H. Zimmermann (eds): Proceedings of the 5th Meeting of the German Neuroscience Society; Thieme, Stuttgart, 627.
- Jastorff J, Giese MA, Kourtzi Z (2004) Neural correlates of the learning of biological motion: An fMRI adaptation experiment. In: Bülthoff HH, Mallot HA, Ulrich R, Wichmann FA (eds): Beiträge zur 7. Tübinger Wahrnehmungskonferenz; Knirsch, Kirchentellinsfurt 137.
- Jastorff J, Kourtzi Z, Giese MA (2005) Learning of biological motion: Combining fMRI and theoretical modeling. In: Bülthoff HH, Mallot HA, Ulrich R, Wichmann FA (eds): Beiträge zur 8. Tübinger Wahrnehmungskonferenz; Knirsch, Kirchentellinsfurt 131.
- Sarkheil P, Jastorff J, Giese MA, Kourtzi Z (2005) Categorization of complex dynamic patterns in the human brain. In: Bülthoff HH, Mallot HA, Ulrich R, Wichmann FA (eds): Beiträge zur 8. Tübinger Wahrnehmungskonferenz; Knirsch, Kirchentellinsfurt 188.
- Jastorff J, Kourtzi Z, Giese MA (2005) Neural plasticity mechanisms for learning of biological motion. Proceedings of the 5th Annual Meeting of the Vision Sciences Society, 935.
- Sarkheil P, Jastorff J, Giese MA, Kourtzi Z (2005) Categorization of complex dynamic patterns in the human brain. Proceedings of the 5th Annual Meeting of the Vision Sciences Society, 850.

Invited Talks:

- Jastorff J (2004) Funktionelle kernspintomographische Untersuchungen zur Grundlage des visuellen Lernens komplexer biologischer Bewegungen. Deutsche Gesellschaft für Klinische Neurophysiologie und funktionelle Bildgebung, Jena, Germany.

Masters Thesis:

- Jastorff J (2002): Learning of artificial biological motion stimuli based on either local or global image features. *Graduate School of Neural and Behavioural Sciences, University Tübingen*