The Selective Attention for Action Model (SAAM)

An Exploration of Affordances in a Computational and Experimental Study

Christoph Böhme

A thesis submitted to the
University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Psychology
College of Life and Environmental Sciences
University of Birmingham
February 2011

Abstract

In this thesis a connectionist model for affordance-guided selective attention for action is presented. The selective attention for action model (SAAM) is aimed at modelling the direct route from vision to action using Gibson's affordances to describe actions. In the model complex affordances are encoded in grasp postures. It is shown that this is a viable method to realise affordances. By deriving complex affordances from grasp postures which in turn are derived from grasp affordances and invariants in the visual information, the model implements a hierarchical structure of affordances as postulated in Gibson's affordance concept. This is a novel way of implementing affordances in a computational model. Three studies were conducted to explore the model. The grasp postures generated by SAAM are verified in Study 1 by comparing them with human grasp postures. These grasps were collected in an experiment which used the same stimulus shapes as the model. In Study 2 the attentional behaviour of SAAM is investigated. It is shown that the model is able to select only one of the objects in visual inputs showing multiple objects as well as selecting the appropriate action category (described by a grasp type). Furthermore, it is shown that the bottom-up selection of the model can be influenced by top-down feedback of action intentions. In Study 3 the findings of Study 2 are applied to stimuli showing two hand-tools (pliers and hammer) to clarify how the findings of Study 2 link to complex affordances and action intentions. The results demonstrate that SAAM offers a novel and powerful way to implement complex affordances. Possibilities for further experimentation and extensions of the model are discussed.

Acknowledgements

First of all I would like to thank my supervisor Dr Dietmar Heinke for having given me the opportunity to do my PhD with him and for always being approachable. Without the many discussions we had, many of the ideas in this thesis would not be there.

I am also grateful to Samantha Hopkins for running the experiment presented in Study 1 and doing the tedious work of marking thousands of fingers in photographs.

And, of course, there are my family and my friends whom I owe a big Thank You. They had put up more than once with my grumpiness when SAAM was doing everything but generating human-like grasps.

The research presented in this thesis was supported by the Engineering and Physical Sciences Research Council.

Contents

I Introduction 1

1 Motivation 2

2 Grasping 5
- 2.1 Grasping in a Computational Model 6
- 2.2 Classifying and Describing Grasps 7
  - 2.2.1 Grasp Taxonomies 8
  - 2.2.2 Quantitative Grasp Encodings 14
- 2.3 A Definition of Grasping 16
- 2.4 Conclusions 18

3 From Vision to Action 20
- 3.1 Dual Route Models 23
- 3.2 Affordances 25
  - 3.2.1 On the Nature of Affordances 26
  - 3.2.2 Experimenting with Affordances 29
  - 3.2.3 Micro-Affordances 33

4 Models for Visual Attention and Action 35
- 4.1 Conclusion 44

II The Selective Attention for Action Model 46

5 Study 1: Modelling Human-like Grasp Postures 47
- 5.1 The Selective Attention for Action Model 47
  - 5.1.1 Mathematical Details of the Model 52
- 5.2 Simulating Grasps 58
  - 5.2.1 Method 58
  - 5.2.2 Results and Discussion 61
- 5.3 Where Do Humans Grasp Objects? – An Experiment 72
  - 5.3.1 Method 72
  - 5.3.2 Results and Discussion 77
- 5.4 Comparison of Simulated Grasps and Experimental Results 85

5.4.1 A Similarity Metric for Grasps 85
5.4.2 Results and Disussion 87
5.5 Summary 91

6 Study 2: Grasp Classification and Two-Object Simulations 94
6.1 Extending SAAM: The Grasp Template Network 95
6.1.1 Bottom-Up Path 96
6.1.2 Top-Down Path 97
6.1.3 Mathematical Details 98
6.2 Simulations 101
6.2.1 Study 2a: Classification Abilities of the Model 102
6.2.2 Study 2b: Two Object Scenes 107
6.2.3 Study 2c: Modulating Grasp Selection 112
6.3 Summary 123

7 Study 3: Affordances and Action Intentions 127
7.1 Study 3a: Affordances 128
7.2 Study 3b: Action Intentions 132
7.3 Summary 136

III General Discussion 138

8 General Discussion 139
8.1 Summary of the Model 139
8.1.1 Findings in Study 1: Evaluation of SAAM's Grasps 140
8.1.2 Findings in Study 2: Grasp Classification and Selective Attention 143
8.1.3 Findings in Study 3: Affordances and Action Intentions . . . 146
8.2 The Role of Grasping in SAAM 147
8.3 SAAM in the Context of Visual Attention 148
8.3.1 Visual Feature Extraction 149
8.3.2 SAAM, SAIM and Biased Competition 150
8.3.3 The Top-Down Path 154
8.4 Affordances 155
8.4.1 A Hierarchy of Affordances 156
8.4.2 Invariants as Body-Related Relative Features 158
8.4.3 Micro-Affordances 159
8.5 Comparison with Other Models 160

9 Conclusions and Outlook 162
9.1 Outlook 162
9.1.1 Extensions of the Model and Additional Simulations 162
9.1.2 Further Experimental Verification 165
9.2 Conclusions 171

IV Appendices 174

A Additional Information about the Experiment in Study 1 175

A.1 Object Dimensions . . . 175
A.2 Instructions used in the Experiment . . . 176
A.3 Reaction Times for Button Release . . . 177
A.4 Reaction Times for Object Lift . . . 185

B Simulation Parameters 193

B.1 Study 1 . . . 193
B.2 Study 2a . . . 194
B.3 Study 2b . . . 195
B.4 Study 2c . . . 196
B.5 Study 3a . . . 198
B.6 Study 3b . . . 199

References 201

List of Figures

2.1 Hand postures for grasp types (1) 9
2.2 Hand postures for grasp types (2) 12
2.3 Opposition spaces 13

3.1 The dual route model 24
3.2 Biomechanical model of stair-climbing 30

4.1 FARS model for grasp-execution 37
4.2 Cisek's model for reach-planning 40
4.3 Ward's model for selective action 42

5.1 Overall structure of SAAM 48
5.2 Edge-filtered visual input 49
5.3 Excitatory connections in SAAM 51
5.4 Weight matrices used in Study 1 60
5.5 Complete weight map 61
5.6 3D-plots of finger map activations 62
5.7 Finger positions in Study 1 (1) 63
5.8 Finger positions in Study 1 (2) 64
5.9 Grasp of a very long bar 67
5.10 Finger map activation in Study 1 70
5.11 Hand length measure 72
5.12 Experimental set-up for Study 1 73
5.13 Object used in Study 1 74
5.14 Initial hand position in Study 1 77
5.15 Analysis of experimental data in Study 1 78
5.16 Grasps observed in the experiment (1) 79
5.17 Grasps observed in the experiment (2) 80
5.18 Proportion of horizontal grasps to object size 81
5.19 Comparison of simulation and experiment (1) 89
5.20 Comparison of simulation and experiment (2) 90
5.21 Probability of horizontal grasps 91

6.1 Bottom-up path of the grasp template network 96
6.2 Top-down path of the grasp template network 98
6.3 Template maps 104
6.4 Activation of template units in Study 2a 105

6.5 Finger positions in Study 2b 109
6.6 Finger map activation in Study 2b 110
6.7 Template activation in Study 2b 111
6.8 Simulation results of Study 2c (1) 115
6.9 Simulation results of Study 2c (2) 117
6.10 Simulation results of Study 2c (3) 119
6.11 Simulation results of Study 2c (4) 121
6.12 Simulation results of Study 2c (5) 122

7.1 Template maps for Study 3 129
7.2 Simulation results of Study 3a 130
7.3 Simulation results of Study 3b (1) 134
7.4 Simulation results of Study 3b (2) 135

8.1 Overall structure of SAIM 150

9.1 Bi-manual version of SAAM 164
9.2 Objects with hidden affordances 166

List of Tables

5.1 Stimuli for Study 1 58
5.2 Duration of simulations in Study 1 71
5.3 Width-to-height ratios of stimuli 82
5.4 Analysis of grasp orientation 82

6.1 Duration of simulations in Study 2a 106
6.2 Comparison of simulation duration 112

Part I

Introduction

1 Motivation

Complex interactions with objects in our environment play an important role in our daily lives. Even a seemingly simple task like driving a nail into a wall involves a number of steps: First, we have to find the hammer and a nail amongst all the other tools and materials in our toolbox, then we need to grasp both correctly so that we can hammer effectively without injuring ourselves and, finally, we can drive the nail into the wall. Doing this successfully does not only depend on our intention to drive the nail into the wall but much more crucially on knowledge about the actions which the objects in our toolbox afford. Moreover, we need to know precisely how to perform the chosen action with a specific tool for effective interaction.

Traditionally, it has been assumed that objects in the environment are identified in an object recognition process in the visual system and that actions are linked to categorical representations of objects. Hence, in the example above the hammer would first be identified as a hammer and then a hammer action which is linked to the 'hammer' object category could be triggered. In the last decades, though, research in action perception produced much evidence supporting a different theory (see Chapter 3). This theory proposes that visual perception for action takes place independently of normal object recognition and that two different pathways exist in the human brain for each form of perception (Ungerleider & Mishkin, 1982; Riddoch, Humphreys, & Price, 1989; Milner & Goodale, 1995). Furthermore, the route from vision to action is often connected to Gibson's affordances (J. J. Gibson, 1966, 1979). In the affordance concept opportunities for actions are assumed be derived directly from invariants in the visual information. Gibson argued, that by successively combining invariants, a hierarchy of affordances accrues which expresses increasingly complexer actions.

In my research I set out to develop a connectionist model of the direct route from vision to action based on Gibsonian affordances. In this model affordances of objects are expressed in the grasp postures used to handle the objects. The model discovers affordances of objects by generating such grasp postures based on the information available in the visual field. It will be demonstrated that this is a viable approach for expressing affordances and, furthermore, an implementation of the hierarchical structure of affordances postulated by Gibson. The hierarchical structure is a crucial property of the model and – to my best knowledge – a novel property in computational models of affordances. To verify that the grasp postures produced by the model resemble human grasp postures, data of grasps used by humans is collected in an experiment and compared to the grasps created by the model.

An important aspect in the scenario described earlier is that it requires the involvement of selective attention, e. g. ignoring the irrelevant contents of the toolbox and identifying the action-relevant item. Hence, this aspect also forms a crucial part of the exploration of the model presented here. It will be shown that attention for action can be guided by affordances. Moreover, I will demonstrate that action intentions can influence the discovery of affordances and, thus, yielding a top-down control of affordance-guided visual attention for action.

The aim of the model presented in this thesis is – in a nutshell – to derive complex action categories for object use from simple action parameters extracted from visual information only and to thereby model effects of selective attention for action and action intentions.

Organisation of the thesis In Chapter 2 research on grasping is reviewed. Grasping serves an important role in our interaction with objects and forms the foundation of my model. The review on grasping is intended to provide a better understanding of different grasp postures and methods to represent grasps. Chapter 3 first summarises experimental work on the perception of action which led to the development of dual route models for visual perception (Sec. 3.1). In the second part of the chapter affordances as a theory for the processes in the direct route for action are discussed (Sec. 3.2). Models aiming at grasping, affordances or perception for action are reviewed in Chapter 4. The Selective Attention for Action Model (SAAM) is presented in Chapter 5. Alongside the model, a first assessment of the model results is presented and compared with data from a behavioural study in humans. In Chapter 6 the model is extended to incorporate classification abilities and action intentions. Using simulations of 'real-life' objects it is demonstrated in Chapter 7 how complex affordances are modelled in SAAM. In Chapter 8 the findings of the studies are discussed. The thesis concludes with an outlook in Chapter 9. The two appendices contain additional information about the experiment in Study 1 (App. A) and detailed parameter specifications for all simulations (App. B).

2 Grasping

Interactions with objects can occur in infinitely different ways. A ball, for example, can be kicked or thrown, a chair can be sat on or stood upon, and a jug can be filled or emptied, to name but a few. However, a common way of interacting with an object is to grasp it and then use it in a subsequent action. In my research I concentrate on this type of object use because it has an interesting property: The way in which an object is grasped tells one something about the subsequent action that is performed with this object (Napier, 1956). This permits to encode affordances of objects by describing how an object is grasped. Furthermore, not only opportunities for action can be expressed in this way but also action intentions of agents. By preparing to perform a specific grasp, an agent expresses its intention of executing a specific subsequent action.

According to the Oxford Dictionary 'to grasp' means "to seize and hold firmly" (Soanes & Hawker, 2008). The problem with this definition is, however, that it comprises a myriad of very different grasps. Is the object grasped with only one or with both hands? Is it a grasp to get control of an object or to hold on to something? Is it at all an object that is grasped? – These questions highlight just some of the variations in grasping. To understand how different types of grasps can be distinguished amongst these variations, literature about grasp taxonomies is reviewed in this chapter. Using these classifications and the requirements for a computational model of grasping, it will be defined how 'grasping' is understood in this thesis and ways to encode grasps in my computational model will be explored.

2.1 Grasping in a Computational Model

A computational model of human grasping needs a way of describing and encoding grasps. Furthermore, a model which focuses on the effects of opportunities for action and action intentions on visual attention requires a representation of grasps which has certain properties: The representation must be able to represent actions while at the same time being related to the visual field in order for action to function as a selective factor in visual attention. It is therefore beneficial if the representation of grasps is spatially related to the information in the visual field. In addition to being linked to the visual field, grasps must be encodable using visually available information only. The reason for this is that the grasps need to be encoded *before* they are actually executed. Thus, only visual information is available at that point. The process which generates the encoded grasps furthermore needs to be simple in terms of computational complexity so that it can be performed continuously. This requirement is founded in the experiments summarised in Chapter 3 which showed that affordances have an effect on participants' behaviour even when no action is required in a task. If the extraction of affordances were a difficult – i. e. demanding – task, such effects should not be observed. Since grasps in my model do not only encode hand postures but also affordances and action intentions, their encoding must be detailed enough to include information about the posture that the hand will adopt in a grasp in addition to the grasp location. Finally, the chosen grasp encoding must be implementable in a neuronal network in a biologically plausible way.

To gain a deeper understanding of human grasp postures and their relation to actions, I will review grasp classifications found in the literature in the following section.

2.2 Classifying and Describing Grasps

Grasp classifications and descriptions have been developed in such different disciplines as medicine, psychology and engineering for a long time. MacKenzie and Iberall (1994) provided an extensive overview of grasp classifications found in the literature throughout the 20th century. Their overview will be provide the basis for the summary of grasp classifications in this section. I will, however, put an emphasis on the applicability of the presented grasp descriptions for encoding affordances and action intentions. To my best knowledge, MacKenzie and Iberall's summary – even though it is already 16 years old – still represents the current state of research in grasp classification systems. More recent developments in research on grasping concentrated mainly on solving the problem of grasping in robots and relied on these earlier results for classifications of human grasps.

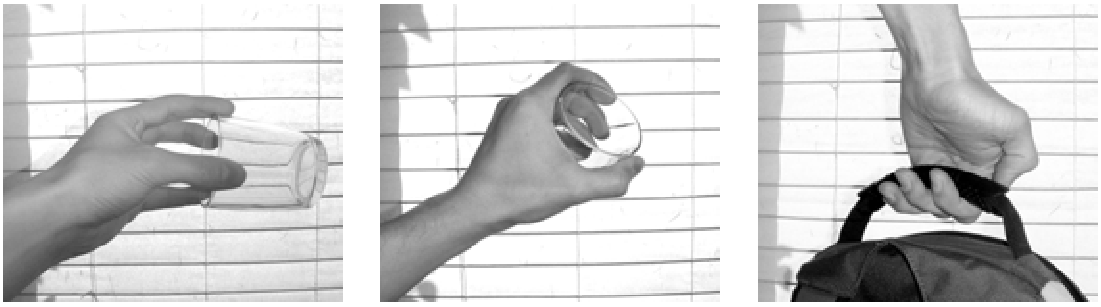
Grasp classifications and descriptions can be divided into two main groups. The first group aims at identifying grasp postures humans are capable of performing and then subdividing this plethora of different grasps into classes of physiologically or functionally equivalent grasps. The second group, on the other hand, seeks to provide quantitative measures to encode individual grasps at a certain level of detail. I will first review grasp taxonomies highlighting how functional features of grasps influence the structure of the classifications. This demonstrates the feasibility of conveying information about subsequent actions with different grasp postures. Additionally, it establishes a terminology for different types of grasps. I will then present quantitative encodings for grasps in order to find an encoding which combines the expressiveness of the qualitative taxonomies with the implementation requirements of my computational model.

2.2.1 Grasp Taxonomies

Early grasp classifications from the first half of the 20th century were mostly motivated by the search for schemes to describe injuries and the resulting losses of grasping capabilities in a standardised way. An important publication from this time is Slocum and Pratt's article in which they introduced the terms 'grasp', 'pinch' and 'hook' to distinguish the main classes of prehensile contact which are still used today (Slocum & Pratt, 1944).[1] They defined "*[g]rasp* [...] as the combined action of the fingers against the opposed thumb and the palm of the hand" and "*[p]inch* [...] as the apposition of the tip of the thumb against the pads of the opposing fingers" (Slocum & Pratt, 1944, p. 535, italics in the original; see Fig. 2.1a and 2.1b). The feature which separates these two classes is the utilisation of the palm in the 'grasp' posture. This involvement of the palm in a 'grasp' facilitates a safe and powerful contact between hand and object. Pinches lack this force component but their reliance on the fingers and the thumb only allows for much more precise and dexterous prehension instead. Interestingly, Slocum and Pratt described the role of the palm in grasps in an almost Gibsonian way: "The breadth of the palm affords a wide base for stability and balance" (Slocum & Pratt, 1944, p. 535). However, it is worth noting that, apart from this statement, Slocum and Pratt defined their prehension classes solely in anatomical and physiological terms.

This is also true for their definition of the 'hook'. In a 'hook', the hand adopts a posture with the "fingers [...] flexed, so that their pads lie parallel and slightly away from the palm" (Slocum & Pratt, 1944, p. 536). This posture can most easily

[1] Slocum and Pratt actually described four types of prehension. Additionally to the three types discussed here, they defined a 'multiple pinch' which is same as the pinch only that instead of the finger pads the finger tips are used to hold the object. The reason for this differentiation is the need for an intact metacarpal arch in order to perform a multiple pinch (Slocum & Pratt, 1944, p. 536). Since this distinction has not been picked up by later authors (probably because it is only relevant in the light of injures and amputations), I also do not differentiate between pinches with the finger tips and pinches with the finger pads but simply consider both as pinches here.



(a) Pinch or precision-grasp (b) Grasp or power-grasp (c) Hook-grasp

Figure 2.1: Hand postures illustrating the three basic grasp types.

be understood if one imagines carrying a (heavy) suitcase with one hand on the handle (see Fig. 2.1c). The ‘hook’ differs from the other two classes fundamentally because control of the object is established only by opposing an external force affecting the object (e. g. gravity when carrying a suitcase) whereas in a ‘grasp’ or ‘pinch’ the object is clamped in a system of forces established between the fingers and the thumb or palm. Because of the absence of an internal system of forces in the ‘hook’ posture, it requires considerably less strength and dexterity in the hand. However, this comes at a great loss in control over the object. From Slocum and Pratt’s points of view as surgeons, though, the low physiological demands of the ‘hook’ posture instead of these functional differences were the most important reason for including it in their taxonomy.

Emphasising Functional Features

It was in 1956 that Napier considered prehension from a new perspective. At first sight his classification appears to be very similar to Slocum and Pratt’s. He called Slocum and Pratt’s ‘grasp’ a ‘power-grip’ and defined it as “[...] a clamp formed by the partly flexed fingers and the palm, counter pressure being applied by the thumb lying more or less in the plane of the palm”. The ‘pinch’ was relabelled as a ‘precision-grip’ meaning that “[t]he object [is] pinched between the flexor aspects of the fingers and the opposing thumb" (Napier, 1956). The 'hook' was classified as a non-prehensile movement which Napier did not consider in his article. While this definition of the classes describes almost the same prehensile hand postures as Slocum and Pratt's classification, the change of names was not merely a stylistic difference but showed a new understanding of prehension: by choosing terms which describe functional features of prehensile actions Napier focused his classification on the question why certain prehensile actions are chosen instead of only asking which prehensile postures the human hand can adopt.

Consequently, Napier looked at factors which influence the choice of a specific type of prehension. He first noted that the shapes and sizes of the objects which are handled during activities are among these factors (Napier, 1956, p. 902). Yet, he also pointed out that the influence of the subsequent action (i. e. 'intended activity' in his words) is most important: "While it is fully recognised that the physical form of the object may in certain conditions influence the type of prehension employed it is clear that it is in the nature of the intended activity that finally influences the pattern of the grip" (Napier, 1956, p. 906). Napier then argued that different actions require different levels of power and precision; by varying the finger positions – especially the orientation of the thumb – the amount of force and dexterity provided by a grip can be adjusted to meet the requirements of an action.

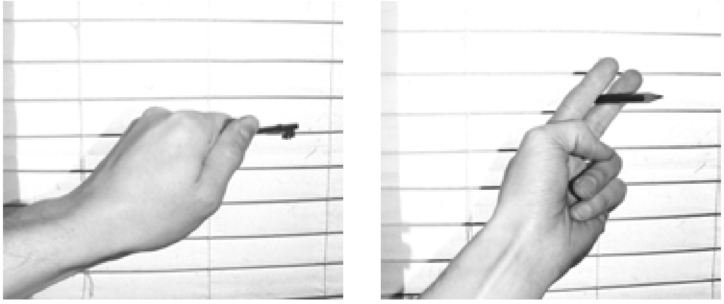
MacKenzie and Iberall argued that Napier's insight was to recognize this relation between requirements for power and precision in actions and the ability of the hand to exercise power and precision grips (MacKenzie & Iberall, 1994, p. 23). Napier's observation bears a noteworthy similarity to Gibson's affordances which, as I will describe in more detail in Sec. 3.2, link abilities of agents with features of objects. In Napier's view, however, interaction with the environment *requires* abilities while in Gibson's notion the environment *offers* opportunities for actions.

Napier's conclusions on the functional aspects of grasping were confirmed by an observational study by Cutkosky and Wright (1986) in which the researchers presented a taxonomy of grasps from a workshop environment. The authors analysed grasps used in workshops and classified them by their power and dexterity properties. The resulting classification showed the existence of a smooth transition from powerful grasps with low levels of dexterity to very dexterous but not powerful grasps. Cutkosky and Wright suggested that this smooth transition comes with the adaptation of generic grasps postures for specific objects. While (generic) task-directed grasps like the 'precision'- or 'power-grasp' possess either power or dexterity properties, the adaptation of a task-directed grasp posture for a concrete object results in the combination of both properties in the final grasp. This reflects Napier's theory that the power and precision properties of grasps depend on the object being grasped. For example, Cutkosky and Wright subdivided Napier's precision grasp class depending on the positions of the fingers. A 'circular precision grip' in which the fingers and the thumb are placed on the outline of a circle allows, for instance, to exert more strength then an 'opposed thumb precision grip' in which the fingers are placed on a straight line opposed by the thumb. Cutkosky and Wright's data also showed that different objects were indeed handled with different grasps as I assume in my approach of encoding action intentions for different objects by describing different grasp postures.

It is notable that Cutkosky and Wright's taxonomy can be linked (informally) to the idea of critical points (see Sec. 3.2.2). Each grasp class in the taxonomy is suitable for a range of object shapes. However, if shapes differ to much – which can be interpreted as a form of critical point – a transition to another grasp class takes place.

Virtual Fingers and Opposition Spaces

So far, we only considered three main classes of grasps. However, it is obvious that these three classes do not cover all types of prehension humans are capable of. Cutkosky and Wright, e. g., also observed a grasp posture called 'lateral pinch' in



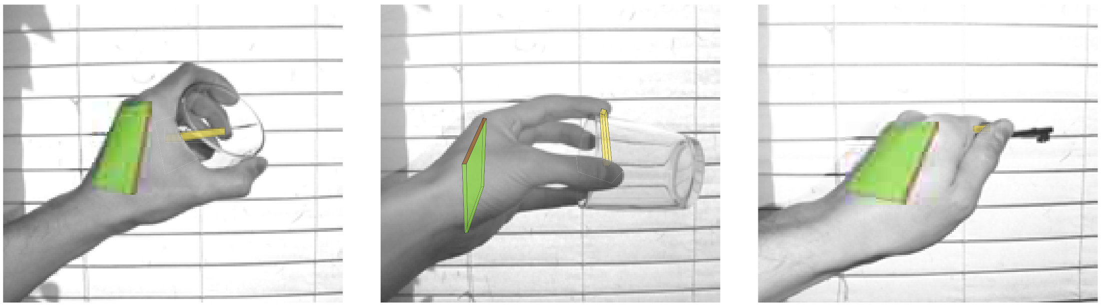
(a) 'Lateral pinch' (b) 'Adduction grip'

Figure 2.2: Hand postures for more specialised grasp types.

which the thumb presses against the side of the index finger, that is used, for instance, when turning a key in a lock (see Fig. 2.2a). Another common hand posture is the 'adduction grip' which is used to hold small objects (most notably cigarettes) between the lateral surfaces of two fingers (Kamakura, Matsuo, Ishii, Mitsuboshi, & Miura, 1980, see Fig. 2.2b). These two types of grasps do not fall into any of the classes defined by the above taxonomies. Furthermore, all the classifications defined their grasp classes based on anatomical descriptions even when the taxonomy was actually aimed at describing the functional properties of grasps (e. g., Napier's classification).

In order to overcome these limitations, Iberall, Bingham, and Arbib (1986) developed the 'opposition space' taxonomy. It builds on the notion of 'virtual fingers' devised by Arbib, Iberall, and Lyons (1985) and uses the direction of the force vector between two virtual fingers to define an 'opposition space' which then classifies the grasp.

The idea of 'virtual fingers' is derived from the observation that the fingers of the hand are often used together as if they were one larger finger, or that the palm acts like a large and powerful finger. Arbib et al. named such groups of fingers or such a utilisation of the palm 'virtual fingers' and defined three types (labelled VF1 to VF3) of them each with a specific functional role. VF1 and VF2 are two virtual fingers which always oppose each other like, for instance, the thumb and the fingers in a



(a) ‘Palm opposition’ (b) ‘Pad opposition’ (c) ‘Side opposition’

Figure 2.3: Hand postures with opposition spaces overlaid. The green plane shows the orientation of the palm, the yellow bar the orientation of the opposition vector between VF1 and VF2.

‘pinch’. VF3 is defined to counteract to external forces (like the ‘hook’). While VF1 and VF2 are always used together VF3 can be used on its own or as part of a grasp involving VF1 and VF2 as well.

Virtual fingers combine parts of the hand which perform a common functional task in a grasp into one virtual effector. Grasps can then be described using virtual effectors instead of directly referring to the anatomy of the hand. In order to describe how the virtual fingers are used in a grasp, Iberall et al. (1986) developed the concept of ‘opposition spaces’. An ‘opposition space’ is defined as the orientation of the force vector between VF1 and VF2 in relation to the surface of the palm. This definition leads to three different types of opposition (see Fig. 2.3): In ‘pad opposition’ the force vector between VF1 and VF2 is parallel to the surface of the palm and perpendicular to the width of the hand. In ‘side opposition’ the force vector is also parallel to the palm but directed parallel to the width of the hand (MacKenzie and Iberall call this “transverse”). Finally, in ‘palm opposition’ the force vector is normal to the surface of the palm. These types of opposition translate to power-grasps, pinches, lateral pinches and adduction grips as follows: If VF1 are the fingers and VF2 is the palm and they are in palm opposition, then this resembles a power-grasp. If the thumb is acting as VF1 and the fingers as VF2 with pad opposition, this results in a pinch, and if the opposition is changed to side opposition, it becomes a lateral pinch. An adduction grip can be described by the index finger as VF1 and the middle finger as VF2 with side opposition.

The comparison of the anatomically defined grasp classes with virtual fingers and opposition spaces showed that it is possible to describe all types of grasps within a single framework using these two concepts. Compared to the other taxonomies, virtual fingers and opposition spaces have the advantage of constructing different grasp classes from a small set of components instead of describing each class separately. However, the major improvement over earlier classifications is the abstraction from the actual anatomy of the hand through the introduction of virtual fingers. By focusing on the functional components of a grasp ('virtual fingers') and their interaction ('opposition spaces') it became possible to describe grasps based on function. It is worth noting, however, that properties like strength and dexterity of a grasp are not defined explicitly any more but must be derived from the type of opposition space and the mapping of the physical fingers to the virtual fingers. Nevertheless, the two concepts provide valuable insights in a compact systematic and non-anatomical description of grasps.

2.2.2 Quantitative Grasp Encodings

While qualitative classifications of grasps allow to distinguish different types of grasps, they cannot describe individual grasp postures and, in particular, where the grasps are directed to in the environment of an agent. These two points, however, are dealt with by many quantitative grasp encodings. Three encodings will be considered here: First, MacKenzie and Iberall's (1994) quantitative description of 'virtual fingers' and 'opposition spaces' will be briefly explained, followed by Jeannerod's (1981), and Smeets and Brenner's (1999) descriptions.

The systematic classification approach of 'virtual fingers' and 'opposition spaces' makes this classification a good candidate for an extension with a quantitative grasp encoding. MacKenzie and Iberall (1994) suggested such an extension. Their encoding constituted of the types of opposition being used in a grasp and the mapping between physical and virtual fingers. The configuration of the virtual fingers was described in great detail by defining length, width and orientation of each virtual finger as well as the orientations of the applied forces (the opposition spaces; Iberall, Torras, & MacKenzie, 1990). Additionally, the amount of force and sensory information available to and from the virtual fingers was given. This quantitative encoding allowed to describe the static postures of the hand and, moreover, the system of forces spanned by the fingers in a very detailed way. For example, a grasp to hold a pen could be described as pad opposition with the thumb as VF1 and the index and middle finger as VF2. The width of VF1 would be one finger (the thumb), its length would be defined as the distance from the first carpometacarpal joint to the tip of the thumb, and its orientation would be perpendicular to the surface of the palm. VF2 would have a width of two fingers (index and middle finger) and its length would be measured from the second and third carpometacarpal joints to the tips of the index and middle finger; its orientation would also be perpendicular to the surface of the palm. Because the fingers are in pad opposition, the forces exerted by the fingers would be parallel to the surface of the palm and opposing each other. The example shows that MacKenzie and Iberall's encoding was based on a description of a system of forces similar to those used in mechanical engineering (see Mason, 2001, e. g.). This and the large number of properties and variables made the description of grasps very complex. The encoding also only described grasps in relation to the hand but not where they are located in the environment of the agent. Thus, it is not possible to link grasp descriptions with locations in the visual field which is one of the requirements of a grasp description for my model.

In contrast to the qualitative classifications and to MacKenzie and Iberall's quantitative description, the Jeannerod's and Smeets and Brenner's descriptions were not developed with the aim of describing grasp postures but the hand and finger positions during reach-to-grasp movements. Hence, these descriptions included the position of the hand in the encoding. However, an accurate description of the final grasp posture is only of minor importance for these encodings since their main application lies in research of hand movements and hand apertures before adopting the actual grasp posture. In particular, Jeannerod's classic description of grasps used only the opening of the hand (grasp width) to describe the grasp posture. This makes this encoding unsuitable to describe grasp postures in my model because it is not possible to express complex affordances and action intentions with it.

Smeets and Brenner's encoding, on the other hand, allows a more detailed description of the grasp posture since it encodes the position of individual fingers. This allows for a more accurate description of precision grasp postures than Jeannerod's description but has, of course, limitations when describing power grasps since the palm is not included in the description. Furthermore, Smeets and Brenner only encoded the positions of the thumb and the index finger which further reduced the expressiveness of the encoding. Nonetheless the idea of describing hand postures with the position of the individual fingers is a simple and elegant approach to the problem of grasp encoding. The original encoding from Smeets and Brenner can easily be extended to include all fingers instead of only the thumb and the index finger. It would even be possible to include other parts of the hand as well; for instance, the position of the centre of the palm could be encoded in the same fashion.

2.3 A Definition of Grasping

The review of the grasp taxonomies showed some of the various different grasp postures humans are capable of performing, many of which considerably differ in the way they are executed. For example, grasps with VF3 are very different to grasps with VF1 and VF2. For the purpose of my research, it is necessary to define ‘grasping’ as it is understood in my thesis more precisely.

All grasp classifications reviewed above showed that grasping is commonly considered an activity which involves only one hand[2]. Hence, my definition is also restricted to single-handed grasps only. Secondly, grasping in the context of this work is used as a means to describe action intentions and affordances of objects. Against this background, my definition can be further restricted to cover only those grasps which are directed towards objects with the aim of using the objects in a subsequent action. Since grasps are performed with the intention of establishing physical control of an object in order to use it, the grasp types covered by my definition can also be constrained by excluding grasps which only use VF3 (i. e. ‘hook’ grasps) because such grasps provide only limited control over the grasped objects. Thus, not many affordances and action intentions can be expressed with them. On the other hand, their execution is quite distinct from grasps performed with VF1 and VF2 which makes it difficult to handle both variants within one model. Moreover, VF3 grasps are not always considered to be a prehensile action (e. g., Napier excludes them). For these reasons only grasps using VF1 and VF2 will be included in my definition.

Based on these assumptions and the resulting restrictions, the term ‘grasping’ as it is used in this thesis is defined as follows:

> *To use one hand to establish physical control of an object so that it can be used in a subsequent action. At least two surfaces of the hand must exert opposing forces in order to hold the object.*

This definition covers all grasp types introduced by the grasp taxonomies presented above except for the hook grasp. It also only defines the minimum requirements for a grasp; in reality there are, however, a number of other properties which can often be

[2] There are publications on two-handed grasping but they form a separate body of literature.

observed during grasping and which improve the stability of a grasp. These properties are the placement of the grasp around the seized object's centre of mass and the placement of the fingers so that object movement in all directions is physically blocked by a finger (this is termed 'form-closure', see Bicchi, 1995 for details). However, since these characteristics may – depending on the subsequent action – not always be desired in a grasp, they are not included in the definition but I will still consider them in the discussion of my model.

2.4 Conclusions

I turned my attention towards grasping in this chapter with the aim of finding a method to encode affordances and action intentions. The review of grasp classifications showed that grasps can express different functional properties of actions which are performed with an object once it has been grasped. Napier in particular was aware of this relation between grasps and subsequent activities. He even described his observations in an almost Gibsonian way. The grasp classifications strongly support the idea of using grasps as a means of encoding affordances and action intentions.

The grasp taxonomies also showed that three main classes of grasps are consistently distinguished which mark various basic functional properties of grasps. These classes are the 'hook', the 'power-grasp' and the 'precision-grasp'. Some more advanced classifications also describe additional types of grasps but these three are common to all presented taxonomies. They emphasise major functional differences like the opposition of the grasp to an external force, or whether a grasp mainly provides strength or dexterity.

While qualitative grasp taxonomies allow to distinguish functionally different types of grasps, these grasp descriptions are not linked to the visual input space which, as I argued, is necessary for my model. Quantitative descriptions are often capable of providing this connection but they also often lack the abilities to describe different types of grasps (especially Jeannerod's model). Smeets and Brenner, however, developed an encoding of grasps which is able to describe grasps in relation to the environment of the agent while keeping a level of detail which allows to describe different types of grasps. When the types of grasps, which can be described in Smeets and Brenner's model, are compared with the three basic grasp classes it becomes clear that it only supports pinch-type grasps. However, these grasps show a high variability (e. g. Cutkosky, 1989) depending on the subsequent action, which makes them suitable to describe a subset of human activities.

This review on grasping provided valuable insights into different types of grasps and their relation to affordances and action intentions. The identification of different grasp classes will help to understand the abilities and limitations my computational model might have. Additionally, Smeets and Brenner's model was identified as a good candidate for an encoding to be used in my model.

3 From Vision to Action

Research in the last decades showed that vision and action systems in the brain are not separate but tightly linked together. Many studies provide diverse evidence for a close interaction between vision and action. For example, Craighero, Fadiga, Umilta, and Rizzolatti (1996) assumed that an object, when seen, automatically evokes motor representations describing actions involving the object. To test this hypothesis the authors used a priming paradigm with a prime indicating the orientation of an object which the participants were asked to grasp. The study showed that congruent primes improved reactions times in the subsequent grasping response. The authors interpreted these data as evidence that visual stimuli can evoke motor representations for objects and that an automatic visuomotor transformation exists. However, the reaction time effect was only observed when the prime was shown before the grasping action was initiated. This suggests that grasps are not planned on-line during the reach-to-grasp movement but that at least the main characteristics of a grasp (like the orientation of the hand) are planned before initiating the action.

In a series of experiments Ellis and Tucker investigated effects of different physical object properties on actions. Effects of left-right orientation of an object and wrist rotation were analysed in Tucker and Ellis (1998). In the experiments participants were asked to decide whether an object was upright or inverted. In one experiment the participants then had to respond with a left- or right-hand key press and in another experiment by rotating their hand left or right. Unknowingly to the participants the objects either had their handles oriented to the left or to the right or they were rotated. The results of the experiments showed faster reaction times when orientation of the handle or the object rotation and the required response were congruent. Tucker and Ellis concluded that these data supported the idea of an automatic priming of visuomotor processes.

In another study, Tucker and Ellis investigated the influence of priming objects compatible with either a power- or a precision grasp on an unrelated response made with a power or precision grasp (Tucker & Ellis, 2001). Again, it was found that reaction times in the congruent condition (e. g., a power grasp response towards an object graspable with a power grasp) were faster than the reaction times in the incongruent condition. These findings were confirmed in an fMRI study by Grèzes, Tucker, Armony, Ellis, and Passingham (2003). The authors found increased activity within the premotor cortex in incongruent trials which they concluded resulted from a competition between the different grasps required for the object and the response. Tucker and Ellis (2004) showed that active object representations were enough to trigger affordance compatibility effects.

In a study using positron emission tomography (PET) Grèzes and Decety (2002) explored neural correlates which may be linked to motor representations arising during visual perception of objects. It was found that cortical areas commonly linked to motor representations were active during visual perception irrespective of the task the participants were asked to perform. Grèzes and Decety interpreted this as evidence for automatic activation of potential actions. Further evidence was found in a study by Grafton, Fadiga, Arbib, and Rizzolatti (1997). In this study PET was used to investigate activation of premotor areas during presentation of real objects. Consistent with Grèzes and Decety's results, the outcome of Grafton et al.'s study showed an activation of premotor areas irrespective of subsequent tasks.

Interestingly, recent experimental evidence does not only indicate that affordances prime actions but also that visual attention is directed towards action relevant locations. Using event-related potentials (ERPs) Handy, Grafton, Shroff, Ketay, and Gazzaniga (2003) observed that tools presented in the right (participants were right-handed) and lower visual fields systematically attracted spatial attention. These results were confirmed using event-related fMRI. In a later study Handy and Tipper (2007) replicated their study with a cohort of left-handed participants. Similar evidence for attentional effects of affordances was found in a study with patients with visual extinction by di Pellegrino, Rafal, and Tipper (2005). Visual extinction is considered to be an attentional deficit in which patients, when confronted with several objects, fail to report the objects on the left side of body space. In contrast, when presented with only one object, patients can respond to the object irrespective of its location. di Pellegrino et al. demonstrated that the attentional deficit could be alleviated when the handle of a cup points towards the left. The authors suggested that an affordance is automatically extracted by visual system and supports attentional selection.

Bekkering and Neggers (2002) investigated effects of action intentions in a visual search task. Participants were instructed to either perfom a grasping or a pointing task towards objects with different orientations and colours. Using eye-tracking it was found that in the grasping condition participants made fewer saccades towards objects with the wrong orientation than in the pointing condition. The number of saccades towards objects with the wrong colour was the same in both conditions. The authors suggested that these results showed that action intentions can influence the perception of action-relevant features of objects. They concluded that this can be best explained as a selection-for-action process.

The idea that action plays a special role in visual search is also supported by Humphreys and Riddoch (2001). They reported a patient who was not able to find objects when their name or a description of their form was given but when it was described what the objects were used for he was able to find the objects. Similar separations between semantic knowledge about objects and action knowledge about objects were found in other experiments as well which led to the development of dual routes models.

3.1 Dual Route Models

Building on earlier work by Ungerleider and Mishkin (1982), Milner and Goodale (1995) suggested a now classical dual route model for vision and action. They postulated the existence of at least two pathways in the human brain along which visual perception is mediated. These pathways are usually referred to by their localisation in the brain as the ‘dorsal route’ and the ‘ventral route’ or by their function as the ‘where’- and the ‘what’-system. The latter names show that the two pathways have been linked with different aspects of visual perception. The dorsal route is thought to be responsible for online visuomotor control of prehensile actions (i. e. reaching and grasping) – hence it is dubbed the ‘where’-system. The ventral stream on the other hand is assumed to be required for object recognition and therefore called ‘what’-system.

Milner and Goodale’s dual-route model only incorporated simple reaching and grasping actions. Riddoch et al. (1989), however, postulated an extended dual route model which included more complex actions as well. The model distinguished between a route via structural knowledge and a route via semantic knowledge about objects. In this model visual input is linked through structural knowledge with action patterns. Auditory input on the other hand is linked through semantic knowledge with action patterns. Finally, structural knowledge and semantic knowledge are linked with each other (see Fig. 3.1). In contrast to the dorsal/ventral stream distinction the action path in this model is not involved in online visuomotor control but activates complex action patterns based on combining structural knowledge with the visual input.

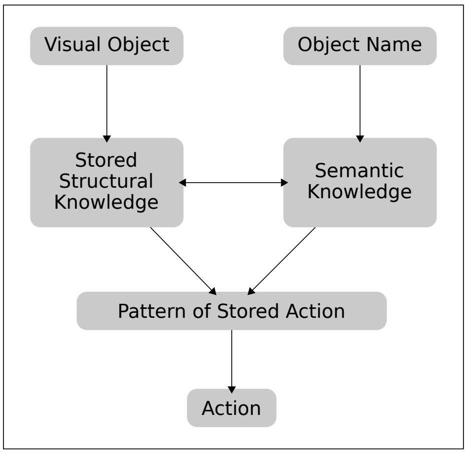


Figure 3.1: The dual route model as it has been suggested by Humphreys and Riddoch (2003).

Evidence for Riddoch et al.'s model stems from experiments with patients with optic aphasia, semantic dementia and visual apraxia. Optic aphasia refers to a condition in which patients are impaired at naming visually presented objects; semantic dementia is linked to a loss of semantic knowledge about objects; visual apraxia, in contrast, describes a condition where perception of semantic object knowledge persists but the performance of actions in response to visually presented objects is impaired. These three disorders can be linked to different parts of the model. Optic aphasia is considered to be caused by a lesion in the link between structural and conceptual knowledge. Semantic dementia has been linked with degeneration of semantic knowledge and visual apraxia with a loss of structural knowledge (see Humphreys & Riddoch, 2003 for a review).

Studies with normal subjects provided further evidence for a direct route from vision to action. Rumiati and Humphreys (1998), for instance, analysed errors which participants made when they had to gesture object use or name a visually presented object under deadline conditions. The authors found that during gesturing participants made more errors due to visual similarities (e. g., making a shaving gesture when a hammer is shown) and less errors due to semantic relations of response and object (e. g., responding with a sawing gesture to a hammer) compared to when they named the object. It was also found that auditory presentation of the object names did not produce visual similarity errors.

3.2 Affordances

The direct pathway from vision to action is often linked to Gibson's affordance theory (J. J. Gibson, 1966, 1979). Affordances can, for a start, be defined as *opportunities for action* which the environment offers to an agent. This definition can be illustrated with a cup being present in the environment so that an agent can drink from it. In this example the cup *affords* a drinking-action for the agent; hence, it can be said that a *drink affordance* exists in this agent-environment-system.

While the general concept of affordances is easy to understand, a closer examination of the literature revealed many differences in the interpretation of the details of the concept. Şahin et al. (2007) argued that these differences are (among other reasons) partly down to the fact that J. J. Gibson did not develop a conclusive definition of affordance but expected it to evolve further over time. Additionally, the authors argued that J. J. Gibson was only interested in the perception aspect of affordances so that he did not consider other aspects of the concept in his research (Şahin et al., 2007). Consequently, the understanding of affordances in the context of my work need to be clarified. In order to do this, J. J. Gibson's original affordance concept will be summarised and experiments will be reviewed which specifically aimed at showing effects which support the affordance concept. I will also look at the question where affordance are located within the agent-environment system to provide a frame of reference for a discussion of the location of affordances in my computational model later. Finally, I will discuss the micro-affordance concept because it is particularly directed at affordance effects in grasping.

3.2.1 On the Nature of Affordances

The affordance concept as it was introduced above focused on describing action opportunities in the environment of an agent. J. J. Gibson, however, defined the term with a much broader meaning as the following quote shows:

> "[T]he *affordances* of the environment are what it *offers* the animal, what it *provides* or *furnishes*, either good or ill." (J. J. Gibson, 1979, p. 127; italics in the original)

In this characterisation affordances appear to be a special type of properties in the environment which express any form of potential active and passive interaction between agents and the environment. The example of the cup's drinking affordance, however, makes clear that affordances cannot merely be a special form of properties of the environment but must also depend on the abilities of agents. If the agent in the example does not have hands to grasp the cup, the cup does not facilitate drinking for this agent. Hence, no drinking affordance exists with regard to this agent. Consequently, J. J. Gibson states that affordances mean "something that refers to both the environment and the animal [. . .]. It implies the complementary of the animal and the environment" (J. J. Gibson, 1979, p. 127).

This rather "hazy" (Şahin et al., 2007, p. 447) definition has sparked a lively debate about the nature of affordances and a number of authors attempted to formalise the concept (Turvey, 1992; Greeno, 1994; Sanders, 1997; Stoffregen, 2003; Chemero, 2003, and others). The first formalisation presented by Turvey (1992) regarded affordances as dispositional properties of the environment combined with dispositional properties of the actor. In his view affordances constitute an ontological category and "[. . .] exist independently of perceiving or conception" (Turvey, 1992, p. 174). A similar viewpoint was taken by Greeno (1994) who also rooted affordances in the environment and complemented them with abilities in the agent.

Because of this separation between affordances and environment on the one side and abilities and actor on the other, Turvey's and Greeno's theorisings were criticised for "[. . .] reinvigorat[ing] the very subject-object dichotomy that the ecological approach so brilliantly overcomes" (Sanders, 1997, p. 97). Sanders was the first to argue that affordances should be seen as ontological primitives and defined with a relativistic approach linking the agent and the environment. Following the relativistic route, Stoffregen (2003) and Chemero (2003) defined affordances as emergent properties of the agent-environment system (Stoffregen) or as the relations between features of the environment and abilities of an agent (Chemero). These two interpretations are similar to Warren, Jr. and Whang's (1987) view who adopted the position that affordances are located in the relation of the agent and the environment: "Perceiving an affordance [. . .] implies perceiving the relation between the environment and the observer's own action system" (Warren, Jr. & Whang, 1987, p. 371). It is important to note here that Warren, Jr. and Whang wrote that the affordance is *perceived*. Thus, they assumed that the affordance is not constructed in the brain from a previously perceived representation but rather perceived *directly*. This understanding is in line with J. J. Gibson's view:

> "The perceiving of an affordance is not a process of perceiving a value-free physical object to which meaning is added in a way that no one has been able to agree upon; it is a process of perceiving a value-rich ecological object." (J. J. Gibson, 1979, p. 140)

This theory of perceiving the meaning of objects directly without mediating pictures or representations has become known as 'direct perception' (J. J. Gibson, 1979, p. 147). It is based on two assumptions: Firstly, in 'direct perception' the environment is assumed to not be perceived as a simple two-dimensional image on the retina[1] but as

[1] I only consider visual perception here. The theory, however, applies to other types of perception as well.

an “ambient optic array” (J. J. Gibson, 1979, p. 114). The optic array is different from an image on the retina as it provides richer information about the environment than a matrix containing only luminance and colour information does. This concept is probably best understood by looking at the analogy of a plain picture of an object compared to a holographic reproduction of same object (Michaels & Carello, 1981, p. 23). Despite being only a flat plate like the normal picture the hologram still contains the complete three-dimensional optical structure of the object. Other authors have linked direct perception to the optical flow, which describes how positions change in two consecutive images (Warren, Jr., 1988). It has been argued that the optical flow is perceived when moving through the environment and that the information encoded in it plays an important role in recognising objects in the environment. Secondly, Gibson postulated that animals do not perceive their environment in absolute measures but rather in terms of relations. This can be illustrated with an example from audio perception: Most people recognize a melody by the intervals between the notes rather than by perceiving the actual frequency of each note when played on a particular instrument (Michaels & Carello, 1981, p. 25). This allows to identify the melody even when environmental conditions and the absolute frequencies change. Because of this insensitivity to changes J. J. Gibson called such sensations ‘invariants’ (J. J. Gibson, 1979, p. 310).

The basic building blocks of ‘direct perception’ are on a much more physical level compared to the concept of affordances. The two concepts are, however, assumed to be linked by applying the concept of invariants to unique combinations of invariants as well. J. J. Gibson argues that such combinations or ‘compound invariants’ are invariants themselves. By combining compound invariants, a hierarchy of invariants can be constructs which finally leads to high-order invariants specifying even complex affordances (J. J. Gibson, 1979, p. 141). Reproducing this hierarchy of affordances will be an important part of my computational model. E. J. Gibson argued that invariants are learned during childhood by interacting with the environment through activities like touching, mouthing, or shaking in order to "[narrow] down from a vast manifold of (perceptual) information to the minimal, optimal information that specifies the affordance of an event, object, or layout" (E. J. Gibson, 2003).

E. J. Gibson's argument shows that knowledge about invariants and subsequently about affordances is acquired through interaction of the animal with its environment. As it was already pointed out earlier in this chapter affordances are seen to relate (Warren, Jr. and Whang) or "impl[y] the complementary" (J. J. Gibson) between the animal and the environment. J. J. Gibson pointed out that this is also true for the information that specifies an affordance: "An affordance [. . .] points two ways, to the environment and to the observer. So does the information to specify an affordance" (J. J. Gibson, 1979, p.141). This means that affordances are specified by combining (or relating) knowledge about the agent's own body with information about invariants in the environment (which are a kind of relation, too).

3.2.2 Experimenting with Affordances

At the beginning of this chapter I reviewed experiments which explored the links between vision and action. While the results of these experiments can often be interpreted with the help of the affordance theory, they did not attempt to investigate affordances and invariants as such. To explore these components of the affordance theory, a special experimental paradigm was developed. Based on the postulated perception of relations between properties of the environment and the observer's own action system it has been concluded that information for affordances uses a 'body-related' metric (Warren, Jr., 1984). Such an intrinsic measure describes properties of the environment with respect to a property of the observer's action system. For example, in Warren, Jr. and Whang's (1987) study of walking through apertures the width of apertures in a wall A was set in relation to the shoulder width S as

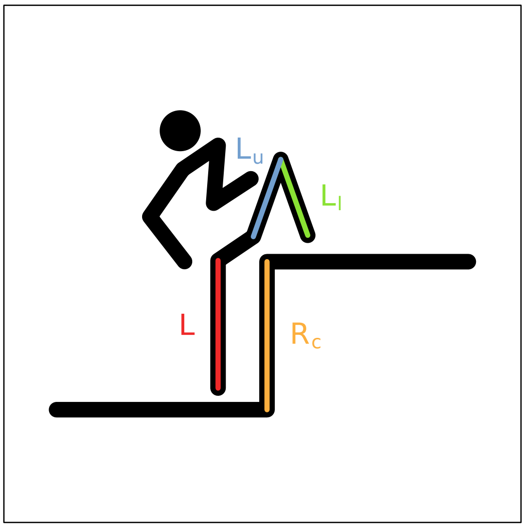


Figure 3.2: Warren, Jr.'s (1984) biomechanical model of stair-climbing. He assumes that the maximal riser height $R_c = L + L_u - L_l$.

$\pi = A/S$. This combination resulted in unit-less numbers which are known as pi-numbers (Warren, Jr., 1984). Because of their relation to a body measure the pi-numbers were independent of the actual size of observers.

In the experimental paradigm the pi-numbers were used to establish two points within the range of possible pi-values: Firstly, *critical points* needed to be identified. At these points an affordance appears, disappears or changes into a different one. This is, e. g., the size of an object at which a person grasping it switches from a single handed grasp to a bimanual grasp. Secondly, *optimal points* need to be identified. These are the points in the pi-number range where the energy necessary to perform an action is minimal (Warren, Jr., 1984). After establishing critical and optimal points the paradigm required that it is verified whether critical points and optimal points can be perceived visually prior to action execution. This is obviously necessary to rule out the possibility that these points are identified only during action execution.

Warren, Jr.'s (1984) seminal work on the perception of a stair-climbing affordance illustrates how the paradigm outlined above is accomplished in a concrete study. In the first of the three experiments in Warren, Jr.'s study participants were chosen so that the group could be split in a short one and tall one representing the 2nd and 98th percentile of normal adult male heights. Then the participants' upper, lower and overall leg length was measured. Using a simple biomechanical model Warren, Jr. calculated the maximum riser height for stair-climbing without support of the hands (see Fig. 3.2 for details) and related it to the leg length of his participants. This resulted in a mean pi-number of $\pi = 0.88$, meaning that the critical riser height was $R_c = 0.88L$. To check if this theoretical value corresponded with the participants' visual perception of the maximal riser height, they were shown photographs projected on a well depicting a two step stair with different riser heights. Subjects were then asked to fill in an answer sheet judging whether the stairs were climbable without using the hands. Furthermore, for each judgement participants were asked to provide a confidence rating on a scale from one to seven.

The results of the experiment showed that the group of tall participants perceived higher risers as still climbable than the 'short' group did. However, when looking at riser height relative to participants' leg length the differences between the two groups disappeared. Moreover the pi-values where the percentage of climbable judgements dropped below 50% (so the majority of subjects regarded higher risers as not-climbable) were 0.88 for the short group and 0.89 for the tall group. These values corresponded well with the previously theoretically established pi-value. In addition to the analysis of judgements, Warren, Jr. also compared the confidence ratings of the participants to riser heights and pi-numbers because he expected to see a drop in confidence of judgement around the critical points. His analysis showed this effect and confirmed the outcomes of the judgement's analysis.

After demonstrating that critical points can be perceived visually, the second experiment of Warren, Jr.'s study aimed at discovering the optimal riser heights for stair climbing. In order to do this subjects climbed steps with different riser heights on a stairmill while oxygen consumption was measured. Based on this data Warren, Jr. calculated the total energy per step cycle divided by the vertical work performed. This was set in relation to $\pi = R/L$ like before and yielded a mean pi-value of 0.26 for both groups showing the optimal point is the same for short and tall subjects.

With the optimal point known Warren, Jr. investigated in the last experiment of his study whether optical riser height can be visually perceived as well. In this final experiment participants were asked to first judge which of two stairs would be easier to climb and then to rate the climbability of stairs on a scale from one to seven. The stairs were presented on photographs like in the first experiment. Again, the ratings and the judgements were compared against $\pi = R/L$ showing that preferred riser heights coincided closely with the pi-numbers for the optimal point determined in experiment two. Thus, evidencing that optimal riser heights can be perceived visually.

Warren, Jr.'s study demonstrates how body-related measures are used to verify affordances experimentally. Many studies have been conducted which look at different types of actions and affordances. These studies investigated affordances for reaching and grasping (Hallford, 1983; Solomon, Carello, Grosofsky, & Turvey, 1984), sitting and stair climbing (Mark, 1987), passing through apertures (Warren, Jr. & Whang, 1987) and other behaviours (see Şahin et al., 2007; Mark, 1987 for summaries). However, despite the high number of studies the underlying methodology using pi-numbers to define body-related measures is very similar in all of them (Şahin et al., 2007).

In summary the example of the stair climbing study demonstrated that affordances can easily be described with pi-numbers and that one affordance covers a range of values with critical and optimal points marking the boundaries and maxima. Thus, we can expect to see variations in the execution of an affordance which should vary around an optimal point but should not stretch over critical points.

3.2.3 Micro-Affordances

In the discussion of the affordance concept so far actions – even complex ones – were regarded as entities, without considering the huge variations which can be observed within a single type of action. An approach which provides a way to account for these variations are Ellis and Tucker's (2000) micro-affordances. The concept picks up the idea of a direct link between actions or action opportunities and objects; apart from this, though, micro-affordances are very distinct from Gibson's affordances.

The most palpable difference is the level at which micro-affordances are defined. Instead of affording entire sequences of actions like Gibson's affordances, micro-affordances facilitate only particular realisations of specific components of an action. In Ellis and Tucker's (2000) experiments, for example, these components are grasp type (power or precision grasp) and wrist rotation. In the grasp type experiment all objects afforded grasping but some of them required a precision grasp and others a power grasp. When participants were presented with one of these objects and then asked to respond to a high or low pitch tone with either a precision or a power grasp on a special response device, Ellis and Tucker found a compatibility effect between the grasp type required for the object and the response required for the tone. Since the objects were not relevant for the task, Ellis and Tucker concluded that the grasp type which was compatible with the objects was facilitated because the objects *afforded* this particular type of grasping. The same was found for the wrist rotation component of grasping. Micro-affordances explain these compatibility effects as the result of a bottom-up process which extracts specific execution parameters for actions. Gibson's affordances do not capture such sub-action affordance effects.

In another contrast to Gibson's affordances Ellis and Tucker emphasised that micro-affordances are associated with a representation of an object and are not directly perceived in a Gibsonian sense. However, the authors argued that action components potentiated by a micro-affordances are intrinsic to the representation of an object. Thus, it is not a completely arbitrary symbolic representation of an action opportunity but a representation which "[has] direct associations between vision and action" (Ellis & Tucker, 2000, p. 468). While this appears to be contradicting Gibson's understanding of affordances, Gibson's emphasis on direct perception must be seen in the context of his time when reasoning in symbolic categories was popular. Against this background it becomes clear that direct perception meant above all the absence of symbolic representations but not the absence of representations at all. Processing of information in the brain obviously needs some form of representation to encode the perceived information in electrical signals. However, such representations can be either symbolic or more low-level patterns directly linked to the perceived pattern of information. Essentially, Gibson argued that knowledge about actions is not derived from symbolic representations but from patterns describing the visual information more directly and in relation to action execution. Hence, the role of representation in Gibsonian affordances and micro-affordances is not that different at all. Micro-affordances share with Gibson's affordances the notion of linking objects and the actions they *afford*. However, while Gibson's affordances focus on the direct perception of action opportunities in an the environment, Ellis and Tucker pursue the idea that micro-affordances are a motor-oriented representation of actions constructed within the brain.

Finally, Ellis and Tucker see micro-affordances as dispositional properties of an observer's nervous system rather than as properties of the environment or the relation between environment and observer. The authors speculated that micro-affordances develop in the brain due to the involvement of motor components in the representation of visual objects.

4 Models for Visual Attention and Action

The problem of identifying opportunities for action and the planning of grasps has been tackled in a number of computational models. Foremost, the problem is regularly encountered in robotics research (e. g., Lam, Ding, & Liu, 2001; Montesano, Lopes, Bernardino, & Santos-Victor, 2007; Saxena, Driemeyer, & Ng, 2008; Bohg & Kragic, 2010). However, the designs of the models and algorithms developed in this area are mostly driven by the technical requirements of robots and not by biological plausibility of the processes. For instance, the grippers of robots are usually much simpler than the human hand; most have only two or three "fingers" with limited flexibility and dexterity. This heavily restricts the number of different grasp postures that these models can adopt. Visual attention is also not normally taken into consideration in these models.

However, modelling is also used as a tool for understanding action and grasping in psychology. Research in this area has produced different models ranging from models for grasp execution and grasp postures to models for recognition of affordances in the visual input and visual attention for action. In the remainder of this chapter these models are reviewed whereby the focus lies on their suitability for modelling complex affordances and action-guided visual attention since these are the main aims of the model presented in this thesis.

Smeets and Brenner's Model Smeets and Brenner (1999) presented a model which built on their idea of using the individual positions of the finger tips to describe grasps (see Sec. 2.2.2). Their model aimed at modelling the characteristics of the finger movement during reach-to-grasp movements, though, and not at modelling how grasp postures are generated for different objects. This is reflected by the input their model required: It consisted of the initial and final finger positions and the time required for the movement. Additionally, only movements of the thumb and the index finger were modelled which limited the number of grasps that could be described. To calculate the position of the thumb and the finger at each point in time, the model used a minimum-jerk approach (Flash & Hogan, 1985). Because of its focus on reach-to-grasp movements Smeets and Brenner's model is obviously only of minor interest within the context of my work. However, the description of hand postures during grasping with the position of the individual fingers is an interesting aspect of the model which I already discussed in Chapter 2. Additionally, there is the possibility of using my model to generate the final finger positions required as input by Smeets and Brenner's model and thus linking both models together. I will come back to this point in the general discussion (Chapter 8).

Iberall and Fagg's Model In contrast to Smeets and Brenner's model, the model developed by Iberall and Fagg (1996) was specifically aimed at modelling hand postures of human grasps. The model was strongly based on the concept of virtual fingers and opposition spaces (see Sec. 2.2.1). The input into the model was a parametric description of the object to be grasped (e. g., the length and diameter of a cylinder), a numerical measure of task difficulty, and a description of the characteristics of the hand (hand width and length). The model was constructed from three neural networks each of which computed a specific parameter of the grasp postures. These parameters were the physical fingers used in virtual finger 2 (VF2), the type of

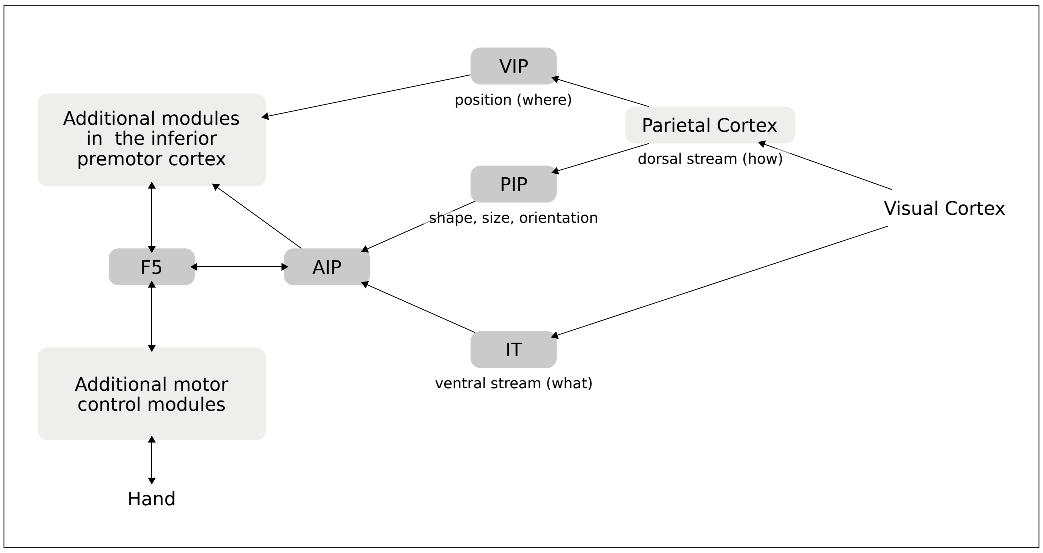


Figure 4.1: Overview of the FARS model (Fagg & Arbib, 1998). The graphic shows a simplified version of the model in which the modules in the inferior premotor cortex and the ones responsible for motor control are not shown individually. The component of most relevance here is the AIP module which received input from the visual input via the dorsal and the ventral stream and combined it with motor information to produce a grasp posture. Fagg and Arbib argrued that affordances manifested in this combination.

opposition space and the opening size of the hand. The neural networks in the model were backpropagation networks with one hidden layer and real human grasp postures were used as a training set. Iberall and Fagg's model showed that the input and output of the model are highly structured but contain no information about the spatial location of the object to be grasped and the produced grasp. This makes the model unsuitable for modelling visual attention for action: Only single objects can be presented to the model and these already need to be preprocessed to extract structural information about the object dimensions. Furthermore, this information about the object is not linked to the visual field any more so that spatial attention based on action-relevant features of objects cannot be modelled. It is also important to note that Iberall and Fagg only concentrated on modelling grasping and did not link their model to Gibson's affordance theory.

The FARS Model Fagg and Arbib (1998) developed the FARS model (Fagg-Arbib-Rizzolatti-Sakata-Model) which not only simulated grasping but also incorporated Milner and Goodale's dual route theory and links to Gibson's affordances. The model focused more on motor control than on visual perception for action. It was linked directly to brain areas and its objective was to predict the firing patterns of brain areas as well as of individual neurons.

The model used a complex coding for input objects involving the object type (sphere, cylinder, or block) and dedicated sets of neurons for each object type to encode the dimensions of the different objects. The extraction of these parameters from the visual input was not modelled. Grasps were described by encoding grasp type, grasp orientation and aperture. The authors constructed their model from a number of modules which they directly linked to brain areas for vision and motor control; this is reflected in the names of the modules (Fig. 4.1). The model focused not on the early stages of visual perception for action but on the transition from visual information to motor control. In the model this transition was located in an AIP (anterior intraparietal area) module where information about objects from the dorsal and the ventral stream was associated with units in the AIP which represented grasp parameters. A central element of the model was the loop between the AIP module and an F5 (part of the inferior premotor cortex) module which was involved in grasp execution. This loop ensured that the grasp specification in the AIP module was updated during grasp execution. The F5 module contained units describing the current state of the grasp (hold, release, flexion, etc).

Fagg and Arbib's model was geared more towards motor control of grasping then towards visual perception for grasping. In contrast to the model presented in this thesis, the FARS model did not simulate effects of the grasps planned in the AIP on visual attention. Moreover, the connection from the visual cortex via the dorsal and ventral streams was only uni-directional in the model and, thus, preventing the AIP from influencing visual attention. The extraction of affordances from the visual input was not of included in the model. Fagg and Arbib argued that affordances manifest in the AIP module where the object features are associated with the grasp parameters. This is in line with Stoffregen and Chemero's definitions of affordances as relations between the environment and the abilities of an animal.

Cisek's Model of the Affordance Competition Hypothesis Cisek (2007) introduced the affordance competition hypothesis to explain the processing of action relevant visual information in the brain. The hypothesis suggests that decisions for actions are the result of a competition between opportunities for action in the environment (affordances) and task demands of the observer. This competition process is assumed to take place continuously and resulting in selective attention to action relevant visual information. The competition processes are supposed to occur not only within cortical areas in the dorsal stream but also between different cortical systems in order to identify the most promising target locations for action. Cisek pointed out that the representations of actions at the end of the dorsal route in the fronto-parietal cortex are neither explicit representations of objects or motor plans but instead a mixture of both. These representations could include for instance the direction of the grasp movement or saccade targets. This shows a remarkable similarity to the micro-affordance concept (Sec. 3.2.3) which aimed at describing representations of components of actions by including both visual and motor information.

In a computational model of a cued reach-decision task (Fig. 4.2) Cisek linked the affordance competition hypothesis to the neural substrate (Cisek, 2007). The model was designed to explain data from an experiment in which a primate was first shown two differently coloured potential target locations positioned on a circle followed by a colour cue in the centre of the screen. The colour of this cue indicated which of the two potential target locations shown before was the actual target. The cue was then

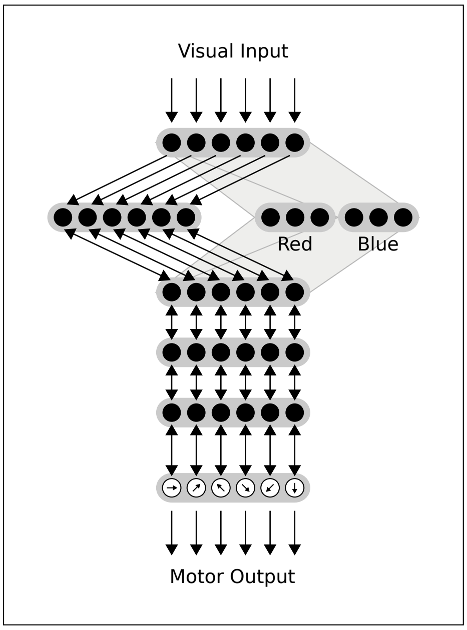


Figure 4.2: Overall structure of Cisek's model for reach- planning (adapted from Cisek, 2007). Populations of neurons are organised in layers which transform information from the visual input into motor commands. Only a subset of the neurons in each layer is shown. The layers which show six neurons react to any stimuli while the layers labelled with 'red' and 'blue' denote layers which only respond to red and blue stimuli. They colour sensitive layers also have a lower spatial resolution than the other layers; hence, only three neurons are shown in these layers. The layers are linked by topographical connections. The layers for red and blue stimuli also project topographically; this is indicated by the grey area linking these layers with the previous and following layers.

followed by an arrow which indicating the direction of the reach movement towards the target location and served as a 'GO' signal. After the 'GO' signal the monkey had to make a movement in direction of the target location (Cisek & Kalaska, 2005). Neural activity in the dorsal premotor cortex and in the primary motor cortex was recorded. The model was constructed from layers of populations of leaky-integrator neurons which were linked topographically and related to brain regions along the dorsal stream in the brain. In each population every neuron had a preferred direction, and stimuli lying in this direction triggered its activity. Within each population the neurons had excitatory connections to neurons which responded to similar directions and inhibitory connections to neurons tuned to different directions. Additionally, the neurons had a passive decay of activity and received noise. During reach-planning the coloured stimuli observed in the visual input activated the neurons which preferred the direction in which the stimuli were located. These activations propagated through the layers and the competition between inhibitory and excitatory connections resulted in a decision for a direction to reach to. Using this model Cisek was able to replicate patterns of neural activation from experimental data.

Cisek's model is notable for demonstrating how a direct link between visual perception and motor control can be established by using a consistent encoding along the dorsal route. This can be construed as an implementation of Gibson's direct perception. However, the understanding of affordances in the model is very restricted – especially in contrast to Gibson's broad definition. The model focused only on the direction of a reach-movement and did – in contrast to the model developed in my work – not include more complex actions like grasping. Nonetheless, Cisek's model is a good example that demonstrated how parameters for actions can be derived directly from visual input.

Ward's Model Modelling visual attention was the main aim of the model developed by Ward (1999). His model did, however, not only simulate visual attention for action but also incorporated the selection of action parameters. In order to simulate attention for action, the input into the model consisted of the contents of the visual array and of a specification of the intended action. Furthermore, Ward integrated the dorsal stream as well as the ventral stream in his model. This is reflected in the existence of 'what', 'where' and 'how' systems in his model (Fig. 4.3). The 'what' system represented the ventral stream and recognised object shapes and colours, the 'where' and 'how' systems stood for the dorsal stream with the 'where' system identifying target locations and the 'how' system encoding action intentions. These

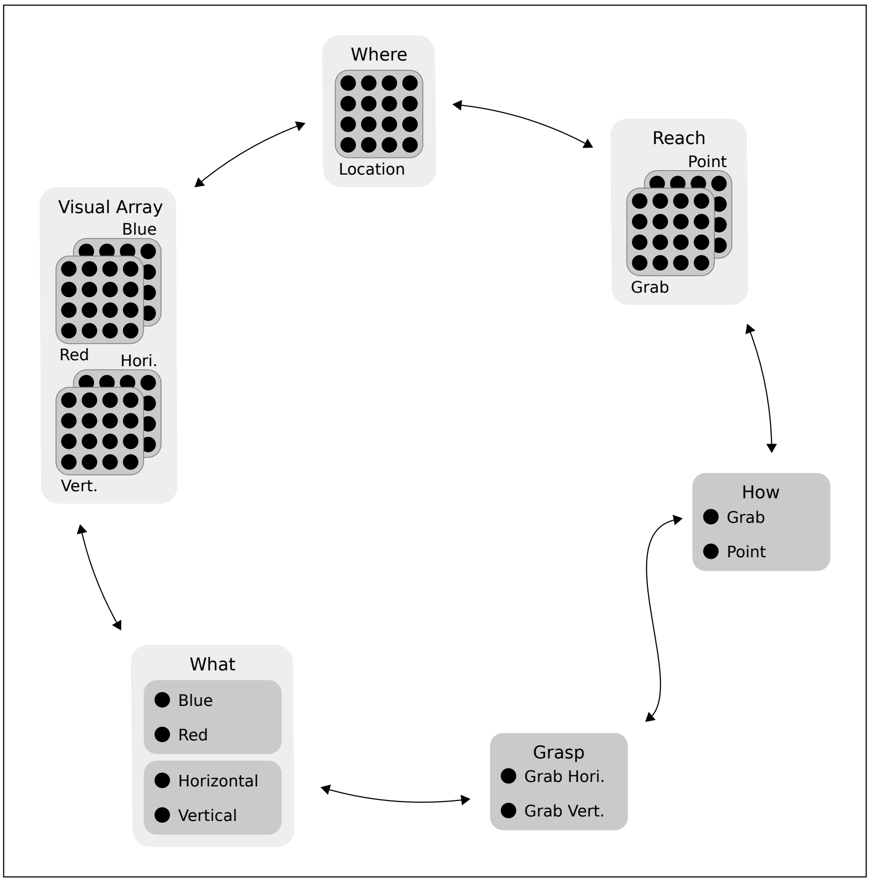


Figure 4.3: Overview of Ward's model for selective action (adapted from Ward, 1999). Units in grey boxes have inhibitory connections (except for the feature maps) and the different subsystems are linked with excitatory connections.

three systems were linked by three additional systems each of which combined the information from two of the former systems in a 'conjunctive code': The 'what' and 'where' systems were connected via feature maps in the visual array, the 'what' and 'how' systems were linked via a 'grasp' system which allowed to describe different types of grasps, and 'where' and 'how' systems were linked via feature maps which encoded action target locations in the 'reach' system. All six systems were linked to each other with excitatory connections. Within each system units were linked with inhibitory connections. The input into the model was specified by setting the contents of the feature maps in the visual array to represent the visual input. The intended action was fed as top-down information from other brain systems (not included in the model) into the model by activating one of the action type units (grab or point) in the ‘how’ system and one of the colour units in the ‘what’ system. During the simulation the excitatory and inhibitory connections in the model triggered a selective process which after a while produced a stable state of the system. In this state each of the systems described a parameter for executing the intended action.

To illustrate how his model operated, Ward provided an example explaining how the model specified an action and selected an object for it when it was asked to grab the red object while being presented with a stimulus showing a red vertical bar on the left and a blue horizontal bar on the right. First, the feature maps in the visual array of the model were initialised according to the input stimulus. The intended action was fed into the model by activating the ‘red’ unit in the ‘what’ system and the ‘grab’ unit in the ‘how’ system . All initial values were constantly applied during the simulation. At the beginning of the simulation the activation in the feature maps of the visual array propagated to the location map in the ‘where’ system and to the colour and orientation units in the ‘what’ system. Since two objects with different colours, orientations and in different locations were described in the visual array the activations balanced each other. However, the additional input that was applied to the ‘red’ unit produced an imbalance in the model. The stronger activation of the ‘red’ unit led to an increased activation of the colour map for red in the visual array. This in turn resulted in a stronger activation of the left side in the location map in the ‘where’ system which fed back its activation into all feature maps of the visual array. Hence, objects in the left half of the feature maps received more support than the ones on the right side. Since the red vertical bar was on the left side the corresponding units in the ‘what’ system gained an advantage over the units representing blue and horizontal. At the same time the sustained initialisation of the ‘grab’ unit in the ‘how’ system activated the units in the ‘grasp’ system and supported the units in the ‘grab’ feature map of the ’reach’ system. Combined with the orientation selection from the 'what' system this resulted in the activation of a vertical grab in the 'grasp' system; in the 'reach' system the activity from the 'where' system led in conjunction with the activation from 'grab' unit in the 'how' system to a selection of the left half of the 'grab" feature map. Thus, at the end of the simulation the action parameters were specified in the 'grasp' system and the object was selected in all feature maps. While this example showed that action-based object selection and action specification is possible with Ward's model, it is interesting to note that the model is not able to model bottom-up action selection without specifying an intended action (grab or point). In particular it is notable that selection in situations like in the stimulus above would not work because the model had no intrinsic biases for neither colour nor orientation. This limits the possibilities of modelling visual attention for action with this model in a bottom-up fashion.

However, Ward used his model mainly to investigate the effects of visual extinction in patients and to determine how functionality was dissociated between the different systems of the model. Of most interest in relation to my research is, however, his investigation of effects of action intentions on perception. Ward explored these effects in simulations in which the 'where' system was lesioned on the left side (by applying an inhibitory bias). He found that the 'grab' action was less sensitive to the damage then the 'point' action. He suggested that this was caused by the additional support from the 'grasp' system in the 'grab' simulation.

4.1 Conclusion

The models reviewed here approached the route from vision to action from very different perspectives. Some models focused exclusively on motor control and others only looked at visual attention for action. Interestingly, most models made no or only a few references to affordances. Most notably, no model combined visual attention, affordances and complex actions. Also, most models required structured descriptions of the objects in the visual input which makes it difficult to interpret them in a Gibsonian sense considering J. J. Gibson's understanding of the direct perception of affordances.

Part II

The Selective Attention for Action Model

5 Study 1: Modelling Human-like Grasp Postures

In the following three chapters the Selective Attention for Action Model (SAAM) is presented. The model generates finger positions for grasping objects which are presented in a visual input field. SAAM is introduced in three steps: First, a version of the model is presented which only utilises bottom-up information from the visual input to produce grasps. Then, in the next chapter, the model is extended to also classify grasp types (e. g., power or precision grasps) and to integrate top-down information about desirable grasps. Finally, in Study 3, I demonstrate how SAAM models affordances and action intentions for tool-use.

This chapter starts with a qualitative and a mathematical description of the initial version of the model, followed by a presentation of first simulation results. Then an experiment is reported in which information about human grasp postures were collected. Finally, the simulation results and the experimental data are compared.

5.1 The Selective Attention for Action Model

The Selective Attention for Action Model follows ideas developed in the Selective Attention for Identification Model (SAIM; Heinke & Humphreys, 2003). Like SAIM it uses a soft-constraint satisfaction approach to select an area in the visual field.

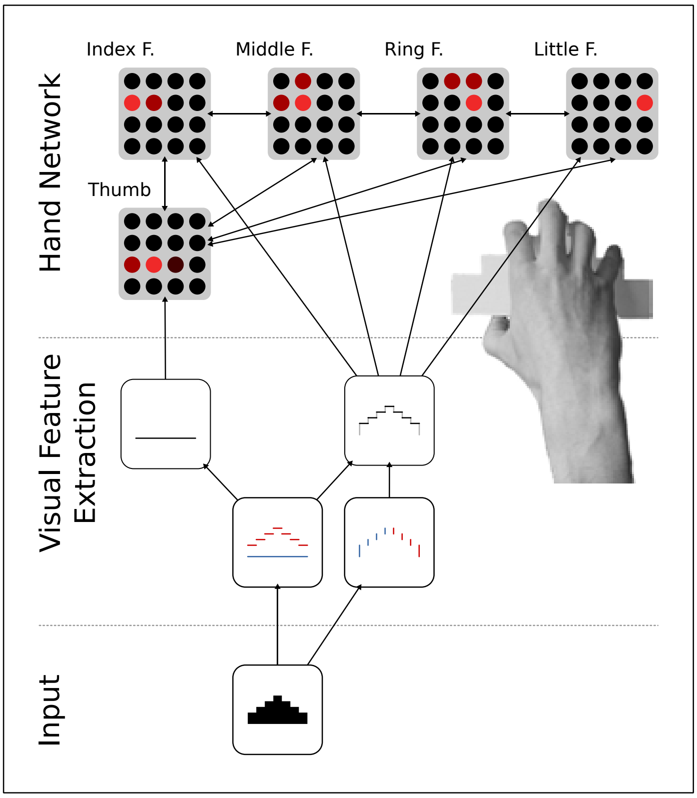


Figure 5.1: Overall structure of the Selective Attention for Action Model. The model has two main components: the visual feature extraction stage which preprocesses the visual input and the hand network which does the actual computation of grasp postures. In the hand network two-dimensional arrays of neurons encode the position of each finger. These maps are connected with each other and receive their input from the visual feature extraction stage.

However, the constraints used and the output which is generated are very different in both models.

Figure 5.1 shows an overview of SAAM. The input consists of black&white images. The output of the model is generated in five 'finger maps' which are part of a 'hand network'. The finger maps encode the finger positions which are required for producing a stable grasp of the object in the input image. At the heart of SAAM's operation is the assumption that stable grasps are generated by taking into account two types of constraints: 'geometrical constraints' derived from the object shape and 'anatomical constraints' imposed by the hand anatomy. It is important to note that SAAM has

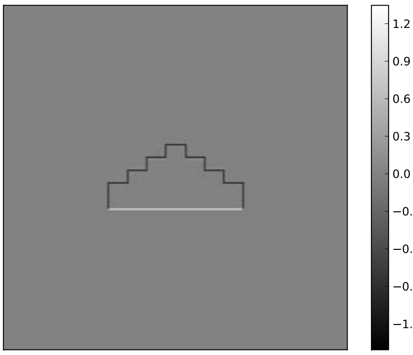


Figure 5.2: Example of the visual input after the geometric constraints were applied. The image shows the output of the horizontally oriented filter combined with the weighted output of the vertically oriented filter. The sign of the activation on the edges is used to separate edges at the bottom side of the stimulus object (white lines) from edges at the top and vertical sides (black lines).

no notion of the forces and torques which need to be balanced in a stable grasp but it relies solely on its constraints to construct a stable grasp posture (see Mason, 2001 for examples on the calculation of forces in grasps). This point will be discussed further in the general discussion in Chapter 8.

In order to ensure that the hand network satisfies the constraints, SAAM follows an approach suggested by Hopfield and Tank (1985). In this soft-constraint satisfaction approach, constraints define activity patterns in the finger maps that are permissible and others that are not. Then, an energy function is defined for which the minimal values are generated by just these permissible activity values. To find these minima, a gradient descent procedure is applied resulting in a differential equation system. The differential equation system defines the topology of a biologically plausible network. The mathematical details of this energy minimisation approach are given in the next section. Here, I focus on a qualitative description of the two types of constraints and their implementation.

The geometrical constraints are extracted from the shape of the object in the visual feature extraction stage. To begin with, obviously only edges constitute suitable contact points for grasps (see Fig. 5.2). Furthermore, edges have to be perpendicular to the direction of the forces exerted by the fingers and the thumb. Hence, only edges with a horizontal orientation make up good contact points since only horizontal hand orientations are considered in the model. This restriction to horizontal edges does however not need to be met by all fingers strictly. For a stable grasp it is usually sufficient if the thumb and two other fingers are on horizontal edges. The remaining fingers can be placed on the vertical edges without loosing stability. On the contrary, this can even contribute to a firmer grasp of an object because the fingers on the vertical edges can prevent sideways movement of the object (thus, two-dimensional form-closure[1] can be achieved; e. g., Bicchi, 1995). Hence, the restriction to horizontal edges is loosened for fingers (not for the thumb, though) to also include vertical edges but with reduced activation compared to the horizontal edges. The edge detectors are implemented using Gabor filters. These filters have been shown to resemble the processes in the primary visual cortex (Daugman, 1985; Field, 1987; Petkov, 1995). To exert a stable grasp, thumb and fingers need to be located at opposite sides of an object. This requirement is realized by separating the output of the horizontally oriented Gabor filter according to the direction of the gradient change at the edge. In fact, the algebraic sign of the response differs at the bottom of a 2D-shape compared to the top of a 2D-shape. Now, if one assumes the background colour to be white and the object colour to be black, the signs of the Gabor filter responses indicate appropriate locations for the fingers and the thumb (see Fig. 5.1 and 5.2 for an illustration). In the output of the vertically oriented Gabor filter the gradient change is ignored and all edges are added to the input of the finger maps. The results of the visual feature extraction feed into the corresponding thumb and finger maps providing the hand network with the geometrical constraints. Note that, of course, the assumptions about the object- and background-colours represent a strong simplification. On the other hand, this mechanism can be interpreted as mimicking the result of stereo vision. In such a resulting 'depth image' real edges suitable for thumb or fingers could be easily

[1] 'Form-closure' means that an object is held by blocking each possible direction of movement with a finger. In a form-closure grasp no force needs to be exerted to hold the object in place. A two-dimensional form-closure grasp of a square, for example, can be achieved by placing one finger at each side of the square.

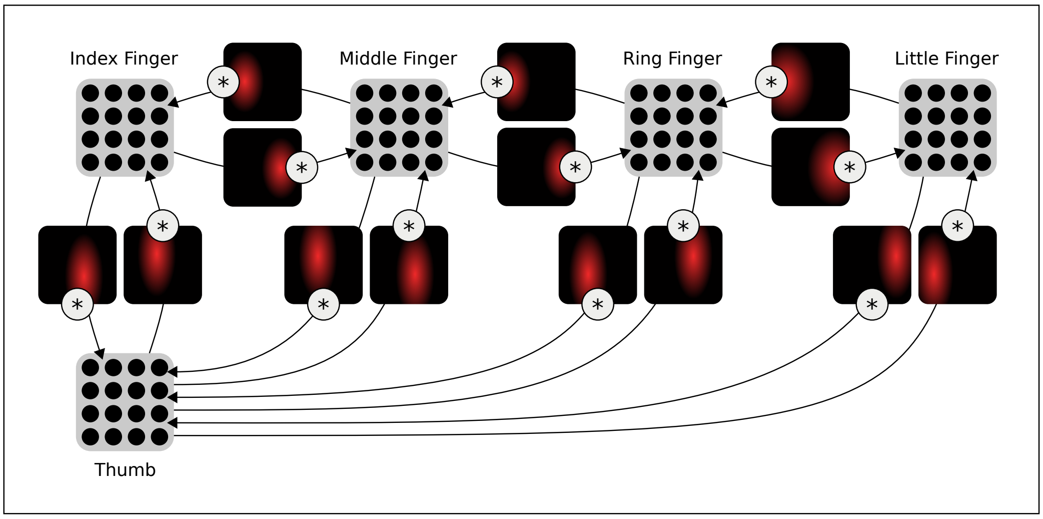


Figure 5.3: Excitatory connections between the finger maps in the hand network. Connections exist only between neighbouring fingers and between each finger and the thumb. Further connections between the fingers do not exist because the weight maps between the fingers effectively form a transitive chain of constraints linking all fingers and making such connections redundant. The excitatory connections project in both directions using a transposed version of the weight map for the feedback projection. The $*$-operator indicates a convolution of the source finger map with the weight map. For clarity the finger maps show only 4×4 neurons; in reality each map consists of at least 512×512 neurons (depending on the study). The activation shown in the weight maps is also only exemplary.

identified.

The anatomical constraints implemented in the hand network take into account that the human hand cannot form every arbitrary finger configuration to perform grasps. For instance, the maximum grasp width is limited by the size of the hand and the arrangement of the fingers on the hand makes it impossible to place the index, middle, ring, and little finger in another order than this one.[2] After applying the energy minimisation approach, these anatomical constraints are implemented by

[2] The anatomical constraints can be linked to Warren, Jr.'s pi-numbers (see Sec. 3.2.2). The hand network contains an implicit description of grasping related properties of the hand anatomy: During grasp generation this description is matched with the geometrical constraints extracted from the visual input. Thus, a relation between the object in the visual input and the hand anatomy is established. This means that grasp generation in SAAM is implicitly guided by a body-related measure which can be described with pi-numbers. This aspect will be discussed further in Sec. 8.4.

excitatory connections between the finger maps in the hand network (see Fig. 5.1 and 5.3). Figure 5.3 also illustrates the weight matrices of the connections. Each weight matrix defines how every single neuron of one finger map projects into another finger map. The direction of the projection is given by the arrows between the finger maps. For instance, neurons in the thumb map feed their activation along a narrow stretch into the index finger map, in fact, encoding possible grip sizes. Each neuron in the target map sums up all activation fed through the weight matrices. Note that all connections between the maps are bi-directional whereby the feedback path uses the transposed weight matrices of the feed-forward path. This is a direct result of the energy minimisation approach and ensures an overall consistency of the activity pattern in the hand network since, for instance, the restriction in grip size between thumb and index finger applies in both directions. Finally, since a finger can be positioned at only one location, a winner-takes-all mechanism was implemented in all finger maps. In the next chapter I will show that this selection mechanism also implements global selection mimicking selective attention.

5.1.1 Mathematical Details of the Model

This section documents the mathematical details of the Selective Attention for Action Model.

Visual Feature Extraction

The filter kernels in the visual feature extraction process are two Gabor filters which are horizontally and vertically oriented. For edge detection only the real part of the complex filter response is required (Daugman, 1985; Field, 1987; Petkov, 1995):

$$K = e^{\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}} \cos(2\pi\frac{x}{\lambda} + \psi) \tag{5.1}$$

with $x \ = x \cos\theta + y \sin\theta$ and $y \ = x \sin\theta + y \cos\theta$.

In the response of the horizontally oriented Gabor filter the top edges of the object are marked by negative values while the bottom edges are marked by positive values in the filter response[3]. This filter characteristic is used to feed the correct input with the geometrical constraints applied into the finger maps and the thumb map. The finger maps receive the negated filter response after all positive activations have been set to zero. The thumb map, on the other hand, receives the filter response with all negative activations set to zero:

$$R_{IJ}^{(\text{f, h})} = \begin{cases} -R_{IJ}^{(\text{h})} & \text{if } R_{IJ}^{(\text{h})} \leq 0, \\ 0 & \text{else.} \end{cases} \tag{5.2}$$

$$R_{IJ}^{(\text{t})} = \begin{cases} R_{IJ}^{(\text{h})} & \text{if } R_{IJ}^{(\text{h})} \geq 0, \\ 0 & \text{else.} \end{cases} \tag{5.3}$$

with $R_{IJ}^{(\text{h})} = I_{IJ} * K_h$ whereby I_{IJ} is the visual input image and K_h a horizontally oriented Gabor filter kernel.

The output $R_{IJ}^{(\text{v})} = I_{IJ} * K_v$ of the vertically oriented Gabor filter K_v is added to the horizontal filter response that feeds into the finger maps. The direction of the gradient change is ignored:

$$R_{IJ}^{(\text{f})} = R_{IJ}^{(\text{f, h})} + \begin{cases} rR_{IJ}^{(\text{v})} & \text{if } R_{IJ}^{(\text{v})} > 0, \\ -rR_{IJ}^{(\text{v})} & \text{else.} \end{cases} \tag{5.4}$$

The factor r controls the activation of the vertical edges in relation to the horizontal ones.

[3] The filter parameters need to be chosen appropriately.

Finally, in order to normalize the amplitude of the filter response across different input images, a sigmoid function is applied to $R_{IJ}^{(\mathrm{f/t})}$:

$$I_{IJ}^{(\mathrm{f/t})} = \frac{1}{1 + e^{-sR_{IJ}^{(\mathrm{f/t})} + m}} \tag{5.5}$$

The parameters s and m are chosen so that the value of the sigmoid function is $0 + \epsilon$ at position 0 and $1 - \epsilon$ at position 1.

The Hand Network

An energy function approach is used to satisfy the anatomical and geometrical constraints of grasping. Hopfield and Tank (1985) suggested this approach where minima in the energy function are introduced as a network state in which the constraints are satisfied. In the following derivation of the energy function, parts of the whole function are introduced, and each part relates to a particular constraint. At the end, the sum of all parts leads to the complete energy function, satisfying all constraints.

The units $y_{IJ}^{(f)}$ of the hand network make up five fields. Each of these fields encodes the position of a finger. $y_{ij}^{(1)}$ encodes the thumb, $y_{ij}^{(2)}$ encodes the index finger, and so on to $y_{ij}^{(5)}$ for the little finger. For the anatomical constraint of possible finger positions, the energy function is based on the Hopfield associative memory approach (Hopfield, 1982):

$$E_{\mathrm{mem}}(y_I) = -\sum_{\substack{ij \\ i=j}} T_{ij} \cdot y_i \cdot y_j. \tag{5.6}$$

The minimum of the function is determined by the matrix T_{IJ}. For T_{ij}s greater than zero, the corresponding y_is should either stay zero or become active in order to minimize the energy function. In the associative memory approach, T_{IJ} is determined by a learning rule. Here, T_{IJ} is chosen so that the hand network fulfils the anatomical constraints. These constraints are satisfied when units in the finger maps that encode finger positions of anatomically feasible postures are active at the same time. Hence, the T_{ij} for these units should be greater than zero, and for all other units, T_{ij} should be less than or equal to zero. This leads to the following equation:

$$E_{\text{exc}}(y_{IJ}^{(G)}) = -\sum_{f=1}^{5} \sum_{\substack{g \\ g=f}} \sum_{ij} \sum_{\substack{s=-L \\ s=0}}^{L} \sum_{\substack{r=-L \\ r=0}}^{L} T_{sr}^{(f \quad g)} \cdot y_{ij}^{(g)} \cdot y_{i+s,j+r}^{(f)}. \tag{5.7}$$

In this equation $T_{SR}^{(f \quad g)}$ denotes the weight matrix from finger f to finger g.

The fact that each finger map should encode only one position is a further constraint. The implementation of this constraint is based on the energy function proposed by Mjolsness and Garrett (1990):

$$E_{\text{WTA}}(y_I) = a \cdot (\sum_i y_i - 1)^2 - \sum_i y_i \cdot I_i. \tag{5.8}$$

This energy function defines a winner-takes-all (WTA) behaviour where I_i is the input and y_i is the output of each unit. This energy function is minimal when all y_i are zero except one, and when the corresponding input I_i has the maximal value of all inputs. Applied to the hand network where each finger map requires a WTA-behaviour, the first part of the equation turns into:

$$E_{\text{inh}}(y_{IJ}^{(F)}) = \sum_f (\sum_{ij} y_{ij}^{(f)} - 1)^2. \tag{5.9}$$

The input part of the original WTA-equation is modified to take the geometrical constraints into account:

$$E_{\text{inp}}(y_{IJ}^{(F)}) = -\sum_f \sum_{ij} w_f \cdot y_{ij}^{(f)} \cdot I_{ij}^{(\text{t/f})} \tag{5.10}$$

These terms drive the finger maps towards choosing positions at the input object which are maximally convenient for a stable grasp. The w_f-factors were introduced to compensate the effects of the different number of excitatory connections in each finger map.

The Complete Model To consider all constraints, all energy functions are combined, leading to the following complete energy function:

$$E_{\text{hand}}(y_{IJ}^{(F)}) = a_{\text{exc}} \cdot E_{\text{exc}}(y_{IJ}^{(F)}) + a_{\text{inh}} \cdot E_{\text{inh}}(y_{IJ}^{(F)}) + a_{\text{inp}} \cdot E_{\text{inp}}(y_{IJ}^{(F)}). \tag{5.11}$$

The parameters $a_{\text{exc/inh/inp}}$ weight the different constraints against each other. These parameters are chosen in a way that SAAM successfully selects contact points at objects in single-object and multiple-object input images. The second condition is particularly important to demonstrate that SAAM can mimic affordance-based guidance of attention. Moreover, and importantly, SAAM has to mimic human-style contact points. Hereby, not only the parameters $a_{\text{exc/inh/inp}}$ are relevant but also the weight matrices of the anatomical constraints which strongly influence SAAM's behaviour.

Gradient Descent The energy function defines minima at certain values of y_i. To find these values, a gradient descent procedure can be used:

$$\tau \dot{x}_i = -\frac{\partial E(y_I)}{\partial y_i}. \tag{5.12}$$

The factor τ is anti-proportional to the speed of descent.

In the Hopfield approach, x_i and y_i are linked together by the sigmoid function:

$$y_i = \frac{1}{1 + e^{-m \cdot (x_i - s)}}. \tag{5.13}$$

Additionally, the energy function includes a leaky integrator so that the descent turns into

$$\tau \dot{x}_i = -x_i - \frac{\partial E(y_I)}{\partial y_i}. \tag{5.14}$$

Using these two assertions, the gradient descent is performed in a dynamic, neural-like network where y_i can be related to the output activity of neurons, x_i the internal activity, and $\partial E(y_I)/\partial y_i$ gives the input to the neurons. Applied to the energy function of SAAM, it leads to a dynamic unit (neuron) which forms the hand network:

$$\tau \dot{x}_{ij}^{(f)} = -x_{ij}^{(f)} - \frac{\partial E_{\text{hand}}(y_{IJ}^{(F)})}{\partial y_{ij}^{(f)}}. \tag{5.15}$$

To execute the gradient descent on a computer, a temporarily discrete version of the descent procedure was implemented using the CVODE-library (Hindmarsh et al., 2005).

Stop Criterion

The model includes a stop criterion for the gradient descent procedure. The iteration process finishes when

$$f \quad F, \max(y_{IJ}^{(f)}) \geq t^{(f)} \tag{5.16}$$

becomes true. The parameters $t^{(f)}$ define the required minimum activation in each finger map. If a simulation does not reach these levels of activation, it is aborted after

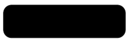
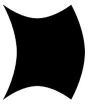
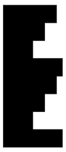
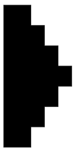
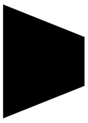
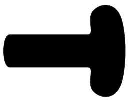
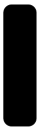
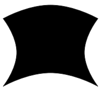
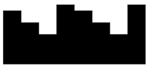
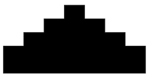
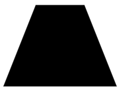
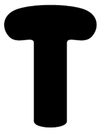
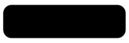
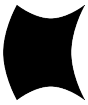
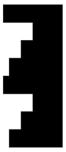
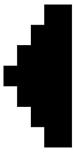
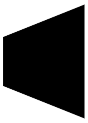
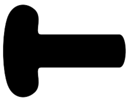
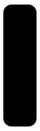
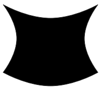
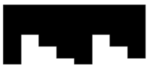
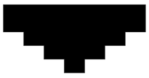
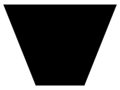
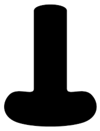
	bar	concave	irregular	pyramid	trapezium	t-shape
0°						
90°						
180°						
270°						

Table 5.1: The objects and the orientations in which they were presented in Study 1. The conditions bar 0° and bar 180° as well as bar 90° and bar 270° are obviously encoding the same stimuli. However, to keep the experiments and the analyses simple, all four orientations were treated as separate conditions.

the differential equation solver reached time t_{max}. The stop criterion enables SAAM to simulate reaction times.

5.2 Simulating Grasps

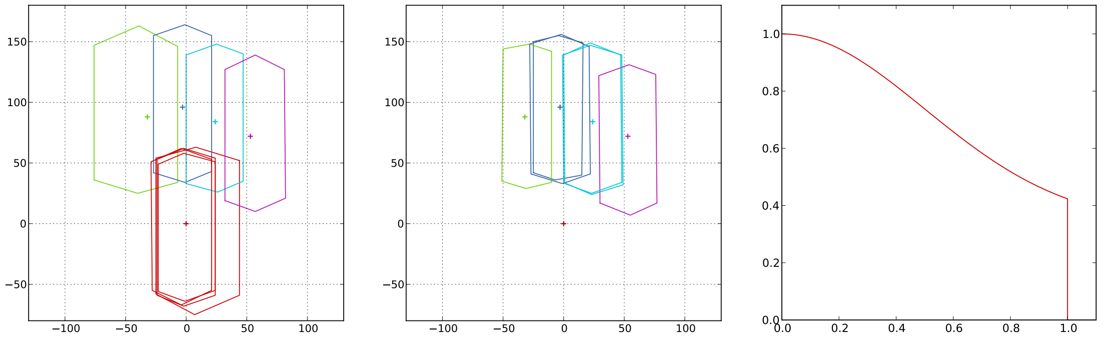
The simulations in this first study aimed at exploring the basic abilities of SAAM to produce finger positions suitable for grasping objects. Of particular interest was the question how the generated grasps compare to grasps performed by humans.

5.2.1 Method

Six objects were designed and presented in four different orientations (Table 5.1). The objects were chosen so that they did not resemble any real, familiar tools. This was done to prevent an influence of knowledge about tool use when the objects were later used in experiments with human participants. Despite this, the objects did, of course, resemble familiar shapes to varying extends; especially shapes similar to the bar and the t-shape can be encountered often in everyday life (e. g., in the letters 'I' and ''T'). Nonetheless, the shapes were considered not to be strongly linked to graspable objects and hence deemed suitable for inclusion in the study. Even, if this assumption proofed wrong, it would result in an interesting difference between the grasps created by the model and the grasps observed in humans due to the model not having this link between the shapes and real-life graspable objects.

The shapes were also designed to be graspable in more than one way and with varying difficulty. Difficulty was not formally assessed but it was assumed that objects with short and irregular edges would be harder to grasp than ones with long, straight edges. Finally, the objects were not only presented in orientations which allowed SAAM to comfortably grasp the objects but also in orientations in which grasping the objects with horizontal grasps (the only grasp orientation supported by the model) was difficult or impossible. These experimental conditions were included to test how the model would handle situations where no good grasp was possible. Additionally, these conditions allow for assessing how humans, who can change the orientation of their hands, handle these objects in contrast to the model.

Simulation Parameters The parameters for the simulations were chosen as follows: $a_{\text{exc}} = 0.4$, $a_{\text{inh}} = 3.3$, and $a_{\text{inp}} = 13.0$. In the gradient descent procedure the parameters were $\tau = 0.7$, $m = 80.0$, and $s = 1.0$. The threshold for the stop criterion of the simulations was set to $t^{(F)} = 0.995$. The maximum simulation time was $t_{\max} = 4.0$. A complete list of all parameters including the Gabor filter parameters can be found in appendix B.1.



(a) Centre points and polygons for the weight maps between thumb and fingers. The different dimensions of the polygons reflect the different flexibility of the fingers in relation to the thumb.

(b) Centre points and polygons for the weight maps between pairs of fingers. The fingers were not arranged in a straight line to mirror the finger positions in a relaxed hand posture.

(c) Modified Gaussian function which was mapped onto the polygon radii. The function had a minimum activation within the polygon area and dropped to zero outside of this area.

Figure 5.4: The weight matrices were constructed from polygons defining an area within a finger can move in relation to another finger. The activation in these areas was defined by rotating a Gaussian function around the centre points of the polygons and scaling the x-axis so that $x = 0.0$ was at the centre points of the polygons and $x = 1.0$ on the outlines of the polygons. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

The Weight Matrices The weight matrices $T_{SR}^{(f\quad g)}$ were designed based on the author's right hand. The matrices formed a crucial part of SAAM as the encoded the anatomical constraints of the grasping hand. The design of these maps was required on the one side to match the hand anatomy and on the other side to be suitable for the energy minimisation approach in the hand network. The latter required in particular that the maps for different pairs of fingers formed an at least roughly consistent grasp posture overall. To practically achieve this, the hand anatomy was encoded in the hand maps in a slightly simplified way. The activation in the weight matrices was derived from a Gaussian function to ensure a smooth transition from the strongly activated central regions of the map to the less strongly activated margins. Since the area in which two fingers can be positioned relatively to each other does not match

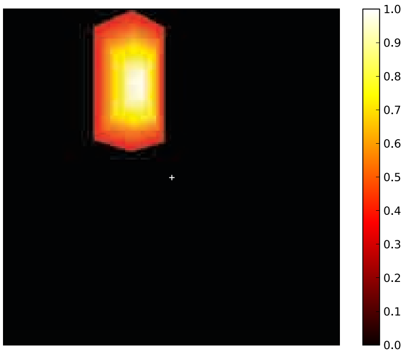


Figure 5.5: Example of a complete weight map showing the activation projected from the thumb (at the centre of the image) to the index finger.

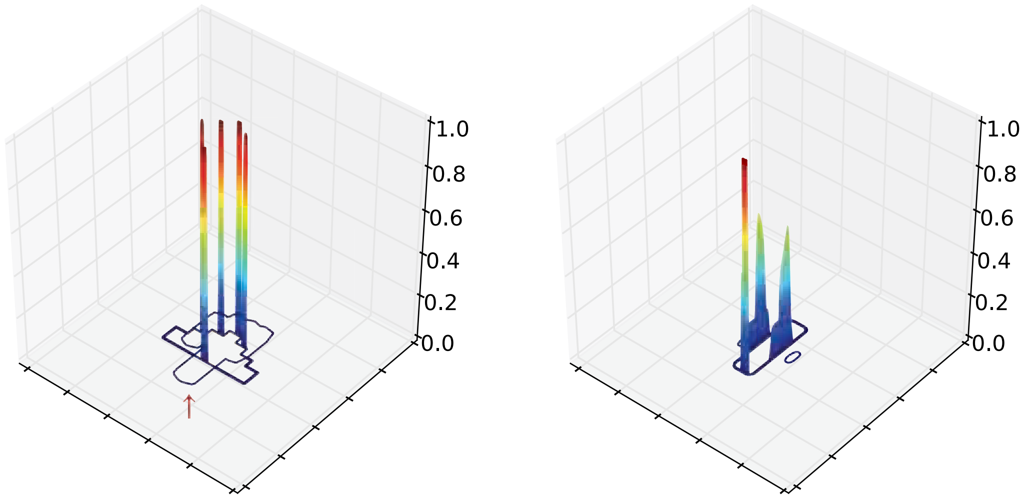
the elliptical outline of a simple two-dimensional Gaussian function, the weight maps were created by rotating a one-dimensional Gaussian function around the centre of a polygon while scaling the function so that its maximum was always at the centre point and the minimum at the edges of the polygon. The centre point described the most comfortable finger position, and the outline of the polygon delimited the area within the finger could be placed at all. Figure 5.4 shows the centre points and outlines used for the creation of the weight matrices. Figure 5.4c shows the modified Gaussian function which was mapped on the polygon radii. The Gaussian function was modified to produce a minimal activation at all points within the polygon:

$$f(x) = \frac{1}{1+\Delta}(e^{-\frac{x^2}{2\sigma^2}} + \Delta) \tag{5.17}$$

This function was mapped to the radii of the polygons so that $f(0)$ was at the centre point and $f(1)$ on the outline of the polygon. All values outside the polygon were set to zero. The values of the parameters were set to $\Delta = 0.5$ and $\sigma = 0.5$. Figure 5.5 shows an example of what the final weight maps looked like.

5.2.2 Results and Discussion

Simulations were run for all 24 experimental conditions shown in Table 5.1. First, the finger positions created by SAAM are reported, followed by a discussion of the reaction times of the model.



(a) Pyramid 90°

(b) Bar 90°

Figure 5.6: Two examples of the activation in the finger maps at the end of the simulations. For each simulation, the activations in all five maps were added and plotted in one graph. Additionally, the outline of the object in the visual field was drawn on the bottom plane of the coordinate system. In both graphs the position of the thumb is indicated by the peak on the bottom-left facing edge of the object. In the pyramid 90°-plot the other fingers followed a normal right-handed grasp. In the bar 90°-plot index and middle finger were on the left side of the object and the other two on the right side. Furthermore, the plot of the pyramid 90° shows an interesting artefact: The blue line pointed to by the red arrow indicates how the activation was projected between the finger maps by the weight maps. While the most activation concentrated on the selected finger position in each map, the inhibition process did not fully suppress the other positions which also received activation through the weight maps; this resulted in the blue line indicating a slight difference in activation of these positions in comparison to the positions which received no activation at all. Hence, the graph shows a weakly activated area in the shape of the activation pattern in the weight maps. This demonstrates how the weight maps project between the finger maps.

Finger Positions Figures 5.7 and 5.8 depict the activation in the finger maps at the end of the simulations. In addition, Fig. 5.6 shows 3D-plots of the final activations in the finger maps for two of the simulations. These plots demonstrate how the activations were distributed within the maps in successful simulations where clear finger positions developed (Fig. 5.6a) and in less successful ones where SAAM did not generate a clear grasp posture (Fig. 5.6b). Finally, Figure 5.10 shows the time course

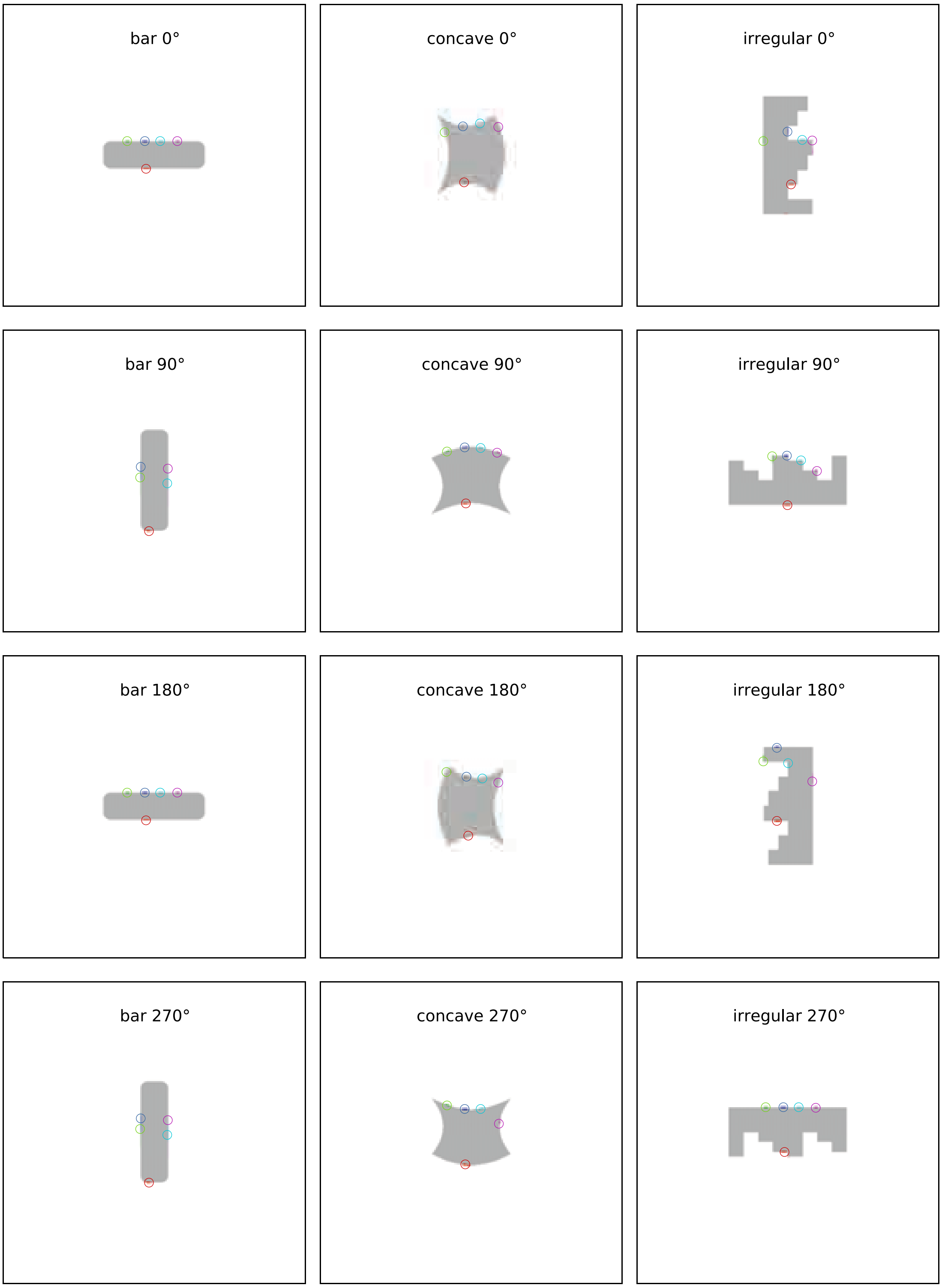


Figure 5.7: Finger positions produced by SAAM (part 1; thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink). The positions with maximum activation are highlighted with circles; the maxima are at the centre of the circles.

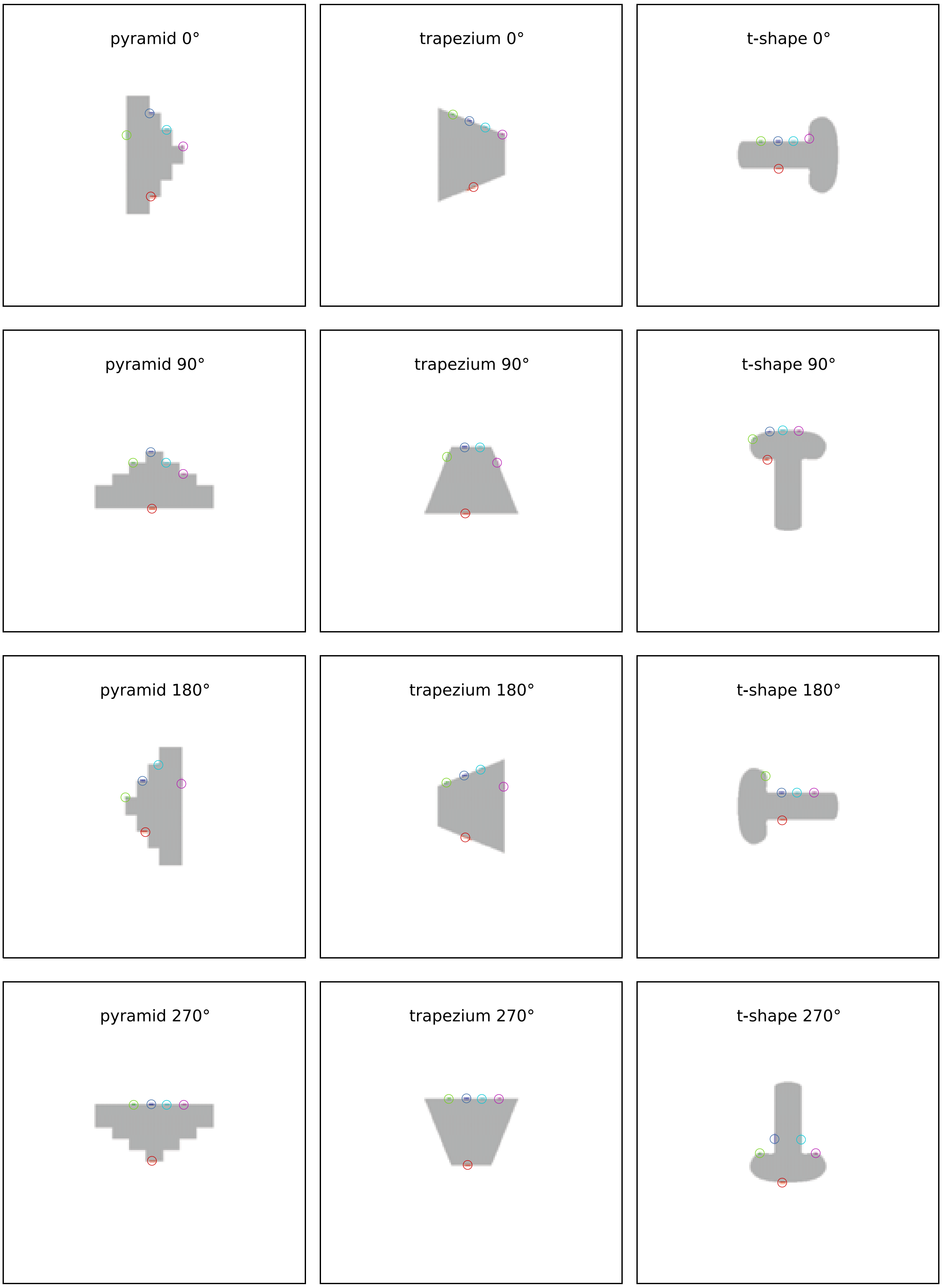


Figure 5.8: Finger positions produced by SAAM (part 2; thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink). The positions with maximum activation are highlighted with circles; the maxima are at the centre of the circles.

of the maximum level of activation in the finger maps for all simulations.

First, the simulation results are discussed with a close focus on the activation patterns which were observed in the finger maps. Scrutinisation of these patterns helped to understand the crucial role that the anatomical constraints in the weight maps play for grasp generation in SAAM. The weight maps are responsible for some characteristic behaviours of the model. To start with, the model did not select a unique position for each finger in the finger maps but a small region along the edge of the objects instead (see Fig. 5.6a, for example). One could now conclude that the WTA process in the hand network was not functioning as intended. However, a close examination of the factors involved in the calculation of the activations offers a different explanation. Due to the method with which the weight maps were constructed, they contained lines of constant activation which were parallel to the edges of their polygon-outline. If such an isoline of activation coincided with an edge of the object, then a whole area in the finger maps received a constant level of activation. This was, of course, additionally affected by overlaps of weight maps projecting from different positions and finger maps. This could alleviate the effect but this did not necessarily happen since many of the isolines in different weight maps had the same direction. Obviously, the WTA process can not select a winner in such a situation where no neuron has an advantage over all other neurons. The activations of the fingers on bar 90°/bar 270° (Fig. 5.7) are particularly good examples of coinciding isolines and edges. Figure 5.5 shows that the weight maps had long vertical isolines due to the shape of the polygons which defined their pattern of activation. This led to long lines of activation on the sides of the edge of the bar object which meant all these positions were equally feasible for grasping. This behaviour of the model provides additional information about the generated grasp postures: If a range of positions in the finger maps has equal activation, this means that there is not just one optimal finger configuration for grasping an object but a number of configurations which fit the geometric and anatomic constraints of the model equally well. By activating all these positions, the WTA process informs us that there are different ways to grasp an object, none which is preferable to the others.

In addition to not selecting unique finger positions, the vertical bar object illustrates another behaviour which emerged from the design of the weight matrices and edge detectors. The example weight map (Fig. 5.5) shows that the longer the isolines were the smaller was their activation; combined with the reduced activation of vertically oriented edges in the input this led to long stretches of small activation in the finger maps which are clearly visible in the output of the simulation. These indicate that with the given geometric and anatomic constraints the model was not able to generate a good grasp of the object. The soft-constraint satisfaction approach of SAAM allows the model to produce only weak activations in such conditions to express that with the given constraints the object may be grasped but that the grasp is not a good one.

The activation of regions instead of unique positions made it necessary to define how a unique finger position can be derived from the finger maps for further analysis. First, all positions with maximum activation were selected in each map. This reduced the size of the regions already heavily because, despite looking completely homogeneous, there still were slight differences in the activation of each position with the centre of the activated area having the maximum activation. The exact reason for this distribution of the activation remains unclear but it is most likely due to the modified Gaussian function which has a very small gradient at the top as well. For most objects this selection of positions with maximum activation resulted already in a single finger position. In simulations where more than one position had maximal activations the centre of gravity of all these activations was chosen as the actual finger position. The selected positions are marked by circles in Figures 5.7 and 5.8; the midpoints of the circles are at the selected finger positions. The activation at these positions is shown in Fig. 5.10.

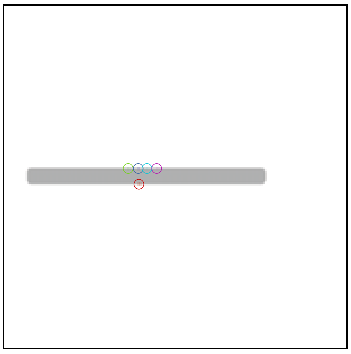


Figure 5.9: Grasp of a very long bar. This simulation shows that SAAM's preference for grasps around the centre of gravity is not due to the length of the bar. Please note that the plot has been resized for printing. The visual input was twice as large as in the simulation of the normal bar.

Overall, the results demonstrate that SAAM is able to select finger positions on a wide range of different objects. The positions seem to be anatomically feasible, i. e. the fingers were not swapped and the thumb was always placed below the fingers. Exceptions from this good performance are the outputs for bar 90°, bar 270° and, to a lesser extent, for irregular 0°, irregular 180°, pyramid 0°, pyramid 180° and t-shape 270°. These objects were, however, not well suited for the horizontal grasp postures which SAAM used. The distance between the objects' top and bottom edges was greater then the maximum grasp width of SAAM's 'hand' which is implicitly encoded in the weight maps between thumb and fingers. In the simulations of the pyramidal and irregular shapes as well as the t-shape the effects of the object length were smaller compared to the bar; small horizontal edges on these objects still presented opportunities for the model to create a horizontal grasp. The slowly increasing activations for some of the fingers in these simulations (see Fig. 5.10) indicate though that the model had difficulties finding finger positions on these objects.

In the bar 0° and bar 180° simulations, the grasps were placed around the centre of gravity (Fig. 5.9 shows that this also works with longer bars). Even though SAAM has no dedicated mechanism to extract the centre of gravity from an object. However, the activation of finger positions in the hand network increases the more distant a position is from the ends of an edge because this allows the weight maps to accumulate more activation from other finger positions. If, for instance, the thumb was placed directly at the bottom left corner of the bar, it would not receive any input from the index finger because the finger would no longer be placed on the bar at all. Hence, positions located closer to the centre of the bar are more desirable for the thumb (and also for the index finger which receives part of its activation from the thumb). Therefore, finger positions near the centre of edges are preferred. This is an emergent property of the model.

The grasps produced in the simulations for concave 0°, concave 180°, concave 270°, and trapezium 90° are good examples of how SAAM utilised the vertical edges for finger placement when the top edge of an object was not wide enough to place all four fingers along it comfortably. In such cases the model positioned the index and little finger on vertical edges and only left the middle and ring finger on the top edge. This behaviour automatically emerged when the top edge was too short to place all fingers in a straight line because the lateral fingers would then be placed in empty space where no activation from visual input would be supporting them. Thus, the vertical edges provided a better location for these fingers and, consequently, the model selected these positions. A similar result could be observed in the trapezium 180° simulation. There the top edge provided enough space for all fingers but the distance such a grasp would require between the thumb and the little finger would result in a less favourable grasp than placing the little finger at the side of the trapezium. Interestingly, this did not happen in the mirror-inverted simulation with the trapezium 0°. Here, the index finger was placed on the top edge and not on the left-hand side of the object. The reason for this difference becomes clear when one looks at the weight maps in figure 5.4a: In these maps the optimal vertical distance between the little finger and the thumb is much less then the optimal vertical distance between the index finger and the thumb. Furthermore, the finger/finger weight maps have a preference for edges with a slight slant towards the right. Together these two properties are responsible for the differing grasp postures in the two orientations of the trapezium.

These examples demonstrate how the anatomical constraints implemented in the weight maps guide the selection of finger positions in SAAM. I also showed that the inhibition introduced by WTA process plays a major role in the selection process. It has already been mentioned that this also affects the temporal behaviour of the model. This will be discussed in greater detail now.

Reaction Times Because of the stop criterion in the model, simulations terminated as soon as they had generated a grasp with a minimum level of activation for each finger. Since the activation can be understood as a measure of how the model judges the quality of a finger position, the stop criterion ensures that the model terminates as soon as it has produced a *good-enough*[4] grasp. Consequently, the duration of a simulation can be interpreted as its reaction time which is required in order to produce a grasp which can be potentially performed. Table 5.2 lists the duration of each simulation and Fig. 5.10 illustrates how the maximum activation in the finger maps developed over time. The table shows that the simulation duration varied considerably in the 24 conditions. Most simulations reached the termination threshold well within $t_{\max}$ but there are also six simulations (denoted by '> 4.0' in the table) which did not exceed the threshold level within the time limit. Except for the trapezium 180° condition, these are all simulations which were already discussed in the previous section because of the poor grasp postures they produced. The failure to produce reaction time results for these objects confirms the point that the model was not able to generate good grasps for these objects with the given the geometrical and anatomical constraints. This does, however, not apply to the trapezium 180° condition in which both, finger maps and maximum activation plots, indicate a good

[4] The meaning of *good-enough* depends, of course, on what the produced grasp is used for. If, for instance, a fragile piece of china is to be grasped, then a *good-enough* grasp must certainly meet a much stricter threshold than a grasp towards a plastic cup.

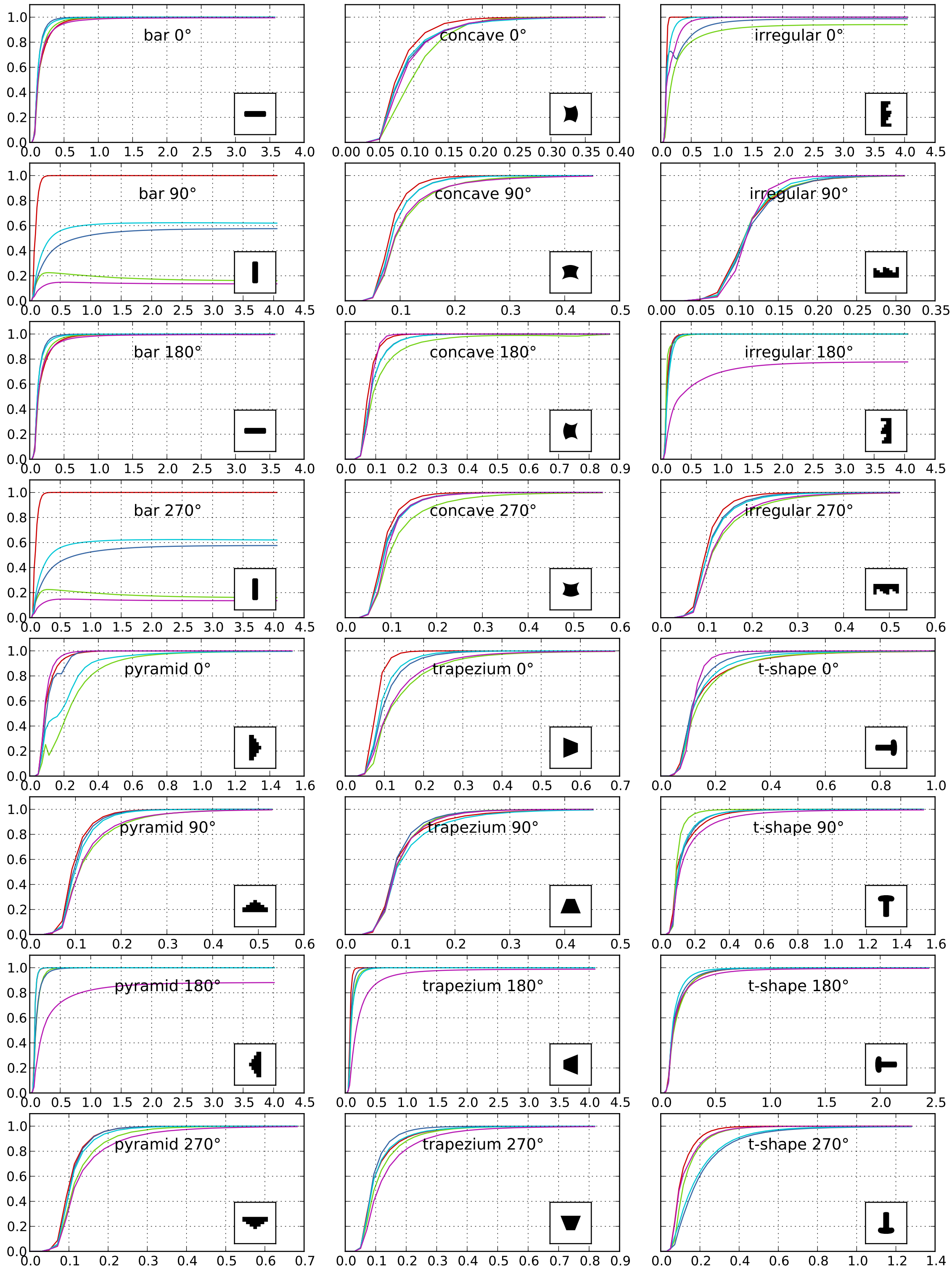


Figure 5.10: Time course of the maximum activation in the finger maps. Please note the different scales on the x-axes. The y-axes have the same scale in all plots. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

activation in all finger maps. Nonetheless, the activation of the little finger (which is 0.9885) stayed slightly under the threshold level and thus prevented the simulation from finishing. A similar behaviour can be observed in the bar 0° and bar 180° conditions; here the little finger stayed also below the threshold but other than in the trapezium 180° simulation, the activation finally increased enough to pass the threshold. A close examination of the time course of the activation for the little finger in the trapezium 180° simulation indicates that the same might happen there if the simulation ran longer.

	bar	concave	irregular	pyramid	trapezium	t-shape
0°	3.574	0.378	> 4.0	1.530	0.687	0.998
90°	> 4.0	0.451	0.311	0.530	0.452	1.532
180°	3.574	0.867	> 4.0	> 4.0	> 4.0	2.442
270°	> 4.0	0.562	0.522	0.682	0.818	1.279

Table 5.2: Duration of the simulations. Times written as > 4.0 denote simulations which were aborted because they reached the maximum simulation time $t_{\max}$.

The two horizontally oriented bars also highlight a characteristic of the reaction times: Thin objects like the two bars and the two horizontally oriented t-shapes tended to be grasped more slowly than larger objects. The plots of the time course of the activation show that the initial activation increase was quick and comparable to other objects but it did not reach the threshold levels. Only after some time the activation concentrated enough in one position so that the threshold was passed. This is likely to be an effect of small activations in the areas of the weight maps which encode small grasps.

Apart from slower reaction times for smaller objects, the durations of the simulations show no obvious patterns. It will be interesting to see if – given the relative simplicity of the objects – the same reaction time differences can also be observed in humans.

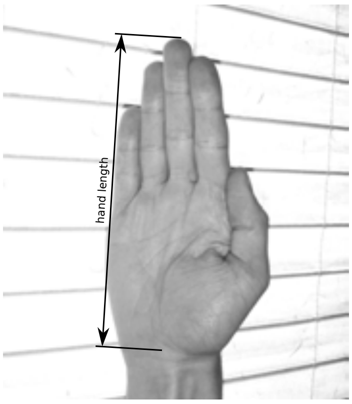


Figure 5.11: The length of the participants' hands was measured from the ball of the thumb to the tip of the middle finger.

5.3 Where Do Humans Grasp Objects? – An Experiment

We have seen that SAAM is capable of creating finger positions for grasping an object in the visual field. But do these finger positions match the grasps produced by humans? To answer this question and to gain a deeper insight into the positioning of individual fingers on simple objects during grasping, an experiment with human participants was conducted. In the experiment participants were asked to grasp objects resembling the stimuli used in the simulations and move them onto a glass plane where a picture of their grasp was taken. The finger positions were then extracted from the photographs, analysed and finally compared with the grasps produced by the simulations.

5.3.1 Method

Participants Seventeen participants (eleven female) took part in the experiment. Mean age was 24.7 years and the age range was 19 to 37 years. All participants reported to be right handed. The length of the hand of each participant was measured from the thenar to the tip of the middle finger to give an indication of hand size (see Fig. 5.11). The mean hand length was 16.71 cm and lengths ranged from 13 cm to 20 cm. All participants were postgraduate or undergraduate students at the

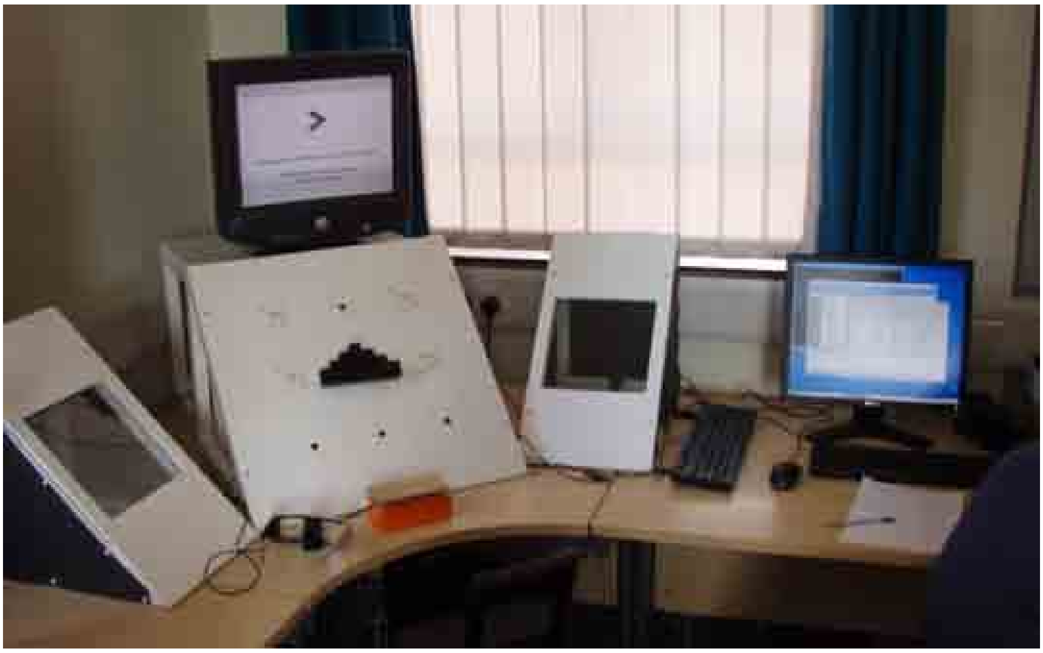


Figure 5.12: The experimental set-up. In the centre was a large board on which the object stimuli were presented; on both sides of it were two smaller boards with glass panes and cameras behind to take pictures of the grasps.

University of Birmingham. They received either course credits or £5 in exchange for their participation.

Apparatus Figure 5.12 shows the experimental set-up. Participants were seated on a stool with an adjustable seat in front of a white wooden board which was 60 *cm* wide and high and tilted at 45°. In the centre of the board was a base plate located which had a big centre pin to hold an object and a small adjustment pin to fixate the object in one orientation. Additionally, a micro-switch was embedded in the plate which was triggered when an object was placed on the board or removed from it.

A button box was positioned in front of the board. Participants were asked to place their hand on this button at the start of each trial. Above the board was a computer screen installed that displayed information to guide the participant and the experimenter through the experiment.

On both sides of the board were smaller boards (30 *cm* × 60 *cm*) set-up which had

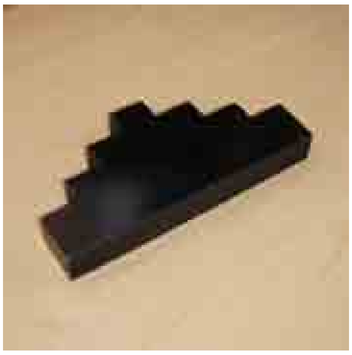


Figure 5.13: Example of the objects used in the experiment.

a glass plane in the centre which was 25 *cm* wide and high. The boards were also positioned at a 45° angle so that the subjects could comfortably place their hands on the glass planes. Behind the glass planes were web cams installed which allowed the experimenter to take pictures of the finger positions on objects placed on the glass plane.

The participants wore liquid crystal glasses which could be toggled between a transparent and an opaque state to control the participants' visual perception (PLATO glasses; Milgram, 1987). When being opaque the glasses appeared milky but did not shade the eyes. Hence, the perceived luminance did not change when the glasses changed state.

The computer screen, the micro-switch on the board, the button box, the cameras and the liquid crystal glasses were connected to a computer running E-Prime 2.0. A second screen and a keyboard were connected to the computer and placed next to the experimental set-up up so that the experimenter was able to comfortably control the experiment.

The experimental set-up allowed the experimenter to control when participants saw the object, measure when they started their grasp movement and when the object was picked up. The cameras recorded the finger positions used in the grasp of the object.

Materials Based on the shapes designed for the simulations (see Table 5.1) six wooden objects were made. The objects were 2.2 cm thick and their size was scaled to retain the relationship between the size of the hand and the size of the objects in the simulations. Due to different hand sizes in participants this could, of course, only be done approximately. The exact dimensions and weights of all objects are listed in Appendix A.1. They were painted black and had a number of holes at the back to allow mounting them on the board in different orientations. An example of the objects is shown in Fig. 5.13.

Design The experiment was a within-subjects design with 24 experimental conditions (six objects in four orientations, see Table 5.1). Each condition was presented four times resulting in 96 trials per participant. At the beginning of the experiment were 4 practise trials to familiarize the participants with the experimental procedure. There were no planned breaks in the experiment but participants could ask for a break at any time.

Procedure Participants completed a short questionnaire specifying their age, gender, handedness and hand length. Then they read instructions explaining the experiment and their task (see Appendix A.2 for the exact instruction text). Additionally, the experimenter ensured that participants understood that the objects should be grasped with the fingers and not clenched using the palm.

At the beginning of each trial a message appeared on the screen asking the participants to focus on the board and warning them that the experimenter will close the glasses. Once the glasses were closed the experimenter placed one of the objects on the board in front of the participants. Afterwards the participants were asked to press down the button with the palm of their right hand (see Fig. 5.14). After a 2000 ms delay the glasses opened. As soon as the participants lifted their hand from the button the glasses were closed again so that the grasp movement had to be performed blind. This was done in order to prevent participants from letting the button go first and only then decide which grasp to perform. Furthermore, it allowed to distinguish two components in the participants' reaction: pre-grasp planning and grasp-execution. This is important because SAAM only models grasp planning based only on information available prior to grasp execution. By closing the glasses after the grasp movement was initiated, corrections of the planned grasp posture based on online updates of the grasp plan during grasp execution should be minimized in the experiment. Once the participants had lifted the object of the board the glasses were opened again and an instruction to place the object in front of one of the two cameras appeared on the screen. The left and right cameras were chosen randomly to ensure that participants did not know beforehand to which board they would have to move the object because this could have affected their grasp. This is known as the 'end-state comfort'; Rosenbaum observed that grasps of objects can be influenced by the final position to which an object is moved to in a subsequent move-action after the object has been grasped. He argued that grasp postures are chosen so that the grasps are most comfortable at final position of the movement (see Rosenbaum, Halloran, & Cohen, 2006 for details about this effect). Since the focus of this study (and SAAM in general) was not on the effects of action plans on grasp postures but on the effects of object affordances on grasps, the experiment was designed to prevent the participants from planning sequences of actions. Finally, at the end of the trial the experimenter took a photograph of the grasp used by the participants and they gave the object back to the experimenter.

The experiment lasted approximately 45 min. After they completed the experiment, participants were debriefed and received course credits or money.

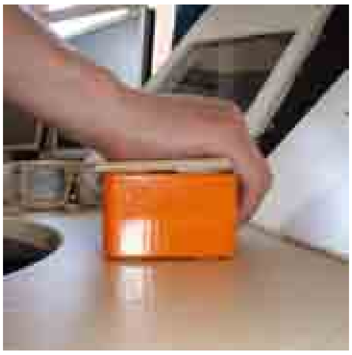


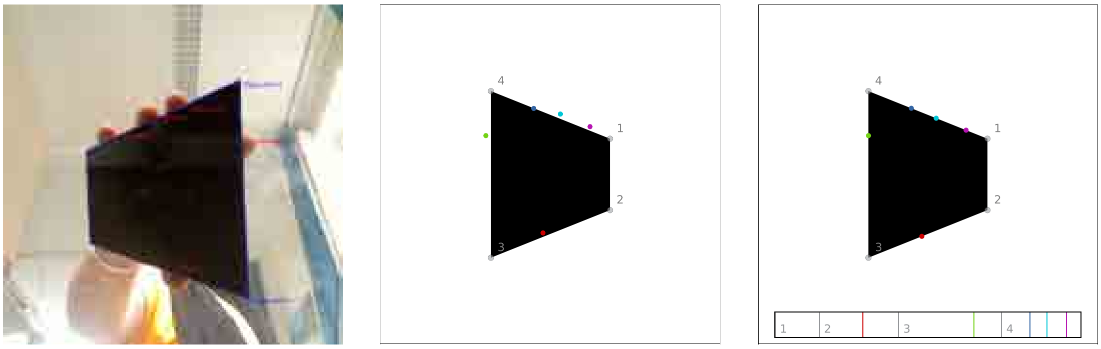
Figure 5.14: Initial hand position at the beginning of each trial in the experiment.

5.3.2 Results and Discussion

The experiment was analysed in two ways: First, an analysis of the grasps used for the different objects was performed. Second, pairwise comparisons of the reaction times in the different conditions were calculated.

Analysis of Finger Positions In order to perform an analysis of the grasps, the positions of the fingers and the objects were marked in the photographs taken during the experiment (see Fig. 5.15a). Using two marked locations on the object, the finger positions were then transformed[5] so that they could be drawn on a standardised version of the object (see Fig. 5.15b). Finally, the finger positions were mapped to the nearest edge of the object. This step was introduced because in the simulations the finger position is (implicitly) defined as the position where the finger touches the object. The mapping step harmonises the meaning of finger position in the simulations and the experiment and eliminates variations introduced by inaccurate marking of fingers in the experiment. Moreover, the mapped finger positions can be represented on a one-dimensional scale by measuring their position along the outline of the object. This representation was developed because it simplifies the comparison of grasps which will prove particularly helpful for comparing the experimental data with the

[5] The transformation included only translation, scaling and rotation. Thus, complex distortions could not be corrected for. However, due to the set-up of the cameras and the position of the objects on the glass plane these distortions were minimal and not deemed a problem for further analysis.



(a) Picture showing the marked positions of the fingers and the object. The pictures were taken from behind the object; therefore, the photo and the marked finger positions are mirror-inverted.

(b) The positions extracted from the photograph were mirrored and transformed so that they are aligned to a standardised representation of the object. The object is presented viewed from above in the same orientation in which the participants saw it.

(c) The Finger positions were then mapped to the outline of the object. The inset shows the position of the fingers on a linear plot of the object outline. The numbers indicate how the linear plot maps to the outline of the object. They were arbitrarily assigned to the corners of the object.

Figure 5.15: Analysis of the photographs collected in the experiment. The object shown in this example is trapezium 0°.

simulation results. Figure 5.15c shows an example for a completely processed trial; the inset illustrates how the finger positions are distributed along a linear plot of the outline of the object.

Finger Positions The finger positions for all 24 conditions are shown in Fig. 5.16 and 5.17. Overall, the results showed that there were one or two preferred ways of grasping each of the objects. Since SAAM was restricted to horizontal grasps only, the grasps in the experiment were grouped into horizontal and vertical grasps. The grasps were classified as horizontal if the thumb was placed on a downwards facing edge and as vertical if the thumb was placed on a leftwards facing edge. In the t-shape conditions eleven grasps were excluded from the analysis because the thumb was

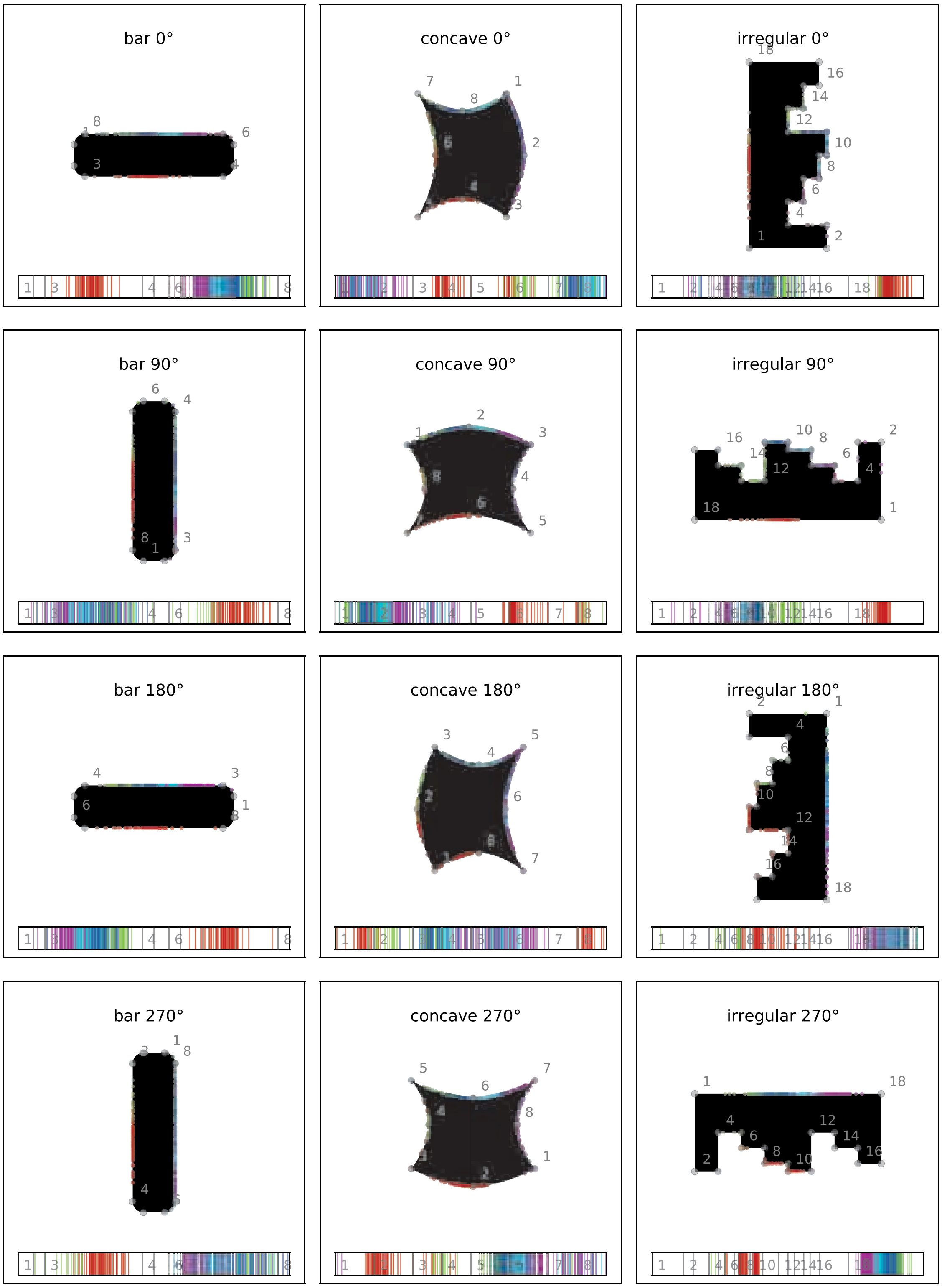


Figure 5.16: Grasps observed in the experiment (part 1). See Fig. 5.15c for an explanation of the linear plots. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

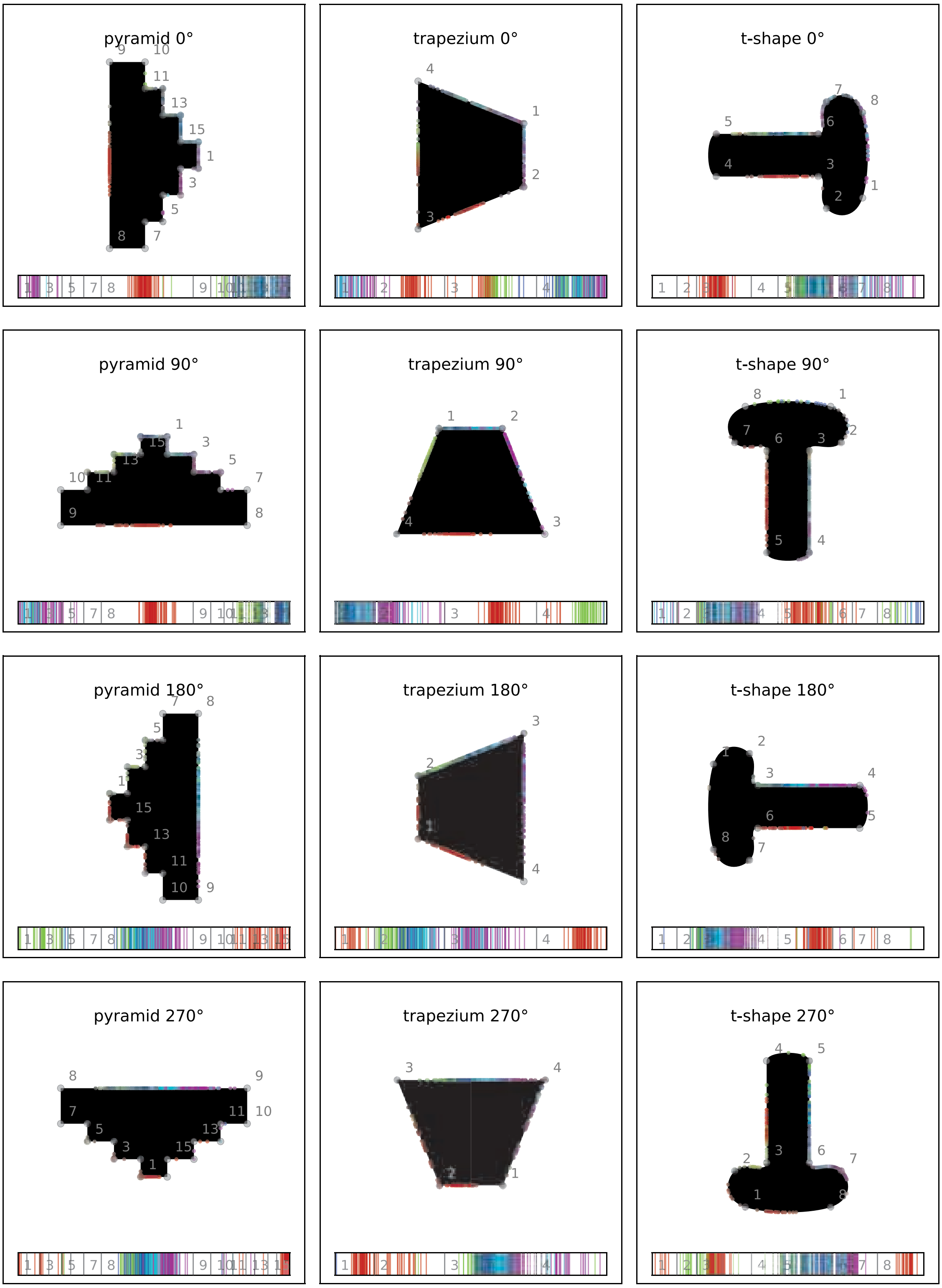


Figure 5.17: Grasps observed in the experiment (part 2). See Fig. 5.15c for an explanation of the linear plots. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

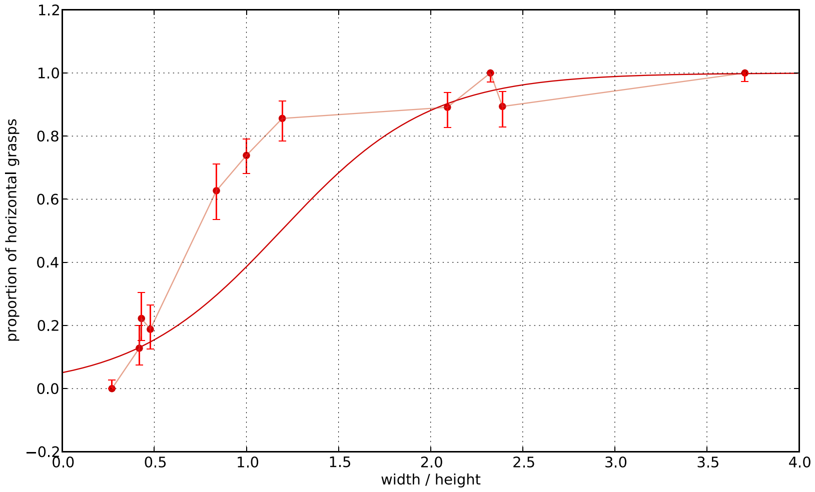


Figure 5.18: This graph shows the proportion of horizontally oriented grasps versus the ratio of object width to height. If an object was not grasped horizontally, it was grasped vertically (other grasps were excluded from the analysis). Objects with ratios less than one were presented in a vertical orientation, objects with a ratio greater then one were horizontally oriented.

placed on the rounded parts of the shape (i. e. the line segments between nodes 1–2 and 7–8 in the plots of the t-shape in Fig. 5.17) which can be classified neither as horizontal nor as vertical. The data showed that in each condition one of the two grasp orientations was preferred depending on the height and width of the objects. To further investigate this relationship between the likelihood of a horizontal grasp and the ratio of object width to height, a logistic model was fitted to the data. Logistic regression analysis was performed yielding

$$\ln\left(\frac{Y}{Y-1}\right) = -2.933 + 2.469r \tag{5.18}$$

with Y being the predicted probability for a horizontal grasp and r being the ratio of object width to height (Table 5.3; see A.1 for details of the definitions of object width and height).

According to the logistic model, the log of the odds for a horizontal grasp is positively related to the width/height ratio of the grasped object ($p < 0.001$; Table 5.4).

	bar	concave	irregular	pyramid	trapezium	t-shape
0°	3.710	0.840	0.420	0.480	1.000	2.320
90°	0.270	1.190	2.390	2.090	1.000	0.430
180°	3.710	0.840	0.420	0.480	1.000	2.320
270°	0.270	1.190	2.390	2.090	1.000	0.430

Table 5.3: Width-to-height ratios of all objects and orientations. Values less than one indicate vertically oriented objects; values greater than one denote horizontally oriented objects.

This means larger ratios (i. e. horizontally oriented objects) were linked with higher proportions of horizontal grasps while smaller ratios (vertically oriented objects) were linked with higher proportions of vertical grasps. This relationship is shown in Fig. 5.18; detailed results of the analysis are summarised in Table 5.4. The analysis of grasp orientation confirmed the observation that the grasp orientation depended on the object width and height. This is an important result considering that SAAM only implemented horizontal grasp postures. I will return to this in the comparison of simulation and experiment.

Predictor	β	Wald's χ^2	*df*	*p*
Constant	−2.933	286.914	1	< 0.001
W/H-Ratio	2.469	258.106	1	< 0.001
Test		χ^2	*df*	*p*
Likelihood ratio test		814.949	1	< 0.001

Table 5.4: Logistic regression analysis of width-to-height ratio and grasp orientation. The analysis was conducted with SciPy 0.8.0 and statsmodels 0.2.0.[6]

Apart from the discrimination between vertical and horizontal grasps, the results give an impression of the variation observed in grasping. The fingers were clearly not positioned arbitrarily but there was still a fair amount of variation in the choice of finger positions. For example, in the pyramid 90° condition the index finger was not

[6] SciPy is available on `http://www.scipy.org/`. The statsmodels toolkit can be retrieved from `http://scikits.appspot.com/statsmodels`.

only placed on varying locations on one step of the pyramid but even on different steps altogether. Similar variations can be found in the other orientations of the pyramid as well as in the conditions which showed the irregular object. Since the other objects had longer edges, variations in finger positioning on these objects only affected the position of the fingers along the edges but not the chosen edge segment. For example, in the bar 0° condition the fingers were always placed on the same two edges and only the exact position varied. Though, to some extent an exception from this were the concave and trapezium objects because they had a width-to-height ratio close to 1.0. The analysis of horizontal and vertical grasps showed that such objects were grasped horizontally as well as vertically. This obviously produced finger positions on at least two edges depending on whether the participants used a horizontal or vertical grasp. The grasps on the concave 180° and the trapezium 180° exemplified this. The thumb, for instance, was placed at the bottom edge as well as on the left edge of the object. However, on each edge the variation of the finger positions was similar to those in the other objects.

An interesting result is the location of the fingers on the concave and trapezium objects. In the description of the visual feature extraction stage of the SAAM (Sec. 5.1) I discussed the possibility of placing fingers not only on the top edges of an object but also on the sides. The finger positions on the trapezium and the concave objects prove that this way of positioning the fingers is not merely a theoretical option but is actually used by humans. Trapezium 90° and trapezium 180°, in particular, illustrate how fingers were moved to vertical edges of the object if the space on horizontal edges was not sufficient for all fingers.

Reaction Times Two reaction times were measured. The first measure was the time from when the glasses became transparent until the participants released the orange button (RT 1). The second measure was the time between releasing the button and lifting the object (RT 2). The first reaction time measure can be interpreted as preparation time or planning time for the grasp while the second one is essentially the duration of the reach-to-grasp movement.

Both reaction time measures were analysed separately. For both datasets pairwise comparisons of all conditions were carried out using related-samples t-tests with Bonferroni adjustments to compare reaction time differences between the different objects and orientations. The reaction times from each participant were averaged using the median. The full results can be found in Appendix A.

The comparison of the RT 1 times revealed no significant pairs, and the RT 2 times showed only two significant pairs (of 276). A reason for this could be the relatively low number of samples per condition (4 per participant × 17 participants). Furthermore, it needs to be considered that all objects had a good size for grasping and clearly identifiable surfaces for placing the fingers on which made grasping them easy despite their different shapes. Thus, the effect of the different orientations and objects on the reaction times might be very small. Since I mainly focused on the analysis of finger positions in my work, the comparison of reaction time differences in the experiment and the simulations was not pursued any further. However, it needs to be noted that the absence of significant effects in the reaction time data is not an indicator that the data describing finger positions are not significant. While the planning and execution times for grasps might only show very little differences for various objects, the final grasp posture can still be very different depending on the object shape as the analysis in the previous section showed. Therefore, the finger positions can safely be analysed despite the lack of significant reaction time data.

5.4 Comparison of Simulated Grasps and Experimental Results

SAAM aims at reproducing human grasp postures. Hence, the grasps of objects it produces should be similar to the grasps humans make on the same objects. To verify this, the simulation results were compared with the experimental data. In order to do this, a metric for the similarity of grasps was developed. This metric was then used to measure how similar SAAM's grasps were to the grasps observed in the experiment. The comparison is accompanied by a comparison of the significant reaction time results from the experiment with the reaction time differences observed in the simulations.

5.4.1 A Similarity Metric for Grasps

It is reasonable to assume that two grasp postures are similar if the fingers are in similar positions in both grasps. This means, the smaller the distances between all pairs of corresponding fingers are in two grasps, the more similar are these grasps. Here, the distances between the fingers are measured along the perimeter of the objects (these are effectively the distances between the lines plotted on the insets of Fig. 5.16 and 5.17). A Gaussian function is then used to normalise the distances measured between pairs of corresponding fingers. This function yields one if two fingers are in exactly the same position and declines towards zero with increasing distance between two fingers:

$$q(f_1, f_2) = e^{\frac{-h(f_1 - f_2)^2}{2\sigma^2}} \tag{5.19}$$

with

$$h(x) = \begin{cases} x & \text{if } x < l/2, \\ l - x & \text{else.} \end{cases} \tag{5.20}$$

The parameters f_1 and f_2 denote the positions of two fingers on the object perimeter and l the length of the object perimeter. Since the distance between two points on a closed outline can be measured in two directions, the helper function $h(x)$ was introduced to always select the shorter of the two distances. The parameter σ of the Gaussian bell curve defines the range of acceptable distances.

Since *all* corresponding fingers should be in similar positions, the scores of all finger pairs are multiplied to calculate the similarity metric for two grasps g_1 and g_2:

$$Q(g_1, g_2) = \prod_{f=1}^{5} q(g_1^{(f)}, g_2^{(f)}). \tag{5.21}$$

To compare a grasp g with a set S of other grasps, e. g., when comparing a simulation result with the grasps made in the experiment, the mean of the similarity scores between g and each individual grasp in S is calculated:

$$Q(g, S) = \frac{1}{S} \sum_{s} Q(g, s) \tag{5.22}$$

The similarity metric rates each grasp on a scale between 0 and 1. A score of zero means that a grasp is unique and no grasp in S is in similar to it. A score of one, on the other hand, means that a grasp perfectly matches all grasps in S. This essentially implies that all grasps in S and the grasp g have all their fingers at exactly the same positions.

It is not only possible to compare a single grasp with a set of other grasps but also to compare all grasps within the set with each other. This is achieved by calculating the similarity of each grasp in the set with the rest of the grasps in the set. Hence, for each $g \quad S$ the score $Q(g, S \quad g\ )$ is calculated. This yields a list of scores characterising the distribution of the grasps in S on the object. Many high scores in this list indicate a high similarity of the grasps in the set. The absence of any high scores on the other hand indicates a high variation of grasp positions because no grasp has many others in its vicinity.

In the analysis the parameter σ of the Gaussian bell curve in the similarity measure was derived from the variations of the finger positions the experiment. In each condition the standard variation of the finger positions of each finger was calculated; σ was then set to the mean of the standard variations. The standard variation was chosen as the basis for the calculation of σ because it ensured that the similarity measure accepted grasps in a vicinity which had a width that corresponded to the width of the distribution of the grasps in the experiment. The reasoning behind this is that if grasps in the experiment varied widely then similarity should be less strictly defined compared to a set of grasps with very small variation. To make the similarity score comparable across conditions the mean of the standard variations was chosen as the final value for σ. It shall be noted that the similarity scores were robust for different values of σ. The scores differed in size but the relation between two scores did not change.

5.4.2 Results and Disussion

Finger Positions Figures 5.19 and 5.20 show how the grasps produced by SAAM compared to the grasps observed in the experiment. To create these charts, the grasps from the experiment were used to create a reference set for each experimental condition. Then, the grasps in each set were scored compared to the other grasps in the same set as described in the previous paragraph. Finally, the scores of the SAAM's grasps were calculated using the experimental reference sets. The scores of the simulation results are coloured in red and highlighted with an arrow if necessary. All scores were sorted by value so that the grasps with the highest similarity scores are at the right side of the graphs. Thus, the further a simulated grasp is located on the right side of a chart the more similarity it has to the grasps from the experiment. The two charts below each plot of grasp scores show the finger positions of the simulated grasp (top plot) and the finger positions from the experiment (bottom plot).

SAAM can only perform horizontally orientated grasps. However, the experimental results revealed that humans preferred to grasp the objects in some of the conditions with vertically oriented grasps. The comparison charts (Fig. 5.19 and 5.20) show that the SAAM performed particularly bad on objects which were vertically oriented. This coincides with the predictions made by the grasp orientation model (see Eq. 5.18). Figure 5.21 demonstrates how the probability for a horizontal grasp as predicted by the logistic model (see Sec. 5.3.2) is linked to the performance of SAAM which was defined as the similarity score of the simulation divided by the observed maximum score per condition. The graph clearly shows that a poor performance of SAAM is linked to objects which humans preferred to grasp with a vertical grasp. However, SAAM's limitation to horizontal grasps forced it to some form of horizontal grasp which was obviously different from the vertical grasps humans used. As it has been pointed out already, the model in such cases often only produced low activations thereby indicating that the generated grasp posture was not very good.

Apart from providing an explanation for the poor similarity of some of SAAM's grasps, Fig. 5.21 also shows that the simulation results in the remaining conditions were very good. The simulated grasps consistently scored high in relation to the observed maximum score.

Reaction Times Since the pairwise comparisons of RT 1 times from the experiment showed no significant results, the temporal behaviour of the model was not compared

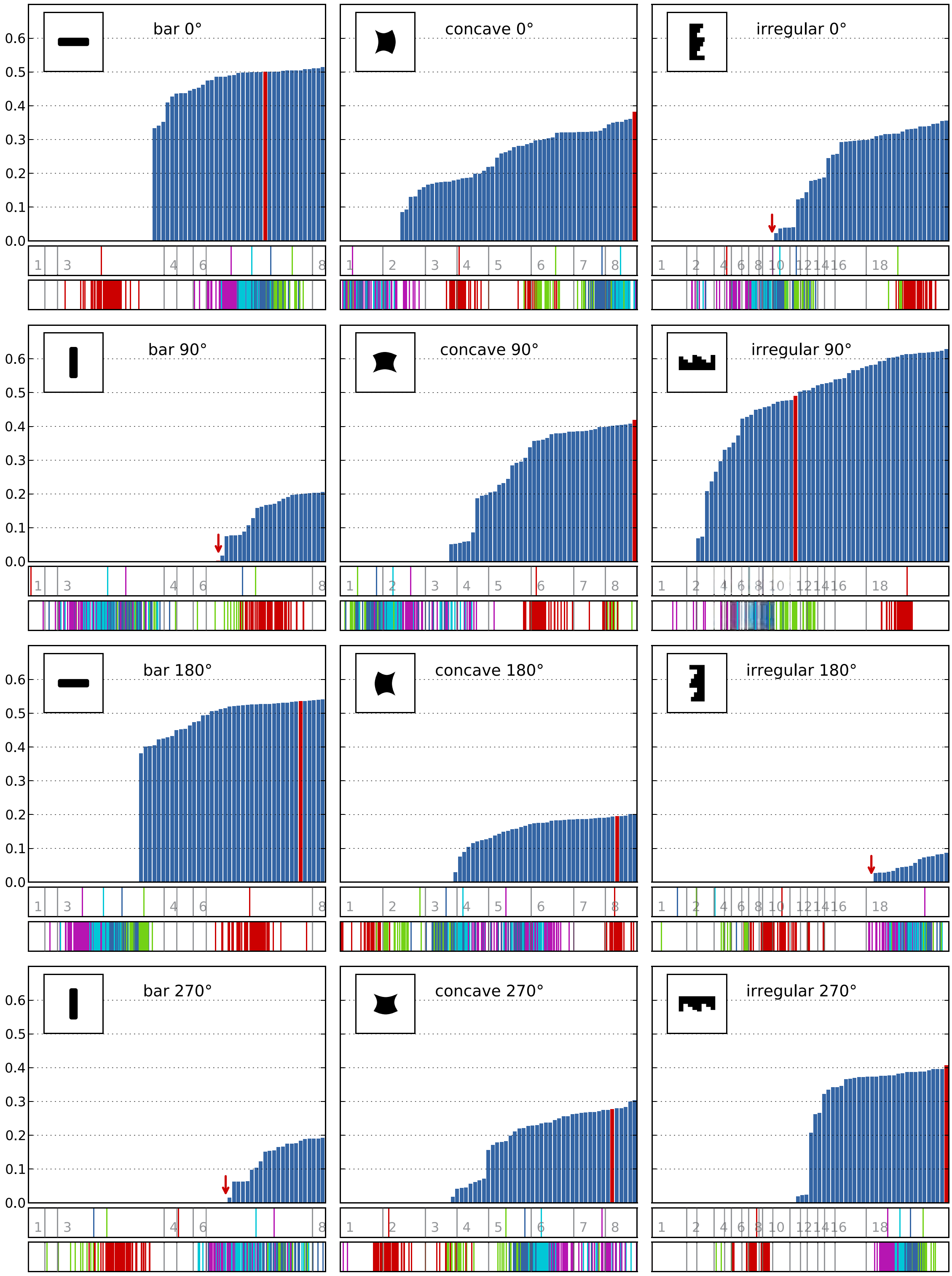


Figure 5.19: Comparison of experimental data and simulation results (simulation results are marked red; part 1). The linear plots below the main charts illustrate how the simulated finger positions (top) compared to the finger positions in the experiment (bottom).

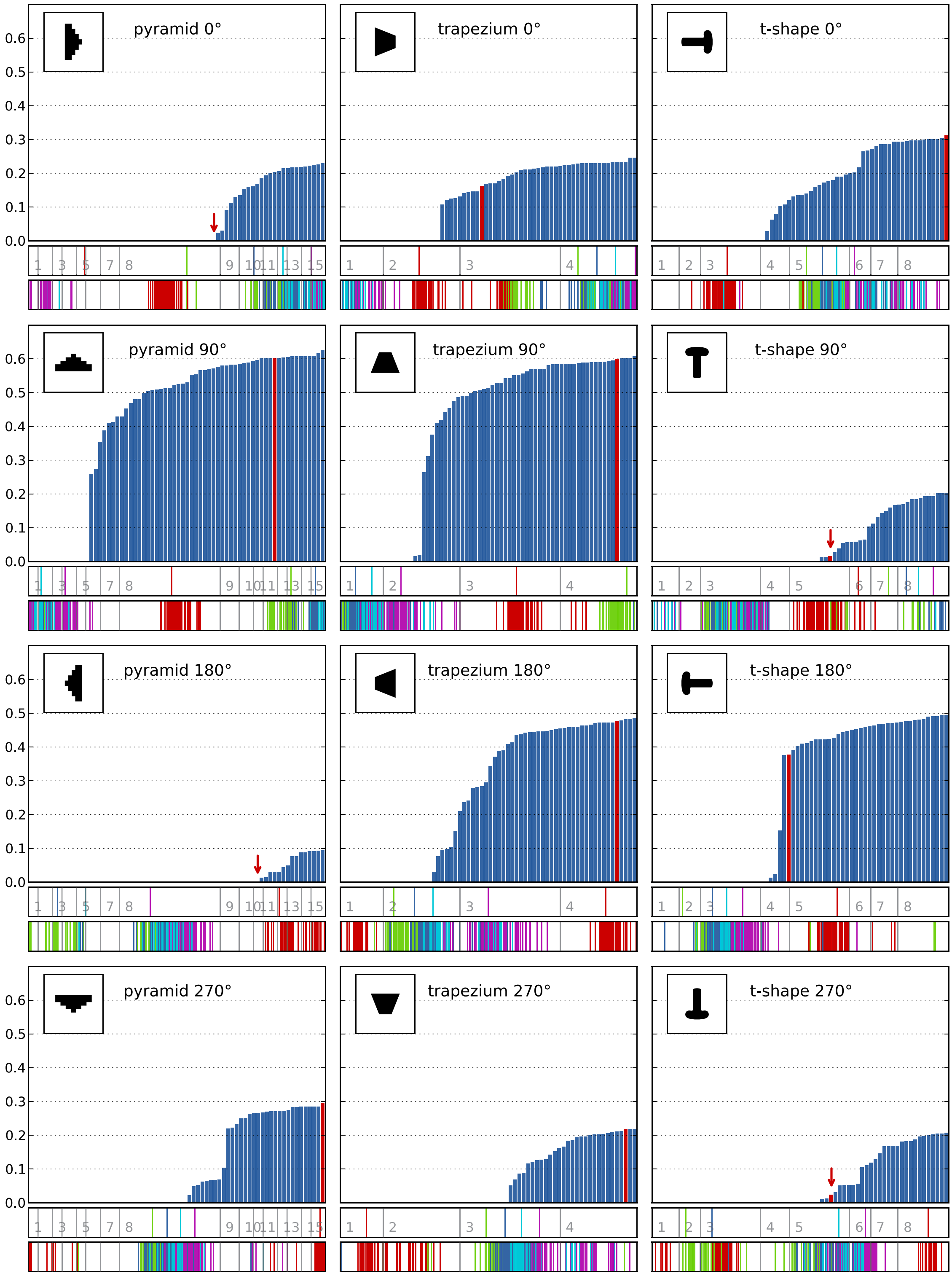


Figure 5.20: Comparison of experimental data and simulation results (simulation results are marked red; part 2). The linear plots below the main charts illustrate how the simulated finger positions (top) compared to the finger positions in the experiment (bottom).

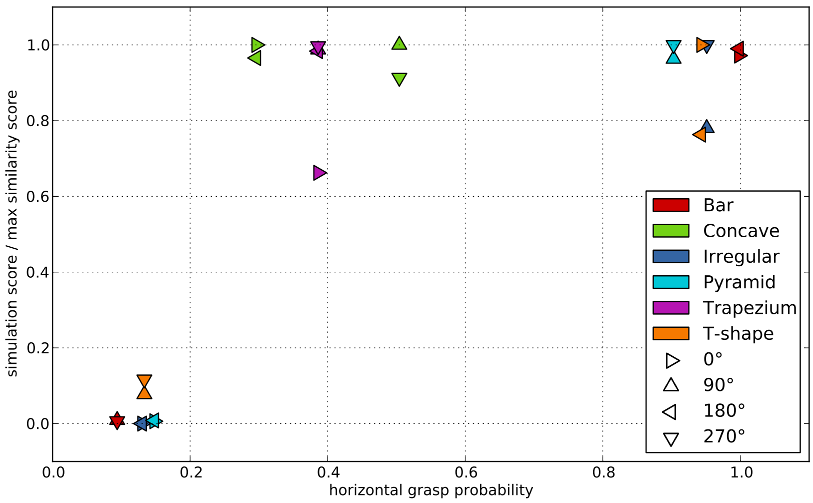


Figure 5.21: This graph compares the probability for a horizontal grasp as it is predicted by the logistic model (see Sec. 5.3.2) with the similarity score that the simulation yielded in relation to the observed maximum similarity score.

with the experimental data.

5.5 Summary

Study 1 showed that grasps created by SAAM were, in general, anatomically feasible. The comparison with the experimental data confirmed this; SAAM produced grasps which had a high similarity to the grasps made in the experiment. On vertically oriented objects, however, SAAM's results differed from the experimental data. While humans were able to rotate their hand and grasp such objects with a vertical grasp, SAAM was restricted to horizontal grasps. Hence, the grasps it produced on vertical objects had not much similarity with human grasp postures. Despite this, SAAM's grasps were not impossible for humans to make, though, and still anatomically feasible. A possible experiment could investigate whether a restriction to horizontal grasp postures would lead humans to make similar grasps on vertically oriented objects as SAAM did.

In the simulations with vertically oriented objects, the activation in the finger maps also developed more slowly than in simulations with horizontally oriented objects. Additionally, these simulations showed how the model handled inputs where a number of equally feasible grasps was possible. In these cases the model resorted to activating all positions which might be part of a grasp. In simulations where no good grasps could be found on the input, the activation of the selected positions was reduced, though, indicating that the model judged these positions not as optimal for grasping because they did not match the anatomical constraints very well. This can be understood as a rating of grasp quality: A grasp which requires an awkward grasp posture is more likely to fail (i. e. the seized object is dropped) than one which is comfortable.

SAAM's ability to express how good it considered a grasp to be, allowed to use a threshold-criterion to stop the simulations once a *good-enough* grasp was generated. A *good-enough* grasp is a grasp which satisfies the anatomical constraints *well enough* so that it is unlikely that the grasp would fail during its execution. Depending on the task the acceptable risk of failing can, of course, differ. The durations of the simulations matched the observations described above. When SAAM was able to create a good grasp, the simulation duration was short. When SAAM could not create a good grasp, the simulations took longer and in some cases did not terminate at all before the cut-off time. The simulation durations also showed that thin objects (i. e. the bar and the t-shapes) were grasped more slowly than bigger objects. This is in line with the observation that precision grasps are slower than power grasps (Derbyshire, Ellis, & Tucker, 2006) assuming that objects handled with precision grasps are usually smaller than objects handled with power grasps.

In the experiment that accompanied the simulations no significant reaction time differences were observed for grasp planning (RT 1). The reaction times for grasp execution (RT 2) showed only two significant pairs of reaction times. Because of this low number of significant results, the reaction time data were not analysed any further.

The simulations revealed two notable properties which emerged from the constraint satisfaction process: First, SAAM's grasps of the horizontally oriented bar were placed around its centre of gravity. This is remarkable since SAAM does not contain any mechanism to recognise an object's centre of gravity. This behaviour emerged solely from the constraint satisfaction process in the hand network. Second, the model was able to place fingers on the vertical edges of the object when the space on the top edges was not sufficient to place all fingers there or when the grasp span would be to large. As a side effect this also improved the stability of the grasp postures because object movement was restricted in all four directions (two-dimensional form-closure; Bicchi, 1995). The experiment showed that this behaviour of the model was not artificial but matched the behaviour of humans who also produced form-closure grasps on the same objects.

It is important to note that SAAM has no notion of the forces and torques which need to be balanced to achieve a stable grasp of an object (see, for instance, Mason, 2001 for examples on the calculation of forces in grasps). In SAAM the geometrical and anatomical constraints implicitly ensure that the forces exerted by the thumb and the fingers form a stable opposition space. The geometrical constraints, for example, ensure that thumb and fingers are located on opposite edges of the object. The ability of the hand network to grasp objects around their centre of gravity is another example of how SAAM implicitly constructs a stable grasp. This method of finding stable grasp points is much simpler than an accurate calculation of the positions by taking into account all forces exerted by fingers. The comparison with the experimental data showed that it is nonetheless sufficient to identify stable grasp points.

6 Study 2: Grasp Classification and Two-Object Simulations

In Study 1 I demonstrated that SAAM can generate finger positions for grasps of single objects and that these grasps are similar to the grasps made by humans. However, a central aspect of my work is selective attention for action. Hence, I will now turn to exploring SAAM's abilities to produce grasps for objects in multi-object situations. By producing a grasp for one of multiple objects in the visual input SAAM performs a selection which can be interpreted as attentional behaviour.

In addition to investigating the emergence of selective attention, I will also extend SAAM to classify the grasps it produces. This extension enables SAAM to recognise different types of grasps, for example, precision and power grasps. Grasp classification makes attentional selection in SAAM twofold: Different grasp locations compete as well as different grasp types. I will first demonstrate with single-object simulations that grasp type selection works in SAAM. Then I will present results of multi-object simulations showing that selection of grasp location and selection of grasp type also work together.

Grasps can be chosen by intentions. In SAAM the grasp classification module has a top-down path which can be used to model this. By preferring one grasp type over others an intention of grasping an object in a specific way can be expressed. Study 2c explores this behaviour.

6.1 Extending SAAM: The Grasp Template Network

The classification of grasps in SAAM is implemented with templates describing the characteristic finger positions for specific types of grasps. Grasp types can be derived from the grasp taxonomies discussed in Chapter 2. SAAM's restriction to two dimensions however limits the choices for different grasp types. Nonetheless, grasp postures like the precision grip or power grip can be approximated by creating grasp types for grasps with either a small or a wide grasp span. The templates describe the characteristic finger positions of specific grasps types. Obviously, the templates need to be independent of the location of the absolute finger positions in the visual input. Hence, the templates need to define the positions of the fingers in a certain grasp type relatively to each other. This is achieved by constructing template maps in the same way as the weight maps for the anatomical constraints. However, the template maps do not encode the anatomically feasible spatial relationships between pairs of fingers but the relationships which are characteristic for a certain grasp type. In other words: the templates define characteristic configurations for pairs of fingers. A template is selected if the activation in the finger maps matches the finger configuration defined in the template.

The templates are part of a grasp template network. This network is linked to the hand network through a bottom-up path and a top-down path. The bottom-up path from the hand network to the grasp template network implements the classification capabilities of the model. It can be used independently of the top-down path. The top-down path provides the reverse connection. It can be used to manipulate the choice of finger positions in the hand network based on the activated grasp template. I will first describe both pathways qualitatively before going into the mathematical details of their implementation in the model.

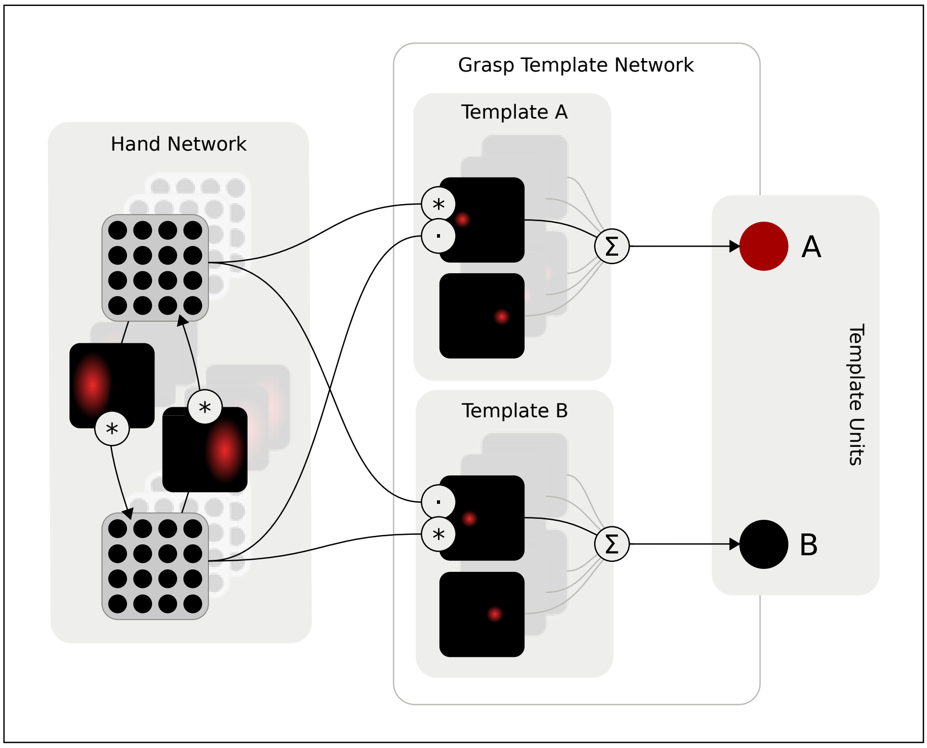


Figure 6.1: Integration of the bottom-up path of the grasp template network into the original model. For clarity only two of the five finger maps in the hand network are shown with the excitatory connections between them. Furthermore, only one connection into the grasp template network is shown per finger map. In the real model each finger map has one link to the grasp template network for each excitatory connection it has with another finger map. In the figure only two templates are shown. This can, of course, be changed to as many templates as necessary.

6.1.1 Bottom-Up Path

Figure 6.1 illustrates how the bottom-up path of the grasp template network connects with SAAM's other components. The new network has one output unit per grasp type. Each grasp type is defined by a template consisting of a set of template maps. These maps describe those relative positions of the fingers which constitute a grasp of the encoded type. Thus, the template maps are similar to the weight maps in the hand network; only that they describe the relative positions of the fingers allowed in a certain grasp type instead of all anatomically feasible finger positions. The template maps are defined for the same pairs of fingers as the weight maps in the hand network (see Fig. 5.3). The activations of the output units of the grasp template network depend on how well the grasp postures described in the templates match with the finger positions generated by the hand network. This matching is calculated in three steps: First, each template map is applied to one of the finger maps it refers to in the same way as the weight maps in the hand network project activation between the finger maps. Thence, the result of this calculation has the same size as a finger map and describes the positions for the second finger in relation to the first one which are permitted by the grasp type template. Second, this result is compared to the other finger map to which the template map refers. This is done by multiplying the result from step one with the finger map to select only those positions which are permitted by the grasp type. Third, the result of step two are summed and added across all template maps of grasp type to yield the activation of the output unit. Thus, a unit is strongly activated when all its template maps match the grasp posture in the finger maps and weakly activated if only a small number of its template maps match the grasp posture generated by the hand network. Finally, a winner-takes-all mechanism ensures that eventually only the best-matching template is selected.

6.1.2 Top-Down Path

The grasp template network can not only classify grasps but the classification result also feeds back into the hand network to guide grasp generation. Figure 6.2 shows how this feedback path is embedded in the model. The template maps in each template are weighted with the activation of output neuron for the template set and then added to their corresponding weight maps in the hand network. This enables the grasp template network to modify the anatomical constraints (which are encoded in the weight maps) dynamically during grasp generation. Thus, the definition of what constitutes a grasp is not fix any more but can be adapted either depending on the

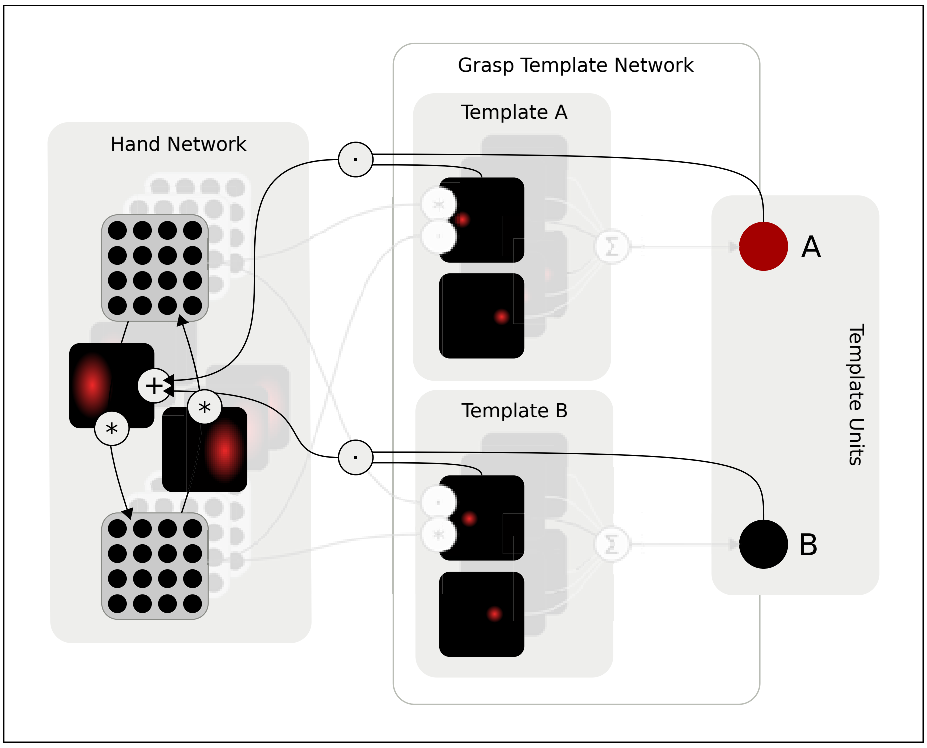


Figure 6.2: Illustration of the top-down path which links the output of the grasp template network with the hand network. Only one feedback path is shown; in the actual model all template maps feed back into the corresponding weight maps.

objects currently present in the visual input or based on external intentions[1]. In order to allow the model to adapt the weight maps independently of their current contents, the template maps are added to instead of multiplied with the weight maps in the hand network. To my best knowledge this implementation of a feedback path is a novel way of implementing top-down feedback in attentional models. The implications this has will be discussed in Chapter 8.

6.1.3 Mathematical Details

The visual feature extraction stage of SAAM is the same as before. The grasp template network is integrated in the original model by adding a new term to the

[1] In the simulations this is realised by using different initial values for different templates to express a preference for a grasp type.

energy function of the hand network. This modification turns the energy function into a function with one input variable for the output of the hand network and one for the output of the grasp template network. When the energy function is derived in order to perform the gradient descent procedure, it splits into a separate equation for each network.

The Grasp Template Network

Equation 5.7, which defines the excitatory connections in the hand network, is modified to include the activation of the template units:

$$E_{\text{exc}}^{\text{tpl}}(y_{IJ}^{(G)}, y_K^{\text{tpl}}) = -\sum_{f=1}^{5} \sum_{\substack{g \\ g=f}} \sum_{ij} \sum_{\substack{s=-L \\ s=0}}^{L} \sum_{\substack{r=-L \\ r=0}}^{L} T_{sr}^{(k:f \quad g)} \cdot y_{ij}^{(g)} \cdot y_{i+s,j+r}^{(f)}, \tag{6.1}$$

with

$$T_{sr}^{(k:f \quad g)} = T_{sr}^{(f \quad g)} + a_{\text{tpl}} \sum_k w_{sr}^{(k:f \quad g)} \cdot y_k^{\text{tpl}}.$$

The term $w_{SR}^{(k:f \quad g)}$ denotes the weight matrix which describes the allowed positions of finger g in relation to finger f in the k-th template. The weight matrices in the template map are multiplied with the activation of the template (y_k^{tpl}) and added to the weight maps of the hand network. Only one grasp type should be selected once a grasp has been generated. This constraint is implemented with the same winner-takes-all (WTA) approach as in the hand network (see Eq. 5.8 and 5.9). However, only the WTA-part of the equation is used and the input part is ignored:

$$E_{\text{inh}}^{\text{tpl}}(y_K^{\text{tpl}}) = (\sum_k y_k^{\text{tpl}} - 1)^2. \tag{6.2}$$

The Complete Model With these changes and additions the new complete energy function for the model becomes

$$E^{\text{tpl}}_{\text{total}}(y^{(F)}_{IJ}, y^{\text{tpl}}_{K}) = a^{\text{hand/tpl}}_{\text{exc}} \cdot E^{\text{tpl}}_{\text{exc}}(y^{(F)}_{IJ}, y^{\text{tpl}}_{K}) + a_{\text{inh}} \cdot E_{\text{inh}}(y^{(F)}_{IJ}) + \\ a_{\text{inp}} \cdot E_{\text{inp}}(y^{(F)}_{IJ}) + a^{\text{tpl}}_{\text{inh}} \cdot E^{\text{tpl}}_{\text{inh}}(y^{\text{tpl}}_{K}). \quad (6.3)$$

Please note that I diverge here slightly from the strict energy function by allowing different excitation factors $a^{\text{hand}}_{\text{exc}}$ and $a^{\text{tpl}}_{\text{exc}}$ for the hand network and the grasp template network. Simulations produced better results when the excitations into both networks were not symmetric.

Gradient Descent The gradient descent procedure is the same as in the first version of the model (see Eq. 5.12 and 5.13) but with the new energy function. This leads to two equations describing the hand network and the grasp template network. The gradient descent in the hand network is

$$\tau \dot{x}^{(f)}_{ij} = -x^{(f)}_{ij} - \frac{\partial E^{\text{tpl}}_{\text{hand}}(y^{(F)}_{IJ}, y^{\text{tpl}}_{K})}{\partial y^{(f)}_{ij}}. \quad (6.4)$$

The gradient descent of the hand network is only different from the original version of SAAM when the top-down path is enabled (i. e. $a_{\text{tpl}} = 0$). If factor a_{tpl} in the energy function is zero, then the variable y^{tpl}_{K} has no effect on the gradient descent in the hand network.

The equation for the gradient descent in the grasp template network is

$$\tau^{\text{tpl}} \dot{x}^{\text{tpl}}_{k} = -x^{\text{tpl}}_{k} - \frac{\partial E^{\text{tpl}}_{\text{hand}}(y^{(F)}_{IJ}, y^{\text{tpl}}_{K})}{\partial y^{\text{tpl}}_{k}}. \quad (6.5)$$

Here, the template factor a_{tpl} is merged with the $a^{\text{tpl}}_{\text{exc}}$ factor and is not assigned a separate value. This also evades problems when a_{tpl} is set to zero to disable the top-down path in SAAM. Otherwise the derivation for the grasp template network would become zero and grasp classification would not be possible.

Stop Criterion

A winner in the grasp template network can only develop once finger positions in the hand network begin to emerge. To prevent the model from stopping before the grasp template network has chosen a winner, the stop criterion was modified. The new stop criterion checks if one of the output neurons in the grasp template network is above a certain threshold t^{tpl} instead of relying on the activations in the finger maps. Additionally, simulations are aborted when the threshold activation level is not exceeded after time t_{max}.

6.2 Simulations

Three sets of simulations were run with the extended version of SAAM. The results are presented in this section. First, I will demonstrate in Study 2a that the classification mechanism described above can indeed identify different grasp postures. The limits of this classification method will be also pointed out. Next I will show in Study 2b that the model is able to select one of two objects presented in the visual input. Moreover, I will explain that in conjunction with the classification abilities of SAAM this selection process can be understood as visual attention for action. Study 2a and 2b only investigated the bottom-up path of the grasp template network; the top-down path was disabled in these simulations. The top-down path will be explored in Study 2c. It will be shown how the grasp template network can utilise preferences for specific grasp types to guide object selection to suitable objects. Additionally, simulations will demonstrate that the top-down path also allows to select different grasps on one object.

6.2.1 Study 2a: Classification Abilities of the Model

This study aimed at testing whether the grasp template network is able to classify different grasp postures. The grasp template network contained three templates in this study which were designed based on two criteria: First, the grasps encoded in the templates obviously needed to be suitable for the objects used in the simulations. To achieve this, the design of the templates was based on the grasp postures observed in Study 1. Additionally, it was ensured that the different templates represented different types of grasps as they were observed in the simulations in Study 1. Second, the template definitions were derived from the grasp taxonomies discussed in Chapter 2. The two grasp types most consistently distinguished in all classifications are the power and the precision grip. While SAAM's restriction to two dimensions prevented encoding the different spatial orientation of the opposition space in these grasp types, they could still be approximated by designing a template with a wide grasp span for power grasps and one with a small grasp span for precision grasps. In addition to these two types, the previous simulations showed grasps which aimed at achieving form-closure of the grasped object (e. g., trapezium 90° or concave 0°). The 'circular precision grip' from Cutkosky's (1989) grasp taxonomy offers a good foundation for such a template. This posture describes a grasp were the fingers are placed roughly on a semicircle. The three types of grasp postures were termed: a 'small grasp', a 'large grasp' and 'circular grasp'. The small and large grasp were not called precision and power grasp since the were only approximations of these grasps. The template maps are described in detail in the next section. The grasp templates were used to classify the grasps which SAAM produced in the Study 1.

Method

Stimuli The same objects and orientations as in Study 1 were used. However, the six stimuli[2] which were shown to be more suited for vertical grasps then for horizontal ones were removed from the stimuli set because of SAAM's restriction to horizontal grasps.

Also, bar 180° and bar 0° were not treated as separate conditions any more. Only bar 0° remained in the stimuli set.

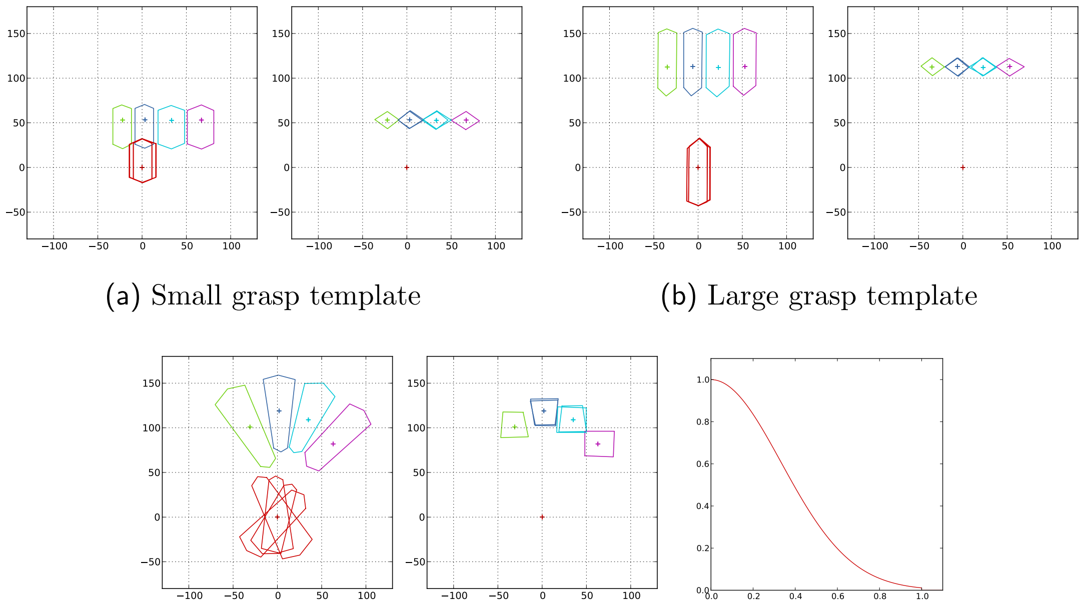
Parameters The parameters and weight maps of the hand network were not changed compared to Study 1. The parameters for the grasp template network were $a^{\mathrm{tpl}}_{\mathrm{exc}} = 0.002$ and $a^{\mathrm{tpl}}_{\mathrm{inh}} = 6.0$. In the gradient descent procedure for the grasp template network the parameters were $\tau^{\mathrm{tpl}} = 6.0$, $m^{\mathrm{tpl}} = 300.0$, and $s^{\mathrm{tpl}} = 0.3$. The output neurons of the grasp template network were initialised with 0.5.

The template activation threshold was set to $t^{\mathrm{tpl}} = 0.999$. The stop criterion based on finger activation was removed. The maximum simulation time was set to $t_{\max} = 4.0$. A complete list of parameters can be found in appendix B.2.

Templates Three grasp templates were added to the grasp template network. One for the classification of large grasps, one for small grasps and one for circular grasps.

The template maps were constructed in the same way as the weight maps between the fingers (see Sec. 5.2.1). Figure 6.3 shows the polygons which were used to define the template maps and the Gaussian function which was used to fill the polygon areas. While the polygons for the weight maps mark a hard border for the finger movement (due to anatomical constraints) this is not the case for the template maps. There is no hard threshold for deciding if a grasp is of a certain type or not but

[2] Bar 90°, bar 270°, irregular 0°, irregular 180°, pyramid 0°, pyramid 180°, t-shape 90°, and t-shape 270°.



(a) Small grasp template

(b) Large grasp template

(c) Circular grasp template

(d) Activation

Figure 6.3: Defintion of the template maps. The left hand-side plots show the polygons which were used to define the template maps between thumb and fingers. The right hand-side ones the polygons for the finger-to-finger template maps. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink. Figure d shows the Gaussian curve which defined the activation within the polygons in the template maps.

instead there is a soft threshold for the transition between different types of grasps. A grasp does not suddenly belong to a certain grasp type just because one finger moved past a polygon border. Therefore, the Gaussian function in the template maps did not include a base level of activation throughout the polygon but smoothly dropped to zero at the edges of the polygons.

The polygons for the different grasp types were defined so that the small grasp template covered all grasps which were roughly as wide as the bar object and where the fingers were on a horizontal line; the large grasp covered everything from there to the maximum hand width; and the circular grasp template covered most other grasps but in particular those which achieved form-closure. The greater coverage of the circular grasp is because the template allows more variation between the fingers

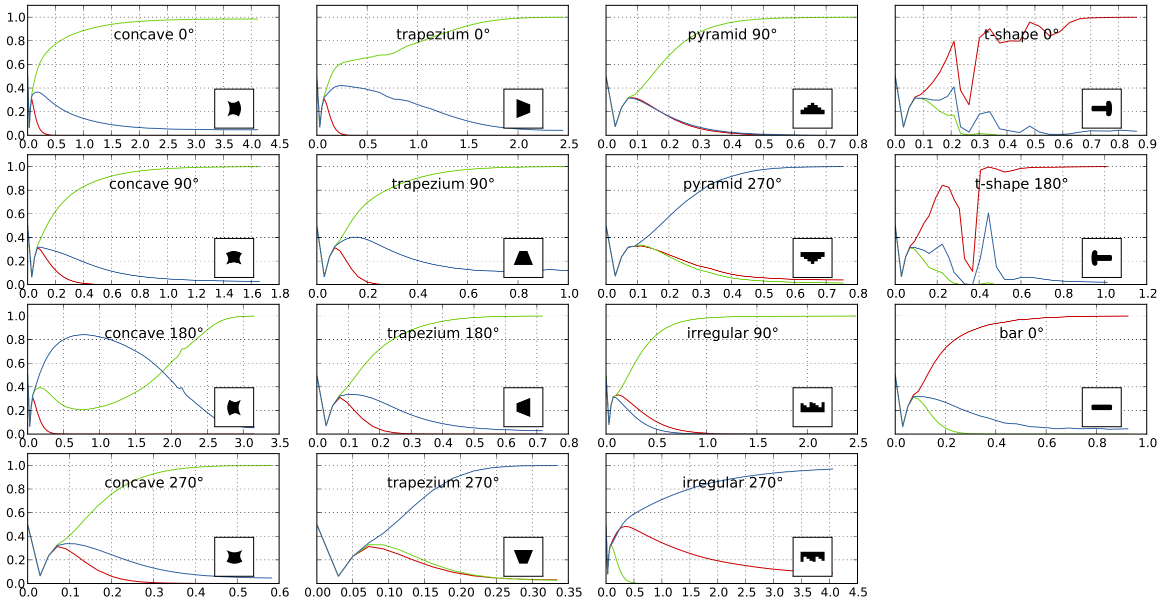


Figure 6.4: Time course of the activation in the template units. The different templates are coloured as follows: red–small grasp, green–circular grasp, blue–large grasp. Please note that the time axis is scaled differently in each plot.

compared to the other two templates. Many grasps were covered by the large grasp template as well as the circular grasp template. It will be interesting to see how the model handled conflicts between the two templates.

Results and Discussion

The finger positions and time courses of the activations in the finger maps were the same as in the previous simulations because the feedback loop was not activated. Therefore, only the template activations in grasp template network are presented in Fig. 6.4. The simulation times are listed in Table 6.1. Due to the modified stop criterion these changed compared to the timings in Study 1.

The plots of the time courses of the template activations showed that the classification worked in general and that the grasps of the different objects were classified appropriately. A close examination of the plots revealed that the competition in the grasp template network depended on the similarity of the different templates. The

	bar	concave	irregular	pyramid	trapezium	t-shape
0°	0.925	> 4.0	–	–	2.444	0.863
90°	–	1.657	2.479	0.789	0.997	–
180°	–	3.151	–	–	0.717	1.011
270°	–	0.582	> 4.0	0.754	0.334	–

Table 6.1: Duration of the simulations. Times written as > 4.0 denote simulations which were aborted because the reached the maximum simulation time t_{max}.

small grasp and the large grasp templates differed only in the grasp size (i. e. distance between fingers and thumb) and not in the relative placement of the fingers. In simulations with objects where all fingers were placed on a straight edge this led to strong competition between the two templates (e. g., in condition bar 0° and irregular 270°). The circular grasp template, in contrast, differed much more and therefore received much less activation in such simulations. In simulations where the generated grasp matched the circular grasp template the opposite behaviour was observed. The circular grasp template was quickly activated and not much competition occurred between between the large and small grasp templates. Prime examples of this were pyramid 90° and irregular 90°. However, some of the other simulations in which the circular grasp template was selected showed that the large grasp template tended to decline more slowly than the small grasp template. The reason for this was a relatively similar grasp size in the large grasp and circular grasp templates. This effect can be seen in the simulations of concave 90°, trapezium 90° and trapezium 180°.

Another very interesting simulation is that of the concave 180°. In this simulation the large grasp first appeared to emerge as the winner from the WTA in the grasp template network but after some time the circular grasp template gained activation and finally the grasp was classified as circular. This effect was caused by the model first placing the index finger on the top edge of the object, and only after about 0.7 time units it was moved to the side of the object. This shift of the position is marked by the small drop of activation in the time course of the finger map activation. Once the finger position changed, the grasp template network began to shift its selection from the large grasp template to the circular grasp template. This behaviour of the model demonstrated that the grasp template network is able to readjusting its selection during grasp-planning if the output of the hand networks changes.

The time course of the template activation (Fig. 6.4) in the two simulations of the t-shape stood out compared to the other simulations. The sudden drops in the activation were most likely caused by the inhibition in the grasp template network because with stronger inhibition factors ($a_{\mathrm{inh}}^{\mathrm{tpl}}$) similar effects were observed in simulations of other objects, too. The influence of the head of the t-shape made it rather difficult for the model to decide on a grasp and consequently on a grasp template. If one or two of the fingers were placed on the head part of the object the grasp would not be a small grasp any more but also neither a large nor a circular grasp. This ambiguities made it difficult for the grasp template network to classify the grasp. Similar, but much less dramatic effects of ambiguous template matching can be seen in the simulations of trapezium 0° and irregular 270° which were very slowly activated and – in case of the trapezium – showed even a phase of stagnation before the final selection occurred. The grasps for both objects showed that they were not exclusively matched by one template and matched neither template perfectly. Therefore, the grasp template was only able to classify the grasp once the hand network had produced strong and localised outputs.

6.2.2 Study 2b: Two Object Scenes

The simulations in the last section showed that the grasp template network in SAAM can classify different types of grasps. Now, I will demonstrate that this does not only work with single-object inputs but with multiple-object inputs as well. As before, the results showed that SAAM selected a grasp-type template. They also showed that all fingers were placed on only one of the objects in the visual input. Together these two selection processes for grasp type and grasp location can be interpreted as attentional selection for action.

Method

Stimuli The stimuli were designed using a subset of six objects from the previous simulations. The stimuli contained two objects each which were placed randomly in the visual field with roughly the same space between the border and the object as before to avoid border effects. Two objects per grasp template set were chosen based on the results from the previous simulation set. Three of the stimuli were devised to test selection in SAAM when both objects were graspable with grasps of the same type. Bar 0° and t-shape 0° activated the small grasp template, pyramid 90° and trapezium 90° the circular grasp template, and trapezium 270° and pyramid 270° the large grasp template. The other three stimuli presented all possible combinations of two objects matching different grasp types. Trapezium 270° and bar 0° matched the large and small grasp templates, t-shape 0° and pyramid 90° the small and circular grasp templates, and trapezium 90° and trapezium 270° the circular and large grasp templates.

Parameters and Templates The same parameters and templates as in the previous study (Sec. 6.2.1) were used. The inputs were, however, twice as wide and high as in the previous simulations to accommodate all objects without reducing their size. Remarkably, the increased input size required no modifications to the simulation parameters.

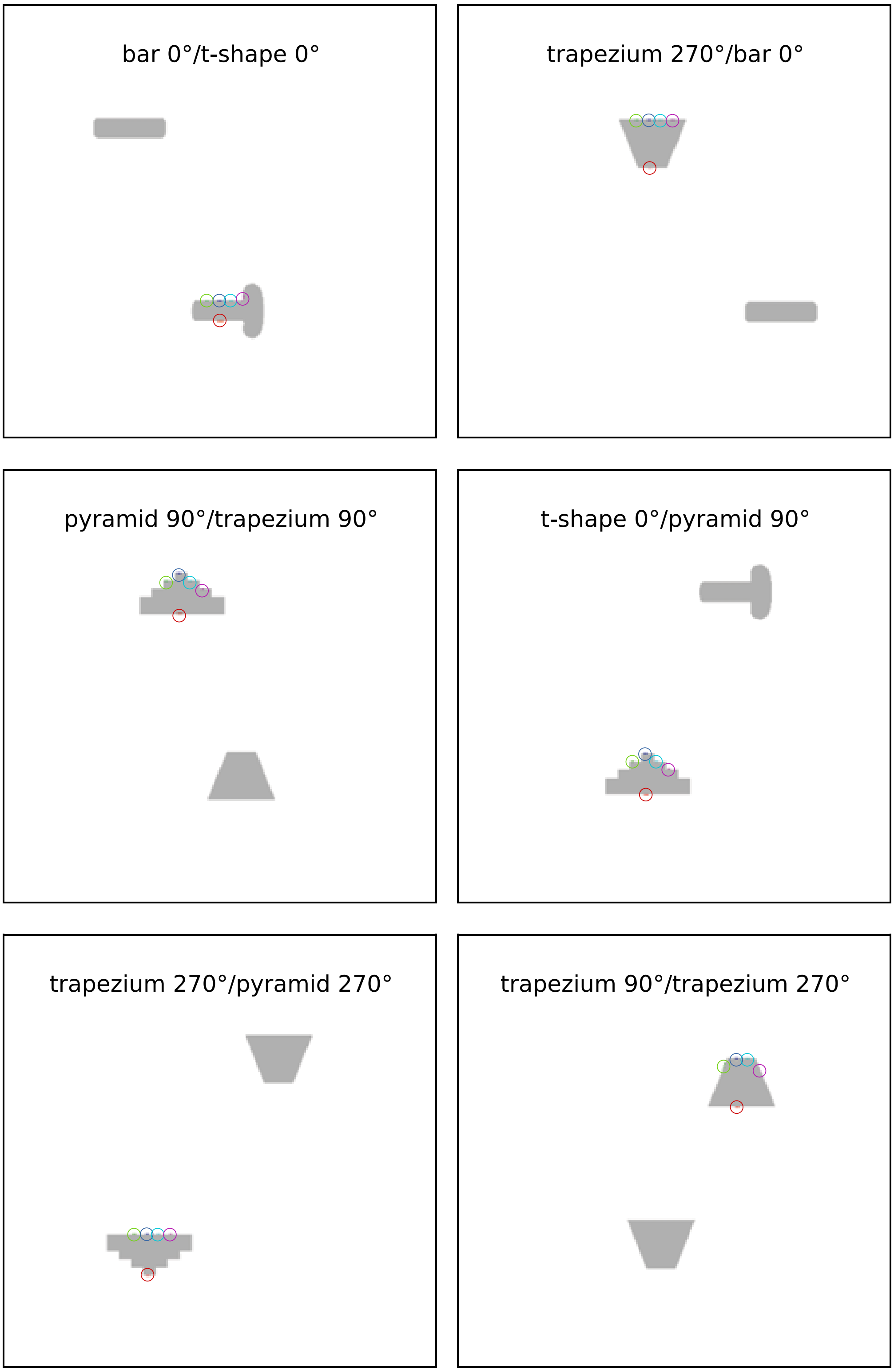

bar 0°/t-shape 0°
trapezium 270°/bar 0°
pyramid 90°/trapezium 90°
t-shape 0°/pyramid 90°
trapezium 270°/pyramid 270°
trapezium 90°/trapezium 270°

Figure 6.5: Output of the bottom-up simulations with two-object stimuli. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

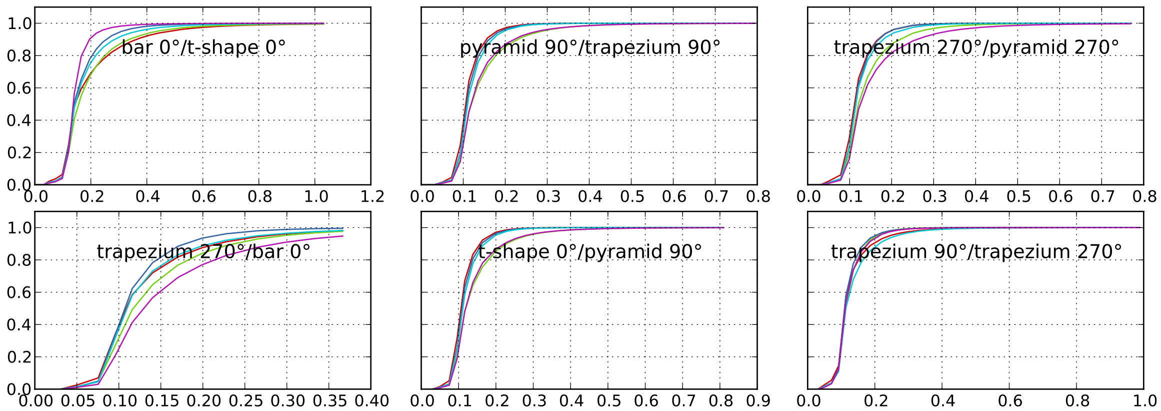


Figure 6.6: Maximum activation in the finger maps of the bottom-up simulations with two-object stimuli. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

Results and Discussion

The results (Fig. 6.5) showed that SAAM was able to perform a selection of one of two objects. Hence, selection of grasp location does work. The finger positions generated here were not different from the finger positions which were created in the single object simulations before. In addition to selecting a grasp location, Fig. 6.7 shows that a grasp type was selected as well. The correct templates for the chosen objects were selected. The time courses of the template activations showed that the final grasp type was selected early during the simulations, and no shifts of activation between different grasp types occurred later during the simulations. This indicates an early selection of the object. This is confirmed by the time courses of the maximum activation in the finger maps (Fig. 6.6) which also showed a quick and smooth rise of the activation without any drops which would indicate a major change in the selected finger positions.

Table 6.2 compares the simulation times of the multi-object simulations and the previous single-object simulations. These results showed that the multi-object simulations were generally slower than the simulation of the selected object on its own. Only in the last simulation (trapezium 90°/trapezium 270°) the multi-object

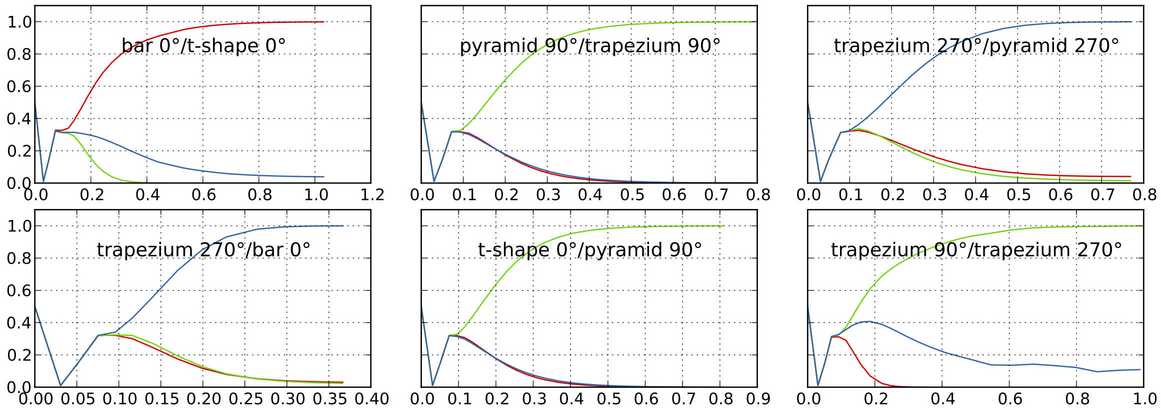


Figure 6.7: Template activation in the bottom-up simulations with two-object stimuli. The templates are coloured as follows: red–small grasp, green–circular grasp, blue–large grasp. Please note that the time axis is scaled differently in each plot.

simulation was marginally faster for unknown reasons. Not surprisingly, the selected object in the multi-object simulations was mostly the one which was grasped faster in the single-object simulations. In the two simulations with trapezium 270° was, however, the opposite behaviour observed: Despite being grasped much faster than the pyramid 270° and the trapezium 90°, the trapezium 270° was not selected for grasping in the two-object simulations. The reason for this was the size of the objects. While it was still easily graspable, the activation of the trapezium 270° was reduced compared to smaller objects in early phases of the simulation because the activation in the thumb-finger weight maps decreased with increasing grasp size (please refer to Fig. 5.4 for details of the distribution of activity in the weight maps). Positions along edges which were further apart received therefore less activation then ones which were closer together. Hence, the pyramid 270° and the trapezium 90° had initially a small advantage which led to their selection even when it eventually took longer to properly select the final finger positions than for trapezium 270°.

Stimulus (Object 1/Object 2)	Both Objects	Object 1	Object 2
bar 0°/**t-shape 0°**	1.031	0.925	0.863
trapezium 270°/bar 0°	0.366	0.334	0.925
pyramid 90°/trapezium 90°	0.794	0.789	0.997
t-shape 0°/**pyramid 90°**	0.808	0.863	0.789
trapezium 270°/**pyramid 270°**	0.769	0.334	0.754
trapezium 90°/trapezium 270°	0.989	0.997	0.334

Table 6.2: Duration of simulations with two objects in comparison to the times from the single object simulations presented in Sec. 6.2.1. The objects which were selected in the two-object simulations are printed in bold.

6.2.3 Study 2c: Modulating Grasp Selection

The last study showed that SAAM is able to classify grasps and also to perform selection of objects in visual inputs with multiple objects. These properties of the model emerged from the bottom-up path alone. The top-down path of the grasp template network was explored in this final study. First, the two-object stimuli from the previous study were simulated again. However, this time the grasp template network preferred one grasp type over the others. The results showed that this affected the selection processes in SAAM producing different results depending on the grasp type preferences.

In addition to the two-object stimuli, two new inputs were designed. One stimulus contained two objects which were positioned so that they could be grasped separately or together depending on the grasp size. This design of this stimulus was motivated by research on action relations within groups of objects. For example, Riddoch, Humphreys, Edwards, Baker, and Willson (2003) showed that pairs of objects which were arranged correctly for using them together (e. g., a bottle with the corkscrew placed at its top end) increased performance in patients with extinction[3] compared

[3] Extinction is a neurological disorder in which patients cannot perceive a stimulus presented in one hemisphere if another stimulus is shown in the other hemisphere at the same time. However, if only a single stimulus is presented, it is perceived in either hemisphere.

to incorrectly positioned objects (see also Humphreys et al., 2010 for a review on research on action relations between objects). The simulation of two objects which can be grasped either separately or together is, of course, a simplification but it still shows that affordances and action intentions are not restricted to single objects. In Chapter 9 I will make suggestions how this could be extended to more complex arrangements of objects. The other new stimulus showed an object which could be grasped in two different ways. The simulations with these stimuli demonstrated that grasp type preferences are not restricted to selecting different objects in multi-object inputs but can also be utilized to guide grasp generation on single objects or groups of objects.

Method

Stimuli Five stimuli were used. Three displayed multi-object inputs. The stimuli from the previous study which displayed objects suitable for different grasp types were chosen. Additionally, two new stimuli were designed. One contained two horizontal bars which were positioned close together so that the could be either grasped together or separately (see Fig. 6.11). This stimulus was included to test whether the top-down path in SAAM can be used to model decisions for grasping one or all objects of a group. The other one showed an object constructed by combining a bar 0° and a trapezium 90° (see Fig. 6.12). The intention of this object was to test whether the top-down path in SAAM can be utilised to generate different grasps on a single object.

Parameters and Templates The simulation parameters are the same as in the last study. Only the template factor a_{tpl} was changed to 4.0 to activate the top-down path. The output neurons of the grasp template network were initialised with 0.3 for the two non-preferred grasp type templates and 0.7 for the template of the preferred grasp type.

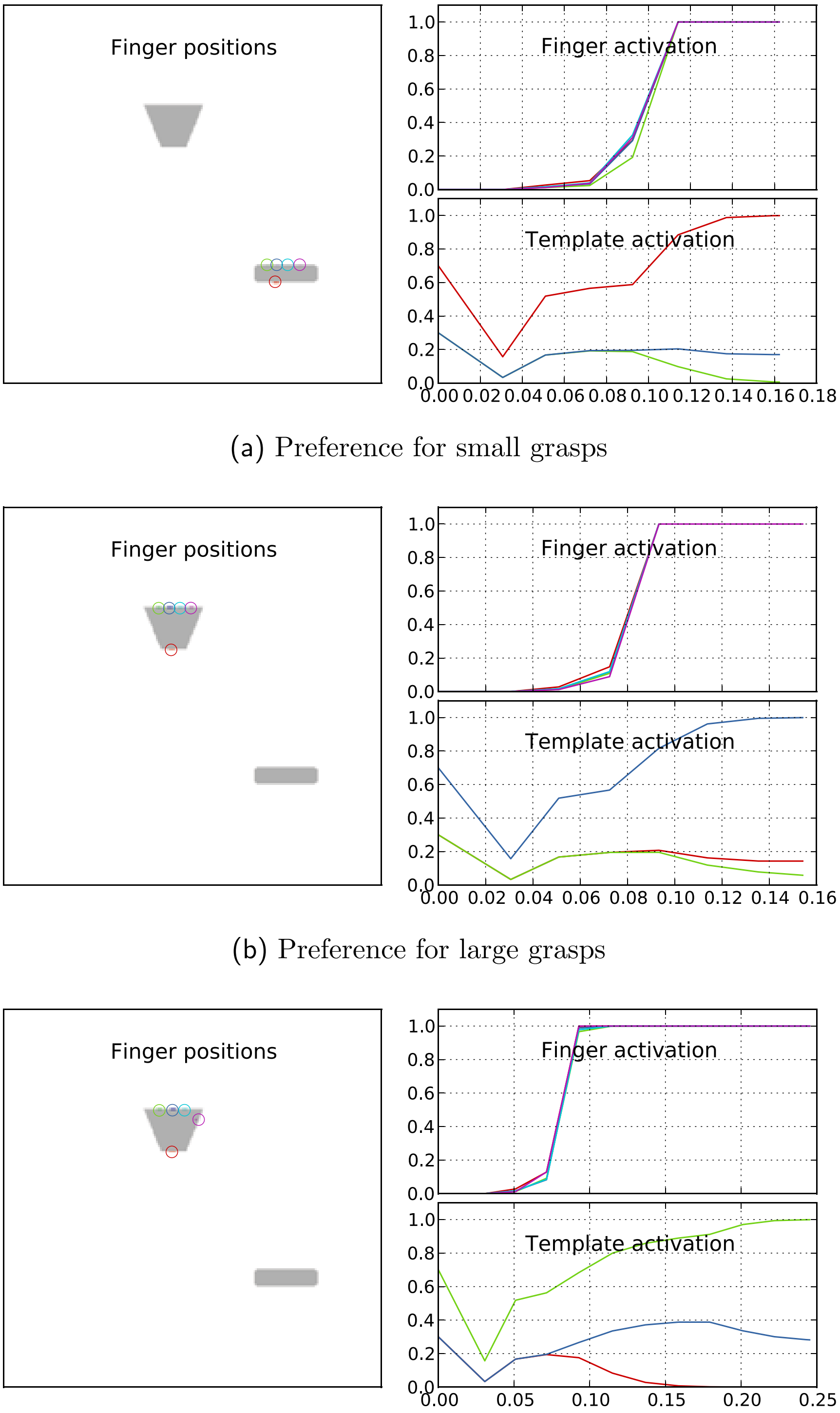
The template maps were also the same as before (Fig. 6.3). The activation of the weight maps in the hand network was, however, reduced to account for the additional activation from the top-down path. This was achieved by multiplying the weight maps in the hand network with 0.3. The polygon outlines of the activated areas were not changed.

Results and Discussion

The simulations were divided in five sets. Each set contained the simulations of one stimulus with preferences for the different grasp types. Before discussing the simulations in detail, it shall be noted that the finger positions in the simulations were different from the ones produced by the bottom-up-only simulations. This was obviously caused by the addition of the template maps to the combined weight maps. Therefore, the generated grasp postures in most simulations closely resembled the activated template. Notably deviations will be pointed out during the discussion.

Trapezium 270° and Bar 0° Figure 6.8 shows the results of the three simulations with the trapezium 270°/bar 0° stimulus. The previous simulations demonstrated that these objects were grasped with a large and a small grasp, respectively. By providing a stronger initial activation for one these two templates, object and grasp selection in SAAM could be directed towards either of the two objects (see Fig. 6.8a and 6.8b). The finger activation plots show that a grasp location was selected quickly. The template activation on the other hand developed much more slowly. Overall however, the simulations ran much faster compared to the simulations without the top-down path. This was due the stronger activation in the combined weight maps compared to the original maps in the hand network.

The template activation plots also show that the template which matched the non-selected object still had a higher activation (≈ 0.2) compared to the third template



(a) Preference for small grasps

(b) Preference for large grasps

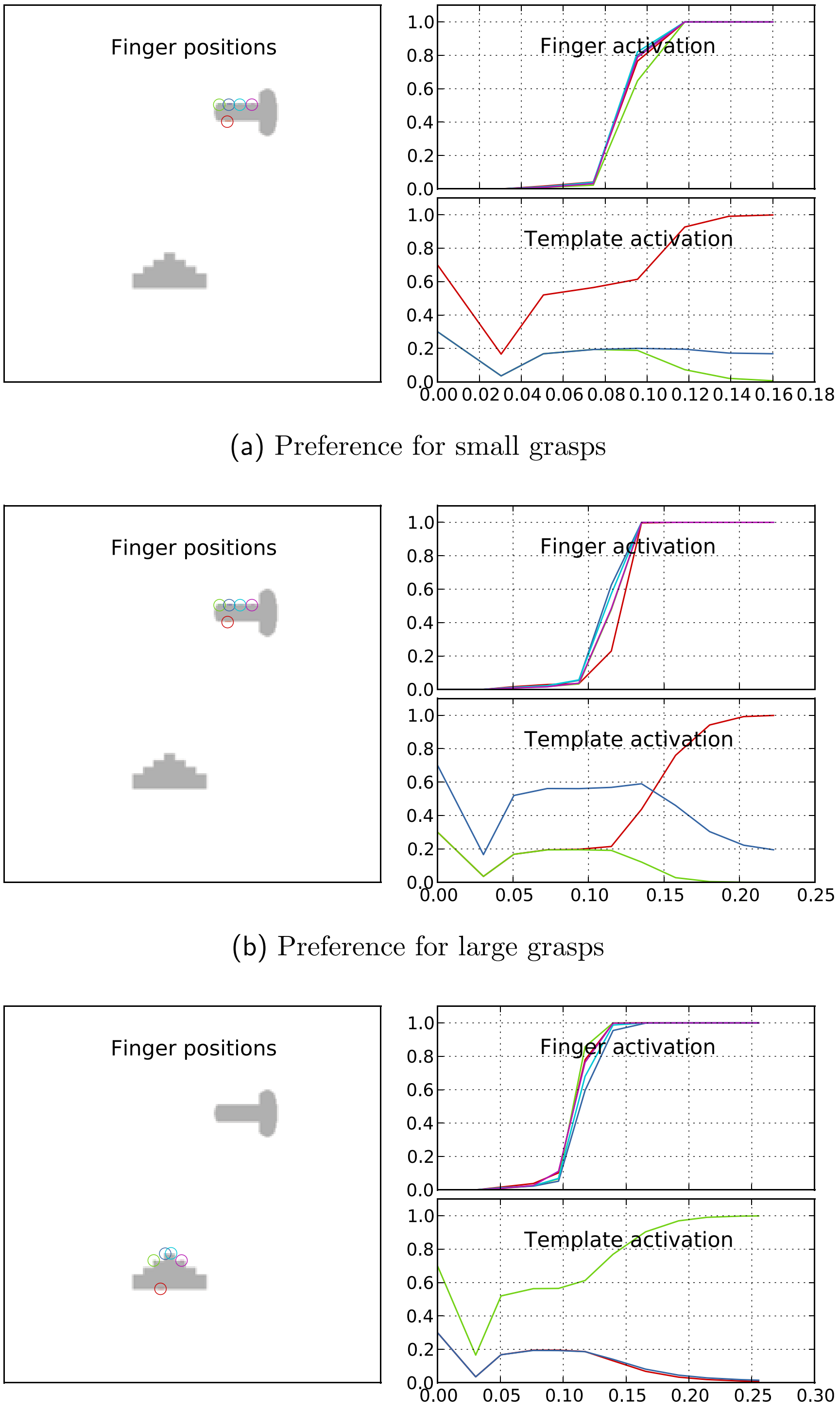
(c) Preference for circular grasps

Figure 6.8: Simulation results for trapezium 270°/bar 0°. The templates are coloured as follows: red–small grasp, green–circular grasp, blue–large grasp. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

which matched neither of the objects.

The last of the three simulations (Fig. 6.8c) had a preference for the circular grasp type. This type did not describe a grasp which was really suitable for either of the two objects. Hence, SAAM could not simply follow the top-down preference and select the object which was best matched by the preferred template; instead, SAAM had to rely on the grasp preferences of the bottom-up path to select one of the objects. Reflecting the result from the simulation with only the bottom-up path enabled, SAAM chose to grasp the large object. However, the generated grasp posture differed. Because of the preference for circular grasps, SAAM placed the little finger at the right-hand edge of the trapezium 270° instead of on the top edge. Consequently, the grasp was classified as a circular grasp. The large grasp template still stayed relatively strongly activated, though. This indicated a partial match of the generated grasp with the large grasp template and a general possibility of grasping the object with a large grasp as well.

T-shape 0° and Pyramid 90° The second set of simulations with the t-shape 0°/ pyramid 90° stimulus showed a behaviour similar to the first three simulations (see Fig. 6.9). A preference for small grasps triggered a grasp of the t-shape 0° while a preference for circular grasps produced a grasp posture for the pyramid 90°. Due to the influence of the circular grasp template, the grasp posture used to grasp the pyramid differed from the one created in the bottom-up-only simulations. While previously each finger was placed on a different step, now the ring and middle finger were both placed on the top of the pyramid. The behaviour in the simulation with a preference for large grasps, which did not fit either of the objects, was very different from the corresponding simulation in the first set, though. The preferred large grasp template had the strongest activation for some time before the selection switched over to the small grasp template which finally became the winner. The t-shape 0° was selected because the large grasp template reinforced postures where all fingers were



(a) Preference for small grasps

(b) Preference for large grasps

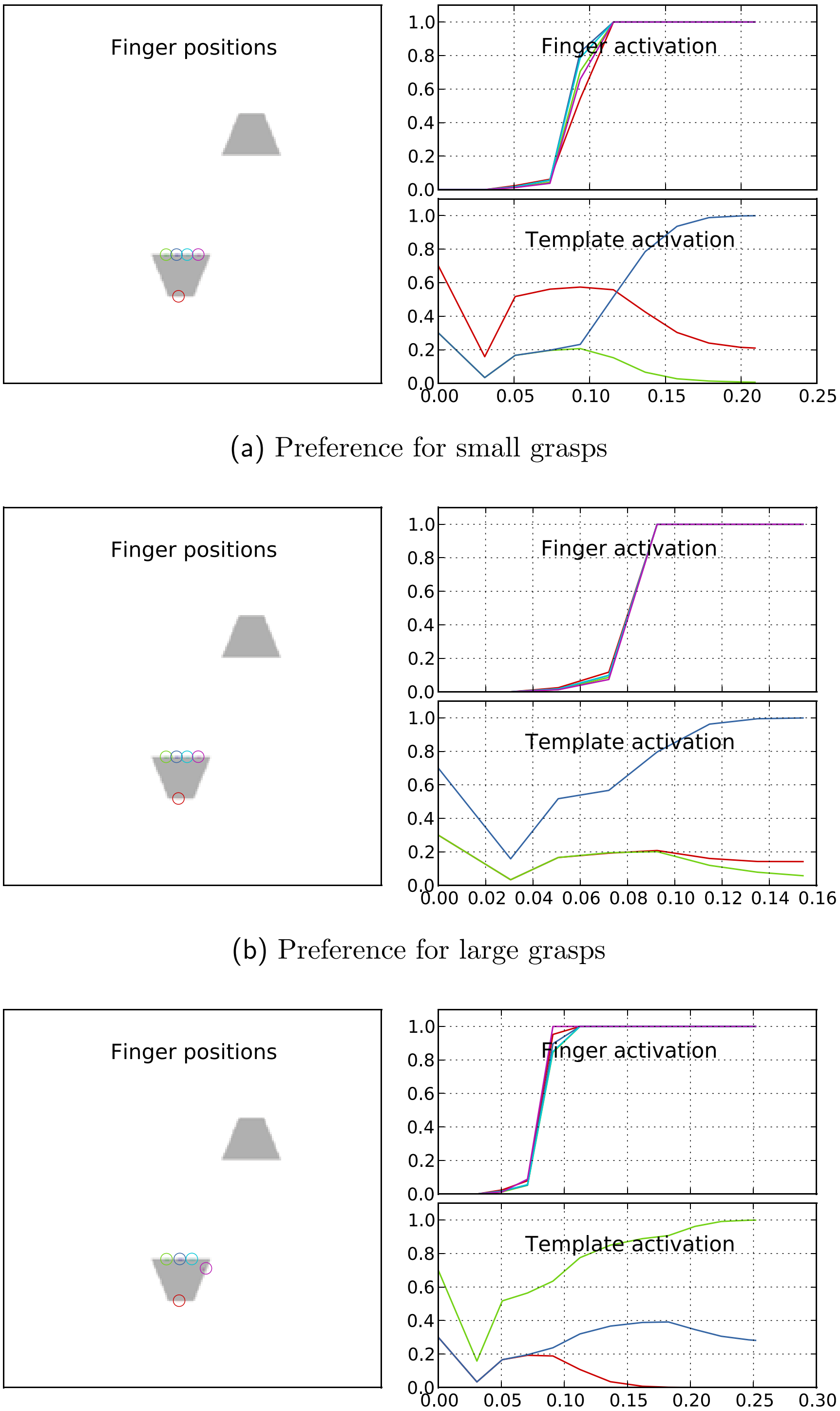
(c) Preference for circular grasps

Figure 6.9: Simulation results for t-shape 0°/pyramid 90°. The templates are coloured as follows: red–small grasp, green–circular grasp, blue–large grasp. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

on a straight line. This caused the selection to initially focus on the t-shape 0°. Once the fingers were activated strong enough the thumb was placed on the bottom edge of the object which then caused the template activation to switch over to the small grasp template. The delayed activation of the thumb can be seen in the time course of the finger activation.

Trapezium 90° and Trapezium 270° The same effect – a reinforcement of horizontally aligned finger positions – was observed in the last set of simulations (Fig. 6.10). When a small grasp was preferred, the trapezium 270° was selected because of its straight edge suitable for positioning the fingers horizontally aligned. The template activation plot shows the same change from the small grasp template to the large grasp template as the pyramid 90°/t-shape 0° simulation only that the curves for the large and small grasp template activations were swapped. The time course of the finger activation also showed the same delay in the increase of the maximum activation in the thumb map as before.

While the selection of either of the two objects worked well for the other two stimuli, this was not the case in this set of simulations. Preference for the large grasp template generated a grasp of the trapezium 270° as expected. Preference for the circular grasp template, on the other hand, did not lead to a grasp of the trapezium 90° as one would assume because of the congruence of the circular grasp template and the grasps produced on trapezium 90° in the earlier simulations. Instead a circular grasp on the trapezium 270° was produced similar to the grasp posture generated in the simulation with the trapezium 270°/bar 0° stimulus with a preference for circular grasps. This is a surprising result because in the previous bottom-up-only simulations the trapezium 90° was preferred over the trapezium 270° although the latter was grasped faster. One would expect that a preference for circular grasps would support the bottom-up path in selecting finger positions at the trapezium 90° making it even



(a) Preference for small grasps

(b) Preference for large grasps

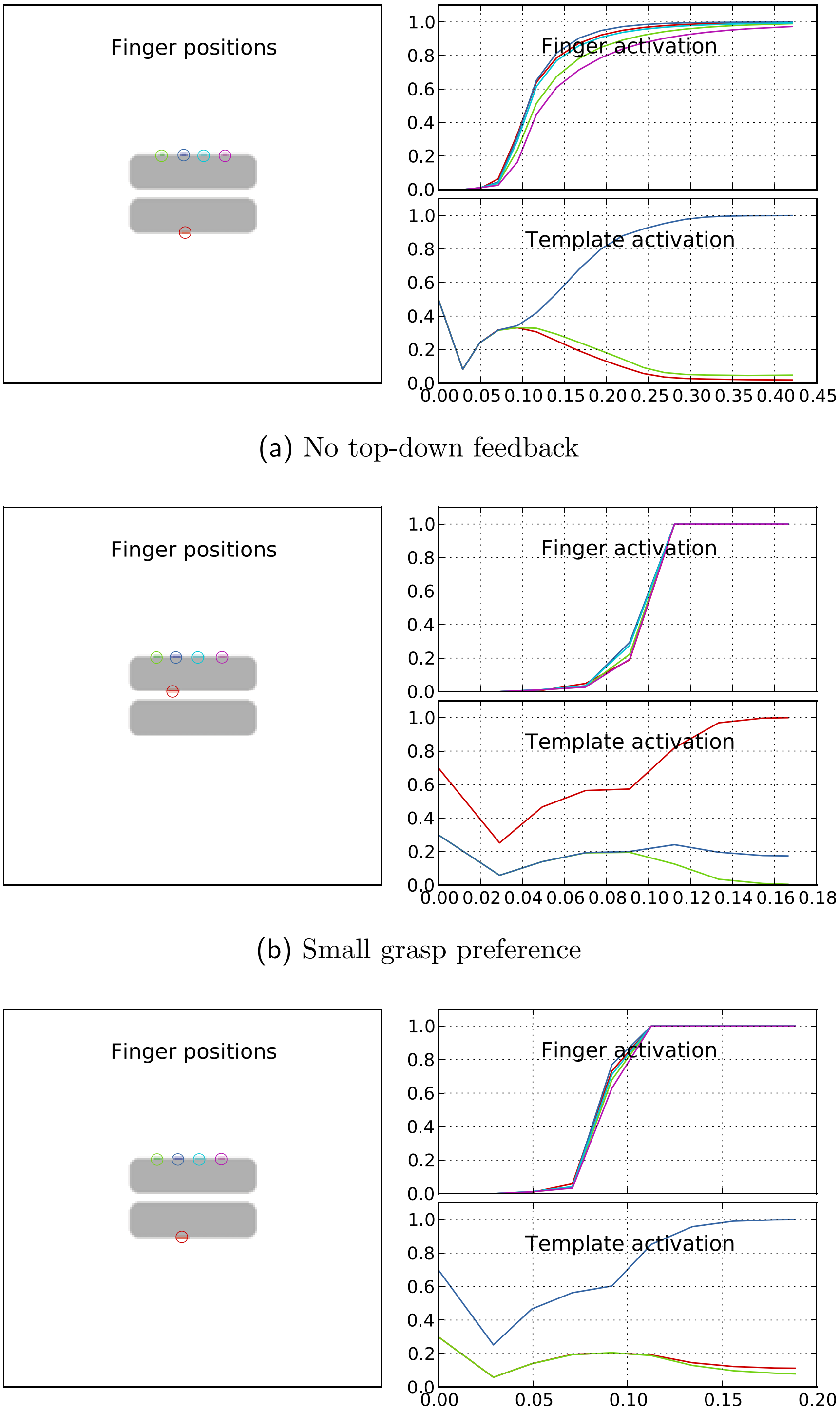
(c) Preference for circular grasps

Figure 6.10: Simulation results for trapezium 90°/trapezium 270°. The templates are coloured as follows: red–small grasp, green–circular grasp, blue–large grasp. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

easier for the model to select the object. However, it must be considered that the combined weight maps were different from the original weight maps. Firstly, the activation of the original weight maps was not as strong as before, and secondly, all three template maps – including the ones for the non-preferred grasp types – were added to the combined weight maps. Even though they were only weakly activated, the large grasp and the small grasp template introduced support for grasp postures with horizontally aligned fingers positions in the combined weight maps. This led to an initial selection of the trapezium 270° which could not be overwritten by the circular grasp template despite it being much stronger activated. The circular grasp template was, however, still strong enough to enforce its grasp type leading to a circular grasp of the trapezium 270°.

Two Bar 0° Objects Another multi-object simulation was conducted to test if grasp preferences can also be used to differentiate between grasping one object from a group of objects and grasping all of them. To simulate this, two bar 0° objects were presented close together so that a large grasp could embrace both objects. Figure 6.11a shows a bottom-up preference to grasp both objects together when no top-down feedback was applied.[4] When top-down preferences for either large or small grasps were enabled though, the model could be controlled to grasp either both or only one of the objects. Figure 6.11c shows that a preference for large grasps produced a grasp which included both bars, and a small grasp preference produced a grasp of only one of the bars (Fig. 6.11b). Which of the two bars was grasped could not be controlled since grasp selection is translation-invariant in SAAM. In order to control this, an extension of the model would be required which inhibits grasp-planning for certain positions in the visual input.

[4] The bottom-up simulation used weight maps with non-reduced activity like the other bottom-up-only simulations in Study 2b did because of the absence of additional top-down activity.

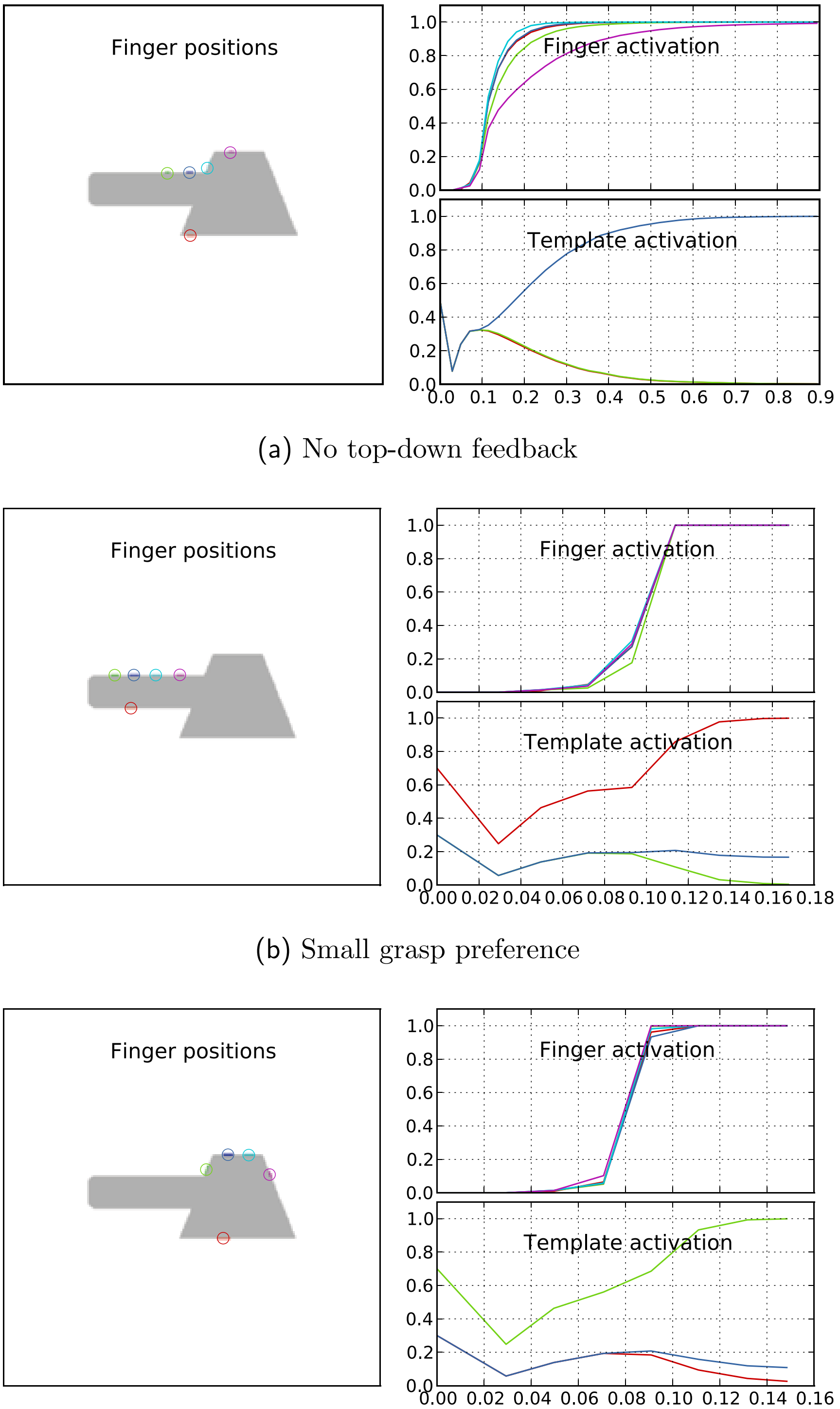


(a) No top-down feedback

(b) Small grasp preference

(c) Large grasp preference

Figure 6.11: Simulation results for two bar 0° objects. The templates are coloured as follows: red–small grasp, green–circular grasp, blue–large grasp. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.



(a) No top-down feedback

(b) Small grasp preference

(c) Circular grasp preference

Figure 6.12: Simulation results for bar&trapezium object. The templates are coloured as follows: red–small grasp, green–circular grasp, blue–large grasp. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

Bar&Trapezium Object These simulations showed that the top-down path in SAAM cannot only be used to guide selection in multi-object simulations but also to control the grasp postures which are used on a single object. Figure 6.12 shows how SAAM grasped an object which could be grasped either with a circular or a small grasp. The bottom-up simulation (Fig. 6.12a) of the object showed that neither side had a bottom-up preference and the grasp was placed close to the centre of gravity of the object instead. The top-down path, however, allowed to control the grasp posture: When a small grasp was preferred, the fingers were placed on the bar-shaped end of the object. When a circular grasp was preferred then SAAM selected finger positions on the trapezium-shaped side of the object. This demonstrates how the top-down path can be utilised to grasp objects in different ways.

6.3 Summary

In Study 2a and 2b the classification abilities of SAAM's grasp template network were explored. The three templates used in this study were derived from power, precision, and circular precision grip postures which were commonly identified in different grasp taxonomies (see Chapter 2). Study 2a confirmed that the grasp template network was able to classify grasps in single object inputs. The simulations revealed that competition between templates which shared properties (e. g., horizontally aligned finger positions or grasp size) increased when an object congruent with these properties was grasped. The simulation of the concave 180° stimulus demonstrated that the grasp template network could also re-adjust to changes of finger positions during the simulation.

The simulations with the t-shape in Study 2a, however, highlighted some difficulties of the grasp template network to select a template. The grasp of the t-shape did not fit the templates very well. Hence, the grasp template network required strong and localised finger positions in the hand network before it was able to select one of the templates. Interestingly, this behaviour disappeared in the two-object simulations. It is not exactly clear why the effect disappeared in the two-object simulations. Most likely, the increased input into the grasp template network – due to the larger visual field – reduced the inhibition in the network. Test simulations showed that greater inhibition values increased the observed fluctuations of template activation.

In Study 2b stimuli with two objects in the visual input were tested in order to assess object selection as well as grasp type selection in SAAM. Grasp type selection confirmed the results from Study 2a. The presence of a second object in the visual input had no effect on the created grasp posture; grasps did not differ from the ones observed in single object simulations. The simulations also showed that the model was able to select one of the objects presented in the visual input. This behaviour can be interpreted as visual attention for action because preparing a movement towards an object means selecting the object for action. Furthermore, the selection of a grasp posture can be understood as selection of an action. This will be further explored in Study 3.

In most two-object simulations the object which was faster in single-object simulations was the one which was selected. However, the simulations with the trapezium 270° showed that this was not always true. Despite being the fastest object in the single-object simulations, the trapezium 270° was not selected in the two-object simulations because the other objects in the input scenes better matched the anatomical constraints of the hand. This showed that fast grasp planning is not necessarily linked to anatomically good grasps and that selection in multiple-object simulations does not only depend on the speed of grasp-planning but also on the anatomical constraints.

The reaction times in the two-object simulations were – except for one result – slower than in the single object simulations. The time course of the activation in the finger maps showed a delayed increase of the activation in the two-object simulations compared to the single object ones. This was probably caused by more activation being fed into the hand network from the visual feature extraction stage (because there were more edges in the visual input) which slowed the WTA process down.

In Study 2c the top-down path of the model was enabled and used to simulate preferences for certain grasp postures. Overall the top-down path produced an increased activation in the combined weight maps compared to the simulations using only the bottom-up path. This increased the overall simulation speed. The results of the simulations showed that preferences for a certain grasp posture enabled the model to guide selection in the hand network towards objects suitable for the preferred grasp posture. The generated grasps were, however, different from the grasps generated by the bottom-up-path alone due to the influence of the template maps on the selection process. This resulted in grasp postures which resembled the selected grasp template as closely as the selected object allowed.

The overlap of the different templates in the combined weight maps turned out to cause unexpected effects in some simulations. For example, two of the templates positioned the fingers similarly (the small and the large grasp). In some simulations this resulted in preference for objects which allowed such placement of the fingers even when a different template was preferred because the added activation of the non-preferred templates reached almost the same levels as the preferred template. Because of this effect, the simulation failed to grasp the trapezium 90° despite a preference for circular grasps in simulations with trapezium 90° and trapezium 270°. This also demonstrated that object selection in SAAM is only driven by matching anatomical constraints and geometrical constraints and not by identifying the objects as such.

If a grasp template was preferred which was not suitable for either object in the visual input then the model exhibited one of two behaviours: It either kept the preferred template activated and grasped one of the objects with an adapted grasp posture not normally used on this object, or it overwrote the grasp template preference and grasped one of the objects while selecting the template matching the generated grasp best. The choice of the model depended on whether one of the objects could be grasped with an alternative grasp which matched the preferred template as it was the case in trapezium 270°/bar 0° simulation, for example. If none of the objects could be grasped with such an adapted grasp then the model grasped one of the objects, and the choice depended on which objects matched better the current activation in the weight matrices. In response to the changing grasp posture the grasp template network shifted the selection from the preferred grasp template to the best matching one (simulations with trapezium 90°/trapezium 270° and t-shape 0°/pyramid 90°). As in the bottom-up simulations the activation of the non-selected templates indicated how well they matched the current grasp posture in the finger maps.

Apart from two object simulations with a focus on object selection, Study 2c also included two stimuli to investigate whether grasp preferences could also be used for more function-oriented decisions. The first stimulus presented two objects in an arrangement allowing to grasp either one of the objects or both together. The results of the simulations showed that preference for either large or small grasps enabled the model to choose between grasping both objects or grasping only one of them. The second stimulus showed only one object. However, this object could be grasped either with a small grasp or a circular grasp. The simulations showed that the finger positions selected by the hand network could be influenced by the grasp template network to either grasp the part of the object which was suitable for a small grasp or for a circular grasp. This finding is similar to the observation that SAAM was able to find alternative grasps on objects when no object directly matched the preferred grasp template in the two-object simulations.

7 Study 3: Affordances and Action Intentions

The objects used in the studies so far had no specific function linked to them. Grasping actions could be performed with them but none of the objects had an associated function it could be used for. Therefore, apart from affordances for grasping no other affordances were investigated in the first two studies. To overcome this limitation and to demonstrate that SAAM cannot only recognise affordances for grasping but other, more complex affordances as well, simulations of grasping hand tools (pliers and hammer) are presented in this final study. This study did not add any new functionality to SAAM but aimed at showing how the representations of affordances and action intentions in SAAM can be linked to affordances of real objects and intentions to use them. Napier's observation that different objects require different specific grasps to properly use them provides the theoretic background for incorporating complex affordances into the model (Napier, 1956). Therefore, affordances for actions will be described by describing their specific grasps.

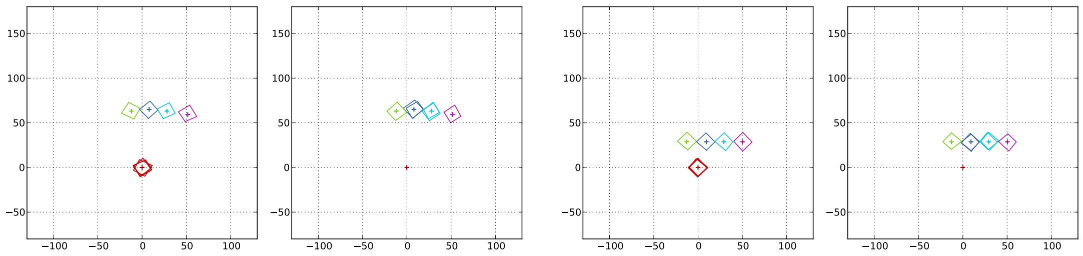
In the first set of simulations the recognition of affordances was investigated when there is no top-down guidance of the selection process. In the second set the ability to guide selection of affordances and objects through SAAM's top-down path was explored. The results showed that the action intentions expressed by the top-down influence allowed to prefer certain affordances over others and to discover new affordances of objects.

7.1 Study 3a: Affordances

The first study in this chapter explored SAAM's ability to recognise affordances in the environment. In the studies presented in chapter 6 the templates in the grasp template network described merely grasp postures suitable for grasping different objects. Now, templates of grasp postures were chosen which additionally encoded actions which could be performed with an object hold with the grasp described by the template. Thus, by classifying a grasp SAAM also discovers what subsequent action could be executed with the object when it was held with this specific grasp. In other words: SAAM found out which complex affordance an object had. Obviously, the affordance which was recognised depended on the grasp posture which was generated by the bottom-up path of the model. The implications this has for the argument about where affordances exist within the object-observer system will be discussed in the general discussion.

Method

Stimuli Two objects were designed which depicted a hammer and pliers viewed in an orientation suitable for right-handed usage. Three different stimuli were prepared with the two hand tools: Two single-object stimuli showing each object separately in the centre of the visual input field and a multi-object stimuli which showed both tools placed randomly in the visual input field. The input in the multi-object simulations was twice as wide and high than in the single-object simulations. The stimuli are shown in Fig. 7.2. Different object dimensions in the figures are because of the



(a) Pliers grasp template

(b) Hammer grasp template

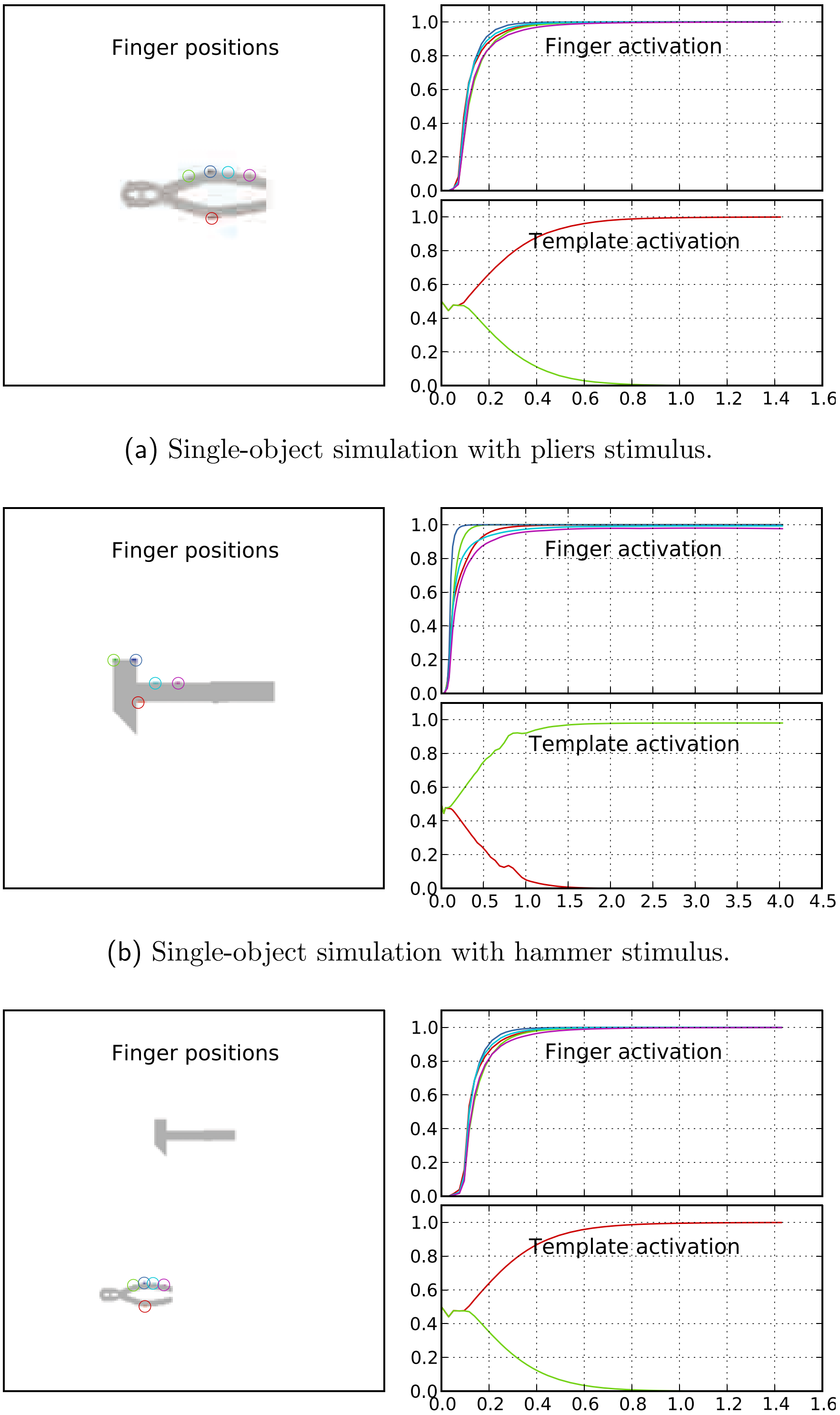
Figure 7.1: Template maps for representing grasps to use a hammer and to use pliers. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

different sizes of the visual input field. The objects had the same size in single and multiple-object simulations.

Parameters and Templates The simulations used the same parameters as the simulations described in Sec. 6.2.1. The top-down path was switched off. The grasp template network contained two templates representing a grasp for using pliers and a grasp for using a hammer. They were constructed in the same way as before. The polygons which defined the finger positions for each grasp type are shown in Fig. 7.1. The same Gaussian function as in the template maps of the first study was used (see Fig. 5.4c).

Results and Discussion

The results of the simulations of bottom-up affordance recognition are shown in Fig. 7.2. The simulation of the pliers stimuli (see Fig. 7.2a) produced no unexpected results. The generated finger positions are sensible for using the pliers and the use-pliers template was activated indicating a selection of use-pliers action. The simulation terminated quickly once the use-pliers template was fully activated. In contrast, the simulation of the hammer stimuli worked less well (see Fig. 7.2b). The finger positions created by SAAM are obviously not suitable for using the hammer.



(a) Single-object simulation with pliers stimulus.

(b) Single-object simulation with hammer stimulus.

(c) Multiple object simulation with pliers and hammer stimulus.

Figure 7.2: Simulations without top-down feedback. Template activation colours are red for pliers-use and green for hammer-use. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

This is however not too surprising, because the simulation of the t-shape 180° stimuli in Chapter 5 already showed a tendency of the model to grasp t- or hammer-shaped objects close to their thick end. The reason for this was the better fit of the wider object parts with the optimal grasp width. On the hammer the thumb could only be placed on the bottom edge of the handle due to the object geometry. While fingers could be positioned at the top edge of the handle this was not the optimal grasp width encoded in the hand maps. Therefore, the model preferred to place the index and middle finger on the head of the hammer. There, the two fingers quickly gained activation (see time course of the finger activation) because of the input from the thumb and mutual reinforcement due to their horizontally aligned positioning. The remaining fingers were then positioned as well as possible in relation to the index and middle finger. With more fingers becoming stronger activated the thumb activation also increased which can be seen in the finger-activation plot. Interestingly, SAAM classified the generated grasp as a hammer-use grasp despite it being clearly different from the template and not suitable for using the hammer. The reason for this was that in these simulations only two grasp types were defined neither of which fitted the generated grasp. Hence, the WTA process in the grasp template network could only select the least bad fitting template which happened to be the hammer-use template. This was because the finger configuration with the parallel arrangement of the index and middle finger and the ring and little finger as well as the grasp width had a more similarity to the hammer-use than the pliers-use template. The slow increase of the activation of the hammer-use template – in contrast to the pliers-use template in the previous simulation – and the final level of activation indicated that the selected template did not match well, though. Also, because the hammer-use template was not fully activated it never reached the threshold defined by the stop criterion and the simulation terminated only after reaching $t_{\max}$.

This outcome of the simulation is interesting because it shows that sometimes knowledge is required to use an object properly. If the hammer is just grasped without any intention of using it (as it was done here) a grasp close to the head is perfectly fine and even sensible considering the weight distribution in a hammer. However, when there is the intention of using the hammer a specific grasp at the handle is required. How such an intention can be expressed and modelled with SAAM will be explored in the next section.

After having discussed the results from the single-object simulations it is not surprising that in the multi-object simulation (see Fig. 7.2c) the pliers are selected and the pliers-use template is activated. The finger positions and time courses of finger and template activation do not differ significantly from the simulation of the pliers stimulus alone.

7.2 Study 3b: Action Intentions

After having shown that the grasp template network can extract affordances from the visual input, the final set of simulations explored whether action intentions (i. e. preferences for certain affordances) could also be simulated with SAAM. By preferring one grasp type over the other SAAM can express the intention of executing the action associated with this grasp type.

Method

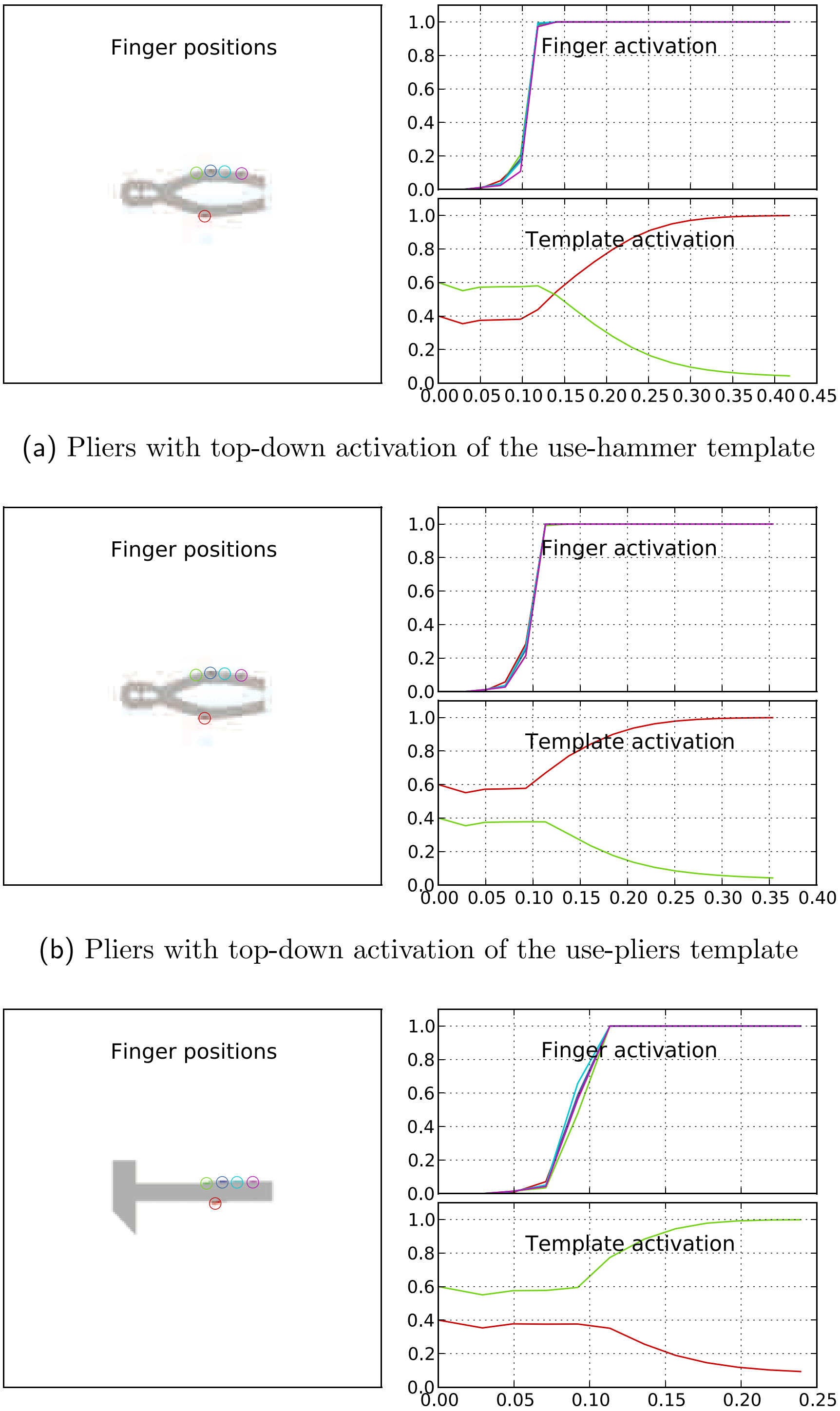
Stimuli The same stimuli as in Study 3a were used.

Parameters and Templates Parameters and templates were the same as in the simulations for modelling affordances. The top-down-path was, however, activated with the template factor set to $a_{\mathrm{tpl}} = 4.0$. The output neurons of the grasp template network were initialised to 0.6 for the preferred grasp and 0.4 for the non-preferred grasps. The weight maps in the hand network were set to the same reduced activity as in the previous top-down simulations (see Sec. 6.2.3).

Results and Discussion

The simulations of the pliers stimulus worked well no matter which grasp type was preferred. In both cases the same grasp posture was produced and the pliers-use template was selected (see Fig. 7.3a and 7.3b). The simulation with a preference for hammer-use ran slightly longer than the other one because the template activation obviously needed to change from the hammer-use to the pliers-use template. Because of the stronger activation of the hammer-use template in the first simulation, the combined weight maps did not fit the pliers handle as well as in the second simulation which resulted in a slower increase of the finger activations. Once the template activation had shifted enough over to the pliers-use template, the increase of the finger activation became faster and comparable to the second simulation where the pliers-use template was preferred from the beginning. The action intentions worked as expected on the pliers stimulus: an intention to use the pliers increases the speed of the grasp generation. If an intention ('use hammer' in this case) could not be realised with the available objects then the bottom-up affordances overwrote the intention.

Compared to the previous bottom-up affordance simulation, the hammer stimulus simulations with action intentions produced interesting results (see Fig. 7.3c and 7.4a). While the hammer had only a (bottom-up) affordance for grasping as it was shown in the previous simulations, the introduction of action intentions produced different grasps than before. The fingers were no longer placed around the head of the hammer but on the handle so that the hammer could be used properly. The development of the time course in the hammer simulations was similar to the time course of the pliers simulations. When a hammer-grasp was preferred, the simulation ran faster than when pliers usage was intended. Like in the pliers simulation which had a preference for the

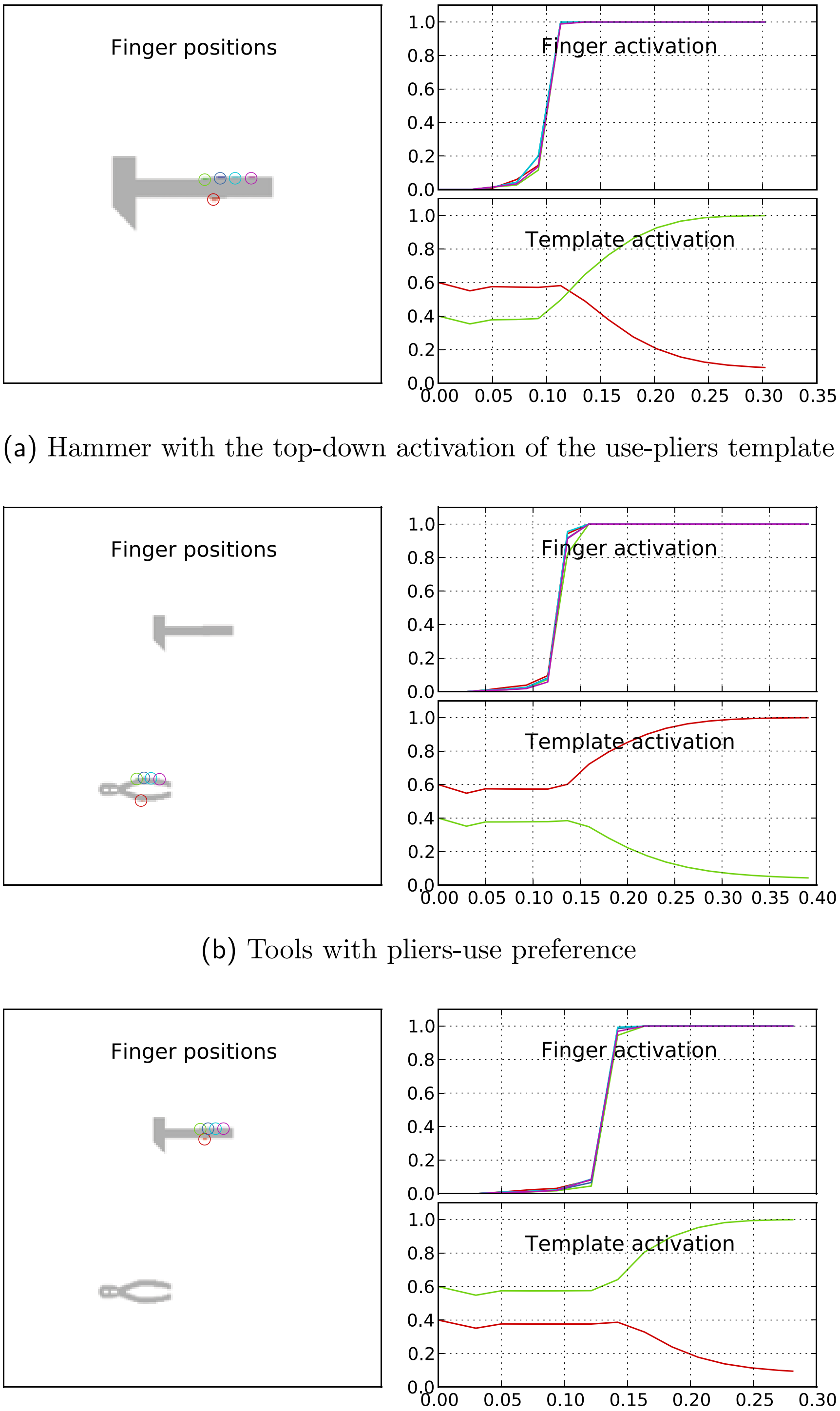


(a) Pliers with top-down activation of the use-hammer template

(b) Pliers with top-down activation of the use-pliers template

(c) Hammer with top-down activation of the use-hammer template

Figure 7.3: Top-down simulations of single-object stimuli. Template activation colours are red for pliers-use and green for hammer-use. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.



(a) Hammer with the top-down activation of the use-pliers template

(b) Tools with pliers-use preference

(c) Tools with hammer-use preference

Figure 7.4: Top-down simulations of multiple-object stimuli. Template activation colours are red for pliers-use and green for hammer-use. Fingers are coloured as follows: thumb–red, index finger–green, middle finger–blue, ring finger–cyan, little finger–pink.

use-pliers grasp the simulation of the hammer stimulus with an initial preference for the pliers-use template took longer because the template activation needed to shift from one template to the other first.

The last two simulations were again simulations with multiple objects. It was tested whether the action intentions can guide object selection. The results shown in Fig. 7.4b and 7.4c confirm this. An intention to use the pliers generated a grasp of the pliers; an intention to use the hammer created a grasp of the hammer which was suitable for using the hammer.

The simulation results confirm SAAM's ability to use action intentions to control on which object the generated grasp is placed and how the object is grasped. This demonstrates how action intentions can guide visual attention for action. It was shown that this works to guide grasp generation in single object stimuli to prepare for specific actions as well as in multi-object stimuli where additionally an object needs to be selected.

7.3 Summary

In Study 3 the findings from Study 2 were applied to stimuli showing 'real' hand tools. These examples illustrate more clearly than the functionless objects used in the previous studies how SAAM's behaviours are linked with affordances and action intentions. First, Study 3a showed how the two hand tools were grasped when only bottom-up affordances were taken into consideration to make a grasp decision. In these bottom-up simulations grasps on both hand tools were produced. The grasp of the pliers was suitable for using the tool. The grasp for the hammer, however, was not suitable for using it for hammering but only for holding the tool. For both tools the appropriate template was activated indicating the action associated with the object. The hammer-use template was not fully activated, though, which indicated that the grasp posture encoded in the template was not deemed to match the created grasp *well-enough*. Considering that the grasp created on the hammer was not suitable for hammering, this was a good result. It must also be considered that only two templates were available. Hence, the model was forced to select one of it and could not say that neither matched at all.

Study 3b explored how action-intentions influenced the grasp decisions of SAAM. First, it showed that planning to make a hammer-use grasp changed the grasp of the hammer to one suitable for using the tool. Second, intentions allowed the model to select the tool which matched the intended action. The single object simulations showed that the model was still able to overwrite these intentions when no suitable object was available. This might seem unnatural at first sight but since the model has no mechanism to *not* create a grasp at all, this result is not surprising. Obviously, some higher level processes are necessary to stop grasping if it is realised that no suitable object for the intended action is available.

Part III

General Discussion

8 General Discussion

The general discussion is organised in four sections: First, the architecture of the model and the findings from the studies are summarised. Then, the model will be discussed in the context of grasping and visual attention, followed by a discussion of the role of affordances in SAAM. This chapter concludes with a comparison of SAAM and the models presented in Chapter 4.

8.1 Summary of the Model

The full version of SAAM has two main pathways: a bottom-up and a top-down path. The bottom-up path starts with a visual extraction stage in which edge detectors mark locations in the visual input which meet the geometrical constraints of grasping – namely horizontal and vertical edges. The output of the horizontal edge detector is separated into edges at the top and edges at the bottom side of objects. The top-side edges are fed into the finger maps of the hand network and the bottom-side edges into the thumb map. The finger maps receive additional input from the vertical edge detector. These have reduced activation, though, compared to the horizontal edge detector. In the hand network weight maps project activation between the finger maps. The weight maps encode the anatomically possible positions for one finger in relation to another finger. By following a gradient descent procedure the model finds the finger positions which best match the geometrical and anatomical constraints.

The grasp posture generated by the model is represented in the activations of the finger maps in the hand network.

In Study 2 SAAM was extended with the grasp template network. This additional network enabled the model to classify the grasp postures described by the finger positions in the output of the hand network. Optionally, a top-down path can be activated to feed information about preferred finger positions from the grasp template network back into the hand network. The grasp template network compares the finger positions from the hand network with grasp postures encoded in template weight maps and selects the best matching template. If the top-down path of the model is activated, the template maps are applied to the original weight maps in the hand network so that the combined weight maps have a bias towards the grasp described by the template.

The abilities of SAAM were explored in three studies which are summarised in the following sections. In Study 1 simple grasp generation in the hand network was tested and the results were compared with experimental data. Study 2 investigated grasp classification and top-down guidance in single and multiple object simulations. Study 3 applied the results from Study 2 to a 'real-life' example to demonstrate how complex affordances and action intentions are handled in SAAM.

8.1.1 Findings in Study 1: Evaluation of SAAM's Grasps

Study 1 explored the bottom-up path of SAAM. The aim of this first study was to demonstrate that SAAM is capable of generating grasp postures in general and, moreover, that these grasp postures are anatomically feasible and similar to the grasps humans use for the same tasks. In order to assess the similarity of simulated grasp postures and experimentally collected grasps, a measure was defined which allowed to estimate how well a grasp fitted in a group of reference grasps. A grasp received a high rating if its fingers were positioned in close proximity to the fingers of many grasps of the reference group. To compare the simulated grasps and the experimental grasps the experimental grasps were used as the reference group. However, in order to interpret the rating score for the simulated grasp, each experimental grasp was rated against all other grasps in the reference group as well. The score for the simulated grasps could then be compared against the ratings of the experimental grasps. The results demonstrated that the simulated grasps rated consistently high compared to the grasps from the experiment. This showed that SAAM was able to produce grasp postures which used finger positions similar to the ones many human participants chose. An exception of this were the grasps of the vertically oriented objects. Due to the restriction to horizontally oriented grasp postures, the model could not create grasps which matched their human counterparts because humans preferred to rotate their hands when grasping vertically oriented objects. This observation was confirmed by fitting a logistic model to the width/height-ratio of the objects in relation to the grasp orientation.

The results of the simulations showed that SAAM was able not only to select a position for each finger but also to convey additional information about the generated grasp: First, the results showed that the size of the activated area for each finger depended on the number of possible finger positions. In cases where more than one grasp posture was equally supported by the geometrical and anatomical constraints, SAAM did not randomly select one of these postures but instead activated all possible finger locations. Second, the activation level of each finger depended on how well the position of the finger fitted the anatomical and geometrical constraints. Grasps which fitted the constraints well were assumed to provide more stable grasps than ones which fitted the constraints less well. The required grasp stability obviously depends on the object that it is grasped and the action the object will be used in. In SAAM this was reflected by the stop threshold criterion. A grasp posture was deemed as suitable if all finger activations were stronger than a pre-defined threshold. Of course, in an advanced version of SAAM this threshold could be adapted based on the requirements of the task that is performed. In Study 1 the threshold was set to a high level to ensure that clear grasp postures were produced. A number of simulations failed to reach this threshold (mainly simulations with vertically oriented objects); these simulations were aborted after a time-out.

The simulation durations showed that grasps for thin objects (i. e. the bar and the t-shapes) were slower than for larger objects. This is in line with experimental data from Derbyshire et al. (2006) who observed that precision grasps were slower than power grasps (see also Tucker & Ellis, 2001). Furthermore, this supports the idea that the templates introduced in Study 2 were approximations of power and precision grasp postures. Hence, the small grasp template which matched the grasps of the bar and t-shape approximated a precision grasp while the large grasp template represented power-grasps.

In the experiment in Study 1 reaction times were measured. However, the analysis revealed almost no significant differences between the different conditions. Therefore, these data were not further analysed and not compared with the simulation data.

The simulations in Study 1 revealed two emergent properties of SAAM. First, SAAM placed the fingers around the centre of gravity of the object despite there being no mechanism to discover the centre of gravity. Second, SAAM was able to place fingers on vertical edges of the objects if the space available on the horizontal edges was not sufficient or if the grasp span were to wide. By doing this the model created grasps which achieved two-dimensional form closure (i. e. movements were blocked in all directions by fingers). The experimental data confirmed that these properties can be observed in grasps made by humans as well.

Finally, an important aspect of SAAM is that it achieved human like grasps without doing explicit calculations of forces and torques. The geometrical and anatomical constraints in the hand network implicitly define the requirements for a stable grasps. This method provides a much simpler way to generate stable grasp postures compared to a calculation based on the forces and torques which occur in a grasp (see, for instance, Mason, 2001 for examples on the calculation of forces in grasps).

8.1.2 Findings in Study 2: Grasp Classification and Selective Attention

After having demonstrated the basic functionality of SAAM in Study 1, I set out in Study 2 to first investigate the grasp classification abilities of the grasp template network and then the new abilities of SAAM which emerged from the activation of the top-down path between the grasp template network and the hand network.

Study 2a tested whether the grasp template network was generally able to classify the grasp postures produced by the hand network in single object bottom-up simulations. The grasp template network contained three different grasp templates which were modelled based on power grasp, precision grasp and circular precision grip postures. These postures were commonly distinguished in the grasp taxonomies which I discussed in Chapter 2. Study 2a demonstrated that the grasp template network was able to select a grasp template based on the grasp produced by the hand network. This showed that the classification of grasps into different grasp type categories works. Since the different grasp types represented different functional categories of grasps, these results can be interpreted as a selection for action. An important aspect of this selection process was that it only relied on functional object features and did not utilise semantic information about the seized object. Hence, the selection of an action in SAAM can be understood as the discovery of an object's affordance. In line with this interpretation, the simulation results revealed that competition in the grasp template network was stronger between templates which shared functional features,

e. g., similar relative finger positions, compared to templates which differed more.

Study 2b aimed at exploring whether SAAM's selection for action mechanism worked only on single object inputs or also on displays containing more than one object. Additionally, this study brought SAAM's ability for object selection into focus. While the grasp template network selected an action, the hand network was responsible for selecting an object in the visual field and generate the parameters required for executing the chosen action with the object. The simulations in Study 2b demonstrated that SAAM was able to perform these tasks by selecting a single object in the visual field as well as selecting the appropriate grasp template at the same time. The selection of an object for action in the hand network can be interpreted as visual attention for action because in order to select an object visually for action it is not important focus on the object as such but instead to specify how it is used in an action. This aspect of the model will be discussed further in Section 8.3.

The object which was selected by SAAM in the two-object simulations could be predicted in most cases by comparing the simulation durations from Study 1. The object with the shorter simulation duration was usually the winner in the two-object simulations. The only exception was the trapezium 270° stimulus which was not selected in simulations despite very short simulation durations in Study 1. This exception is interesting because it showed that object selection depended obviously not only on the speed of grasp generation but also on the conformance of the generated grasp posture with the anatomical constraints. The example of the trapezium 270° demonstrated that an object could be selected for action if it fitted the anatomical constraints better then other objects in the visual field even when it took SAAM longer to find this good match. Therefore, it can be concluded that quickly grasped objects did not necessarily draw the focus of attention (for action) towards themselves. In contrast it is more likely that fast objects were often fast in the first place because they provided an easily identifiable good match with the anatomical constraints.

In the final part of Study 2 the top-down path which connected the grasp template network and the hand network was activated in order to enable SAAM to guide object selection based on preferences for certain grasp postures. The simulation results of Study 2c showed that this guidance through the top-down path worked well and in most simulations SAAM selected the object which fitted the preferred grasp posture best. Additionally, it was observed that the grasp postures in Study 2c differed from the grasp postures created by earlier simulations without top-down feedback. Obviously, this was caused by the modification of the weight maps in the hand network during the simulations. It highlights the capabilities of the top-down path to not only guide object selection towards specific objects but also to directly influence the grasp posture which is produced. In the last two simulations of Study 2c this behaviour was utilised to guide finger placement to different parts of an object.

The modification of the combined weight maps during grasp planning is a powerful method to incorporate feedback into the hand network. However, the combination of the three different template maps and the original weight maps was highly dynamic and did not always produce the expected results. For example, overlapping activations from different template and weight maps could result in very strong regions of activation in the combined weight maps which were difficult for other templates to exceed even when they were preferred over the others. In some simulations this effect proved to have caused unexpected results.

The simulations in Study 2c also demonstrated that SAAM could adapt in situations when no object matched the preferred grasp template. In these cases either one of the objects would be grasped with a grasp not normally used for it, or one of the objects was grasped with a different grasp than the one described by the preferred template. In this case the classification in the grasp template network changed to reflect the actual grasp posture.

Above, I discussed that grasp classification in SAAM can be understood as the discovery of an object's affordance. Consistent with this interpretation the behaviour of the top-down path can be seen as the processing of action intentions in the model. By preferring a certain grasp type, e. g., for power grasps, over the other SAAM indicates an intention to perform a 'power grasp' action.

8.1.3 Findings in Study 3: Affordances and Action Intentions

Study 3 was conducted with the intention to apply SAAM to a 'real-life' example in order to better demonstrate how its behaviour can be explained as affordances and action intentions. This was achieved by presenting two hand-tools (hammer and pliers) in the visual input and by using grasp templates which resembled the grasps used when handling the two tools.

The bottom-up simulations in Study 3a aimed at showing how SAAM discovered the affordances of the tools directly from the visual input. This worked worked very well for the pliers stimulus. The 'pliers-use' grasp template was activated to indicate that the object in the visual input has an affordance for being used as pliers, and the hand network selected finger positions at the handle of the pliers. In contrast, the simulation of the hammer stimulus did not yield the expected result. The grasp template activated the hammer template but this activation was not strong enough to exceed the stop criterion threshold. This meant, SAAM was not able to properly discover the affordance for using the hammer. The reason for this was the grasp posture which was selected by the hand network. Instead of grasping the hammer at the handle, it generated a grasp around the head of the hammer. Obviously, it is not a good idea to use a hammer while holding it like this. Nonetheless the grasp is appropriate for carrying the hammer. These two simulations revealed an important aspect of bottom-up affordances in SAAM: They only included information about the shape of the object and the anatomical constraints; no knowledge about using specific objects was incorporated in the affordance discovery. The case of the hammer showed that SAAM could not derive a hammer-use grasp purely through bottom-up affordances. The fact that a hammer needs to be grasped as far a way from the centre of gravity as possible is something that cannot be derived from the visual input alone.

In order to investigate how grasp creation changed when additional knowledge was available, the top-down path was enabled in the simulations in Study 3b. The results showed that two things can be achieved by including top-down knowledge about object use. First, it enabled the model to discover if an object could be grasped in a way for using it in a specific task. The simulations of the hammer stimulus proved that this worked: Instead of grasping the hammer around the head, the model used its knowledge to guide the hand network to produce a grasp at the handle of the hammer.[1] Second, the top-down knowledge allowed the simulation of action intentions. By preferring a grasp suitable for using the hammer, SAAM could be directed to select the hammer instead of the pliers and vice versa.

The next section explains the special role of grasping in SAAM. In the following sections two aspects of the simulations in Study 2 and Study 3 are discussed further. First, relations between visual attention for action and visual attention for identification are highlighted, then SAAM's connection to affordances is discussed.

8.2 The Role of Grasping in SAAM

Grasping plays an important role for specifying actions and finding affordances in SAAM. Based on Napier's (1956) observation that the way we grasp an object

[1] It must be noted that the extend to which SAAM can include knowledge about functional object use in the grasp templates is, of course, limited. This is partly due to the limitation of SAAM to two dimensions and partly a general restriction of the expressiveness of grasps. Obviously, it would be difficult to guide grasp selection to a specific position on a uniform object which has no variations in its shape (for instance a long pole). This limitation could be overcome by extending preprocessing of the visual input. For example, only edges in a specific area of the object could be fed into the hand network to enforce a grasp generation within this part of the object.

tells us how we use this object in a subsequent action, I developed the idea to use grasp postures as an encoding for affordances and action intentions. The simulations demonstrated that this is viable method to implement affordances. However, the rationale behind this method and its realisation in SAAM still needs to be explained. In the model the connection between grasp postures and complex actions proposed by Napier is made in the grasp template network. This network matches the grasp postures generated by the hand network to prototypical grasp postures for specific actions (grasp templates). For example, in Study 3 the grasp template network identified grasp postures either as a grasp for hammering or as a grasp for using the pliers. Additionally, the hand network makes a connection between grasp postures and objects in the visual input. Together these two connections – in the grasp template network and in the hand network – establish a path linking objects to actions and, thus, realising affordances. Thereby, the grasp postures in the finger maps are the crucial element of this path: They enable SAAM to link objects in the visual input to complex actions defined in the grasp template network. Of course, the link between objects and actions is not uni-directional but also connects actions to objects enabling the model to describe action intentions with grasp postures as well. This was demonstrated in Study 2c and Study 3b.

8.3 SAAM in the Context of Visual Attention

Selective attention for action is, of course, linked to visual attention for identification. Discussing SAAM against this background provides some interesting insights. This involves the visual feature extraction stage, the model in general, and the top-down path in particular.

8.3.1 Visual Feature Extraction

The visual feature extraction stage of SAAM forms a crucial part of the model's architecture. It identifies locations in the visual input where features of the environment suitable for interacting with it might be. In SAAM Gabor-based edge detectors are used to find such locations. As I argued in Chapter 5, Gabor-filters are used because they are a biologically plausible implementation of edge detectors (Daugman, 1985; Field, 1987; Petkov, 1995). This plausibility does, however, not explain for what reason these filters developed in the brain at all. Work by Olshausen and Field suggested that Gabor-type filters developed in the striate cortex (V1) because they are capable of extracting high-order relations between visual structures which are often found in natural images (Olshausen & Field, 1996a, 1996b). While this explains why this specific type of edge detector developed in the mammalian brain, it does still not explain for what purpose edge detectors are required in the first place. Olshausen and Field argued that the filters provided an efficient representation for further processing in the brain. SAAM demonstrates this but, furthermore, it provides a broader explanation for the existence of the edge detectors: Edges need to be identified in the brain because they mark locations in the environment which are suitable for grasping. One might argue now that edge detectors are also found in other mammals which are not able to grasp objects (like cats, e. g.). However, in general it can be observed that edges mark disruptions in the continuous fabrics that make up the environment. Such locations are obviously usable for grasping but also for non-grasping interaction, e. g., manipulation with the mouth or the paws. Hence, it is important for animals to recognise these locations which form the basis for interaction and consequently a basis for affordances as well.

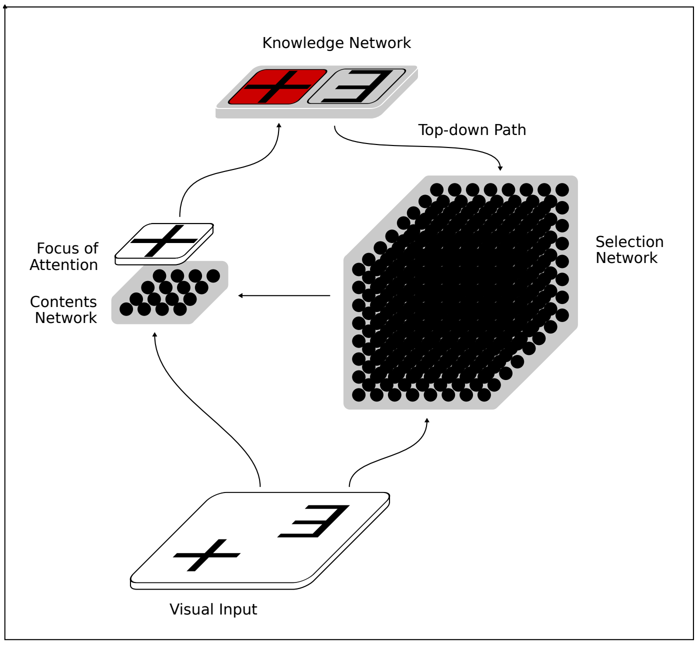


Figure 8.1: Overall structure of the Selective Attention for Identification Model (SAIM; adapted from Heinke & Humphreys, 2003). The visual input feeds into the selection network which selects a region in it. The contents network maps this region into the focus of attention (FOA). The knowledge network compares the contents of the FOA with templates of known objects and directs the selection process in the selection network towards locations showing the object which matches the selected template.

8.3.2 SAAM, SAIM and Biased Competition

SAAM's architecture was inspired by the Selective Attention and Identification Model (SAIM; Heinke & Humphreys, 2003). Both models are similar in their aim of modelling selection in multiple object scenes. However, the factors that guide the selection process and the encoding of the results are very different: SAAM selects objects for action based on affordances and expresses its selection by producing the parameters needed to execute a grasping action. SAIM, on the other hand, selects objects based on their salience and the selected object is projected into a 'focus of attention' for further processing.

SAIM is constructed from three individual neural networks (see Fig. 8.1). The 'selection network' is at the heart of the model. It selects which parts of the visual input are mapped into the focus of attention. Within the network excitatory connections activate neighbouring positions in the visual field while inhibitory connections enforce a one-to-one mapping between positions in the visual input and the output of the model. The actual output of the model is computed in the 'contents network' which receives the mapping computed by the selection network and uses it to project a portion of the visual input into the focus of attention. Finally, the 'knowledge network' adds template units corresponding to stored knowledge about objects to SAIM. The network matches the contents of the focus of attention with the object templates in the knowledge network. The activation of the template units is then fed back into the selection network forming a top-down path in the model.

The analogue structures of SAIM's selection and knowledge networks in SAAM are the hand network and the grasp template network. A counterpart to the contents network does not exist in SAAM because the hand network directly produces the desired output. The similarities between SAIM and SAAM exist not only on the architectural level but continue within the neural networks: The networks in both models use a combination of excitatory and inhibitory connections between neurons to guide the bottom-up selection process (in the hand and the selection network) as well as the top-down selection process (in the two template networks).

In SAAM the excitatory connections in the hand network activate neurons which satisfy the anatomical constraints defined in the weight maps. SAIM's weight matrices in the selection network, in contrast, define a neighbourhood constraint which encodes properties of objects. For example, neurons become active when the neurons representing neighbouring input locations are active thus encoding the spatial layout of objects. In both networks inhibitory connections enforce a selection process which selects the neurons which best satisfy the constraints defined in the weight matrices of the excitatory connections. It is because of the different weight matrices that the hand network in SAAM selects finger positions and the selection network in SAIM produces a mapping from the visual field to the focus of attention.

The excitatory connections of the two template networks define matchings between the model output and the templates. The templates are constraints for specific output patterns. Hence, the process can again be interpreted as a constraint satisfaction problem. As in the other networks of SAAM and SAIM, inhibitory connections exist in both template networks which select the template which is matched (or satisfied) best by the model output.

This system of two counteracting processes – an excitatory one which activates neurons satisfying domain-specific constraints and an inhibitory one which selects a subset of these neurons – may represent a general principle for information processing in attentional systems. Thereby, the weight matrices play a crucial part, because they define the constraints that stimuli need to satisfy in order to draw attention, and they consequently define what is behaviourally important within a specific attentional system.

Interestingly, the architecture of both models, SAIM and SAAM, follows the biased competition model which Desimone and Duncan (1995) posited to explain mechanisms of selective visual attention (Desimone & Duncan, 1995; Desimone, 1998). In the biased competition model the visual input is processed in parallel and all stimuli in the visual field compete for attention. In order to guide attention to the behavioural relevant locations the competition is influenced by bottom-up and top-down biases. Bottom-up biases guide the competition towards the selection of salient stimuli; top-down biases enable an animal to direct the competition towards behavioural relevant stimuli. Depending on the behaviour these top-down biases can vary widely. The description of the top-down biases for a certain behaviour has been termed “attentional template” (Duncan & Humphreys, 1989). In SAIM and SAAM the competition is distributed across two networks: The bottom-up part of the competition is computed in the selection and hand network while the top-down part is processed in the knowledge and grasp template network.

The individual networks in SAIM and SAAM are also consistent with the biased competition model. Work by Desimone and Duncan argued that two synaptic mechanisms can be distinguished in brain areas for visual perception. The first mechanism specifies which inputs of a neuron cause it to fire and the second one specifies which inputs allow a cell to fire (Desimone & Duncan, 1995). While Desimone and Duncan have not linked these two mechanisms directly to excitatory and inhibitory connections between neurons, these connections can be made in SAAM: The first mechanism reflects the functionality of the excitatory connections and the second can be associated with the inhibitory connections in SAIM and SAAM. This further supports the idea that the opposing excitatory and inhibitory processes in SAIM and SAAM signify a general principle.

The same functional separation of the two types of neural connections can also be observed in Cisek's computational model (see Chapter 4; Cisek, 2007). In this model excitatory connections propagate activation in certain locations between different layers of neurons while within each layer inhibitory connections generate a process to select one of these locations. In line with Desimone and Duncan's biased competition hypothesis Cisek interprets these interactions in and between different cortical areas as a competitive process (between affordances in this case). This means that the basic principles behind SAAM and Cisek's model are quite similar. However, while Cisek's model explores competition between affordances at a low level looking only at very simple affordances, SAAM attempts to investigate competition between high-level affordances in a connectionist framework.

8.3.3 The Top-Down Path

The discussion of the bottom-up path showed that the weight matrices play a crucial role in the implementation of a model's specific functionality. This observation leads to another interesting point of SAAM's architecture: The weight maps of the hand network in SAAM are not static but dynamically modulated with the top-down feedback from the grasp template network. To my best knowledge this is a novel method of applying a top-down bias in a connectionist model for visual attention. In SAIM and other models, feedback from the top-down path is applied to the visual input in order to facilitate bottom-up selection of stimuli which satisfy the top-down bias. For example, Deco and Zihl (2001) presented a model in which top-down information feeds back into feature maps through excitatory connections to activate specific locations in these maps (see Heinke & Humphreys, 2005, for a review of connectionist models for visual attention). In contrast, the visual input in SAAM does not change during execution; top-down feedback is only applied to the weight maps of the hand network.

By modulating its weight maps SAAM alters the definition of what make an anatomically feasible grasp dynamically during grasp generation. This means, it changes the constraints which stimuli must satisfy in order to be considered as behavioural relevant. As I argued above, these constraints in the weight matrices of a network are a model's defining characteristic. Hence, defining the constraints of a model essentially means implementing a specific selection process in the model. So, when SAAM modifies its weight maps during execution, it effectively dynamically adapts its selection process.

8.4 Affordances

SAAM was developed with Gibson's theory of affordances in mind. Study 3 demonstrated how the processes in the model can be related to affordances and action intentions. This topic requires further discussion. In Sec. 3.2 two propositions concerning the localisation of affordances within the agent-environment system were discussed: Turvey (1992) and Greeno (1994) located affordances in the environment, attached to objects, and existing independently of an agent's perception. The other school of thought argued that affordances only emerge from the relation between features in the environment and abilities of an agent (Stoffregen, 2003; Chemero, 2003). SAAM provides support for the latter theory. The weight maps in SAAM encode the constraints imposed on possible grasp postures by the anatomy of the hand. Thus, they specify the abilities of the agent to grasp objects. These abilities are then matched in the hand network with the edges which were detected as features of the environment. Only if these edges are in a configuration which fits the anatomical constraints, finger positions may be produced, classifying the object to which the edges belong to as graspable. Hence, 'graspability' emerges from the conjunction of features and abilities in the hand network; because, in order "[t]o be graspable, an object must have opposite surfaces separated by a distance less than the span of the hand" (J. J. Gibson, 1979, p. 133).

An important aspect of the affordance theory is the *direct* perception of affordances in the 'ambient optic array' (J. J. Gibson, 1979, p. 114). Obviously, SAAM does not actually retrieves its input from an ambient optic array in the Gibsonian sense, but uses a plain image containing only luminance information instead. This simplification was made because of the complexity involved in creating a realistic image of the ambient optic array in the computer. The images used by SAAM can be seen as a very simplified version of the ambient optic array. However, they still provide enough information for SAAM to find locations suitable for placing the fingers in a grasp. A more advanced version of the model could, of course, use a better approximation of the ambient optic array and extract additional information from it in its visual feature extraction stage. For example, binocular input images would allow to retrieve depth information from the environment. This information could be used to discover additional surfaces on which the fingers could be placed. Such improvements of the visual extraction stage would not change the hand network. It would still receive an activity map indicating locations where fingers could be placed, only that this map is now compiled from different sources of information in the ambient optic array. Interestingly, this map is very similar to the 'retinotopic neural map' which forms the first layer of the 'affordance competition framework' by Cisek (2008). The neurons in the retinotopic neural map represent the retinal space and their activity "reflects the likelihood that something of interest occupies [their] location in retinal space" (Cisek, 2008, p. 213). Since SAAM models grasping, locations of interest are obviously those that are suitable for finger placement.

8.4.1 A Hierarchy of Affordances

The affordance theory postulates a hierarchy of increasingly complex affordances constructed from 'invariants' and 'compound invariants' (see Sec. 3.2). Modelling this structure in a connectionist model was one of the main targets of my work. I will now show how SAAM's architecture matches this hierarchy of affordances. Within the affordance framework the information which is used to build the map of finger positions is known as invariants of the ambient optic array (J. J. Gibson, 1979, p. 310). As I explained earlier, invariants describe fixed (i. e. *invariant*) relations within the ambient optic array. Edges mark such a relation: a sharp boundary between two visually distinct regions. The edge detectors in SAAM's visual feature extraction stage discover these invariants and encode them in the map of locations suitable for placing the fingers. The process of generating a grasp from this information in the finger maps in the hand network is then, again, a process of discovering invariants. This time, however, fixed relations between invariants are explored. J. J. Gibson termed such relations 'compound invariants'. In the hand network the weight maps define which relationships can exist between the invariants in the network's input with regard to the hand of the actor. E. J. Gibson argued that these relationships are learned during childhood through interaction with the environment (E. J. Gibson, 2003). A future version of SAAM could explore the possibility of generating the hand maps through learning. This would, however, require feedback during learning which not only indicates the general feasibility of a grasp posture and its ability to hold an object firmly but also the comfort of performing the grasp. This information could be acquired either with an accurate computer model of the physiognomy of the human hand or by using an actual robot arm with appropriate sensors which provide information whether a grasp was successful and how comfortable[2] the posture for the robot was.

Recognition of compound invariants cannot only be observed in the hand network but also in the grasp template network. This network represents the processing of the highest, most complex, manifestation of affordances in SAAM. The templates in the grasp template network define invariants of finger positions in the output of the hand network which represent grasp postures for specific actions. Therefore, activation of the corresponding template units represents the existence of affordances for these actions in the visual field. Interestingly, the term affordance can be applied to the output of every stage of SAAM: The visual feature extraction stage detects locations

[2] 'Comfort' is, of course, not an intuitive measure to collect in a technical system. One way of defining such a measure could be based on the energy which is required to adopt a specific posture with a robot arm. Such an approach would also reflect Warren, Jr.'s experiments in which he measured the energy consumption of humans to find the most comfortable riser height of steps and the optimal points for a stair-climbing affordance (Warren, Jr., 1984).

with an affordance for finger placing, the hand network finds configurations of these locations which are graspable, and the grasp template network discovers affordances for complex actions.

8.4.2 Invariants as Body-Related Relative Features

A final aspect of invariants which shall be discussed here, is their relation to the body of the actor. Warren, Jr. explored this aspect in his experiments in which he related the size of bodily parts and objects to each other (e. g., hand length related to object width) by defining pi-numbers (Warren, Jr., 1984; Warren, Jr. & Whang, 1987; see Sec. 3.2.2 for details). The design of SAAM's hand network shows a striking similarity to the rationale behind Warren, Jr.'s experiments: The weight maps in the hand network relate the anatomy and physiognomy of the (modelled) actor's hand to the objects the actor sees. The effects of this relation on the generated grasps were described in Studies 1 and 2; depending on how object shapes and hand anatomy matched, the outcome of the different simulations varied widely. The multi-object simulations in Study 2 and 3 showed how some objects were preferred over others because they better matched the activity patterns in the weight maps. In Warren, Jr.'s experiments optima for an affordance and changes from one affordance to another were termed 'optimal points' and 'critical points' in the pi-number range. It is an interesting question, if optimal and critical points can be determined for the hand encoded in SAAM. An experiment which could provide a baseline for such an study was presented by Newell, Scully, Tenenbaum, and Hardiman (1989). In this study the number of fingers used to grasp different sized cubes was investigated. Changes in the number of fingers used to grasp a cube would mark the critical points in this task.

8.4.3 Micro-Affordances

Above, all outputs of the different networks in SAAM were interpreted as descriptions of increasingly complex affordances. However, in Sec. 3.2.3 the concept of micro-affordance (Ellis & Tucker, 2000) was introduced to explain the facilitation of components of an action in contrast to complete actions like Gibsonian affordances do. For instance, while Gibson would argue that an object has an affordance for grasping, in the micro-affordance concept an object would have a micro-affordance for a specific orientation of the hand in a grasp. Micro-affordances offer another interpretation of the output of the visual feature extraction stage and, in particular, the hand network of SAAM. Both components recognise features of objects in the visual input which give rise to specific grasps. The visual feature extraction finds, for instance, locations in the visual field which afford finger placings suitable for horizontally oriented grasps. Hence, locations representing a ‘horizontal grasp’ micro-affordance are selected. The hand network is slightly different because it does not explicitly define and encode different feature dimensions representing micro-affordances. Rather, micro-affordances for object features like potential grasp widths or grasp types are implicitly defined in the weight maps and indirectly encoded in the output of the hand network. Instead of outputting, for instance, a grasp type and a grasp span, the model outputs the exact position of each individual finger. Thus, grasp span and grasp type are provided implicitly and indirectly in the output.

Tucker and Ellis pointed out that micro-affordances have – in contrast to Gibson’s affordances – a representation, and this representation is directly associated with vision and action. The finger positions generated in the hand network can be seen as such a (combined) representation of micro-affordances. They are directly extracted from the visual input and still linked to it by the same frame of reference. However, at the same time the representation contains information for the motor system to plan a grasping action (e. g., by linking SAAM to Smeets and Brenner's model as suggested in Sec. 8.5). Micro-affordances are assumed to be "dispositional properties of a viewer's nervous system" (Ellis & Tucker, 2000, p. 466). Hence, they are fully specified in the observer and do not depend on the object or the environment. This is different from the understanding of affordances in SAAM which are situated in the conjunction between object features and the abilities which are encoded in the hand matrices.

8.5 Comparison with Other Models

In Chapter 4 models for visual attention and action were reviewed. None of the models discussed there combined visual attention, affordances and complex actions. One of the aims of the development of SAAM was to build a model which combined these three functions.

SAAM modelled the perception of visual affordances. The model worked directly on the information in the visual field instead of relying on preprocessed information about objects. This enabled SAAM to handle arbitrary objects and inputs containing multiple objects. This is essential for modelling visual attention. In contrast to some of the reviewed models (Fagg and Arbib's model and Ward's model), SAAM did not include a semantic route for object recognition.

Similar to Cisek's (2007) model, SAAM used a consistent encoding for actions throughout the model. This was achieved by using finger positions for the descriptions of grasp postures similar to the representation suggested by Smeets and Brenner. However, instead of modelling only the position of the thumb and the index finger, the positions of all five fingers were encoded in the finger maps. Complex actions were described in SAAM by specifying their characteristic grasp postures as Napier (1956) suggested. The similarity of the grasp encoding in SAAM (vision-oriented) and Smeets and Brenner's (1999) model (action oriented) demonstrate that this representation can link vision and action, thus, describing affordances for grasping and possibly more complex actions. An extended version of SAAM may be able to be linked to a computational version of Smeets and Brenner's model so that not only grasp planning but also grasp execution can be modelled.

9 Conclusions and Outlook

I set out to build a connectionist model for complex affordances and action intentions. In this thesis I explored many properties of this model. However, there are more behaviours in SAAM which would be interesting to explore, especially in connection with further experimental work. Starting points for such explorations are outlined in the next section before I conclude my work with some final considerations.

9.1 Outlook

In the general discussion I already mentioned some opportunities for further research based on SAAM. These ideas are summarised here and additional experiments are suggested and outlined.

9.1.1 Extensions of the Model and Additional Simulations

Currently, SAAM only relies on black and white differences in the visual input. This is only a poor approximation of the ambient optic array. This could be improved by using stereo images as input and computing a disparity image. Sudden changes of depth in this image mark potential borders of objects. Hence, an edge detector applied to the disparity image should be able to identify locations which may be suitable for grasping. These information could be combined with the colour based edge detectors. Such an enhanced visual feature extraction stage would, of course, also permit new experiments to investigate how different features like depth-based edges and colour-based edges contribute to grasp-planning in humans.

On the motor-control end SAAM could be extended to also execute the grasps it planned; therefore, simulating the complete route from vision to action and not stopping after grasp planning as it does at the moment. One way to extend the model into motor control would be to link it with Smeets and Brenner's model. The similarity of the encoding for grasps in both models would make this relatively easy.

Having a model which is able to model visual perception for action as well as action execution offers new possibilities: For example, such a model can be used to learn the anatomical constraints in the weight maps instead of using predefined patterns of activation. To learn the anatomical constraints the model would start producing random grasps and then use the feedback from the grasp execution to evaluate whether the chosen positions constituted a successful and comfortable grasp and save this information in the weight maps. Over time this should produce weight maps which represent the anatomic abilities of the hand closely.

An extended evaluation of the model could take Warren, Jr.'s critical and optimal points into account (see Sec. 3.2.2). This would be particularly interesting in combination with learnt weight maps because their structure might not be as clear as in the manually designed maps used here. One experiment to investigate this, is Newell et al.'s (1989) experiment which assessed the relation between the size of a cube and the number of fingers used to grasp it. However, other experiments could also assess the relationship between hand length, object height and vertical and horizontal grasps. Such experiments could also take the initial orientation of the hand into account to test whether this affects the critical point at which a grasp changed from horizontal to vertical orientation or vice versa.

In one of the simulations in Study 2c (Sec. 6.2.3) I tested SAAM's abilities to model grasp generation and action selection based on action relations within groups

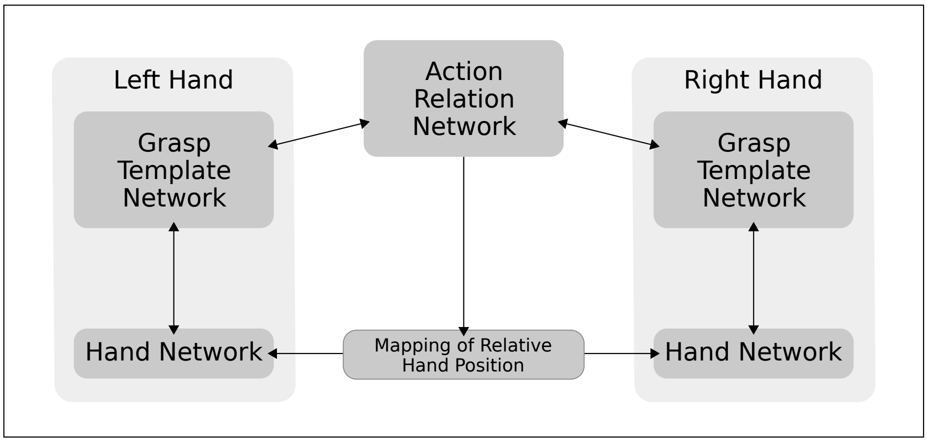


Figure 9.1: Overview of a bi-manual version of SAAM which could be used to model action relations between groups of objects in greater depth.

of objects. This was motivated by research by Riddoch et al. who showed that action relations between objects can support perception (see Sec. 6.2.3 for more details). However, the action relations investigated by Riddoch et al. were much more complex than the one simulated in Study 2c. For example, one of the object pairs they used was a bottle and a corkscrew while SAAM only used two bar-shaped objects. In order to explore such action relations between objects further, SAAM needs to be extended (see Fig. 9.1). Actions involving two objects are typically performed using both hands. Hence, SAAM would need to be duplicated so that each hand is represented by a separate hand network and grasp template network. Of course, the weight maps and template maps would need to be mirrored in the networks for the left hand. The left and right hand version of SAAM would be linked together by a network which connects to the two grasp template networks and encodes combinations of grasp types for both hands which constitute an action relation (e. g., a hammer-use grasp for the right hand and a small grasp to hold the nail for the left hand). Additionally, the hand networks of each hand would be connected to each other. Through these connections each hand could support locations for the other hand which constitute a good position for performing a combined action. These links would be similar to the connections between fingers in each hand network in so far as they describe the relative positions of the hands to each other. Since suitable positions for each hand depend on the action relation selected by the new network introduced above, this network would need to modify the weight maps used to project activation between the hand networks to fit to the current action. This is, of course, similar to the top-down feedback from the grasp template network in the current version of SAAM.

Finally, SAAM could be linked to a model for visual attention for identification (e. g., SAIM; see Sec. 8.3.2). Such a combination would constitute a complete implementation of Humphreys and Riddoch's (2003) dual route theory (see Sec. 3.1). Consequently, the architecture of the extended model could be similar to the Naming and Action Model (NAM, see Sec. 3.1). In this extension SAAM would implement a direct route for action and SAIM a semantic route. Both models could be linked at two points: First, the visual fields of both models could be connected so that a selection in one model increases the saliency of the selected area in the visual field of the other model. This connection would correspond to the link between stored structural knowledge and semantic knowledge in NAM. Second, the grasp templates in SAAM could be connected to templates of objects in SAIM for which the provide a suitable grasp. This would give SAIM access to stored action patterns. The connection would therefore correspond to the link between semantic knowledge and stored action patterns in NAM. Such a model would not only simulate the complete dual pathway including the semantic route in addition to the action route but would also allow to design experiments to examine the interaction and relationship of the two routes.

9.1.2 Further Experimental Verification

The grasps generated by SAAM in single-object simulations were verified with experimental data in Study 1. One possibility for further verifications would a comparison

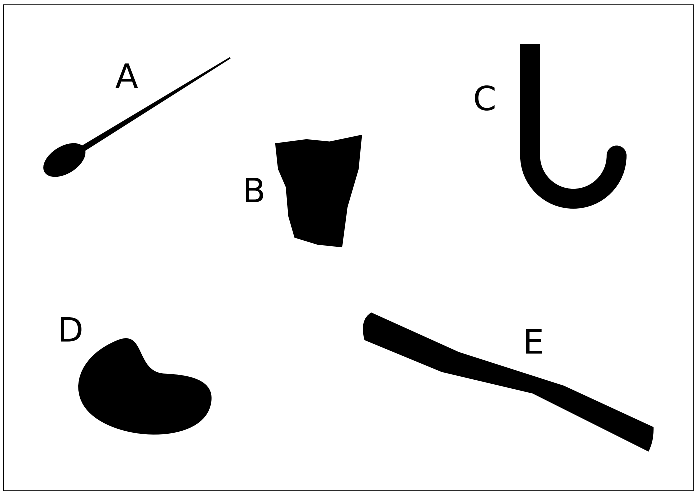


Figure 9.2: Examples of objects which might at first sight be perceived as functionless but when action intentions are present, they may be perceived as tools. For example, object A can be used as a baton if held at the rounded end and as a drumstick if held at the thin end; object B can be used a primitive hammer if held around its upper side; when object C is seen as the handle of an umbrella, it should be grasped at the straight part of the handle; object D can be used as a grinder when the hand is placed in the coving on the top edge of the object; and object E can be used as club when the hand is placed at one of the ends.

of the temporal behaviour of the model in single-object simulations to reaction time differences observed in humans. However, another – wider – field for experimental verification are the multi-object simulations. In this section experiments are outlined which can verify SAAM's results in such simulations.

Grasp Postures

Study 1 showed that on vertically oriented objects SAAM created grasps which differed from human grasp postures due to its restriction to horizontal grasp postures. An experiment could apply the same restriction to human participants by asking them to only perform horizontal grasp postures. It would be interesting to see whether the resulting grasps match the grasps SAAM created on vertical objects.

In Study 3 SAAM predicted that grasp postures change when objects are grasped with the intention of using them instead of simply holding them. An experiment to verify this prediction could use objects like the ones shown in Fig. 9.2. These objects do not have an intrinsic function, and a naïve participant would be likely to produce a grasp which is purely chosen for its comfort and stability. If participants are however told that the objects have a function, then the grasps should change to reflect the grasp posture needed for using the objects according to their function. This change would reflect the activation of the top-down path in SAAM which leads to the incorporation of complex affordances into the generation of the grasp posture.

Attention

SAAM showed a clear preference for some objects in multiple object simulations. A comparison of the choice made by the model and the choices humans would make could ameliorate the verification of the behavioural plausibility of the model. An experimental set-up to identify objects which humans prefer is however not straight forward. A pilot study aiming at investigating this question by presenting participants with two objects and asking them to choose one of them showed that the objects are perceived equally graspable and choices were made at random. This is not too surprising considering the amount of practise humans have on grasping. The simplicity and low demand of the task thus may have led the participants to entertain themselves during the experiment by exploring all objects. Choosing objects which vary more widely in their perceived graspability is not a solution for this problem, because this would require knowledge of the graspability (which is the target measure of the investigation) beforehand. The approach to compare reaction time differences for grasps in single object displays has also not proven successful as I discussed in Chapter 6. This is most likely also because of the general simplicity of the grasping task for humans. One approach to make the task more difficult would be to use a search task with present/absent trials in which multiple objects are presented and the participants need to find the target object first before grasping it. The target in this task could be defined, for instance, either by its shape or by its colour. A odd-one out search is also possible to further increase the difficulty of the task. Reaction times for different target objects could then be compared to establish the graspability of each object. An supplemental experiment in which participants respond with key-presses to whether the target object is present or absent would allow to account for reaction times effects attributed to different visual saliency of the objects. A different approach to solve the problem of task difficulty would be to increase the cognitive demand of the experiment with an unrelated task. Participants could, for instance, be instructed to play a simple computer game with a joystick held in one hand. The game would then require them to hold down a button with the other hand but at certain points lifting one of the objects would be required to perform an action in the computer game (e. g., picking some item up). In such a difficult task minor differences in the graspability of an object may become apparent. Another approach to find differences in the graspability of the objects would be to ask participants to rate all objects on their perceived graspability. This might, however, lead the participants to judge the objects based on analytic reasoning about graspability instead of assessing the actual graspability of the objects.

Priming

Another approach to investigating action and affordances in experiments is the presentation of a prime which might be associated with an action (through affordances or other means) prior to the execution of an action task by the participant. Researchers tried different types of stimuli as primes (e. g., Craighero et al., 1996; Bekkering & Neggers, 2002; Borghi et al., 2007; Vingerhoets, Vandamme, & Vercammen, 2009). These stimuli can be broadly categorised into those showing an object that affords a specific action, those showing the execution of a specific action directly, and those priming a certain characteristic of the action, for instance the orientation of a grasping action. Object primes can be construed as a form of priming based directly on visual affordances. In contrast, priming by showing the execution of an action can be interpreted as a trigger for a mental simulation of the action which causes the motor system to prepare for this action which then in turn affects perception for action. The priming of certain properties of actions can be interpreted similar to object primes only that it relies on concrete feature dimensions instead of complex affordances. This links this form of priming closely to micro-affordances. All three priming methods can be – to varying extents – simulated with SAAM.

Object priming can be achieved in SAAM in two ways: One way is to initialise the templates in the grasp template network with different activations to simulate a behavioural preference caused by a prime. The other way is to place the prime in the visual input of the model and replace it during the simulation with the actual target thus simulating the actual perception of the prime during the trial. During prime presentation the model should produce a grasp for the prime and select a matching template. When the target is then presented the grasp posture should be adapted to fit the target. If prime and target are compatible, this process should benefit from the preselected finger positions as well as the preselected grasp template. Depending on the similarity of the shapes of prime and target and their position in the visual input, the preselected finger positions might, however, not support grasp generation but delay it because the target requires the fingers in different positions than the prime even when both objects are still compatible with the same template. Using these methods, different experiments can be simulated with SAAM. For example, the prime could show an object compatible with a large or a small grasp and afterwards an object suitable for either grasp could be shown (an object similar to the one used in Study 2c). The generated grasps should then differ depending on the prime. This can, of course, be also combined with a search display instead of a single object. Considering cognitive demands the latter might also be more likely to produce successful results in an experiment with human subjects.

The method of displaying a prime in the visual input cannot only be used to model object priming but also to model the priming of specific feature dimensions. Since features are implicitly described in the hand network, this is very similar to object priming, though. The main difference between object priming and feature priming in SAAM is in the level of specification: While an object prime is expected to affect the grasp template activated in the grasp template network, a feature prime should only prime components of specific grasp posture in the hand network. This could be, for instance, a wide or small distance between the fingers and the thumb. Hence, feature prime might best be explored in a version of SAAM without top-down feed-back in order to dissociate priming-effects of complex affordances from the effects of micro-affordances.

Priming by showing the execution of an action cannot be directly modelled in SAAM since the model can only perceive affordances and has no means of recognising action gestures. Though, it would be possible to simulate experiments which showed static hand postures used during grasping as primes (e. g., Borghi et al., 2007). Such an experiment would, however, still require to prime certain grasp postures directly in the grasp template network by initialising them accordingly due to SAAM's restriction to affordance perception. Alternatively, a priming experiment in humans could explore how humans react when they are primed for a certain grasp posture but cannot execute this grasp in the subsequent action. It is an interesting question if the grasp postures or the reaction times are affected like SAAM predicts.

Effect of the Initial Hand Position

Finally, SAAM could also be used to investigate effects of initial hand positions on grasping. To achieve this, the finger maps could be initialised with specific finger positions to indicate an initial posture of the hand. Simulations could then test how different start positions affect the simulation duration and the generated grasp posture. Such simulations could easily be verified with an experiment in which participants are instructed to place there hands differently in each trial. For example, they could be asked to adopt either a power or precision grasp posture.

9.2 Conclusions

The aim of my thesis was to develop and explore a connectionist model which used an affordance-based approach to find and select opportunities for action in the environment as well as to generate the parameters required for executing the selected action. The result of this work was the Selective Attention for Action Model (SAAM). This model simulated a direct route from vision to action. It detected complex affordances and the associated action parameters directly from the information available in the visual input field. The model implemented this process through a hierarchy of affordances for increasingly complex actions. This was demonstrated in three studies.

In Study 1 I showed that SAAM's grasp postures for objects resembled the grasps humans produced when grasping these objects. The comparison was based on data collected in an experiment in which human grasp postures were recorded while participants grasped the same stimulus shapes as those which were used in the simulations.

In addition to the bottom-up path which was explored in Study 1, SAAM also had a top-down path to incorporate stored knowledge into the model. This extension was introduced and explored in Study 2. I showed that the top-down process in SAAM not only allowed control of object selection but also of grasp posture and grasp placement on an object. Furthermore, I demonstrated that the top-down path even allowed to make choices between grasping one object or a whole group of objects by preferring small or large grasps.

Study 2 demonstrated that SAAM exhibited behaviour which can be understood as selective attention for action. The model was able to select one of two objects in the visual input as well as the appropriate action category (defined by the grasp type) for using the selected object. Moreover, this bottom-up choices could be influenced through top-down feedback based on action intentions. In addition to the attentional behaviour, the first two studies revealed two notable emergent properties: First, it emerged that SAAM was able to create form-closure grasps on appropriate objects. Second, SAAM was also capable of positioning the fingers around the centre of mass of objects without explicitly identifying the centre of mass first.

In Study 3 simulations with hand-tools were presented to demonstrate more clearly how affordances and action intentions were handled in the model. It was shown that SAAM was able to identify affordances of objects as well as to guide visual selection towards objects suitable for an intended action. This showed how invariants can be combined to create affordances describing high-level actions. It also showed how basic action-parameters (grasp location) can be implicitly included in the affordance description. This demonstrated that the model fits well into Gibson's affordance framework. Additionally, Study 3 showed that grasp postures are a viable method to encode affordances for complex actions as Napier (1956) suggested.

SAAM combined excitatory and inhibitory processes. I argued that the specific implementation of these two processes in SAAM might show a general principle of information processing in attentional systems whereby the excitatory process finds salient locations and the inhibitory process performs the selection. Additionally, SAAM also showed a novel way of integrating top-down information in a model for selective attention by modifying the weight maps of the excitatory connections instead of the visual input.

In addition to the work presented here, the possibilities for further experimentation and extensions of the model described in the outlook underline SAAM's potential and versatility for researching the route from vision to action in a computational model.

Part IV

Appendices

A Additional Information about the Experiment in Study 1

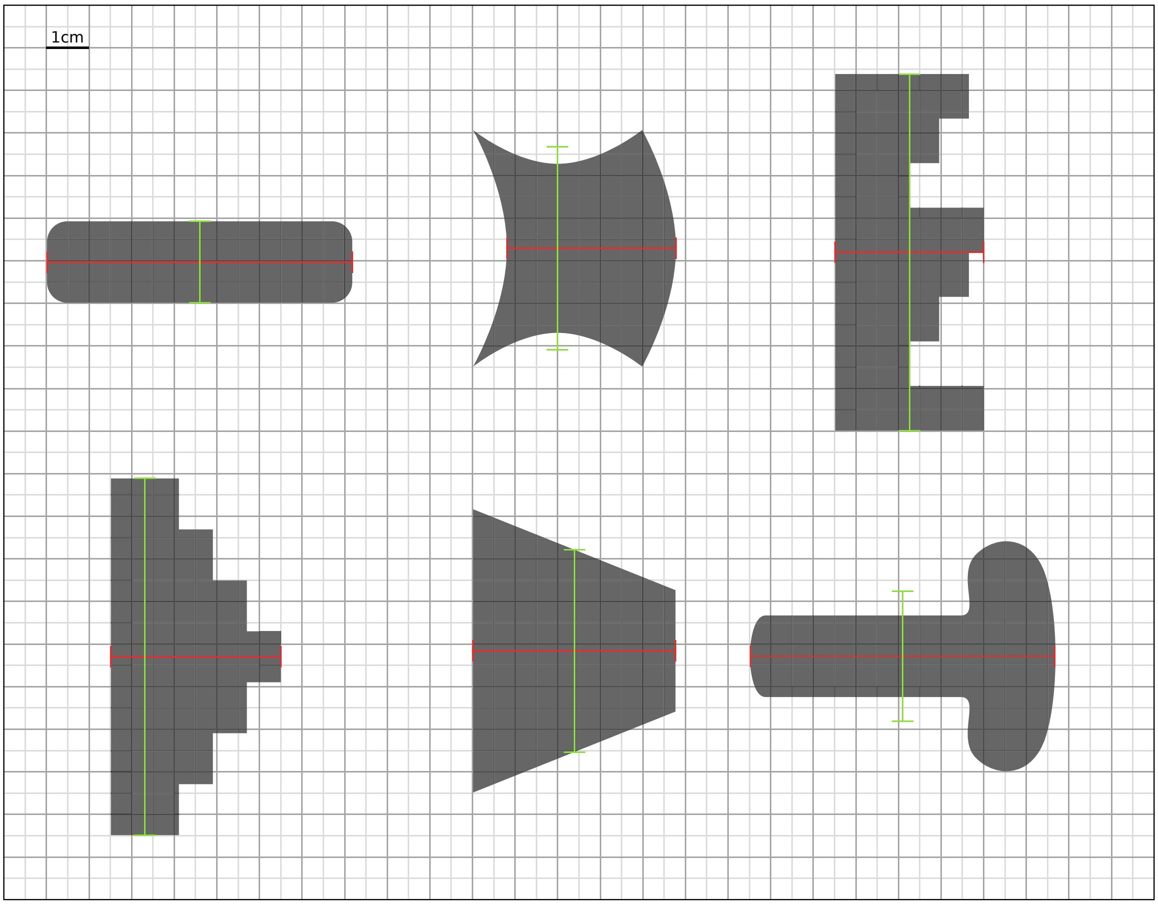
This appendix contains additional information about the materials which were used in Study 1 (Chapter 5). Additionally, it contains lists of all pairwise comparisons of both reaction times (RT 1 and RT 2) which were collected in the experiment.

A.1 Object Dimensions

The objects were cut from a plywood board with plastic veneer and painted in silk black. The weights of the objects are listed in the following table:

Object	Weight
Bar	97 g
Concave	170 g
Irregular	170 g
Pyramid	156 g
Trapezium	162 g
T-shape	140 g

The objects were 2.2 cm thick; their other dimensions are shown in the figure below. The coloured lines indicate how object width (red) and height (green) were measured to calculate the width-to-height ratio for the logistic regression analysis.



A.2 Instructions used in the Experiment

Participants received the following instructions:

> Hi!
>
> Thank you for taking part in this study.
>
> It is an experiment about grasping. You are asked to grasp a number of objects and in each trial a picture of your grasp is taken.
>
> During the experiment you are wearing special glasses which can be turned opaque by the computer. While you cannot see the experimenter will

place an object on the board in front of you. Then she will ask you to hold down the large orange button with the palm of your hand. Two seconds later the glasses will become transparent again and you should pick up the object quickly. However, as soon as you release the orange button the glasses will close again and you have do your grasp blindly.

After you have picked up the object the glasses become transparent again and arrow appears on the screen to tell you on which camera panel you should place the object. Place the object onto the camera panel and hold it on the glass pane. Try NOT TO CHANGE your grasp while doing this. The experimenter will take a picture of your hand posture and ask you to return the object to her.

You can withdraw from this experiment at any time. If you want a break during the experiment – just ask for one!

A.3 Reaction Times for Button Release

These reaction times were measured from the point when the glasses became transparent until the participants released the orange button. The following table lists the results of pairwise comparisons of all objects. Significant reaction time differences are highlighted with a grey background.

Comparison	t-Test Result
bar 0° ($M = 476.03$ ms, $SD = 137.57$)	
> bar 90°	$t(16) = 0.490$, $p = 0.63$
> bar 180°	$t(16) = 1.820$, $p = 0.09$
> bar 270°	$t(16) = 0.840$, $p = 0.41$
< concave 0°	$t(16) = -0.610$, $p = 0.55$
< concave 90°	$t(16) = -0.410$, $p = 0.68$

$<$ concave 180°	$t(16) = -0.310$, $p = 0.76$
$<$ concave 270°	$t(16) = -0.460$, $p = 0.65$
$<$ irregular 0°	$t(16) = -1.030$, $p = 0.32$
$>$ irregular 90°	$t(16) = 3.240$, $p = 0.005$
$>$ irregular 180°	$t(16) = 0.140$, $p = 0.89$
$>$ irregular 270°	$t(16) = 0.080$, $p = 0.93$
$<$ pyramid 0°	$t(16) = -0.930$, $p = 0.37$
$>$ pyramid 90°	$t(16) = 2.750$, $p = 0.01$
$>$ pyramid 180°	$t(16) = 0.410$, $p = 0.69$
$>$ pyramid 270°	$t(16) = 0.460$, $p = 0.65$
$>$ trapezium 0°	$t(16) = 0.700$, $p = 0.49$
$>$ trapezium 90°	$t(16) = 0.050$, $p = 0.96$
$<$ trapezium 180°	$t(16) = -0.620$, $p = 0.55$
$>$ trapezium 270°	$t(16) = 1.420$, $p = 0.17$
$>$ t-shape 0°	$t(16) = 0.420$, $p = 0.68$
$>$ t-shape 90°	$t(16) = 0.220$, $p = 0.83$
$>$ t-shape 180°	$t(16) = 0.060$, $p = 0.95$
$<$ t-shape 270°	$t(16) = -0.670$, $p = 0.51$
bar 90° ($M = 467.62$ ms, $SD = 143.97$)	
$>$ bar 180°	$t(16) = 0.830$, $p = 0.42$
$>$ bar 270°	$t(16) = 0.420$, $p = 0.68$
$<$ concave 0°	$t(16) = -0.800$, $p = 0.43$
$<$ concave 90°	$t(16) = -0.680$, $p = 0.50$
$<$ concave 180°	$t(16) = -0.900$, $p = 0.38$
$<$ concave 270°	$t(16) = -1.200$, $p = 0.25$
$<$ irregular 0°	$t(16) = -1.590$, $p = 0.13$
$>$ irregular 90°	$t(16) = 3.080$, $p = 0.007$
$<$ irregular 180°	$t(16) = -0.380$, $p = 0.71$
$<$ irregular 270°	$t(16) = -0.420$, $p = 0.68$
$<$ pyramid 0°	$t(16) = -1.540$, $p = 0.14$
$>$ pyramid 90°	$t(16) = 2.190$, $p = 0.04$
$<$ pyramid 180°	$t(16) = -0.170$, $p = 0.86$
$>$ pyramid 270°	$t(16) = 0.130$, $p = 0.90$
$>$ trapezium 0°	$t(16) = 0.190$, $p = 0.85$
$<$ trapezium 90°	$t(16) = -0.710$, $p = 0.49$
$<$ trapezium 180°	$t(16) = -1.300$, $p = 0.21$
$>$ trapezium 270°	$t(16) = 0.680$, $p = 0.50$
$<$ t-shape 0°	$t(16) = -0.010$, $p = 0.99$
$<$ t-shape 90°	$t(16) = -0.160$, $p = 0.87$
$<$ t-shape 180°	$t(16) = -0.280$, $p = 0.78$
$<$ t-shape 270°	$t(16) = -1.820$, $p = 0.09$
bar 180° ($M = 454.38$ ms, $SD = 127.83$)	
$<$ bar 270°	$t(16) = -0.130$, $p = 0.90$
$<$ concave 0°	$t(16) = -1.800$, $p = 0.09$

< concave 90°	$t(16) = -1.170$, $p = 0.26$
< concave 180°	$t(16) = -1.460$, $p = 0.16$
< concave 270°	$t(16) = -1.660$, $p = 0.12$
< irregular 0°	$t(16) = -2.420$, $p = 0.03$
> irregular 90°	$t(16) = 2.110$, $p = 0.05$
< irregular 180°	$t(16) = -0.950$, $p = 0.36$
< irregular 270°	$t(16) = -1.130$, $p = 0.27$
< pyramid 0°	$t(16) = -1.680$, $p = 0.11$
> pyramid 90°	$t(16) = 2.160$, $p = 0.05$
< pyramid 180°	$t(16) = -0.900$, $p = 0.38$
< pyramid 270°	$t(16) = -0.420$, $p = 0.68$
< trapezium 0°	$t(16) = -0.370$, $p = 0.71$
< trapezium 90°	$t(16) = -1.190$, $p = 0.25$
< trapezium 180°	$t(16) = -1.770$, $p = 0.10$
< trapezium 270°	$t(16) = -0.120$, $p = 0.90$
< t-shape 0°	$t(16) = -0.640$, $p = 0.53$
< t-shape 90°	$t(16) = -0.880$, $p = 0.39$
< t-shape 180°	$t(16) = -0.980$, $p = 0.34$
< t-shape 270°	$t(16) = -1.790$, $p = 0.09$
bar 270° ($M = 456.91$ ms, $SD = 116.82$)	
< concave 0°	$t(16) = -1.340$, $p = 0.20$
< concave 90°	$t(16) = -0.700$, $p = 0.50$
< concave 180°	$t(16) = -1.100$, $p = 0.29$
< concave 270°	$t(16) = -1.070$, $p = 0.30$
< irregular 0°	$t(16) = -1.370$, $p = 0.19$
> irregular 90°	$t(16) = 1.720$, $p = 0.11$
< irregular 180°	$t(16) = -0.620$, $p = 0.54$
< irregular 270°	$t(16) = -0.620$, $p = 0.54$
< pyramid 0°	$t(16) = -1.200$, $p = 0.25$
> pyramid 90°	$t(16) = 1.500$, $p = 0.15$
< pyramid 180°	$t(16) = -0.440$, $p = 0.66$
< pyramid 270°	$t(16) = -0.230$, $p = 0.82$
< trapezium 0°	$t(16) = -0.180$, $p = 0.86$
< trapezium 90°	$t(16) = -0.660$, $p = 0.52$
< trapezium 180°	$t(16) = -1.210$, $p = 0.24$
> trapezium 270°	$t(16) = 0.020$, $p = 0.98$
< t-shape 0°	$t(16) = -0.340$, $p = 0.74$
< t-shape 90°	$t(16) = -0.550$, $p = 0.59$
< t-shape 180°	$t(16) = -0.790$, $p = 0.44$
< t-shape 270°	$t(16) = -1.160$, $p = 0.26$
concave 0° ($M = 485.56$ ms, $SD = 144.88$)	
< concave 90°	$t(16) = -0.010$, $p = 0.99$
> concave 180°	$t(16) = 0.110$, $p = 0.91$
< concave 270°	$t(16) = -0.020$, $p = 0.99$

< irregular 0°	$t(16) = -0.540$, $p = 0.59$
> irregular 90°	$t(16) = 2.770$, $p = 0.01$
> irregular 180°	$t(16) = 0.480$, $p = 0.64$
> irregular 270°	$t(16) = 0.450$, $p = 0.66$
< pyramid 0°	$t(16) = -0.530$, $p = 0.60$
> pyramid 90°	$t(16) = 2.870$, $p = 0.01$
> pyramid 180°	$t(16) = 0.830$, $p = 0.42$
> pyramid 270°	$t(16) = 0.710$, $p = 0.49$
> trapezium 0°	$t(16) = 0.820$, $p = 0.42$
> trapezium 90°	$t(16) = 0.400$, $p = 0.69$
< trapezium 180°	$t(16) = -0.200$, $p = 0.85$
> trapezium 270°	$t(16) = 1.440$, $p = 0.17$
> t-shape 0°	$t(16) = 0.630$, $p = 0.54$
> t-shape 90°	$t(16) = 0.500$, $p = 0.62$
> t-shape 180°	$t(16) = 0.520$, $p = 0.61$
< t-shape 270°	$t(16) = -0.130$, $p = 0.90$
concave 90° ($M = 485.94$ ms, $SD = 179.98$)	
> concave 180°	$t(16) = 0.080$, $p = 0.94$
< concave 270°	$t(16) = -0.000$, $p = 1.00$
< irregular 0°	$t(16) = -0.420$, $p = 0.68$
> irregular 90°	$t(16) = 2.100$, $p = 0.05$
> irregular 180°	$t(16) = 0.420$, $p = 0.68$
> irregular 270°	$t(16) = 0.420$, $p = 0.68$
< pyramid 0°	$t(16) = -0.520$, $p = 0.61$
> pyramid 90°	$t(16) = 1.730$, $p = 0.10$
> pyramid 180°	$t(16) = 0.810$, $p = 0.43$
> pyramid 270°	$t(16) = 0.700$, $p = 0.49$
> trapezium 0°	$t(16) = 1.290$, $p = 0.21$
> trapezium 90°	$t(16) = 0.400$, $p = 0.69$
< trapezium 180°	$t(16) = -0.140$, $p = 0.89$
> trapezium 270°	$t(16) = 1.200$, $p = 0.25$
> t-shape 0°	$t(16) = 0.880$, $p = 0.39$
> t-shape 90°	$t(16) = 0.480$, $p = 0.64$
> t-shape 180°	$t(16) = 0.290$, $p = 0.78$
< t-shape 270°	$t(16) = -0.120$, $p = 0.91$
concave 180° ($M = 483.00$ ms, $SD = 171.13$)	
< concave 270°	$t(16) = -0.170$, $p = 0.86$
< irregular 0°	$t(16) = -0.850$, $p = 0.41$
> irregular 90°	$t(16) = 2.500$, $p = 0.02$
> irregular 180°	$t(16) = 0.430$, $p = 0.68$
> irregular 270°	$t(16) = 0.370$, $p = 0.72$
< pyramid 0°	$t(16) = -1.010$, $p = 0.33$
> pyramid 90°	$t(16) = 2.600$, $p = 0.02$
> pyramid 180°	$t(16) = 0.600$, $p = 0.56$

> pyramid 270°	$t(16) = 0.610$, $p = 0.55$
> trapezium 0°	$t(16) = 0.570$, $p = 0.57$
> trapezium 90°	$t(16) = 0.410$, $p = 0.69$
< trapezium 180°	$t(16) = -0.580$, $p = 0.57$
> trapezium 270°	$t(16) = 1.030$, $p = 0.32$
> t-shape 0°	$t(16) = 0.490$, $p = 0.63$
> t-shape 90°	$t(16) = 0.320$, $p = 0.75$
> t-shape 180°	$t(16) = 0.310$, $p = 0.76$
< t-shape 270°	$t(16) = -0.300$, $p = 0.76$
concave 270° ($M = 485.97$ ms, $SD = 166.89$)	
< irregular 0°	$t(16) = -0.790$, $p = 0.44$
> irregular 90°	$t(16) = 2.870$, $p = 0.01$
> irregular 180°	$t(16) = 1.100$, $p = 0.29$
> irregular 270°	$t(16) = 0.660$, $p = 0.52$
< pyramid 0°	$t(16) = -0.850$, $p = 0.41$
> pyramid 90°	$t(16) = 2.620$, $p = 0.02$
> pyramid 180°	$t(16) = 0.930$, $p = 0.37$
> pyramid 270°	$t(16) = 1.000$, $p = 0.33$
> trapezium 0°	$t(16) = 0.860$, $p = 0.40$
> trapezium 90°	$t(16) = 1.060$, $p = 0.31$
< trapezium 180°	$t(16) = -0.340$, $p = 0.74$
> trapezium 270°	$t(16) = 1.370$, $p = 0.19$
> t-shape 0°	$t(16) = 0.980$, $p = 0.34$
> t-shape 90°	$t(16) = 0.440$, $p = 0.66$
> t-shape 180°	$t(16) = 0.440$, $p = 0.66$
< t-shape 270°	$t(16) = -0.230$, $p = 0.82$
irregular 0° ($M = 499.24$ ms, $SD = 186.31$)	
> irregular 90°	$t(16) = 2.860$, $p = 0.01$
> irregular 180°	$t(16) = 1.230$, $p = 0.24$
> irregular 270°	$t(16) = 1.230$, $p = 0.24$
< pyramid 0°	$t(16) = -0.250$, $p = 0.81$
> pyramid 90°	$t(16) = 2.880$, $p = 0.01$
> pyramid 180°	$t(16) = 1.500$, $p = 0.15$
> pyramid 270°	$t(16) = 1.420$, $p = 0.18$
> trapezium 0°	$t(16) = 1.150$, $p = 0.27$
> trapezium 90°	$t(16) = 1.220$, $p = 0.24$
> trapezium 180°	$t(16) = 0.590$, $p = 0.56$
> trapezium 270°	$t(16) = 1.960$, $p = 0.07$
> t-shape 0°	$t(16) = 1.220$, $p = 0.24$
> t-shape 90°	$t(16) = 0.850$, $p = 0.41$
> t-shape 180°	$t(16) = 1.070$, $p = 0.30$
> t-shape 270°	$t(16) = 0.470$, $p = 0.65$
irregular 90° ($M = 422.85$ ms, $SD = 103.88$)	

< irregular 180°	$t(16) = -2.980$, $p = 0.009$
< irregular 270°	$t(16) = -2.670$, $p = 0.02$
< pyramid 0°	$t(16) = -2.580$, $p = 0.02$
< pyramid 90°	$t(16) = -0.520$, $p = 0.61$
< pyramid 180°	$t(16) = -2.600$, $p = 0.02$
< pyramid 270°	$t(16) = -1.790$, $p = 0.09$
< trapezium 0°	$t(16) = -1.910$, $p = 0.07$
< trapezium 90°	$t(16) = -2.960$, $p = 0.009$
< trapezium 180°	$t(16) = -2.770$, $p = 0.01$
< trapezium 270°	$t(16) = -2.040$, $p = 0.06$
< t-shape 0°	$t(16) = -2.010$, $p = 0.06$
< t-shape 90°	$t(16) = -2.720$, $p = 0.02$
< t-shape 180°	$t(16) = -2.190$, $p = 0.04$
< t-shape 270°	$t(16) = -3.390$, $p = 0.004$
irregular 180° ($M = 473.24$ ms, $SD = 147.59$)	
< irregular 270°	$t(16) = -0.060$, $p = 0.95$
< pyramid 0°	$t(16) = -1.190$, $p = 0.25$
> pyramid 90°	$t(16) = 1.950$, $p = 0.07$
> pyramid 180°	$t(16) = 0.230$, $p = 0.82$
> pyramid 270°	$t(16) = 0.490$, $p = 0.63$
> trapezium 0°	$t(16) = 0.480$, $p = 0.64$
< trapezium 90°	$t(16) = -0.180$, $p = 0.86$
< trapezium 180°	$t(16) = -0.930$, $p = 0.36$
> trapezium 270°	$t(16) = 0.930$, $p = 0.37$
> t-shape 0°	$t(16) = 0.340$, $p = 0.74$
> t-shape 90°	$t(16) = 0.040$, $p = 0.97$
< t-shape 180°	$t(16) = -0.060$, $p = 0.95$
< t-shape 270°	$t(16) = -1.140$, $p = 0.27$
irregular 270° ($M = 474.35$ ms, $SD = 157.22$)	
< pyramid 0°	$t(16) = -0.990$, $p = 0.34$
> pyramid 90°	$t(16) = 2.780$, $p = 0.01$
> pyramid 180°	$t(16) = 0.360$, $p = 0.72$
> pyramid 270°	$t(16) = 0.330$, $p = 0.75$
> trapezium 0°	$t(16) = 0.420$, $p = 0.68$
< trapezium 90°	$t(16) = -0.050$, $p = 0.96$
< trapezium 180°	$t(16) = -0.930$, $p = 0.37$
> trapezium 270°	$t(16) = 0.990$, $p = 0.34$
> t-shape 0°	$t(16) = 0.300$, $p = 0.77$
> t-shape 90°	$t(16) = 0.080$, $p = 0.94$
< t-shape 180°	$t(16) = -0.010$, $p = 0.99$
< t-shape 270°	$t(16) = -0.680$, $p = 0.50$
pyramid 0° ($M = 504.62$ ms, $SD = 209.78$)	
> pyramid 90°	$t(16) = 2.220$, $p = 0.04$

> pyramid 180°	$t(16) = 1.270$, $p = 0.22$
> pyramid 270°	$t(16) = 1.440$, $p = 0.17$
> trapezium 0°	$t(16) = 1.180$, $p = 0.26$
> trapezium 90°	$t(16) = 1.330$, $p = 0.20$
> trapezium 180°	$t(16) = 0.610$, $p = 0.55$
> trapezium 270°	$t(16) = 1.530$, $p = 0.14$
> t-shape 0°	$t(16) = 1.150$, $p = 0.27$
> t-shape 90°	$t(16) = 0.860$, $p = 0.40$
> t-shape 180°	$t(16) = 0.850$, $p = 0.41$
> t-shape 270°	$t(16) = 0.690$, $p = 0.50$
pyramid 90° ($M = 430.38$ ms, $SD = 116.70$)	
< pyramid 180°	$t(16) = -2.070$, $p = 0.06$
< pyramid 270°	$t(16) = -1.150$, $p = 0.27$
< trapezium 0°	$t(16) = -1.170$, $p = 0.26$
< trapezium 90°	$t(16) = -2.310$, $p = 0.03$
< trapezium 180°	$t(16) = -2.740$, $p = 0.01$
< trapezium 270°	$t(16) = -1.480$, $p = 0.16$
< t-shape 0°	$t(16) = -1.460$, $p = 0.16$
< t-shape 90°	$t(16) = -1.860$, $p = 0.08$
< t-shape 180°	$t(16) = -2.130$, $p = 0.05$
< t-shape 270°	$t(16) = -2.610$, $p = 0.02$
pyramid 180° ($M = 469.59$ ms, $SD = 155.73$)	
> pyramid 270°	$t(16) = 0.210$, $p = 0.84$
> trapezium 0°	$t(16) = 0.330$, $p = 0.75$
< trapezium 90°	$t(16) = -0.400$, $p = 0.69$
< trapezium 180°	$t(16) = -0.990$, $p = 0.34$
> trapezium 270°	$t(16) = 0.860$, $p = 0.40$
> t-shape 0°	$t(16) = 0.090$, $p = 0.93$
< t-shape 90°	$t(16) = -0.090$, $p = 0.93$
< t-shape 180°	$t(16) = -0.200$, $p = 0.84$
< t-shape 270°	$t(16) = -1.260$, $p = 0.22$
pyramid 270° ($M = 464.47$ ms, $SD = 147.34$)	
> trapezium 0°	$t(16) = 0.070$, $p = 0.95$
< trapezium 90°	$t(16) = -0.490$, $p = 0.63$
< trapezium 180°	$t(16) = -0.930$, $p = 0.37$
> trapezium 270°	$t(16) = 0.390$, $p = 0.70$
< t-shape 0°	$t(16) = -0.170$, $p = 0.86$
< t-shape 90°	$t(16) = -0.300$, $p = 0.77$
< t-shape 180°	$t(16) = -0.380$, $p = 0.71$
< t-shape 270°	$t(16) = -1.230$, $p = 0.24$
trapezium 0° ($M = 463.06$ ms, $SD = 140.69$)	
< trapezium 90°	$t(16) = -0.550$, $p = 0.59$

< trapezium 180°	$t(16) = -0.810$, $p = 0.43$
> trapezium 270°	$t(16) = 0.360$, $p = 0.72$
< t-shape 0°	$t(16) = -0.360$, $p = 0.72$
< t-shape 90°	$t(16) = -0.460$, $p = 0.65$
< t-shape 180°	$t(16) = -0.360$, $p = 0.72$
< t-shape 270°	$t(16) = -1.240$, $p = 0.23$
trapezium 90° ($M = 475.15$ ms, $SD = 152.58$)	
< trapezium 180°	$t(16) = -0.870$, $p = 0.40$
> trapezium 270°	$t(16) = 1.030$, $p = 0.32$
> t-shape 0°	$t(16) = 0.440$, $p = 0.67$
> t-shape 90°	$t(16) = 0.110$, $p = 0.91$
> t-shape 180°	$t(16) = 0.020$, $p = 0.98$
< t-shape 270°	$t(16) = -1.310$, $p = 0.21$
trapezium 180° ($M = 490.94$ ms, $SD = 177.76$)	
> trapezium 270°	$t(16) = 1.470$, $p = 0.16$
> t-shape 0°	$t(16) = 0.830$, $p = 0.42$
> t-shape 90°	$t(16) = 0.550$, $p = 0.59$
> t-shape 180°	$t(16) = 0.610$, $p = 0.55$
> t-shape 270°	$t(16) = 0.100$, $p = 0.92$
trapezium 270° ($M = 456.38$ ms, $SD = 127.06$)	
< t-shape 0°	$t(16) = -0.610$, $p = 0.55$
< t-shape 90°	$t(16) = -0.840$, $p = 0.41$
< t-shape 180°	$t(16) = -0.860$, $p = 0.40$
< t-shape 270°	$t(16) = -1.640$, $p = 0.12$
t-shape 0° ($M = 467.85$ ms, $SD = 151.00$)	
< t-shape 90°	$t(16) = -0.170$, $p = 0.87$
< t-shape 180°	$t(16) = -0.230$, $p = 0.82$
< t-shape 270°	$t(16) = -1.200$, $p = 0.25$
t-shape 90° ($M = 472.06$ ms, $SD = 110.98$)	
< t-shape 180°	$t(16) = -0.100$, $p = 0.92$
< t-shape 270°	$t(16) = -0.610$, $p = 0.55$
t-shape 180° ($M = 474.65$ ms, $SD = 140.97$)	
< t-shape 270°	$t(16) = -0.500$, $p = 0.62$

A.4 Reaction Times for Object Lift

These reaction times were measured from the release of the orange button until the micro-switch in the object mount was triggered which happened when the object was lifted. Significant reaction time differences are highlighted with a grey background.

Comparison	t-Test Result
bar 0° ($M = 710.15$ ms, $SD = 187.44$)	
> bar 90°	$t(16) = 1.410, p = 0.18$
> bar 180°	$t(16) = 2.780, p = 0.01$
> bar 270°	$t(16) = 4.490, p < 0.001$
> concave 0°	$t(16) = 3.400, p = 0.004$
> concave 90°	$t(16) = 3.400, p = 0.004$
> concave 180°	$t(16) = 3.440, p = 0.003$
> concave 270°	$t(16) = 3.930, p = 0.001$
> irregular 0°	$t(16) = 1.290, p = 0.22$
> irregular 90°	$t(16) = 2.300, p = 0.04$
> irregular 180°	$t(16) = 1.620, p = 0.12$
> irregular 270°	$t(16) = 0.830, p = 0.42$
> pyramid 0°	$t(16) = 2.060, p = 0.06$
> pyramid 90°	$t(16) = 2.330, p = 0.03$
> pyramid 180°	$t(16) = 1.550, p = 0.14$
> pyramid 270°	$t(16) = 1.120, p = 0.28$
> trapezium 0°	$t(16) = 0.880, p = 0.39$
> trapezium 90°	$t(16) = 3.660, p = 0.002$
> trapezium 180°	$t(16) = 2.960, p = 0.009$
> trapezium 270°	$t(16) = 4.580, p < 0.001$
> t-shape 0°	$t(16) = 2.110, p = 0.05$
> t-shape 90°	$t(16) = 3.270, p = 0.005$
> t-shape 180°	$t(16) = 1.990, p = 0.06$
> t-shape 270°	$t(16) = 1.920, p = 0.07$
bar 90° ($M = 681.00$ ms, $SD = 179.68$)	
> bar 180°	$t(16) = 3.310, p = 0.004$
> bar 270°	$t(16) = 4.880, p < 0.001$
> concave 0°	$t(16) = 4.640, p < 0.001$
> concave 90°	$t(16) = 3.090, p = 0.007$
> concave 180°	$t(16) = 4.500, p < 0.001$
> concave 270°	$t(16) = 3.480, p = 0.003$
< irregular 0°	$t(16) = -0.010, p = 0.99$
> irregular 90°	$t(16) = 2.550, p = 0.02$

$>$ irregular 180°	$t(16) = 0.900$, $p = 0.38$
$<$ irregular 270°	$t(16) = -0.360$, $p = 0.72$
$>$ pyramid 0°	$t(16) = 1.540$, $p = 0.14$
$>$ pyramid 90°	$t(16) = 1.930$, $p = 0.07$
$>$ pyramid 180°	$t(16) = 0.820$, $p = 0.42$
$<$ pyramid 270°	$t(16) = -0.040$, $p = 0.97$
$<$ trapezium 0°	$t(16) = -0.170$, $p = 0.87$
$>$ trapezium 90°	$t(16) = 3.240$, $p = 0.005$
$>$ trapezium 180°	$t(16) = 3.160$, $p = 0.006$
$>$ trapezium 270°	$t(16) = 5.570$, $p < 0.001$
$>$ t-shape 0°	$t(16) = 1.770$, $p = 0.10$
$>$ t-shape 90°	$t(16) = 3.760$, $p = 0.002$
$>$ t-shape 180°	$t(16) = 1.480$, $p = 0.16$
$>$ t-shape 270°	$t(16) = 0.860$, $p = 0.40$
bar 180° ($M = 629.47$ ms, $SD = 185.81$)	
$>$ bar 270°	$t(16) = 1.200$, $p = 0.25$
$>$ concave 0°	$t(16) = 0.910$, $p = 0.38$
$>$ concave 90°	$t(16) = 1.220$, $p = 0.24$
$>$ concave 180°	$t(16) = 1.110$, $p = 0.28$
$>$ concave 270°	$t(16) = 0.070$, $p = 0.94$
$<$ irregular 0°	$t(16) = -2.600$, $p = 0.02$
$<$ irregular 90°	$t(16) = -0.350$, $p = 0.73$
$<$ irregular 180°	$t(16) = -1.620$, $p = 0.13$
$<$ irregular 270°	$t(16) = -2.620$, $p = 0.02$
$<$ pyramid 0°	$t(16) = -1.610$, $p = 0.13$
$<$ pyramid 90°	$t(16) = -0.610$, $p = 0.55$
$<$ pyramid 180°	$t(16) = -1.550$, $p = 0.14$
$<$ pyramid 270°	$t(16) = -2.300$, $p = 0.04$
$<$ trapezium 0°	$t(16) = -2.610$, $p = 0.02$
$>$ trapezium 90°	$t(16) = 0.320$, $p = 0.76$
$>$ trapezium 180°	$t(16) = 1.460$, $p = 0.16$
$>$ trapezium 270°	$t(16) = 3.140$, $p = 0.006$
$<$ t-shape 0°	$t(16) = -0.390$, $p = 0.70$
$>$ t-shape 90°	$t(16) = 1.080$, $p = 0.30$
$<$ t-shape 180°	$t(16) = -1.680$, $p = 0.11$
$<$ t-shape 270°	$t(16) = -1.520$, $p = 0.15$
bar 270° ($M = 604.91$ ms, $SD = 174.53$)	
$<$ concave 0°	$t(16) = -0.410$, $p = 0.69$
$>$ concave 90°	$t(16) = 0.470$, $p = 0.64$
$<$ concave 180°	$t(16) = -0.490$, $p = 0.63$
$<$ concave 270°	$t(16) = -1.390$, $p = 0.18$
$<$ irregular 0°	$t(16) = -3.420$, $p = 0.004$
$<$ irregular 90°	$t(16) = -1.560$, $p = 0.14$
$<$ irregular 180°	$t(16) = -2.450$, $p = 0.03$

< irregular 270°	$t(16) = -3.950$, $p = 0.001$
< pyramid 0°	$t(16) = -3.000$, $p = 0.009$
< pyramid 90°	$t(16) = -1.510$, $p = 0.15$
< pyramid 180°	$t(16) = -3.440$, $p = 0.003$
< pyramid 270°	$t(16) = -4.190$, $p < 0.001$
< trapezium 0°	$t(16) = -2.830$, $p = 0.01$
< trapezium 90°	$t(16) = -0.640$, $p = 0.53$
> trapezium 180°	$t(16) = 0.680$, $p = 0.51$
> trapezium 270°	$t(16) = 2.440$, $p = 0.03$
< t-shape 0°	$t(16) = -1.210$, $p = 0.25$
< t-shape 90°	$t(16) = -0.230$, $p = 0.82$
< t-shape 180°	$t(16) = -2.970$, $p = 0.009$
< t-shape 270°	$t(16) = -2.730$, $p = 0.01$
concave 0° ($M = 614.38$ ms, $SD = 192.91$)	
> concave 90°	$t(16) = 0.850$, $p = 0.41$
< concave 180°	$t(16) = -0.030$, $p = 0.98$
< concave 270°	$t(16) = -0.630$, $p = 0.54$
< irregular 0°	$t(16) = -2.660$, $p = 0.02$
< irregular 90°	$t(16) = -1.080$, $p = 0.30$
< irregular 180°	$t(16) = -2.050$, $p = 0.06$
< irregular 270°	$t(16) = -3.230$, $p = 0.005$
< pyramid 0°	$t(16) = -2.840$, $p = 0.01$
< pyramid 90°	$t(16) = -1.310$, $p = 0.21$
< pyramid 180°	$t(16) = -2.870$, $p = 0.01$
< pyramid 270°	$t(16) = -3.050$, $p = 0.008$
< trapezium 0°	$t(16) = -2.710$, $p = 0.02$
< trapezium 90°	$t(16) = -0.370$, $p = 0.72$
> trapezium 180°	$t(16) = 1.170$, $p = 0.26$
> trapezium 270°	$t(16) = 3.010$, $p = 0.008$
< t-shape 0°	$t(16) = -0.880$, $p = 0.39$
> t-shape 90°	$t(16) = 0.250$, $p = 0.80$
< t-shape 180°	$t(16) = -2.060$, $p = 0.06$
< t-shape 270°	$t(16) = -2.340$, $p = 0.03$
concave 90° ($M = 588.97$ ms, $SD = 159.73$)	
< concave 180°	$t(16) = -0.830$, $p = 0.42$
< concave 270°	$t(16) = -1.430$, $p = 0.17$
< irregular 0°	$t(16) = -2.540$, $p = 0.02$
< irregular 90°	$t(16) = -1.380$, $p = 0.19$
< irregular 180°	$t(16) = -2.160$, $p = 0.05$
< irregular 270°	$t(16) = -2.660$, $p = 0.02$
< pyramid 0°	$t(16) = -2.080$, $p = 0.05$
< pyramid 90°	$t(16) = -2.110$, $p = 0.05$
< pyramid 180°	$t(16) = -1.930$, $p = 0.07$
< pyramid 270°	$t(16) = -2.800$, $p = 0.01$

$<$ trapezium 0°	$t(16) = -3.970$, $p = 0.001$
$<$ trapezium 90°	$t(16) = -1.070$, $p = 0.30$
$>$ trapezium 180°	$t(16) = 0.150$, $p = 0.89$
$>$ trapezium 270°	$t(16) = 2.000$, $p = 0.06$
$<$ t-shape 0°	$t(16) = -1.360$, $p = 0.19$
$<$ t-shape 90°	$t(16) = -0.700$, $p = 0.50$
$<$ t-shape 180°	$t(16) = -2.090$, $p = 0.05$
$<$ t-shape 270°	$t(16) = -2.860$, $p = 0.01$
concave 180° ($M = 614.76$ ms, $SD = 171.57$)	
$<$ concave 270°	$t(16) = -0.640$, $p = 0.53$
$<$ irregular 0°	$t(16) = -3.360$, $p = 0.004$
$<$ irregular 90°	$t(16) = -1.060$, $p = 0.31$
$<$ irregular 180°	$t(16) = -2.580$, $p = 0.02$
$<$ irregular 270°	$t(16) = -3.290$, $p = 0.005$
$<$ pyramid 0°	$t(16) = -3.210$, $p = 0.005$
$<$ pyramid 90°	$t(16) = -1.370$, $p = 0.19$
$<$ pyramid 180°	$t(16) = -2.470$, $p = 0.03$
$<$ pyramid 270°	$t(16) = -3.240$, $p = 0.005$
$<$ trapezium 0°	$t(16) = -3.250$, $p = 0.005$
$<$ trapezium 90°	$t(16) = -0.360$, $p = 0.73$
$>$ trapezium 180°	$t(16) = 1.120$, $p = 0.28$
$>$ trapezium 270°	$t(16) = 3.080$, $p = 0.007$
$<$ t-shape 0°	$t(16) = -0.940$, $p = 0.36$
$>$ t-shape 90°	$t(16) = 0.270$, $p = 0.79$
$<$ t-shape 180°	$t(16) = -2.400$, $p = 0.03$
$<$ t-shape 270°	$t(16) = -2.270$, $p = 0.04$
concave 270° ($M = 627.97$ ms, $SD = 179.00$)	
$<$ irregular 0°	$t(16) = -2.290$, $p = 0.04$
$<$ irregular 90°	$t(16) = -0.500$, $p = 0.62$
$<$ irregular 180°	$t(16) = -1.570$, $p = 0.13$
$<$ irregular 270°	$t(16) = -2.820$, $p = 0.01$
$<$ pyramid 0°	$t(16) = -1.770$, $p = 0.10$
$<$ pyramid 90°	$t(16) = -0.810$, $p = 0.43$
$<$ pyramid 180°	$t(16) = -1.510$, $p = 0.15$
$<$ pyramid 270°	$t(16) = -3.540$, $p = 0.003$
$<$ trapezium 0°	$t(16) = -3.060$, $p = 0.008$
$>$ trapezium 90°	$t(16) = 0.300$, $p = 0.77$
$>$ trapezium 180°	$t(16) = 1.260$, $p = 0.23$
$>$ trapezium 270°	$t(16) = 3.190$, $p = 0.006$
$<$ t-shape 0°	$t(16) = -0.380$, $p = 0.71$
$>$ t-shape 90°	$t(16) = 0.820$, $p = 0.43$
$<$ t-shape 180°	$t(16) = -1.740$, $p = 0.10$
$<$ t-shape 270°	$t(16) = -3.190$, $p = 0.006$

irregular 0° ($M = 681.24$ ms, $SD = 185.13$)	
$>$ irregular 90°	$t(16) = 1.450$, $p = 0.17$
$>$ irregular 180°	$t(16) = 1.040$, $p = 0.32$
$<$ irregular 270°	$t(16) = -0.260$, $p = 0.80$
$>$ pyramid 0°	$t(16) = 0.880$, $p = 0.39$
$>$ pyramid 90°	$t(16) = 1.560$, $p = 0.14$
$>$ pyramid 180°	$t(16) = 0.540$, $p = 0.60$
$<$ pyramid 270°	$t(16) = -0.020$, $p = 0.99$
$<$ trapezium 0°	$t(16) = -0.160$, $p = 0.88$
$>$ trapezium 90°	$t(16) = 2.530$, $p = 0.02$
$>$ trapezium 180°	$t(16) = 2.300$, $p = 0.04$
$>$ trapezium 270°	$t(16) = 3.960$, $p = 0.001$
$>$ t-shape 0°	$t(16) = 1.710$, $p = 0.11$
$>$ t-shape 90°	$t(16) = 2.950$, $p = 0.009$
$>$ t-shape 180°	$t(16) = 1.170$, $p = 0.26$
$>$ t-shape 270°	$t(16) = 0.570$, $p = 0.57$

irregular 90° ($M = 637.88$ ms, $SD = 203.88$)	
$<$ irregular 180°	$t(16) = -1.070$, $p = 0.30$
$<$ irregular 270°	$t(16) = -2.730$, $p = 0.01$
$<$ pyramid 0°	$t(16) = -1.480$, $p = 0.16$
$<$ pyramid 90°	$t(16) = -0.270$, $p = 0.79$
$<$ pyramid 180°	$t(16) = -1.200$, $p = 0.25$
$<$ pyramid 270°	$t(16) = -2.600$, $p = 0.02$
$<$ trapezium 0°	$t(16) = -1.740$, $p = 0.10$
$>$ trapezium 90°	$t(16) = 0.680$, $p = 0.50$
$>$ trapezium 180°	$t(16) = 1.860$, $p = 0.08$
$>$ trapezium 270°	$t(16) = 3.260$, $p = 0.005$
$<$ t-shape 0°	$t(16) = -0.040$, $p = 0.97$
$>$ t-shape 90°	$t(16) = 1.430$, $p = 0.17$
$<$ t-shape 180°	$t(16) = -0.970$, $p = 0.35$
$<$ t-shape 270°	$t(16) = -1.230$, $p = 0.24$

irregular 180° ($M = 663.21$ ms, $SD = 193.16$)	
$<$ irregular 270°	$t(16) = -0.900$, $p = 0.38$
$>$ pyramid 0°	$t(16) = 0.060$, $p = 0.95$
$>$ pyramid 90°	$t(16) = 1.030$, $p = 0.32$
$<$ pyramid 180°	$t(16) = -0.070$, $p = 0.94$
$<$ pyramid 270°	$t(16) = -0.910$, $p = 0.37$
$<$ trapezium 0°	$t(16) = -0.920$, $p = 0.37$
$>$ trapezium 90°	$t(16) = 2.160$, $p = 0.05$
$>$ trapezium 180°	$t(16) = 1.940$, $p = 0.07$
$>$ trapezium 270°	$t(16) = 4.210$, $p < 0.001$
$>$ t-shape 0°	$t(16) = 1.340$, $p = 0.20$
$>$ t-shape 90°	$t(16) = 2.020$, $p = 0.06$

$>$ t-shape 180°	$t(16) = 0.190$, $p = 0.85$
$<$ t-shape 270°	$t(16) = -0.130$, $p = 0.90$
irregular 270° ($M = 688.26$ ms, $SD = 196.48$)	
$>$ pyramid 0°	$t(16) = 1.370$, $p = 0.19$
$>$ pyramid 90°	$t(16) = 1.450$, $p = 0.17$
$>$ pyramid 180°	$t(16) = 1.000$, $p = 0.33$
$>$ pyramid 270°	$t(16) = 0.330$, $p = 0.74$
$>$ trapezium 0°	$t(16) = 0.130$, $p = 0.90$
$>$ trapezium 90°	$t(16) = 2.740$, $p = 0.01$
$>$ trapezium 180°	$t(16) = 3.610$, $p = 0.002$
$>$ trapezium 270°	$t(16) = 4.260$, $p < 0.001$
$>$ t-shape 0°	$t(16) = 1.470$, $p = 0.16$
$>$ t-shape 90°	$t(16) = 4.410$, $p < 0.001$
$>$ t-shape 180°	$t(16) = 1.240$, $p = 0.23$
$>$ t-shape 270°	$t(16) = 0.830$, $p = 0.42$
pyramid 0° ($M = 662.06$ ms, $SD = 177.59$)	
$>$ pyramid 90°	$t(16) = 0.830$, $p = 0.42$
$<$ pyramid 180°	$t(16) = -0.180$, $p = 0.86$
$<$ pyramid 270°	$t(16) = -0.980$, $p = 0.34$
$<$ trapezium 0°	$t(16) = -0.860$, $p = 0.40$
$>$ trapezium 90°	$t(16) = 1.810$, $p = 0.09$
$>$ trapezium 180°	$t(16) = 2.560$, $p = 0.02$
$>$ trapezium 270°	$t(16) = 4.260$, $p < 0.001$
$>$ t-shape 0°	$t(16) = 0.800$, $p = 0.43$
$>$ t-shape 90°	$t(16) = 2.810$, $p = 0.01$
$>$ t-shape 180°	$t(16) = 0.140$, $p = 0.89$
$<$ t-shape 270°	$t(16) = -0.180$, $p = 0.86$
pyramid 90° ($M = 644.26$ ms, $SD = 186.92$)	
$<$ pyramid 180°	$t(16) = -0.670$, $p = 0.51$
$<$ pyramid 270°	$t(16) = -1.530$, $p = 0.15$
$<$ trapezium 0°	$t(16) = -1.940$, $p = 0.07$
$>$ trapezium 90°	$t(16) = 1.030$, $p = 0.32$
$>$ trapezium 180°	$t(16) = 1.600$, $p = 0.13$
$>$ trapezium 270°	$t(16) = 3.930$, $p = 0.001$
$>$ t-shape 0°	$t(16) = 0.250$, $p = 0.80$
$>$ t-shape 90°	$t(16) = 1.360$, $p = 0.19$
$<$ t-shape 180°	$t(16) = -0.790$, $p = 0.44$
$<$ t-shape 270°	$t(16) = -0.950$, $p = 0.36$
pyramid 180° ($M = 665.29$ ms, $SD = 201.66$)	
$<$ pyramid 270°	$t(16) = -0.650$, $p = 0.52$
$<$ trapezium 0°	$t(16) = -0.540$, $p = 0.60$
$>$ trapezium 90°	$t(16) = 1.400$, $p = 0.18$

> trapezium 180°	$t(16) = 3.300$, $p = 0.004$
> trapezium 270°	$t(16) = 4.690$, $p < 0.001$
> t-shape 0°	$t(16) = 0.750$, $p = 0.47$
> t-shape 90°	$t(16) = 2.310$, $p = 0.03$
> t-shape 180°	$t(16) = 0.230$, $p = 0.82$
< t-shape 270°	$t(16) = -0.030$, $p = 0.98$
pyramid 270° ($M = 681.62$ ms, $SD = 206.20$)	
< trapezium 0°	$t(16) = -0.140$, $p = 0.89$
> trapezium 90°	$t(16) = 3.220$, $p = 0.005$
> trapezium 180°	$t(16) = 2.630$, $p = 0.02$
> trapezium 270°	$t(16) = 4.470$, $p < 0.001$
> t-shape 0°	$t(16) = 1.640$, $p = 0.12$
> t-shape 90°	$t(16) = 3.170$, $p = 0.006$
> t-shape 180°	$t(16) = 1.250$, $p = 0.23$
> t-shape 270°	$t(16) = 1.020$, $p = 0.32$
trapezium 0° ($M = 684.76$ ms, $SD = 171.51$)	
> trapezium 90°	$t(16) = 3.510$, $p = 0.003$
> trapezium 180°	$t(16) = 2.590$, $p = 0.02$
> trapezium 270°	$t(16) = 4.510$, $p < 0.001$
> t-shape 0°	$t(16) = 1.620$, $p = 0.13$
> t-shape 90°	$t(16) = 3.410$, $p = 0.004$
> t-shape 180°	$t(16) = 1.230$, $p = 0.24$
> t-shape 270°	$t(16) = 0.930$, $p = 0.37$
trapezium 90° ($M = 622.35$ ms, $SD = 199.04$)	
> trapezium 180°	$t(16) = 1.010$, $p = 0.33$
> trapezium 270°	$t(16) = 2.460$, $p = 0.03$
< t-shape 0°	$t(16) = -0.670$, $p = 0.51$
> t-shape 90°	$t(16) = 0.490$, $p = 0.63$
< t-shape 180°	$t(16) = -1.780$, $p = 0.09$
< t-shape 270°	$t(16) = -2.630$, $p = 0.02$
trapezium 180° ($M = 583.21$ ms, $SD = 194.02$)	
> trapezium 270°	$t(16) = 1.450$, $p = 0.17$
< t-shape 0°	$t(16) = -1.220$, $p = 0.24$
< t-shape 90°	$t(16) = -1.060$, $p = 0.31$
< t-shape 180°	$t(16) = -2.140$, $p = 0.05$
< t-shape 270°	$t(16) = -2.150$, $p = 0.05$
trapezium 270° ($M = 541.74$ ms, $SD = 159.33$)	
< t-shape 0°	$t(16) = -2.900$, $p = 0.01$
< t-shape 90°	$t(16) = -2.350$, $p = 0.03$
< t-shape 180°	$t(16) = -3.800$, $p = 0.002$
< t-shape 270°	$t(16) = -4.590$, $p < 0.001$

t-shape 0° ($M = 639.06$ ms, $SD = 219.17$)	
> t-shape 90°	$t(16) = 0.910$, $p = 0.37$
< t-shape 180°	$t(16) = -0.920$, $p = 0.37$
< t-shape 270°	$t(16) = -0.940$, $p = 0.36$
t-shape 90° ($M = 609.97$ ms, $SD = 170.81$)	
< t-shape 180°	$t(16) = -2.170$, $p = 0.05$
< t-shape 270°	$t(16) = -2.080$, $p = 0.05$
t-shape 180° ($M = 659.21$ ms, $SD = 196.22$)	
< t-shape 270°	$t(16) = -0.320$, $p = 0.75$

B Simulation Parameters

B.1 Study 1

Parameter	Value
Size of visual input	512×512
Visual feature extraction (Gabor filter)	
Bandwidth	2
θ	1.57
λ	3.00
γ	0.70
ψ	1.57
Activation of vertical filter	0.61
Hand network	
a_{inh}	3.3
a_{exp}	0.4
a_{inp}	13.0
s	1.0
m	80.0
τ	0.7
w_f	$\begin{bmatrix}4 & 2 & 3 & 3 & 2\end{bmatrix}$
Weight maps (Hand network)	
Gauss function Δ	0.5
Gauss function σ	0.5
Stop criteria	
t_{max}	4.0
y-threshold $t^{(F)}$	$\begin{bmatrix}0.995 & 0.995 & 0.995 & 0.995 & 0.995\end{bmatrix}$

B.2 Study 2a

Parameter	Value
Size of visual input	512 × 512
Visual feature extraction (Gabor filter)	
Bandwidth	2
θ	1.57
λ	3.00
γ	0.70
ψ	1.57
Activation of vertical filter	0.61
Hand network	
a_{inh}	3.3
a_{exp}	0.4
a_{inp}	13.0
s	1.0
m	80.0
τ	0.7
w_f	$\begin{bmatrix}4 & 2 & 3 & 3 & 2\end{bmatrix}$
Weight maps (Hand network)	
Gauss function Δ	0.5
Gauss function σ	0.5
Grasp template network	
$a_{\text{inh}}^{\text{tpl}}$	6.0
$a_{\text{exp}}^{\text{tpl}}$	0.002
s^{tpl}	0.3
m^{tpl}	300.0
τ^{tpl}	6.0
Initial template activation	0.5
Template maps (Grasp template network)	
Gauss function Δ	0.0
Gauss function σ	0.3
Stop criteria	
t_{max}	4.0
y^{tpl}-threshold t^{tpl}	0.999

B.3 Study 2b

Parameter	Value
Size of visual input	1024 × 1024
Visual feature extraction (Gabor filter)	
Bandwidth	2
θ	1.57
λ	3.00
γ	0.70
ψ	1.57
Activation of vertical filter	0.61
Hand network	
a_{inh}	3.3
a_{exp}	0.4
a_{inp}	13.0
s	1.0
m	80.0
τ	0.7
w_f	$\begin{bmatrix}4 & 2 & 3 & 3 & 2\end{bmatrix}$
Weight maps (Hand network)	
Gauss function Δ	0.5
Gauss function σ	0.5
Grasp template network	
$a_{\text{inh}}^{\text{tpl}}$	6.0
$a_{\text{exp}}^{\text{tpl}}$	0.002
s^{tpl}	0.3
m^{tpl}	300.0
τ^{tpl}	6.0
Initial template activation	0.5
Template maps (Grasp template network)	
Gauss function Δ	0.0
Gauss function σ	0.3
Stop criteria	
t_{max}	4.0
y^{tpl}-threshold t^{tpl}	0.999

B.4 Study 2c

Parameter	Value
Size of visual input	1024×1024
Visual feature extraction (Gabor filter)	
Bandwidth	2
θ	1.57
λ	3.00
γ	0.70
ψ	1.57
Activation of vertical filter	0.61
Hand network	
a_{inh}	3.3
a_{exp}	0.4
a_{inp}	13.0
a_{tpl}	4.0
s	1.0
m	80.0
τ	0.7
w_f	$\begin{bmatrix} 4 & 2 & 3 & 3 & 2 \end{bmatrix}$
Weight maps (Hand network)	
Gauss function Δ	0.5
Gauss function σ	0.5
Scale factor	0.3
Grasp template network	
$a_{\text{inh}}^{\text{tpl}}$	6.0
$a_{\text{exp}}^{\text{tpl}}$	0.002
s^{tpl}	0.3
m^{tpl}	300.0
τ^{tpl}	6.0
Initial activation of preferred template	0.7
Initial activation of non-preferred templates	0.3
Template maps (Grasp template network)	
Gauss function Δ	0.0
Gauss function σ	0.3

Stop criteria	
t_{max}	4.0
y^{tpl}-threshold t^{tpl}	0.999

B.5 Study 3a

Parameter	Value
Size of visual input (single object)	512×512
Size of visual input (multiple objects)	1024×1024
Visual feature extraction (Gabor filter)	
Bandwidth	2
θ	1.57
λ	3.00
γ	0.70
ψ	1.57
Activation of vertical filter	0.61
Hand network	
a_{inh}	3.3
a_{exp}	0.4
a_{inp}	13.0
s	1.0
m	80.0
τ	0.7
w_f	$\begin{bmatrix}4 & 2 & 3 & 3 & 2\end{bmatrix}$
Weight maps (Hand network)	
Gauss function Δ	0.5
Gauss function σ	0.5
Grasp template network	
$a_{\text{inh}}^{\text{tpl}}$	6.0
$a_{\text{exp}}^{\text{tpl}}$	0.002
s^{tpl}	0.3
m^{tpl}	300.0
τ^{tpl}	6.0
Initial template activation	0.5
Template maps (Grasp template network)	
Gauss function Δ	0.0
Gauss function σ	0.3
Stop criteria	
t_{max}	4.0
y^{tpl}-threshold t^{tpl}	0.999

B.6 Study 3b

Parameter	Value
Size of visual input (single object)	512×512
Size of visual input (multiple objects)	1024×1024
Visual feature extraction (Gabor filter)	
Bandwidth	2
θ	1.57
λ	3.00
γ	0.70
ψ	1.57
Activation of vertical filter	0.61
Hand network	
a_{inh}	3.3
a_{exp}	0.4
a_{inp}	13.0
a_{tpl}	4.0
s	1.0
m	80.0
τ	0.7
w_f	$\begin{bmatrix}4 & 2 & 3 & 3 & 2\end{bmatrix}$
Weight maps (Hand network)	
Gauss function Δ	0.5
Gauss function σ	0.5
Scale factor	0.3
Grasp template network	
$a_{\text{inh}}^{\text{tpl}}$	6.0
$a_{\text{exp}}^{\text{tpl}}$	0.002
s^{tpl}	0.3
m^{tpl}	300.0
τ^{tpl}	6.0
Initial activation of preferred template	0.6
Initial activation of non-preferred templates	0.4
Template maps (Grasp template network)	
Gauss function Δ	0.0
Gauss function σ	0.3
Stop criteria	
$t_{\max}$	4.0
y^{tpl}-threshold t^{tpl}	0.999

References

Arbib, M. A., Iberall, T., & Lyons, D. M. (1985). Coordinated control programs for movements of the hand. In A. W. Goodwin & I. Darian-Smith (Eds.), *Hand function and the neocortex* (pp. 111–129). Berlin: Springer.

Bekkering, H., & Neggers, S. F. W. (2002). Visual search is modulated by action intentions. *Psychological Science*, *13*(4), 370–374.

Bicchi, A. (1995). On the closure properties of robotic grasping. *The International Journal of Robotics Research*, *14*(4), 319–334.

Bohg, J., & Kragic, D. (2010). Learning grasping points with shape context. *Robotics and Autonomous Systems*, *58*(4), 362–377.

Borghi, A. M., Bonfiglioli, C., Lugli, L., Ricciardelli, P., Rubichi, S., & Nicoletti, R. (2007). Are visual stimuli sufficient to evoke motor information? studies with hand primes. *Neuroscience Letters*, *411*(1), 17–21.

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology*, *15*(2), 181–195.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1485), 1585–1599.

Cisek, P. (2008). The affordance competition hypothesis. In R. L. Klatzky, B. MacWhinney, & M. Behrmann (Eds.), *Embodiment, ego-space, and action* (pp. 203–246). Psychology Press.

Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron*, *45*, 801–814.

Craighero, L., Fadiga, L., Umilta, C. A., & Rizzolatti, G. (1996). Evidence for visuomotor priming effect. *NeuroReport*, *8*(1), 347–349.

Şahin, E., Çakmak, M., Doğar, M. R., Uğur, E., & Üçoluk, G. (2007). To afford or not to afford: a new formalization of affordances toward affordance-based robot control. *Adaptive Behavior*, *15*(4), 447–472.

Cutkosky, M. R. (1989). On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on Robotics and Automation*, *5*(3), 269–279.

Cutkosky, M. R., & Wright, P. K. (1986). Modeling manufacturing grips and correlations with the design of robotic hands. In *Proceedings of the IEEE International Conference on Robotics and Automation* (Vol. 3, pp. 1533–1539).

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, *2*(7), 1160–1169.

Deco, G., & Zihl, J. (2001). Top-down selective visual attention: A neurodynamical approach. *Visual Cognition*, *8*(1), 118–139.

Derbyshire, N., Ellis, R., & Tucker, M. (2006). The potentiation of two components of the reach-to-grasp action during object categorisation in visual memory. *Acta Psychologica*, *122*(1), 74–98.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*, 1245–1255.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

di Pellegrino, G., Rafal, R., & Tipper, S. P. (2005). Implicitly evoked actions modulate visual selection: evidence from parietal extinction. *Current Biology*, *15*(16), 1469–1472.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458.

Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology*, *91*(4), 451–471.

Fagg, A. H., & Arbib, M. A. (1998). Modeling parietal–premotor interactions in primate control of grasping. *Neural Networks*, *11*(7–8), 1277–1303.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*(12), 2379–2394.

Flash, T., & Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *The Journal of Neuroscience*, *5*(7), 1688–1703. Available from `http://www.jneurosci.org/cgi/content/abstract/5/7/1688`

Gibson, E. J. (2003). The world is so full of a number of things: On specification and perceptual learning. *Ecological Psychology*, *15*(4), 283–287.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton-Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton-Mifflin.

Grafton, S. T., Fadiga, L., Arbib, M. A., & Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *NeuroImage*, *6*(4), 231–236.

Greeno, J. G. (1994). Gibson's affordances. *Psychological Review*, *101*(2), 336–342.

Grèzes, J., & Decety, J. (2002). Does visual perception of objects afford action? evidence from a neuroimaging study. *Neuropsychologia*, *40*(2), 212–222.

Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: an fmri study of implicit processing. *European Journal of Neuroscience*, *17*(12), 2735-2740.

Hallford, E. (1983, June). *The specification of an object's size taken with reference to an observer's hand.* Paper presented at the Second International Conference on Event Perception, Vanderbilt University, Nashville, TN.

Handy, T. C., Grafton, S. T., Shroff, N. M., Ketay, S., & Gazzaniga, M. S. (2003). Graspable objects grab attention when the potential for action is recognized. *Nature Neuroscience*, *6*(4), 421–427.

Handy, T. C., & Tipper, C. M. (2007). Attentional orienting to graspable objects: what triggers the response? *NeuroReport*, *18*(9), 941–944.

Heinke, D., & Humphreys, G. W. (2003). Attention, spatial representation and visual neglect: Simulating emergent attention and spatial memory in the selective attention for identication model (SAIM). *Psychological Review*, *110*(1), 29–87.

Heinke, D., & Humphreys, G. W. (2005). Computational models of visual selective attention: A review. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 273–312). Hove & New York: Psychology Press.

Hindmarsh, A. C., Brown, P. N., Grant, K. E., Lee, S. N., Serban, R., Shumaker, D. E., et al. (2005). SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software*, *31*(3), 363–396.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. In *Proceedings of the National Academy of Sciences* (Vol. 79, pp. 2554–2558).

Hopfield, J. J., & Tank, D. W. (1985). "neural" computation of decisions in optimization problems. *Biological Cybernetics*, *52*(3), 141–152.

Humphreys, G. W., & Riddoch, M. J. (2001). Detection by action: neuropsychological evidence for action-defined templates in search. *Nature Neuroscience*, *4*(1), 84–88.

Humphreys, G. W., & Riddoch, M. J. (2003). From vision to action and action to vision: a convergent route approach to vision, action, and attention. *Psychology of Learning and Motivation*, *42*, 225–264.

Humphreys, G. W., Yoon, E. Y., Kumar, S., Lestou, V., Kitadono, K., Roberts, K. L., et al. (2010). The interaction of attention and action: From seeing action to acting on perception. *British Journal of Psychology*, *101*, 185–206.

Iberall, T., Bingham, G., & Arbib, M. A. (1986). Opposition space as a structuring concept for the analysis of skilled hand movements. In H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 158–173). Berlin: Springer.

Iberall, T., & Fagg, A. H. (1996). Neural network models for selecting hand shapes. In A. M. Wing, P. Haggard, & J. R. Flanagan (Eds.), *Hand and brain: The neurophysiology and psychology of hand movements.* Academic Press.

Iberall, T., Torras, C., & MacKenzie, C. L. (1990). Parameterizing prehension: A mathematical model of opposition space. In T. Kohonen & F. Fogelman-Soulié (Eds.), *Proceedings of the Third Cognitiva Symposium.* Madrid, Spain: Elsevier Science.

Jeannerod, M. (1981). Intersegmental coordination during reaching at natural visual objects. In J. B. Long & A. D. Baddeley (Eds.), *Proceedings of the Ninth International Symposium on Attention and Performance* (Vol. IX, pp. 153–169). Hillsdale, NJ: L. Erlbaum Associates.

Kamakura, N., Matsuo, M., Ishii, H., Mitsuboshi, F., & Miura, Y. (1980). Patterns of static prehension in normal hands. *The American Journal of Occupational Therapy*, *34*(7), 437–445.

Lam, M.-L., Ding, D., & Liu, Y.-H. (2001). Grasp planning with kinematic constraints. In *IEEE International Conference on Intelligent Robots and Systems* (Vol. 2, pp. 943–948).

MacKenzie, C. L., & Iberall, T. (1994). *The grasping hand* (Vol. 104). North-Holland.

Mark, L. S. (1987). Eyeheight-scaled information about affordances: A study of sitting and stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 361-370.

Mason, M. T. (2001). *Mechanics of robotic manipulation* (R. C. Arkin, Ed.). MIT Press.

Michaels, C. F., & Carello, C. (1981). *Direct perception.* Englewood Cliffs, New Jersey: Prentice Hall.

Milgram, P. (1987). A spectacle-mounted liquid-crystal tachistoscope. *Behaviour Research Methods, Instruments and Computers*, *19*(5), 449–456.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action.* Oxford, UK: Oxford University Press.

Mjolsness, E., & Garrett, C. (1990). Algebraic transformations of objective functions. *Neural Networks*, *3*(6), 651—-669.

Montesano, L., Lopes, M., Bernardino, A., & Santos-Victor, J. (2007). Modeling affordances using bayesian networks. In *International Conference on Intelligent Robots and Systems.* San Diego, USA.

Napier, J. R. (1956). The prehensile movements of the human hand. *The Journal of Bone and Joint Surgery*, *38B*, 902–913.

Newell, K. M., Scully, D. M., Tenenbaum, F., & Hardiman, S. (1989). Body scale and the development of prehension. *Developmental Psychobiology*, *22*(1), 1–13.

Olshausen, B. A., & Field, D. J. (1996a). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(13), 607–609.

Olshausen, B. A., & Field, D. J. (1996b). Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, *7*(2), 333–339.

Petkov, N. (1995). Biologically motivated computationally intensive approaches to image pattern recognition. *Future Generation Computer Systems*, *11*(4–5), 451–465.

Riddoch, M. J., Humphreys, G. W., Edwards, S., Baker, T., & Willson, K. (2003). Seeing the action: neuropsychological evidence for action-based effects on object selection. *Nature Neuroscience*, *6*(1), 82–89.

Riddoch, M. J., Humphreys, G. W., & Price, C. J. (1989). Routes to action: evidence from apraxia. *Cognitive Neuropsychology*, *6*(5), 437–454.

Rosenbaum, D. A., Halloran, E. S., & Cohen, R. G. (2006). Grasping movement plans. *Psychonomic Bulletin & Review*, *13*(5), 918–922.

Rumiati, R. I., & Humphreys, G. W. (1998). Recognition by action: dissociating visual and semantic routes to action in normal observers. *Journal of*

Experimental Psychology: Human Perception and Performance, *24*(2), 631–647.

Sanders, J. T. (1997). An ontology of affordances. *Ecological Psychology*, *9*(1), 97–112.

Saxena, A., Driemeyer, J., & Ng, A. Y. (2008). Robotic grasping of novel objects using vision. *The International Journal of Robotics Research*, *27*(2), 157–173.

Slocum, D. B., & Pratt, D. R. (1944). The principles of amputations of the fingers and hand. *The Journal of Bone and Joint Surgery*, *26*, 535–546.

Smeets, J. B. J., & Brenner, E. (1999). A new view on grasping. *Motor Control*, *3*(3), 237–271.

Soanes, C., & Hawker, S. (Eds.). (2008). *Compact oxford english dictionary* (3rd edition revised ed.). Oxford: University Press.

Solomon, J., Carello, C., Grosofsky, A., & Turvey, M. T. (1984). *Body scaled information for reaching.* Paper presented at the 55th annual meeting of the Eastern Psychological Association, Baltimore, MD.

Stoffregen, T. A. (2003). Affordances as properties of the animal-environment system. *Ecological Psychology*, *15*(2), 115–134.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 830–846.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition*, *8*(6), 6.

Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, *116*(2), 185–203.

Turvey, M. T. (1992). Affordances and prospective control: an outline of the ontology. *Ecological Psychology*, *4*(3), 173–187.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–558). Cambridge, MA, USA: MIT Press.

Vingerhoets, G., Vandamme, K., & Vercammen, A. (2009). Conceptual and physical object qualities contribute differently to motor affordances. *Brain and Cognition*, *69*(3), 481–489.

Ward, R. (1999). Interactions between perception and action systems: a model for selective action. In G. W. Humphreys, J. Duncan, & A. Treisman (Eds.), *Attention, space and action: Studies in cognitive neuroscience* (pp. 311–332). Oxford University Press.

Warren, Jr., W. H. (1984). Perceiving affordances: visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 683–703.

Warren, Jr., W. H. (1988). Action modes and laws of control for the visual guidance of action. In O. G. Meijer & K. Roth (Eds.), *Complex movement behaviour – 'the' motor-action controversy* (Vol. 50, pp. 339–379). North-Holland.

Warren, Jr., W. H., & Whang, S. (1987). Visual guidance of walking through apertures: body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 371–383.