UNIVERSIDADE TÉCNICA DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

A Computational Approach on the Co-Development of Visual Sensorimotor Structures

Jonas Ruesch

Supervisor: Doctor Alexandre José Malheiro Bernardino

Co-Supervisor: Doctor José Alberto Rosado dos Santos-Victor

Thesis approved in public session to obtain the PhD Degree in
Electrical and Computer Engineering

Jury final classification: Pass with Merit

Jury

Chairperson: Chairman of the IST Scientific Board

Members of the Committee:

Doctor Rolf Pfeifer

Doctor Hélder de Jesus Araújo

Doctor José Alberto Rosado dos Santos-Victor

Doctor Ana Luísa Nobre Fred

Doctor Alexandre José Malheiro Bernardino

2013

UNIVERSIDADE TÉCNICA DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

A Computational Approach on the Co-Development of Visual Sensorimotor Structures

Jonas Ruesch

Supervisor: Doctor Alexandre José Malheiro Bernardino

Co-Supervisor: Doctor José Alberto Rosado dos Santos-Victor

Thesis approved in public session to obtain the PhD Degree in

Electrical and Computer Engineering

Jury final classification: Pass with Merit

Jury

Chairperson: Chairman of the IST Scientific Board

Members of the Committee:

Doctor Rolf Pfeifer, Full Professor, University of Zurich

Doctor Hélder de Jesus Araújo, Full Professor,

Faculdade de Ciências e Tecnologia, Universidade de Coimbra

Doctor José Alberto Rosado dos Santos-Victor, Full Professor,

Instituto Superior Técnico, Universidade Técnica de Lisboa

Doctor Ana Luísa Nobre Fred, Assistant Professor,

Instituto Superior Técnico, Universidade Técnica de Lisboa

Doctor Alexandre José Malheiro Bernardino, Assistant Professor,

Instituto Superior Técnico, Universidade Técnica de Lisboa

2013

Contents

"Joe's blind left eye was not blind. Joe had devised an artificial eye, wired into the optic center, that presented his mind with pictures, often quite at variance with the reports of the right eye. This was especially noticeable when he looked at human and animal subjects, and he came to realize to what extent that which we see is conditioned by what we expect to see – that is, by a habitual scanning pattern, whereas the artificial eye had no scanning pattern. The lens was fixed and Joe had to direct it by movements of his head. On the other hand, the lens could be adjusted to a wide angle, which greatly extended the range of his peripheral vision. He found that he could read motives and expressions with great precision by comparing the data of the good eye, which was picking up what someone wants to project, and the data of the synthetic eye. Sometimes the difference in expression was so grotesque that he was surprised it was not immediately apparent to anyone."

William S. Burroughs – The Western Lands

Abstract

To follow a goal-directed behavior, an autonomous agent must be able to acquire knowledge about the causality between its own motor actions and the corresponding changes induced in sensory feedback. Since the complexity of such sensorimotor relationships directly influences required cognitive resources, this work proposes that it is of great importance to keep an autonomous agent's sensorimotor relationships *simple*. This implies that the agent should be designed in a way such that sensory consequences can be described and predicted in a simplified manner. Living organisms implement this paradigm by adapting sensory and motor systems specifically to their behavior and environment. As a result, they are able to predict sensorimotor consequences with a strongly limited amount of (expensive) nervous tissue. In contrast, most robots are composed of standard components and are not nearly as well adapted to a specific behavior and environment. Implicitly, this lack of adaptation has important implications which are often neglected. In this context, the thesis proposes that advantageous sensory and motor structures can be evolved by rewarding the ability to predict self-induced stimuli through simple sensorimotor relationships. It is shown that this criterion can be employed to conjointly develop layers of visual receptors and motor primitives which together are more efficient with respect to both, computational and physical resources. Different from classical work in sensorimotor learning, the structure of sensory and motor spaces are considered to be variables of the learning problem. The principal focus thus, does not simply lie on learning an appropriate sensorimotor map, but on shaping the morphology of sensors and actuators such that the induced sensorimotor relationship is simplified. Experiments consider a simulated agent recording realistic visual stimuli from natural images. The obtained results demonstrate the ability of the proposed method to i) synthesize visual sensorimotor structures adapted to an agent's environment and behavior, and ii) serve as a computational model for testing hypotheses regarding the development of biological visual sensorimotor systems. In conclusion, the work proposes a unified approach for self-organizing sensory and motor primitives from sensorimotor experience and paves the way towards more advanced autonomous systems.

Keywords: sensorimotor learning, sensorimotor coupling, self-organization of sensorimotor structures, visual receptors, visual motor primitives, visual stimulus prediction.

Resumo

Para um agente autónomo conseguir concretizar objectivos, é antes de mais essencial ser capaz de adquirir conhecimento sobre as suas próprias ações e mudanças resultantes na percepção sensorial. Como a complexidade de tais relações sensório-motoras influencia diretamente os recursos cognitivos necessários, este trabalho propõe que é de grande importância manter *simples* estas relações. Isto implica que o agente deve ser concebido de tal forma que as consequências sensoriais possam ser descritas e preditas de forma simplificada. Organismos vivos implementam este paradigma adaptando os sistemas sensoriais e motores especificamente para o seu comportamento e meio ambiente, consequentemente ganhando a capacidade de prever as consequências sensório-motoras com uma quantidade limitada de tecido nervoso (dispendioso). Em contraste, a maioria dos robôs são compostos de componentes padronizados e não estão de forma alguma tão bem adaptados a um determinado comportamento e ambiente. Implicitamente, essa falta de adaptação tem implicações importantes que muitas vezes são negligenciadas. Neste contexto, esta tese propõe que estruturas sensório-motoras vantajosas podem ser evoluídas premiando a capacidade de prever estímulos auto-induzidos através de relações sensório-motoras simples. Mostra-se que este critério pode ser empregue para conjuntamente desenvolver camadas de receptores visuais e primitivas motoras que em conjunto são mais eficientes no que diz respeito a recursos computacionais e físicos. Contrário a trabalhos clássicos em aprendizagem sensório-motora, a estrutura dos espaços sensoriais e motores é considerada variável. Portanto o foco principal não consiste simplesmente em aprender um mapa sensório-motor adequado, mas em moldar a morfologia dos sensores e actuadores tal que a relação sensório-motora induzida seja simplificada. As experiências realizadas consideram um agente simulado que grava estímulos visuais realistas a partir de imagens naturais. Os resultados obtidos demonstram a capacidade do método proposto para i) sintetizar estruturas sensório-motoras visuais adaptadas ao ambiente e comportamento do agente, e ii) servir de modelo computacional para testar hipóteses sobre o desenvolvimento de sistemas sensório-motores visuais biológicos. Concluindo, o trabalho propõe uma abordagem unificada para a auto-organização sensorial e primitivas motoras a partir de experiências sensório-motoras e abre o caminho para sistemas autónomos mais avançados.

Palavras-chave: aprendizagem e acoplamento sensório-motor, auto-organização de estruturas sensório-motoras, receptores visuais, primitivas motoras visuais, previsão de estímulos.

x

Acknowledgments

First and foremost, I would like to thank both of my supervisors Alexandre Bernardino and José Santos-Victor. Not only did they provide me with the freedom and the time to pursue a project inspired completely on the basis of my own interests and ideas, but they also actively supported me during its entire course. Alexandre, despite all his duties and commitments in various research projects, always found time to discuss my work providing invaluable input by sharing his broad knowledge and extensive practical experience. He accompanied me through all the ideas I explored, including the many which were eventually abandoned. No matter if a problem surfaced as a conceptual mismatch, a setback during an experiment, or a bug in a program, he would always be available and interested to get to the bottom of it. For all of this, I would like to express my heartfelt gratitude.

In the same way I would like to thank José, my co-supervisor and head of VisLab. Thanks to his huge commitment and great support for all of us, VisLab has been a unique and inspiring environment to realize the present work. On a personal level, I very much appreciate his positive way of thinking and the motivation he keeps providing to all the lab members. And last but not least, thank you for all the charutos and home-made ginjinha to celebrate successful project review meetings, Christmas dinners, and other major and minor breakthroughs.

For his immeasurable amount of help and support I owe a great debt of gratitude to Ricardo Ferreira. Without him, this thesis would not have been possible in its present form. Since the day he started to listen to my ideas he kept helping me in countless hours to put an important part of them into concrete and consistent terms. He provided me with input and insights beyond what I could ever have expected from anybody. Thank you!! I learned a lot, and I am very grateful for all the creative insights, the many enlightening perspectives, and all the ever so patient explications. You showed me how to draw an Ockham's razor faster than my shadow. I'll keep exercising...

On a personal level, I would like to thank all the current and former members of VisLab and all its recurrently visiting associates I met during my studies. With many of them I not only worked, but also very much enjoyed going out and spending time in the wonderful city of Lisbon. Thanks to all of you, VisLab always felt much more than just a place to work at. For the many helpful and inspiring scientific discussions, I would particularly like to thank Manuel Lopes, Luis Montesano, Michele Tavella, Giampiero Salvi, Matthijs Spaan and Jesús Capitán.

Equally heartfelt thanks go to the members of the Artificial Intelligence Laboratory at the University of Zurich. In particular I would like to express my gratitude to Rolf Pfeifer for the many ways in which he an his work inspired me during the first year of my PhD studies at the AILab in Zurich. The ideas brought forward by his group and the many fruitful discussions with all the members of AILab have influenced my view on Artificial Intelligence on many different levels in a very constructive manner. In particular I would like to acknowledge Matej Hoffmann, Lijin Aryananda, Nathan Labhart, Shuhei Miyashita, and Max Lungarella.

A huge tribute for emotional support and many happy hours also goes to all my friends who came to visit me during my stay in Lisbon. Big hug to all of you! You brought me a lot of joy and welcome distraction from sometimes gloomy research moods.

And at last, but in many ways in the very first place, a deep-felt thanks and a heartily hug to my parents in an attempt to express my appreciation for their immense support and attention. You are above all responsible for assembling the essential prerequisites which led to the emergence of this thesis.

Chapter 1

Introduction

Contents

Building machines with skills comparable to those of a human is an age-old dream of mankind. The first known written record of an anthropomorphic automaton dates back more than two thousand years. A passage in the Daoist text *Lieh-tzǔ* describes how the craftsman Yan Shi presents a life-sized moving figure to the Chinese King Mu of Zhou [41]. Although this historical record is considered a legend, it reflects ancient man's desire to create autonomous systems.

Today, robots outperform humans in several areas. For example, they handle mathematical operations much faster than we do, and in industrial manufacturing facilities, they assemble parts, weld, or place products with accuracy and speed, way beyond what we can achieve. However, outside the structured environment of assembly lines, artificial systems are still far from reaching the autonomy or versatility of biological organisms. This is mainly due to the fact that, unlike the mechanistic execution of a prespecified task, the interaction with a natural environment requires an artificial system to behave in an adaptive manner and to take appropriate and context dependent actions. Despite remarkable progress in some areas of artificial intelligence, advancements in this direction have been slow. Designing a system able to act under real-world conditions has proven to be extraordinarily difficult. The interaction with an unknown environment requires an artificial agent not only to feature

suitable sensors and actuators, but also to exhibit advanced *sensorimotor coordination*, meaning motor actions are deduced from sensory stimuli in a prompt and adequate manner. While today's hardware enables robotocists to equip artificial agents with sensors and actuators more powerful than ever, the sensorimotor skills of these designs are often lagging behind considerably. Thus, while many robots possess sensory systems which in principle are able to gather sufficient information to solve difficult tasks, they are unable to choose appropriate actions based on recorded stimuli. The following section motivates that this is not primarily because these robots are incapable per se, but to a big part because the structure of their sensory and motor systems make it often unnecessarily hard to couple sensory and motor signals efficiently.

1.1   In Favor of Simple Brains

When a robot fails to select a sensible action despite the potential availability of required sensory information, one could say the system lacks the required cognitive skills to find a solution. However, one might argue conversely, the reason why the systems fails to deduce an appropriate action is simply because the relationship between sensory and motor signals is too complicated in order for the robot to translate recorded sensory stimuli into a motor action adequate in the current context. While the first formulation suggests to increase the robot's processing power, the second formulation implies an alternative approach. Instead of striving to increase cognitive capabilities, one could try to find a design for the robot's sensorimotor apparatus such that less complex operations are required to translate sensory feedback into purposeful actions. Seen from the latter perspective it can be conjectured: if sensory and motor systems of a robot are well concerted and adapted to a task and environment, then the cognitive load imposed on the agent's processing system is reduced. According to this line of thinking, this work considers the two following points of crucial importance for the design of autonomous artificial agents:

1. The sensory system should be adapted to the agent's motor system and environment. This means that the sensory system should provide (only) stimuli which are meaningful with respect to the agent's motor capabilities and environment. Or, conversely, that sensing something which is never relevant for a possible motor action is redundant.

2. The motor system should be adapted to the agent's sensory system and environment. This means that the motor system should support (only) actions which lead to meaningful transformations of sensed stimuli. Or conversely, actions which lead to a discontinuous sensory experience are
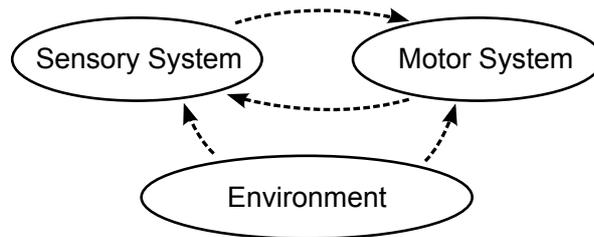
Figure 1.1: Proposed directions of influence for structural adaptations of an artificial sensorimotor system. On one hand, both the sensory and the motor system are influenced by the environment. On the other hand, sensory and motor systems can be expected to influence each other mutually.

unfavorable since they hamper the linkage between consecutive sensory stimuli and thus impede the construction of coherent percepts.

Fig. 1.1 schematically depicts the directions of influence when considering such mutual adaptations to sensory and motor systems with respect to a given environment. In the remainder of this section, the reduction of required brainpower in favor of better adapted physical structures and interaction strategies is first motivated from a biological perspective and subsequently discussed with respect to the design of autonomous robots.

A Biological Perspective. In biological systems, evolutionary pressure has a profound impact on the development of an animal's nervous system [83]. Although at first sight, it seems generally advantageous for an organism to possess more sophisticated cognitive capabilities, from an evolutionary perspective, it is not desirable to have an oversized brain. In fact, the maintenance of an extensive nervous system is expensive. Even at rest, neural tissue consumes a considerable amount of energy [63]. It is therefore reasonable to assume that there is high selective pressure for evolving efficient solutions for stimulus processing. This view is supported by a number of neuroscientific studies. The review provided in [78] includes a number of examples which demonstrate how nervous systems reduce energy consumption by adapting their morphology and physiology. For example, insect photoreceptors have a membrane that filters the light-induced current generated by the receptor. This filter is tuned to discard frequencies which are not relevant for the animal's behavior and provides a reduction of bandwidth for further signal processing circuits [77]. For neural codes where information is represented by a population of neurons, it has been suggested that a sparse coding strategy, where only a small proportion of neurons in the population is active, is frequently used in biological signal processing. Evidence for such sparse codes has been found in different species, see e.g. [79, 133]. Finally, energetic costs are also assumed to be reduced by "saving wire". For example by placing brain regions

3

"Morphological Computation"   "Signal Based Computation"
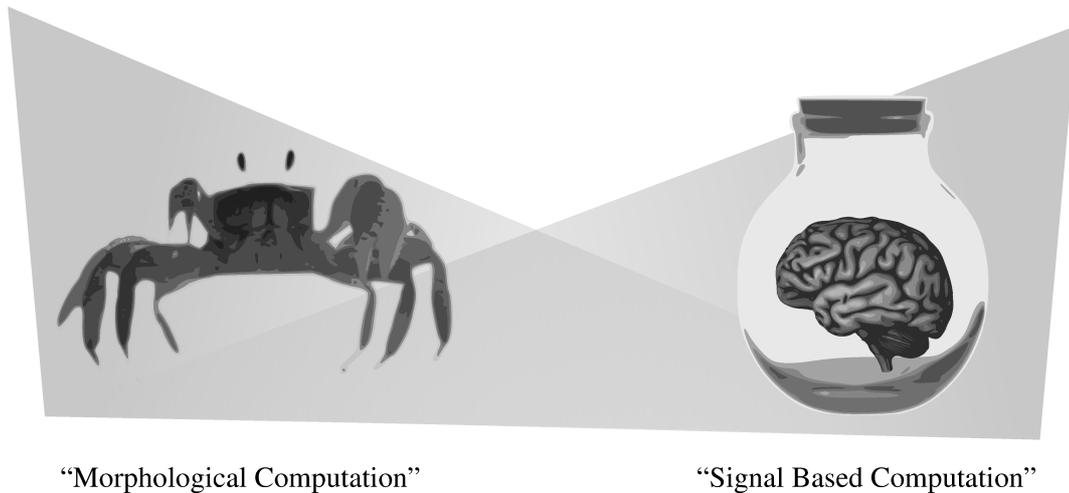
Figure 1.2: Two different, but not exclusive paradigms for an autonomous system to approach a problem i) a system positioned on the left side of the figure tends to solve a task supported by its *physical body*, and ii) a system positioned on the right side of the figure tends to solve a task with an emphasis on analyzing and *processing signals* representing the problem. In most robotic setups "morphological computation" and "signal based computation" are combined to solve a task at hand. It can be argued that suitable adaptations to a system's body can achieve a shift towards the left side, and thus, can reduce signal processing resources required.

with high interconnectivity close to each other [18]. Others have conjectured that the entire layout of neural connections in the mammalian cortex is minimizing wiring costs [19]. Last but not least, animals with a bigger nervous system are subject to a prolonged phase of training during their ontogenetic development. This imposes considerable secondary costs and is a possible threat to offspring fitness.

Eventually, it *is* worth to develop a bigger brain, *if* the complexity of the overall task faced by the organism – survival in the most general sense – becomes more complex. In other words, despite the associated costs, if survival requires a more complex behavior, it is advantageous to develop a more sophisticated nervous system. Examples provided in [78] underline this relationship. E.g., it has been found that some animals living on islands show reductions of specific sensory signal processing structures compared to their mainland conspecifics. It is assumed that such reductions, for example in the visual system, occur due to reduced need for vigilance in the absence of predators on an island [54]. Similarly, fruit flies bred under laboratory conditions experience reduced selective pressure on their visual system and evolve a reduced compound eye due to the abundant availability of food and mating partners [125].

In conclusion, considering the costs associated with the development of neural tissue and studies demonstrating the degeneration of neural structures under relieved selective pressure, it is reasonable
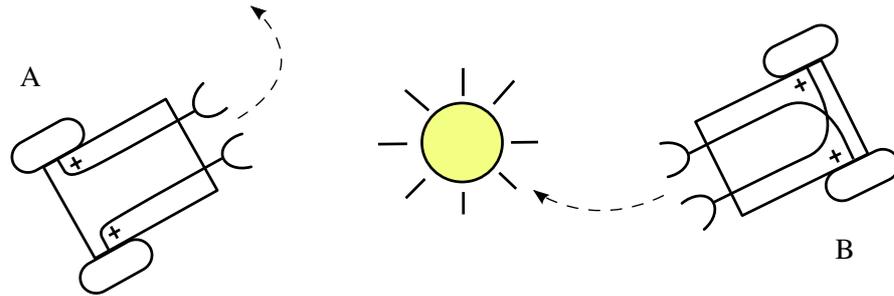
Figure 1.3: Two Braitenberg vehicles. Each of them is equipped with two actuated wheels and two temperature or light receptive sensors. The vehicles have a "minimalist brain" consisting of two wires. In vehicle A, sensors route positive feedback to actuators located on the same lateral side. In vehicle B, sensors are connected to actuators crosswise. Despite this simple setup, both versions exhibit a diverse behavior. Vehicle A avoids a source of heat or light and vehicle B is attracted to it.

to assume that as long as a task can be solved by "cheaper" means, the development of more resource-intensive cognitive abilities is delayed in biological systems. Of course, to what extent cognitive resources can be "saved" by implementing smart adaptations depends on the specific task to be solved and the associated potential for specialization. Furthermore, a highly adapted design always comes at the cost of loosing the ability to address more general problems. A specific survival strategy in a very particular ecological-niche allows for a higher degree of adaptation and might reduce considerably the pressure for more advanced cognitive skills. On the other hand, animals which follow a more generalist surviving strategy typically evolve more complex behaviors and require more versatile sensory systems which in turn demand for more sophisticated stimulus processing systems and cognitive capabilities.

Reducing Required Cognitive Resources in Artificial Systems. In robotics and artificial intelligence, the strategy of solving a problem with less computational power through a specifically designed body has sometimes been referred to as "morphological computation", see for example [84, 87]. A great number of examples demonstrating this approach for artificial systems can be found in [88] and [86]. There, it is shown how the cognitive load imposed on artificial agents can be reduced by taking advantage of the morphological characteristics of an agent's body and the properties of the ecological niche inhabited.

A particularly descriptive example is due to an early but highly influential work by Braitenberg [14]. In this work a number of different vehicle designs are described of which the most famous one consists of just two actuators and two receptive sensor elements, see also Fig. 1.3. Each actuator of this vehicle drives a wheel, and each sensor element records a quantity like temperature or luminance.

(a) Robot *Chica*   (b) Perception-action cycle
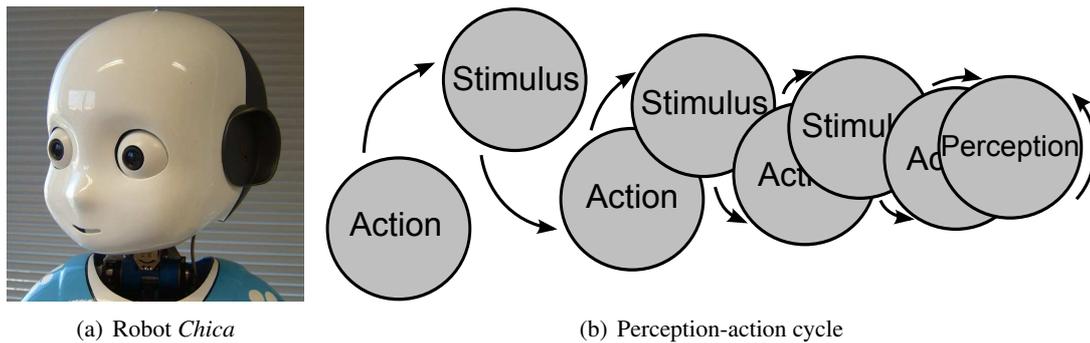
Figure 1.4: Example of a reactive sensorimotor loop. The visual attention system described in [106], shown here running on the robot *Chica*, continuously drives the robot's gaze towards visually salient locations. The execution of this action-perception cycle leads to a reciprocal recursion during which actions are chosen based on recorded stimuli, and in turn stimuli are influenced by the actions taken. In this sense, actions become a part of sensing, and sensing becomes a part of acting.

By connecting the sensor elements in different manners directly to the actuators, an extremely simple relationship between sensors and actuators is established. Despite such a more than simplistic "brain", the vehicle exhibits a quite diverse behavior and can solve a comparatively difficult task. Depending on the wiring, the vehicle can find and follow the source of the measured quantity, or it moves away from such a source. Thus, with a circuit consisting of only two wires, the vehicle establishes a tight connection with its environment and implements a specific behavior.

Other more recent designs for artificial systems making use of "cheap" visual perception include robots with solve more complex tasks like navigation and visual tracking or object recognition [136, 36, 37, 98, 35]. A robot inspired by the praying mantis using "peering movements" of its upper body and head for visual depth perception has been described in [15]. Directly related to adaptive sensorimotor structures as addressed by this work, a robot with an adjustable 1-dimensional visual sensor is presented in [65]. This robot learns to change the distribution of its visual receptors such that projected stimuli undergo a uniform translation during straight locomotion. The proposed optimization of the sensor relates to the idea of reducing cognitive resources through structural changes, in the sense that the resulting receptor distribution facilitates visual distance estimation.

To conclude, this section briefly reviews the visual attention system implemented by the author in previous work [106]. Observations made there directly relate to the tight interdependency of sensory and motor systems in autonomous agents and marked the beginning of the work addressed in this thesis. A short explanation of the system is described next. When running the mentioned attention system on the humanoid robot *iCub* (Fig. 1.4 shows an early copy named "Chica"), then the robot acts according to a simple sensorimotor transformation which combines the selection of visually salient regions

with a rule for temporal stimulus inhibition. According to this rule, the robot explores its surroundings by continuously focusing on conspicuous locations in its field of view. The resulting behavior of the robot is surprisingly diverse and is perceived as anthropomorphic even though the underlying rule is very simple. This is mainly due to the fact that the system establishes a close relationship with its environment by engaging the robot in a continuous action-perception cycle with prompt reactions to external events. Furthermore, it is interesting to observe that the sustained execution of such a action-perception cycle leads to a reciprocal recursion. While the behavior selects actions based on recorded sensory stimuli, sensory feedback in turn is influenced by the actions taken. After following the execution of action-perception cycles for several iterations it becomes apparent that the classical view of a sequential "sense", "process", "act" is blurred. Actions become a part of sensing, and sensing becomes a part of acting. This view on temporal processes in visual perception is also supported by the *premotor theory of attention* which proposes that the planning of motor actions directly influences visual attention [99, 24]. Furthermore, an inspiring introduction to a paradigm which relates perception directly to sensorimotor coordination and *sensorimotor contingencies* can also be found in [81].

In Conclusion. This section motivated that, i) the complexity of the relationships induced between sensory and motor signals by a particular organism or artificial system affects the cognitive resources required by that system, and ii) adaptations to the design of a sensorimotor system can be made in a way such that relationships between sensory and motor signals are simplified.

The next section discusses sensorimotor transformations in particular. These mapping functions are of importance with respect to the above summarized observations. They encode sensorimotor relationships by mapping sensory stimuli to motor signals and vice versa.

1.2 Sensorimotor Transformations

For a purely reactive agent, a simple brain can be considered a function which transforms sensory stimuli into motor actions. This section provides a brief overview on modeling and learning such sensorimotor relationships, in addition, it is outlined how the physical structure of a sensorimotor system influences these transformations.

In mathematical terms, sensorimotor transformations can simply be considered functions mapping an input domain to an output range in one of two directions: either the transformation takes as input a stimulus and provides as output a motor command, or it takes a motor command and provides a stimulus. Since such a mapping between sensory stimuli and motor actions is in most cases initially

unknown, it is typically desirable to learn transformation functions via unsupervised exploration or supervised teaching processes. However, due to nonlinearities and high dimensional sensor and motor signals, this is often a non-trivial task. In general, the problem of finding an approximation of a desired transformation has to be addressed by learning a non-linear function. In robotics a number of techniques from the broad range of applicable methods have found to frequently provide satisfactory results. Two particular methods should be mentioned here. The first approach employs a set of basis functions which are then linearly combined to represent a complete mapping function. A review on models of sensorimotor transformations using this approach can be found in [92]. A second approach uses self-organizing maps to represent the relationship between sensory input and motor actions [55]. Similar to a function constructed by linearly combining basis functions, self-organizing maps attempt to approximate a principal manifold in the input space by adapting the position of discrete nodes according to experienced training samples. Due to their simple use in online and unsupervised learning problems, the use of self-organizing maps is of particular appeal in applications with such requirements.

Sensor to Motor Transformations.   In the *sensor to motor* direction, a sensorimotor transformation associates motor actions to sensory stimuli. This direction is usually considered the "default" direction when referring to sensorimotor maps.

Studies on visual sensorimotor transformations in a sensor to motor direction typically concentrate on learning oculomotor actions required to center a target stimulus on the visual sensor, or on how to translate visual signals into coordinated eye-hand movements for reaching [39, 27].

Motor to Sensor Transformations.   In the *motor to sensor* direction, a sensorimotor transformation estimates future sensory stimuli based on given motor actions. Such a transformation represents a feedforward model of an animal's or artificial agent's sensorimotor system. It is of particular interest in the context of this thesis, since it is expected to represent a "location" for where to implement the ideas outlined in Sect. 1.1. Why is further explained in the remainder of this section.

Chap. 2 provides a detailed overview of feedforward sensorimotor transformations in biological systems and outlines their importance for living organisms. A review of recent work on modelling feedforward transformations can be found in [137]. In developmental robotics, the prediction of self-induced sensory feedback has been previously addressed in work described for example in [62].

8

Measuring Properties of Sensorimotor Transformations. In a study conducted by Lungarella et al. information theoretic measures were used to analyze the causal structure present in the information flow induced by sensorimotor activities of a visual system [67]. The results show that the characteristics of the recorded signals have strong ties to spatiotemporal relationships defined between the physical embodiment and the movement strategies executed by the system. Based on this insight, it is concluded that the physical structure and the behavior of an embodied system mutually influence each other in defining the characteristics of recorded stimuli.

Variable Sensorimotor Structures. Clearly, changes to the design of a sensorimotor system induce changes to a previously established sensorimotor map. This, conversely, also means that, the properties of the associated sensorimotor transformation can be influenced by changing the design of the sensorimotor system. Considering the preference for "simple brains" introduced in the previous section, the relationship between properties of sensorimotor transformations and the design of sensorimotor systems is of principal importance.

To the best of the author's knowledge, the extent of previous work addressing the design of sensorimotor structures depending on properties of induced sensorimotor transformations is limited. A particular example of a robotic implementation directly related to this idea has been described in [65] and was previously reviewed at the end of Sect. 1.1. Other influential work described in [21] will be reviewed as part of Sect. 1.3. Partly related work on explicitly deducing the topology of unknown visual sensor layouts has been published in [80]. There, the retinotopic layout of an unknown visual sensor is reconstructed using an entropy maximization method relying on information distance measures between sensor elements. Comprehensive work on the development of sensorimotor relationships from an internal perspective considering an unknown sensorimotor structure has also been described in [89].

1.3 Self-Similarity in Visual Sensorimotor Systems

At the beginning of this work, different sources of inspiration led to the belief that self-similarity can serve as a relevant property to qualify the adaptation of sensorimotor structures. This section motivates why and reviews related material. First, a general introduction is provided. Subsequently, it is discussed how self-similarity manifests itself in visual sensorimotor systems. Eventually, the section is concluded by reviewing work published by Clippingdale et al. [21]. There, a measure of self-similarity is used to develop abstract geometrical layouts for visual sensor topologies. This

(a) Sierpinski Triangle[*]   (b) Seahorse Tail[*]   (c) Sunflower[*]   (d) Broccoli[*]
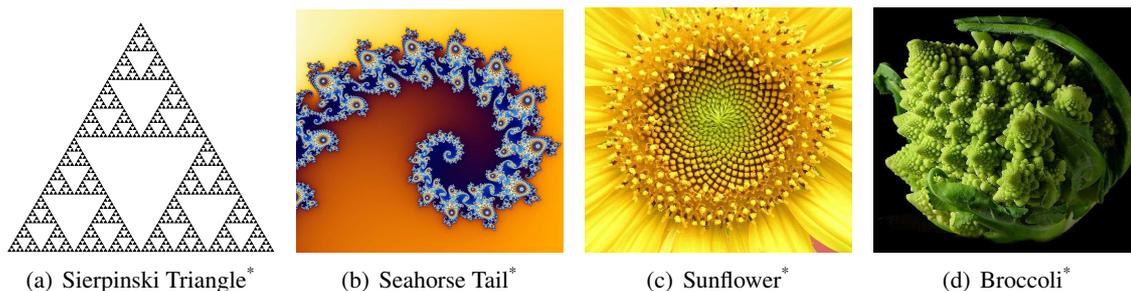
Figure 1.5: Structures with self-similar properties. The Sierpinski Triangle and the so-called Seahorse Tail of the Mandelbrot set are exactly self-similar under combinations of dilation and rotation transformations. The center of the sunflower shown in (c), and the broccoli floret shown in (d) are approximately self-similar under combinations of dilation and rotation transformations.

directly relates to the previous sections 1.1 and 1.2, because it reveals that self-similar properties in sensorimotor systems correlate with simplified sensorimotor transformations.

What Are Self-Similar Structures?   An object is considered self-similar if it is approximately or exactly similar to part of itself. This is for example a principal property of fractals like the Sierpinski triangle shown in Fig. 1.5(a). The shown pattern is scale invariant, meaning it can be scaled by a specific factor any number of times and its shape does not change. Similarly, the Mandelbrot set is self-similar with respect to dilation centered at so called *Misiurewicz* points. A section of the Mandelbrot set centered at a Misiurewicz point can be seen in Fig. 1.5(b). In nature, self-similarity often appears related to growth processes, both in plants as well as in animals. Two representative examples are depicted in Fig. 1.5(c) and Fig. 1.5(d). A more in depth discussion on natural growth processes and how to model them can be found for example in [96] and [8].

For later purposes, it is important to point out that a structure is always self-similar with respect to a specific set of transformations. Hence, self-similarity can be described as a combination of a structure S and a corresponding set M of transformations $\{t_m\}_{m \in M}$ such that

$$S = \bigcup_{m \in M} t_m(S). \tag{1.1}$$

Thus, the fractal shown in Fig 1.5(a) is self-similar under certain dilation transformations and the structures depicted in Fig. 1.5 (b) – (d) are self-similar under certain transformations composed of

(a) Paralejurus (genus), Trilobita (class)[*].   (b) Erbenochile (genus), Trilobita (class)[*].

Figure 1.6: Two examples of fossilized trilobite eyes. Trilobite fossils document the oldest preserved visual system; and, they show that in many trilobite species, the so-called *eyelets* were arranged in patterns of highly regular and self-similar logarithmic spirals. This is especially noticeable for species with eyelets of varying size, compare also Fig. 6 in [20].

dilation and rotation.

In sensorimotor systems, self-similarity can be considered to be present if sensors and actuators are designed in a way such that there are motor actions which can induce self-similar stimulus transformations. Considering a visual sensor composed of a number of discrete receptors, a self-similar stimulus transformation means for example that there exist motor actions which move the sensor in a way such that recorded visual stimuli are displaced exactly or approximately from receptor to receptor. Interestingly, visual sensorimotor systems in living organisms often approximate such self-similar properties for some subset of available motor actions. The next paragraph reviews a few examples and discusses a possible explanation for why this might be advantageous.

Self-Similarity in Early Visual Systems. Many fossils of Trilobite eyes – the oldest preserved visual system – show structures with strong self-similar properties. Similar to modern compound eyes, the eyes of these animals were composed of independent lenses called *eyelets*. In many trilobite species these eyelets were arranged in patterns of intersecting logarithmic spirals [20]. Two examples with moderate spatial variance in eyelet size are shown in Fig. 1.6, other illustrations of fossils with clearly expanding logarithmic spiral patterns and greater variance in eyelet size can be found in [20] (Fig. 6). Non-uniformly sized eyelets arranged in expanding logarithmic spirals are suitable to induce perfectly self-similar stimulus transformations under actions which lead to a dilation of the projected stimulus in radial direction of the spiral. The induced transformation is an affine transformation which contracts

---

[*]Image licenses: (a) and (b) released under the Creative Commons Attribution-ShareAlike (CC BY-SA) by users Pseudomorph and Micha L. Riser (in sequence) of the English Wikipedia project.

towards a point on the sensor surface (far) outside the area covered by the eyelets. For an animal living on the sea floor, locomotion results in a "zoom in" on the soil in front of it. Thus, it could be argued that there might exist naturally occurring actions in Trilobite locomotion which induce self-similar stimulus transformations. Recent evidence discussed in [112] also suggests that the trilobite eye might have implemented a system of muscles which was able to move the receptor areas inside each unit with respect to the fixed lenses, similar to the mobile retina of jumping spiders reviewed in Appendix A. Such a motor system could in principle allow for the execution of actions which induce self-similar stimulus transformations, although this is pure speculation since very little is known about the inner (non-calcite) structure of the trilobite eye and even less is known about the temporal dynamics of these animals.

With regard to research concerned with the morphogenesis of trilobite eyes, it shall be noted that the immediate reason for why trilobite eyes exhibit self-similar properties is with high probability to be found in a genetically coded lens-emplacement program [126]. According to this theory, eyelets are placed via a genetic program iteratively executed during ontogenetic development. This would mean, Trilobite eyelets do not adaptively organize themselves according to experienced stimuli. However, arguing in favor of self-similar stimulus transformations, it is of course possible that selective pressure during phylogenetic development might have preferred eyelet placement strategies which generate self-similar sensor structures because they provide an advantage for the animal.

Self-Similarity in Modern Camera-Type Visual Systems. While receptor topologies of compound eyes can be observed directly at the periphery of an animal, receptor density distributions in camera-type eyes require a closer look at the visual projection surface within the eye. Studies measuring the distribution of retinal ganglion cell layers revealed that there are profound differences between receptor density distributions in camera-type eyes of different species.

Primates and mammals with binocular vision typically feature a *fovea*. This means, they have a small high-resolution area in the center of their retina, and outside this area a in radial direction almost logarithmically decreasing receptor density. In [117], it is pointed out that such a *log-polar*-like receptor distribution corresponds to a mapping function which transforms image rotation and dilation (zoom) into simple coordinate shifts in the log-polar coordinate system. Thus, if an eye featuring such a receptor distribution is focusing on an object, and that object is rotated or scaled, then the object's projection on the eye is merely shifted along the log-polar coordinate axes. In other words, under such actions the image undergoes a transformation with high self-similarity properties with respect to a log-polar receptor distribution. In conclusion, it was argued in [117] that this property results in an

advantage for neural processing since the sensor topology could achieve image invariance for these transformations at a low computational cost by simply shifting the image.

Other mammals, for example sheep, pigs, and horses, feature a horizontally elongated ganglion cell distribution in their fovea called *visual streak* [44]. Such cell distributions are typically found in animals which live in open territory and have a predominantly monocular vision system. According to these observations, it has been suggested that an animal's environment and behavior have a direct influence on visual receptor distributions in camera-type eyes [50]. For example, in [49] the retinal ganglion cell distribution of two different kangaroo species have been compared. It was found that while the kangaroo *Dendrolagus doriana* which lives in the rain forest of New Guinea has a nearly radially uniform ganglion cell distribution, the red kangaroo *Macropus rufus* which lives in open Australian grassland and desert habitats shows a clear visual streak. For a qualitative reproduction of the results presented in [49], see Fig. 1.7. In general, it is assumed that retinal ganglion cell distributions with a visual streak could account for the fact that horizontal image translations are more frequently experienced by the host of such a system. Differently from the eyes of typical predators, animals with a visual streak have limited binocular vision and their behavior is less "object oriented". On the other hand, it is important for these species to observe the horizon, a behavior which induces horizontal image shifts. Thus, it might be assumed that the visual streak improves, on average, the self-similarity of experienced stimulus transformations.

A particularly curious ganglion cell distribution has been recorded in the eyes of African elephants. These animals combine the two previously described receptor topologies in one eye. They feature a visual streak pointing at the horizon, but additionally also have an area of high receptor concentration in the upper temporal area of their retina. Not surprisingly, when the visual streak is aligned with the horizon, the center of the Elephant's fovea points exactly at its trunk [123]. Considering that an elephant most of the time keeps its eye aligned with the horizon while using its trunk at the same time for dexterous object manipulation, its eye seems to reflect the characteristics of experienced stimulus translations which predominantly are: shifts along the horizon, and small movements around the tip of the trunk.

In conclusion, for many stimulus transformations the discussed camera-type eyes are clearly unable to achieve perfect self-similarity. However, their ganglion cell distributions seem to approximate self-similar transformations for a number of commonly experienced transformations in recorded visual stimuli. This means that, for certain types of actions, the error with respect to perfect self-similarity is smaller than for others. The next section reviews work which makes use of such an error measure to

(a) Red Kangaroo (*Megalia rufa*)   (b) Tree Kangaroo (*Dendrolagus doriana*)
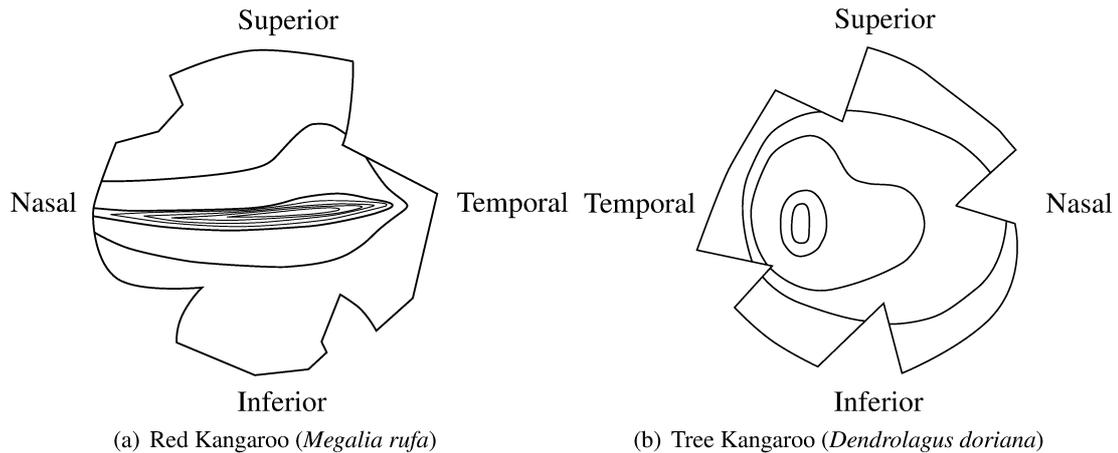
Figure 1.7: Behavior-dependent receptor distributions as found in camera-type eyes. The plots show a qualitative reproduction of retinal ganglion cell densities as measured by Hughes for two different kangaroo species [49]. The Red Kangaroo *Megalia rufa* (on the left) lives in open Australian grassland. Its retinal ganglion cell topography shows a clear *visual streak*. The Doria's tree Kangaroo *Dendrolagus doriana* (on the right) spends most of its life off the ground in treetops. Its retinal ganglion cell distribution is almost radially uniform. The two layouts clearly reflect the specific adaptations of visual sensors to terrestrial and arboreal life-styles.

optimize an abstract sensor topology according to a measure of self-similarity.

Self-Similarity, a General Principle for the Organization of Visual Systems? This section reviews previous work done by Clippingdale and Wilson [21]. Their investigations are of central importance to this thesis since they provide not only theoretical and empirical evidence that self-similarity can be used to structure simplified models of receptor distributions, but also that under appropriate sets of transformations the obtained distributions resemble receptor distributions found in camera-type eyes like the ones discussed in the previous paragraph.

Clippingdale and Wilson propose to organize abstract representations of sensor topologies by minimizing a measure of self-similarity under a given set of transformations. They show that a set of points, initially randomly distributed on a planar disk, converges to a stable configuration with a highly regular structure under the following rules:

- Points are conjointly transformed by rotation, dilation, and translation actions which are applied according to a given probability distribution.

- After each iteration points are moved towards transformed points lying closest.

Interestingly, it was found that under the described conditions, certain action probability distributions induce point distributions which resemble closely receptor topologies found in foveal sensor layouts

14

(a) Trans. h = 1.0, v = 1.0   (b) Trans. h = 0.19, v = 0.2   (c) Trans. h = 0.5, v = 0.2   (d) Trans. h = 0.0, v = 0.0
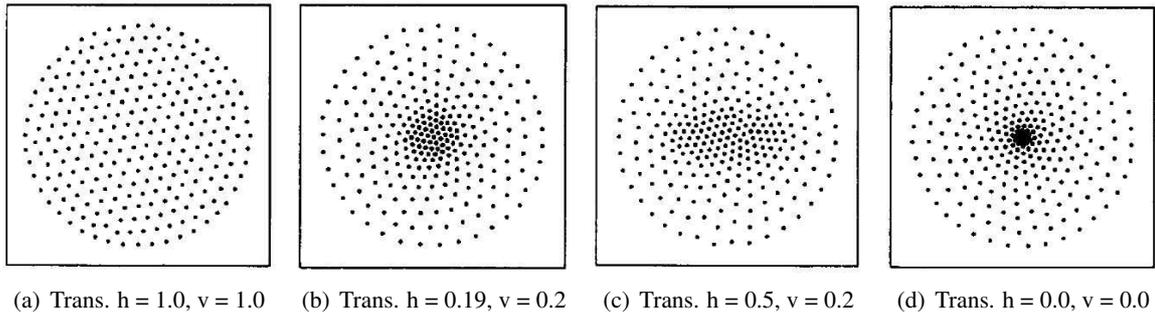
Figure 1.8: Geometrical point layouts obtained by Clippingdale and Wilson according to the algorithm described in [21]. The different plots show layouts obtained under different layout transformation action probability distributions. Actions are sampled from a uniform distribution of *arbitrary* rotation and dilation actions, and a uniform distribution of x- and y-translations of *limited* range. The different ranges of translation actions are denoted below each plot. Reproduction with kind permission of S. Clippingdale.

of camera-type visual systems. The results of Clippingdale and Wilson are shown in Fig. 1.8. The probability distributions of transformations which lead to the configurations shown in these figures are composed of rotation and dilation transformations uniformly distributed over an *arbitrary* range, combined with horizontal and vertical translations distributed over different *limited* ranges.[1]

In summary, Clippingdale and Wilson provide empirical support regarding the validity of the hypothesis that animals optimize their visual sensorimotor systems to achieve "closer-to-invariant" image transformations for certain sets of actions. Their work shows, that with a preference for spatiotemporal self-similarity, abstract sensor layouts can be synthesized which resemble receptor distributions found in biological organisms, and that these layouts are obtained under stimulus transformations which can be expected to be characteristic for the respective organisms. Referring back to feedforward sensorimotor maps discussed in Sect. 1.2, it is important to note that the measure for self-similarity, as introduced by Clippingdale and Wilson, directly relates to a request for simpler feedforward sensorimotor maps. It essentially favors (on average) exact receptor-to-receptor stimulus translations.

A number of restrictions related to the approach proposed by Clippingdale and Wilson are noteworthy in the context of this thesis. Firstly, in [21] it is assumed that there is knowledge about the sensor topology, meaning the algorithm has access to the spatial position of sensory elements (and size, in case of Gaussian sensor elements). Secondly, it is assumed that the displacement of sensor elements with respect to the sensor surface is known a priori from a given transformation; which

---

[1]In [21] Clippingdale and Wilson extended the same algorithm to work with 2-dimensional Gaussians instead of simple geometrical points. To do so, the Euclidean distance measure between two points was replaced with the inner product between two Gaussians in the self-similarity measure.

for a real visual sensor has to be reformulated as: it is assumed that the new locations of a previously recorded stimuli are known from a given transformation. Both assumptions are unrealistic when considering an autonomously developing organism or robot. Rather, it seems appropriate to take an intrinsic perspective of a developing system and to base self-organization of the sensorimotor apparatus on recorded stimuli and motor commands directly available to the system. From this point of view, it cannot be assumed that measurements concerning the spatial topology of the sensory system can be obtained. Also, it is improbable that information about the spatial displacement of visual receptors is readily available since motor commands and stimulus displacements are usually connected via complex transformations. In contrast, this thesis adopts a more natural approach where the sensor topology is considered unknown and visual stimuli are recorded with a realistic sensor model.

1.4   Approach of This Work

The goal of this work is to develop a method which allows for the self-organized synthesis of sensorimotor structures in artificial visual systems. Considering a co-developmental process, it is proposed that sensory and motor systems mutually influence each other such that the structure of a visual sensor organizes according to the characteristics of the given motor apparatus, and, vice versa, the motor system adapts to the structure of the associated sensor. As a common direction for such a joint development the present work proposes that sensor morphology and motor primitives should concurrently adapt such as to *simplify* the relationship between recorded sensory stimuli and executed motor actions. A first indication on how to qualify "simple" for sensorimotor relationships in visual systems has been given in Sect. 1.3.

In Fig. 1.9, a simplified sketch of a general sensorimotor system is shown. Dashed lines indicate components of the system which are initially unknown. Thus, both, the sensory system S, and the motor system M, are considered variables of the proposed problem. The two arrows between S and M indicate the two directions of possible sensorimotor transformations. The arrow denoted *behavioral policy B* represents the map which associates motor actions to experienced stimuli. This map is always considered to be known in this work. In the opposite direction, the arrow denoted *forward model P* represents the sensorimotor map which estimates sensory stimuli based on selected motor actions. This map is initially unknown and represents a variable of the problem which depends on S and M. Having proposed that a perceptual system with favorable properties possesses a simplified feedforward model, the structure of P is of central importance. Accordingly, a first goal of this work is to find a measure which evaluates the simplicity of P. Subsequently, the central idea is to adapt S and M such
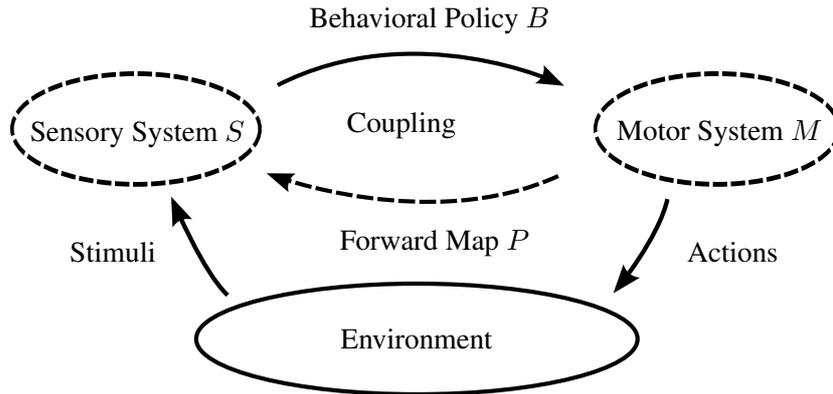
Figure 1.9: A general sensorimotor system with subsystems S and M which are coupled in two directions via sensorimotor maps B and P. Dashed lines denote initially unknown components of the system. The policy B associates actions to recorded stimuli and is considered to implement a given behavior. The sensorimotor system (S, M) and its sensorimotor transformation (P) are variables of the proposed problem which are defined under the assumptions specified later in Chap. 3.

that the previously proposed measure is optimized.

The problem outlined above is addressed in four steps. In a first part, an empirical study investigates feedforward sensorimotor relationships in visual systems with respect to accuracy and the complexity of the associated prediction models. It results a measure which is directly related to the concept of self-similarity as introduced in Sect. 1.3. In a second step, it is investigated how a motor system M should be organized such as to allow for simple forward models P considering a given sensor S and action selection policy B. The third part addresses the inverse problem: how should the sensor S be adapted such as to allow for simple forward models P given M and B? In a last step, part two and three are merged into a unified approach which conjointly develops sensorimotor structures S and M according to a given behavior B.

An important objective of this work is to keep an *intrinsic* perspective, meaning that a process structuring sensory and motor systems should use only information which can be expected to be accessible "from inside" a developing agent. This objective impedes for example the use of information describing the current structure of the sensorimotor apparatus, since from the perspective of a developing nervous system it is not realistic to assume that a developing neural circuit can access information about the physiology of the body it is embedded in. Thus, the principal information available from an intrinsic perspective are recorded sensory stimuli and issued motor commands. Consequently, the proclaimed goal is to adapt sensorimotor structures S and M solely according to experienced sensorimotor signals. In this respect, the organization of sensor topologies as proposed in this work considerably differs from the approach proposed by Clippingdale and Wilson (reviewed in the previous

section) which requires information not available from an intrinsic perspective.

At last, an important constraint which is imposed throughout all four parts of the present work is that, agents are considered to be stateless in the sense that changes in visual stimuli induced by motor actions are independent of the system's current position with respect to the environment. The remainder of this section provides an overview of the four principal parts of this thesis.

1. **A Measure of Visual Sensorimotor Coupling.** Work carried out in a first step addresses the influence of sensor morphology and sensor movements on the structure of an induced stimulus. The objective is to understand how actions and sensor structure interrelate when recording primary visual stimuli and to assess the quality of visual sensorimotor coupling in terms of predictability of visual stimuli and the complexity of the model required to predict stimulus changes. In this sense, pairings of sensor topologies and actions are preferred which firstly lead to a predictable change in sensory stimulus and secondly lead to a change in stimulus which can be described using a simple model. With these criteria an environment-independent measure is developed specifying "how simple" it is to predict the outcome of a given action under a given sensor topology. This measure is then used to evaluate two different sensor layouts using three different types of actions. By visual inspection of the obtained measurements, it is easily possible to detect good motor actions specific to a given sensor topology. Based on the assumption that actions leading to predictable sensor feedback and simpler prediction models are preferable, it is then possible to deduce how an agent's actions and behaviors are coupled with its sensor topology. It is found that the resulting conclusions match well with observations made for biological systems described in Sect. 1.3. This study has been published in [102].

2. **Organizing Motor Topologies.** In a second part, the question is addressed how to spatially organize a number of motor primitives M in a given motor space for a given sensor layout S in order to optimize a criterion similar to the one developed in the first part. The thesis addresses this questions by first proposing a biologically inspired model for an adaptive visual sensorimotor system capable of learning to predict visual stimuli from self-induced motor actions incorporating also structural adaptations to the underlying motor space. This model is then extended by releasing some previously specified constraints.

   First, a visual stimulus prediction circuit is considered where motor signals enter through motor movement fields modeled as multivariate Gaussians of variable size and position integrating input from the motor space. This model is inspired by a neural feedforward pathway found in primates which connects the superior colliculus with the frontal eye field, previously discussed

18

in Sect. 2. According to a functional understanding of visual stimulus prediction in the primate brain, a graphical interpretation of the addressed circuit is deduced first. Subsequently, a learning process optimizing stimulus predictability is introduced. This process adapts at the same time the feedforward circuit as well as the topological organization of a number of corollary discharge neurons or motor movement fields integrating input from the motor space. The obtained results demonstrate the adaptation of the introduced model with respect to different sensor topologies and sensor movement strategies. Two interesting and interrelated properties are observed: the circuit connecting motor and sensor areas converges to a particularly sparse configuration in terms of number of connections; and the geometry of the circuit's connection nodes shows strong self-similarity properties in the sense of [21]. This work has been published in [104].

In an extension, the constraint that motor movement fields must implement a Gaussian model is relaxed and a discretized representation allowing for arbitrary motor movement field functions is introduced instead. This work has been published in [103].

3. **Organizing Sensor Topologies.** Dual to the second part, this part investigates how to spatially organize a number of visual receptors S on a given sensor surface for a given behavior B in order to optimize a criteria similar to the one developed in the first part. The thesis addresses this question in two steps described next. Both consider spatially extended visual receptors which integrate luminance according to receptive field functions described in a discretized form allowing for arbitrary receptive field shapes.

First, it is investigated how prediction and spatiotemporal correlation can be exploited to develop visual sensors well-adapted to an artificial agent's interaction with its environment in the sense of good sensorimotor coupling as introduced in the first part. The result is a computational method for synthesizing visual sensor topologies according to experienced stimulus transformations. This method establishes a relation between a sensor's spatial layout and experienced stimulus transformations by adopting the basic principle of self-similarity as proposed by Clippingdale and Wilson in [21]. Though, instead of considering point-like sensor elements, a realistic visual sensor is simulated with spatially extended receptive fields. Different from [21], it is imposed that the algorithm has no access to information about the topological layout of the sensor being organized. This means, the organization of the sensor layout has to be achieved solely by observing the activation of an orderless array of visual receptors. Hence, the implementation of a rule similar to the one proposed in [21] becomes considerably more

challenging. In particular, the Euclidean distance measure between transformed and original points has to be replaced with a measure related to how activation is transported between visual receptors when the recorded stimulus changes. This issue is addressed by introducing a criterion based on spatiotemporal cross-correlation of receptor activation. This criterion allows for the implementation of an optimization which organizes the layout of visual receptors depending on sensorimotor activity. At the same time, the algorithm is also required to find a suitable shape for the receptive fields of visual receptors. This approach shows that spatially coherent receptive fields can evolve driven only by the predominantly low spatial frequency of natural images [13]. By rewarding spatial correlation within receptive fields, smoothly overlapping clusters organize on the sensor surface without any further constraint on the spatial shape of a receptor's integration area. This work has been published in [105]. Preliminary work appeared in [101].

In an adaptation to the first method for synthesizing visual sensors, the request for self-similar structures is combined with a request for accurate representation of available visual signals. Similar to the previous method, it is proposed that the perceptual system should develop a sensor layout such that the relationship between past and future stimuli is simplified on average in the sense of self-similarity, but additionally, the system should also optimize available resources to accurately perceive an observed phenomena. This additional request for a more accurate representation is accommodated by introducing an error between the available signal and its reconstruction from the recorded stimulus. This work has been published in [103].

4. **Organizing Sensor and Motor Topologies Conjointly.** In a final step, part (2) and (3) are combined into a unified method which allows for the organization of the topology of sensor and motor spaces S and M conjointly according to agent-environment interaction patterns resulting from a given behavior B.

The presented method takes as input experienced efferent (motor commands) and afferent (stimuli) signals, and evolves a spatial layout for light receptive fields and motor movement fields. It is shown that visual receptive fields and motor movement fields can evolve simultaneously when minimizing a simple error measure – assembled from part (2) and (3) – which contemplates the reconstruction error for recorded stimuli with respect to given input signals and the prediction error for stimuli resulting from self-initiated actions. As previously found, the low spatial frequency of natural images induces the development of spatially coherent and smoothly overlapping receptive fields on the sensor side without any further constraint on spatial shape.

At the same time on the motor side, individual movement fields evolve such as to displace the sensor ensuring high temporal coherence of visual stimuli. The presented results demonstrate how the proposed principles can be used to develop sensory and motor systems with favorable mutual interdependencies.

1.5 Summary of Contributions

The present thesis introduces an approach to develop sensorimotor structures of artificial visual systems in a self-organized manner. In a first step, it is proposed that the quality of a sensorimotor system is related to the simplicity of the relationship between sensors and actuators. Subsequently, it is shown that by optimizing a criterion related to this property, it is possible to conjointly structure the layout of sensory and motor systems according to a given higher level behavior. In summary, the principal contributions are:

1. Development of a criterion assessing the characteristics of sensorimotor relationships in visual systems. The proposed criterion is related to the amount of information required to describe stimulus transformations induced by motor actions executed by the considered system.

2. Topological organization of visual motor primitives according to the proposed criterion for a given sensor layout and behavior. Resulting motor primitives are optimized with respect to the physical structure of the given sensor in order to induce simple signal transformations.

3. Topological organization of visual receptive fields according to the proposed criterion for a given motor layout and behavior. Resulting visual receptors are optimized to allow for simple visual stimulus transformations under a given set of motor actions.

4. Given a desired higher level behavior, conjointly organize motor primitives and visual receptors. It is demonstrated that sensor and motor topologies which are tuned to the characteristics of the system's interaction with its environment can be synthesized solely driven by experienced efferent and afferent signals.

Chapter 2

Stimulus Prediction in Biological Systems

Contents

This chapter provides an overview of neuroscientific studies and results related to neural circuits involved in visual stimulus prediction and (forward) sensorimotor mapping. In an introductory part, a number of commonly used terms are introduced and an important concept, the *reafference principle*, is reviewed. In addition, the last paragraph of the introduction reviews the *superior colliculus*, a well studied brain area where visual sensory stimuli are topographically mapped onto underlying motor layers which control eye and body orientation. From a biological perspective, this area is of particular interest, since it can be considered a biological prototype for the sensorimotor mapping addressed in this work. Subsequently, the two principal sections of this chapter review a number of neural circuits dedicated to stimulus prediction. Sect. 2.1, reviews stimulus prediction at different levels along the visual pathway, and Sect. 2.2 discusses neuroscientific findings related to adaptive processes in visual neural circuits.

Inspired by the results reviewed in Sect. 2.1, this chapter also proposes a simplified model of a visual feedforward circuit. This model is visualized in Fig. 2.3 and is later revisited and integrated as a part of the approach proposed for the organization of sensorimotor topologies described in Chap. 4.

Common Terminology in Neuroscience.   In neuroscientific terms, *afferent* signals are stimuli coming from the periphery of the central nervous system, *efferent* signals are typically motor related signals travelling towards the periphery of the nervous system. Depending on their origin, afferent signals can
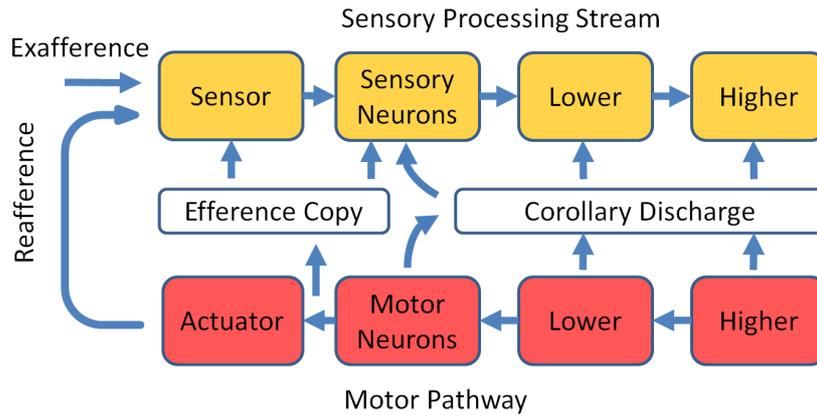
Figure 2.1: Efference copy and corollary discharge circuits along the sensory processing stream and motor pathway. Adapted from [25].

be further decomposed into a component of the signal which is generated by events external to the organism (*exafferent*), and a component which is induced by self-initiated movements (*reafference*).

The Reafference Principle. In the middle of the last century, von Holst and Mittelstaedt formulated the *reafference principle* which proposes that it is an important property of nervous systems to be able to distinguish reafference from exafference [134]. It was suggested that a copy of a motor command, the *efference copy (EC)*, is used to distinguish the reafferent part of the signal from the exafferent part. At the same time, similar conclusions led Sperry to coin the term *corollary discharge* (CD) in his work [121]. Nowadays, despite the conceptual similarity between EC and CD, the commonly used terminology makes a slight difference between the two. CD is in general used to refer to signals which are transmitted along feedforward connections from the motor pathway to the sensory processing stream. These circuits can connect from any tier of the motor pathway to any other tier in the sensory processing stream. EC on the other hand is considered to be a motor signal affecting sensory channels close to the effector / sensor periphery. Fig. 2.1 illustrates these different levels of feedforward connections. For further reading on corollary discharge circuits including this definition and an attempt to classify them, see [25, 26].

The Superior Colliculus. The optic tectum, or in mammals the superior colliculus (SC), is a phylogentically ancient structure in the vertebrate midbrain, often studied in relation to eye and head movements triggered by visual stimuli, but also known to be responsible for full body reorientation as a response to multimodal sensory input [33, 43]. The tectum is organized as a layered structure where the superficial layers receive somatosensory input while the deeper layers are concerned with attention

24

orienting movements. Depending on the somatosensory system of an animal, the sensory modalities processed by the tectum may be balanced differently. In diurnally active mammals relying mainly on vision as one of their principal senses, the superficial layers of the superior colliculus receive primarily visual input. These optic input layers are retinotopically organized, meaning the topological map of the recording sensor (the retina) is still present in the superior colliculus. This topology is preserved through several layers and as far as down to the deeper motor areas. As a result, neural activity in the motor layers of the SC code eye saccades in a gaze-related retinotopic reference frame. This layout of movement fields in the SC was first revealed in studies done by [100]. Using microstimulation the spatial layout of the motor map was deduced in terms of relative gaze orientation angles. It was found that when stimulating locations in the motor layers of the SC, *relative* saccades are triggered, no matter where the eye was in the orbit previously to the saccade [111, 100]. Another influential study later presented further evidence which supports the so called *population coding theory* for motor layers in the SC [64]. According to this theory, a blob-like activation of a number of neurons in the motor layer triggers a saccade to a target location which is encoded as the vector sum of all active motor neurons.[1] This theory on the working of motor layers is revisited later when formulating the definition for motor movement fields as used in this work, see Chap. 3.

In conclusion, the optic tectum is of interest to this work in the sense of a biological prototype since it appears to implement a general design for a forward sensorimotor map applicable to different input modalities, generating output activity used for a great variety of behaviors.

2.1 Neural Circuits That Predict Sensory Stimuli

Biological systems predict stimuli at a multitude of different levels. Neural structures adapted to the spatiotemporal properties of visual signals appear for example as early as in the ganglion cell layer of the vertebrate retina. Also, at later processing stages, corollary discharge circuits predicting exafferent stimuli based on efferent signals have been found to be ubiquitous in neural tissue.

This section first provides an overview of neural circuits concerned with prediction by following a classification of corollary discharge circuits proposed in [25]. Subsequently, two particular neural circuits are reviewed in more detail. The first is directly implemented in the vertebrate eye, just after recording the visual signal. It provides a statistical estimate of the spatiotemporal characteristics of visual stimuli at a very first processing stage. The second is found in primates located between the

---

[1]Note, to eventually move the eyes, retinotopic motor signals have to undergo a non-trivial transformation while travelling from the SC to the oculomotor nuclei [71].

superior colliculus and the frontal eye field. It is used to predict visual stimuli based on oculomotor signals. For the latter, this section also proposes a simplified model which is later instantiated in Sect. 4.5.

Corollary Discharge Circuits at Different Levels.   On a first level, corollary discharge circuits are used to distinguish between external signals (with origin in the environment) and signals induced by self-initiated actions. Male crickets for example filter their self-produced bursts of sound by generating a neural signal which anticipates the auditory stimulus [135, 93]. By doing so, the animal is able to suppress its own chirping while focusing on the response of female crickets. The same filtering strategy has been discovered in a number of other species which elicit escape reactions depending on whether a sensory signal is self-generated or results from an event in the environment [28, 32].

On a higher level, stimulus prediction is believed to be a basic mechanism used to achieve stable perception. For any organism, the acquisition of a coherent percept of the environment is not a passive one-step action but is the result of a continuous process of sensorimotor interactions which take place over a number of iterations. Rats for example explore objects by tactile whisking; bats "see" the world by listening to the echo of their self-produced ultrasonic waves, and many animals relying on vision continuously have to move their eyes and body to sweep their visual field over an observed scene. The motor actions involved in these exploration movements often induce drastic changes in sensory stimuli which can be understood when trying to focus for example on the image stream recorded by an abruptly moved camera. The question arises, how is the brain able to assemble a stable and coherent percept in light of such radical stimulus changes? Clearly, this task becomes considerably easier if information about the planned sensor repositioning action is assumed to be available. Given the appropriate neural circuitry, the expected sensory stimuli could be predicted and used to relate sensory input over a sequence of sensor movements. Neuroscientific studies on the neural mechanisms underlying coherent perception in rats (tactile whisking) and bats (echolocation) have been described by [2, 72].

From a dynamical point of view, stimulus prediction is also important for fast action sequences. If a signal providing sensory feedback for a previous action reaches the motor system only *after* the next action has to be initiated, then prediction is indispensable to plan accurate motor commands. This is for example the case for a sequence of fast eye movements (saccades). Studies with primates showed that the brain actually relies on a predictive control strategy: motor commands for subsequent saccades are issued before proprioceptive or visual sensory feedback from the previous movement is available to the motor system [129, 69, 132, 94, 11]. Furthermore, considering that primates execute

up to three visual saccades per second during normal behavior and knowing that neurons in the frontal and parietal visual areas of the primate brain record afferent signals with a latency of at least 60 ms, relying on purely passive afferent signals would mean that the cortical representation of the visual world is inaccurate during almost 20% of the time [17, 40, 97]. Such delay and accuracy is likely to be compensated by neural feedforward pathways like the one discussed in the next paragraph.

Early Predictive Circuits in Retinal Cell Layers. From early studies of vertebrate and insect visual systems, it is known that visual receptive fields which feature antagonistic center and surround areas play an important role at the first levels of stimulus processing. [2] In a number of influential studies described by Srinivasan, Laughlin, Dubs, and Barlow, it was proposed that bipolar and ganglion cells in the vertebrate retina implement a form of spatiotemporal stimulus prediction [122, 6]. In a framework termed *predictive coding theory* it was suggested that the (negative) surround area of such a cell's receptive field generates a statistical estimate of the signal expected at a certain point in space or time. This prediction is then subtracted from the signal that was actually recorded at the (positive) central area of the receptive field. A number of advantages resulting from such a mechanism were proposed, among them e.g. deblurring or edge enhancement [122]. In addition it was proposed in [122] that local, statistical prediction mechanisms might remove redundant signal components and thus contribute to a more efficient coding of visual signals. Referred to as the *efficient coding theory*, the latter proposal has subsequently been investigated over several decades, and remains an active research topic in neuroscience [131, 5, 91]. Recent studies suggest that the receptive fields of ganglion cells in the vertebrate retina are highly adaptive with respect to the spatiotemporal characteristics of observed stimuli, see also Sect. 2.2.

An Exemplary Corollary Discharge Circuit: the SC-MD-FEF Pathway. An important and well studied neural feedforward circuit is the SC-MD-FEF pathway in the primate cortex leading from the superior colliculus (SC) via the medial dorsal nucleus (MD) to the frontal eye field (FEF). This circuit is responsible for visual stimulus prediction during eye saccades.

Input to the SC-MD-FEF pathway is motor activation generated by the deep layers of the superior colliculus. These motor signals ascend the SC-MD-FEF pathway as corollary discharge in a feedforward direction and eventually reach the frontal eye field, see also Fig. 2.2. There, these signals are integrated with visual signals which reach the FEF through the main sensory processing stream. A prediction of future visual stimuli is achieved by modulating the receptive fields of visual neurons

---

[2]The integration function of receptive fields with antagonistic center and surround areas can for example be approximated by the second derivative of a multivariate Gaussian.
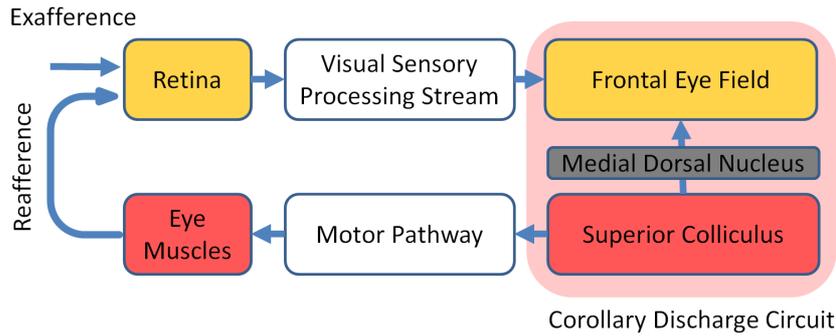
Figure 2.2: The SC-MD-FEF corollary discharge circuit connecting the superior colliculus (SC) via the medial dorsal nucleus (MD) with the frontal eye field (FEF). Note the inverted (feedforward) direction with respect to the motor pathway and sensory processing stream.

depending on the corollary discharge arriving from the superior colliculus.[3] That is, when a saccade is executed, the RFs of the visual neurons in the FEF are modulated to integrate visual signals from the target location of the saccade. The shifted integration profile of such a modulated visual neuron is then called *future field (FF)*[120]. Consequently, the presaccadic FF and the postsaccadic RF sample the same absolute location in visual space. Comparison of presaccadic and postsaccadic FEF neuron activation can therefore in principle be used for both, stabilization purposes and to distinguish exafferent from reafferent stimuli (filtering). This hypothesis that neurons with shifting receptive fields are able to perform comparative operations is supported by a number of studies, among them [31, 23, 119]. Work describing the SC-MD-FEF pathway in general can be found in [118, 119, 120, 129, 130, 74]. Other areas in the visual system where neurons with shifting RFs have been found include the lateral intraparietal sulcus (LIP) described in [10, 22, 31], and extrastriate visual areas like V4 discussed in [76, 128].

In Fig. 2.3, a graphical interpretation of the topological and functional relationships reviewed here is introduced. Signals traveling along the SC-MD-FEF pathway originate from a peak of activation in a layer of SC motor neurons coding eye movements in a retinotopic reference frame which is denoted (a) in Fig. 2.3. Along the corollary discharge pathway, this activation is integrated by corollary discharge neurons (CDNs, (b) in Fig. 2.3). The CDNs project in turn through feedforward connections (c), to visual neurons and their connections (d). We will resume on this interpretation when proposing the computational model of the SC-MD-FEF circuit in the next section.

---

[3]Neurons with this type of variable receptive fields are said to have *shifting* receptive fields. In general, the receptive field of a neuron is considered to be spatially fixed with respect to the underlying input neurons. However, neuroscientific studies showed that in several areas in the visual system there are neurons which feature shifting receptive fields.
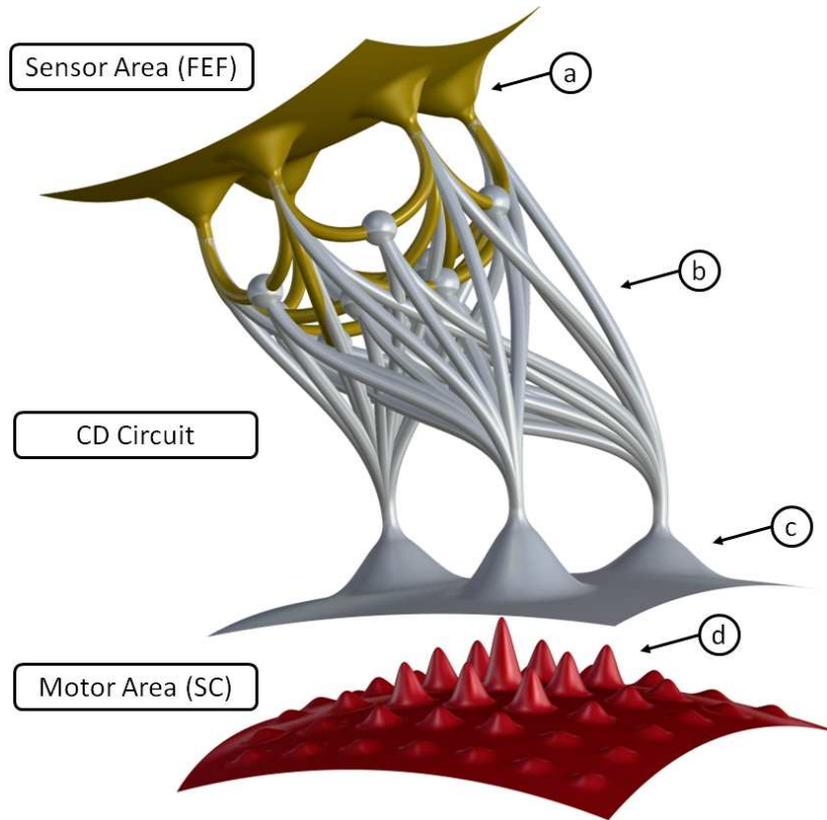
Figure 2.3: Model of a visual corollary discharge circuit. A population of motor neurons codes visual saccades in a retinotopic reference frame (d). An intermediate layer of corollary discharge neurons (c) collects activation from the underlying motor layer and projects through feedforward connections (b) to the sensor area (a). The corollary discharge signals modulate the activation of visual receptive fields and their connections such as to predict a future visual stimulus resulting from an activation in (a).

2.2 Plasticity in Visual Neural Circuits

Adaptations in visual neural circuits have been studied since the early days of modern neuroscience. As an inspiration for the structural optimization of artificial visual sensorimotor circuits developed in Chap. 5, this section reviews a number of studies providing long-standing evidence for such adaptive processes in living organisms. Some of these adaptive processes take place during the development of an organism and result in subsequently stable structures, others remain dynamic and occur even at the timescale of seconds. The section concludes by revising the degrees of freedom of the model presented in Fig. 2.3 with respect to an adaptive process which optimizes the prediction of visual stimuli.

Early influential work by Hubel and Wiesel showed that ontogenetic development of the visual areas V1–V5 involves a high degree of adaptivity in higher vertebrates [46, 47]. Based on single-unit recordings, the formation of orientation sensitive cells in striate and extrastriate cortices could be ob-

29

served, clearly influenced by experienced stimuli [48]. Since then, a great body of work confirmed that plasticity is present along the entire visual sensory processing stream, for an overview see e.g. [90]. Moreover, plasticity also plays a major role in the superior colliculus. In [53], King et al. describe for example how visual and auditory maps in the SC are topographically aligned during early life. And in recent work, the alterations of the visual map in the superior colliculus due to lacking neuronal activity in the early retina have been studied [73]. It was found that without previous visual stimuli, the projection layers in the superior colliculus are a coarse retinotopic map given by morphogenetic development. Subsequently, during growth, the spatially correlated firing of retinal ganglion cells refines the organization of the retinotopic layers in the superior colliculus. This organization is defective if natural neural activity due to lacking or disturbed visual input is not present after birth. Hence, like in the striate and extrastriate cortices, the fine tuning of the superior colliculus clearly depends on sensorimotor contingencies experienced while the animal interacts with its environment. This form of reafference exploration is a commonly found learning strategy in nature [113]. An equivalent methodology termed *motor babbling* can be applied in artificial embodied systems

Adaptive processes also occur at the level of early predictive circuits implemented in the vertebrate retina. It was recently found that ganglion cells adapt their receptive fields within a timescale of seconds to the statistical properties of recorded stimuli [45]. This remarkable plasticity has been confirmed with respect to contrast and intensity but is also clearly related to the spatiotemporal characteristic of observed patterns. It could be shown that the ganglion-cell receptive fields change dynamically depending on the correlation structure of the visual environment. The result of the adaptation is that predictable stimulus features are suppressed compared with novel features. Also it is conjectured that ganglion cells might contribute to an early decorrelation of visual stimuli [91].

The circuit proposed in Fig. 2.3 can learn to predict visual stimuli and minimize the prediction error by adapting its network topology in two different ways. Firstly, similar to the adaptation of receptive fields reviewed above for ganglion cells, corollary discharge neurons can adapt the shape and position of their receptive fields (b) with respect to the underlying motor layer; and secondly, feedforward connections (c) can alter their amount of discharge. A solution approach where these adaptation mechanisms will be implemented is described in Sect. 4.5.

Chapter 3

Problem Formulation

Contents

This chapter formulates the problem addressed in this work. A brief overview of the considered artificial system is given first. Subsequently, Sect. 3.2 provides the reader with a dictionary of the most important symbols used to describe the problem. An in depth formulation of the considered visual sensorimotor system follows in Sect. 3.3, Sect. 3.4, and Sect. 3.5. In Sect. 3.6 a general form of a forward sensorimotor map for visual stimulus prediction is introduced. Eventually, Sect. 3.7, identifies the input to the considered problem, and Sect. 3.8 outlines a general criterion for the organization of sensorimotor structures. The chapter is concluded by a summary of assumptions and the principle questions addressed in this work.

31

3.1 Overview

An artificial agent is considered consisting of a sensor S, composed of an array of light sensitive receptors s, and a motor layer M, composed of a number of motor primitives m. The agent lives in a static environment E and works as a closed sensorimotor loop which is found at each timestep in a particular state x. It observes the stimulation of its receptor array and takes actions by activating its motor primitives depending on the recorded stimulus. To choose actions based on recorded stimuli the agent possesses a given behavioral policy B which associates to any given stimulus an action. Thus, for a sequence of discrete time steps, the agent records at each time step a stimulus as the activation of its receptors $\mathbf{o}$ and triggers an action by activating its motor primitives via weights $\mathbf{a}$. Each selected action can induce a change in the observed stimulus which leads to the selection of a new action in the next time step.

3.2 Nomenclature

Before the considered sensorimotor system is detailed further in the next section, this section provides an overview of the most important symbols used throughout this work. All symbols are explained ordered by topic with references to their definition in the text where applicable. This section may be skipped by the casual reader.

Agent State & Environment

E **Environment.** A given world in which the agent lives. The world is assumed to be static (Assumption (1) in Sect. 3.9).

x **Agent State.** The agent's current state from an external perspective. A state can consist e.g. of a global position, orientation or velocity of the agent. A state x fully describes the agent's situation within the environment E. The state assumed at the next timestep is denoted $x + 1$. (Sect. 3.5).

$\mathcal{X}$ **State Space.** The agent's state space containing the set of possible states $x \in \mathcal{X}$. A more in depth description of an external perspective on the considered agent and its environment is given in Sect. 3.5.

Sensory System

Φ_s **Sensor Area.** A physical area of the agent's body in which signals are recorded. This area is d_s-dimensional and has a given parametrization in $\mathbb{R}^{d_s}$. It represents the domain on which visual signals $i \in \mathcal{I}$ are defined (Sect. 3.3, Embodiment, Fig. 3.2(a)).

d_s **Dimension of the Sensor Area.** The dimensionality of the physical structure of the sensor recording a signal. While d_s is typically considered to equal 2 for a 2-dimensional surface onto which an image of the environment is projected, other values for d_s could be assumed. For example a 1-dimensional sensor can be imagined, see e.g. the sea snail Oxygyrus described in Appendix A. Also $d_s = 3$ might be desirable to model sensors with a number of light sensitive cell layers located on top of each other, as featured for example by the visual system of jumping spiders, Appendix A (Sect. 3.3, Embodiment).

i **Visual Signal.** A function $i : \Phi_s \to \mathbb{R}$, which assigns to each location y in Φ_s a luminance value. A function i can be thought of as an image projected onto the visual sensor under consideration, see also Sect. 3.4 and Fig. 3.3(a). For example, in a camera-type eye, i represents an activation profile of the retinal surface – i.e. a projected (gray scale) image. Analogously, in a digital camera, i represents an activation of the image sensor. In this work the range of visual signals i is considered to be $[0, 1]$ if not otherwise specified (Sect. 3.4).

$\mathcal{I}$ **Sensor Space.** The function space describing a set of possible visual signals $i \in \mathcal{I}$.

y **Location in the Sensor Space.** Denotes a point in the sensor area Φ_s.

s **Receptive Field.** A function $s : \Phi_s \to \mathbb{R}$ describing the profile according to which a visual receptor integrates input from the sensor area. The function s assigns to each location y in Φ_s a weight determining how much a visual signal $i(y)$ contributes to the value recorded by a visual receptor with receptive field s (Eq. (3.1), Eq. (3.2)).

n_s **Number of Visual Receptors.** A given number of visual receptors, each described by a receptive field function s (Assumption (6) in Sect. 3.9).

S **Sensor Topology.** A set of n_s visual receptive fields s describing the topology of the visual sensor. The topology S is the principal variable of interest on the sensor side of

the sensorimotor system considered in this work. Any instance $S = \{s_1, s_2, \ldots, s_{n_s}\}$ is a valid sensor layout (Eq. (3.1)).

$\mathcal{S}$ **Sensor Topology Space.** A n_s-dimensional function space describing a set of possible sensor topologies S. An element $S \in \mathcal{S}$ is a set of n_s functions, where each function $s : \Phi_s \to \mathbb{R}$ describes a visual receptive field.

N_s **Resolution of a Discretized Domain for Visual Signals.** The number of discrete elements when considering a discretized domain for visual signals $i \in \mathcal{I}$. A discretization of the domain of visual signals is used to represent otherwise continuous functions i or s on a computer (Sect. 4.1).

$\mathbf{s}$ **Discretized Receptive Field.** A discretized representation of s. In this representation, a receptive field s is described as a real-valued vector of size N_s (Sect. 4.1).

$\mathbf{S}$ **Discretized Sensor Topology.** A discretized representation of S. In this representation, a sensor topology S is described as a real-valued matrix $\mathbf{S} = [\mathbf{s}_1{}^\top; \mathbf{s}_2{}^\top; \ldots; \mathbf{s}_{n_s}{}^\top]$ of size $n_s \times N_s$ where each row represents a receptive field (Sect. 4.1, Eq. (4.1)).

$\mathbf{i}$ **Discretized Visual Signal.** A discretized representation of i. In this representation, a visual signal i is described as a real-valued vector of size N_s (Sect. 4.1, Eq. (4.1)).

$\mathbf{I}$ **Batch of Discretized Visual Signals.** A batch of visual signals where each column of the matrix $\mathbf{I}$ describes a visual signal $\mathbf{i}$ (Sect. 4.1, Eq. (4.3)).

$\mathbf{o}$ **Visual Stimulus.** A visual stimulus recorded by the visual sensor through a given sensor topology S. A stimulus $\mathbf{o}$ is a real-valued vector of size n_s and entries $o_j \geq 0$ (Sect. 3.3 Internal Perspective, Sect. 3.4, Eq. (3.1)).

$\mathbf{O}$ **Batch of Visual Stimuli.** A batch of visual stimuli where each column of the matrix $\mathbf{O}$ describes a visual stimulus $\mathbf{o}$ (Sect. 4.1, Eq. (4.3)).

$\mathcal{O}$ **Stimulus Space.** A set of possible visual stimuli $\mathbf{o} \in \mathcal{O}$ recorded by a visual sensor through a given sensor topology S (Sect. 3.5).

ι **Image Projection Function.** Returns the resulting visual signal i in a given state x. (Sect. 3.5 Fig. 3.4).

Motor System

Φ_m **Motor Area.** A physical area in the agent's body in which motor signals are encoded. This area is a topological map of the d_m-dimensional domain of the functions $q \in \mathcal{Q}$ and has a parametrization in $\mathbb{R}^{d_m}$ which is assumed to be given (Sect. 3.3, Embodiment, Fig. 3.2(b)).

d_m **Dimension of the Motor Area.** The dimensionality of the domain of the functions $q \in \mathcal{Q}$, or in other words, the dimensionality of a motor command of the considered agent. The value of d_m equals the number of actuators or degrees of freedom available to the agent (Sect. 3.3, Embodiment).

q **Motor Signal.** A function $q : \Phi_m \to \mathbb{R}$ returning the activation for each location in the motor area Φ_m. A function q can be thought of as an activation profile according to which the motor system of the agent selects a particular action. For a more in depth description of the action selection process (Sect. 3.4, Eq. (3.3), Fig. 3.3(b)).

$\mathcal{Q}$ **Motor Space.** A function space describing a set of possible motor signals $q \in \mathcal{Q}$.

w **Location in the Motor Space.** Denotes a point in the domain of the functions $q \in \mathcal{Q}$. Because of the given mapping of the domain of the functions $q \in \mathcal{Q}$ and the motor area Φ_m, a location w also denotes a point in Φ_m.

m **Movement Field.** A function $m : \Phi_m \to \mathbb{R}$ describing the influence area of a motor primitive. In an efferent direction (triggering motor signals), a function m is defined as a probability density function which defines for each location w in Φ_m a probability of activation, see also Eq. (3.3). In an afferent direction (routing corollary discharge), the same function m assigns to each motor space coordinate w in Φ_m a weight determining how much a motor signal $q(w)$ contributes to the value recorded by the motor primitive (Eq. (3.1), Eq. (3.7)).

n_m **Number of Motor Primitives.** A given number of motor primitives, each described by a movement field function m (Assumption (7) in Sect. 3.9).

M **Motor Topology.** A set of n_m movement fields m describing the topology of the motor system. The topology M is the principal variable of interest on the motor side of the sensorimotor system considered in this work. Any instance $M = \{m_1, m_2, \ldots, m_{n_m}\}$ is a valid motor layout (Eq. (3.4)).

$\mathcal{M}$ **Motor Topology Space.** A n_m-dimensional function space describing a set of possible motor topologies M. An element $M \in \mathcal{M}$ is a set of n_m functions, where each function $m : \Phi_m \to \mathbb{R}$ describes a motor movement field.

N_m **Resolution of a Discretized Domain for Motor Signals.** The number of discrete elements when considering a discretized domain for motor signals $q \in \mathcal{Q}$. A discretization of the domain of motor signals is used to represent otherwise continuous functions q or m on a computer (Sect. 4.1).

$\mathbf{m}$ **Discretized Movement Field.** A discretized representation of m. In this representation, a movement field m is represented as a real-valued vector of size N_m (Sect. 4.1).

$\mathbf{M}$ **Discretized Motor Topology.** A discretized representation of M. In this representation, a motor topology M is described as a real-valued matrix $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_{n_m}]$ of size $N_m \times n_m$ where each column represents a movement field (Sect. 4.1, Eq. (4.1)).

$\mathbf{q}$ **Discretized Motor Signals.** A discretized representation of q. In this representation, a motor signal q is described as a real-valued vector of size N_m (Sect. 4.1, Eq. (4.1)).

$\mathbf{Q}$ **Batch of Discretized Motor Signals.** A batch of motor signals where each column of the matrix $\mathbf{Q}$ describes a motor signal $\mathbf{q}$.

$\mathbf{a}$ **Motor Activation.** A motor command triggered by the agent on top of a given motor topology M, see also Eq. (3.4). An activation $\mathbf{a}$ is a real-valued vector of size n_m and entries $a_k \geq 0$ describing a weighted activation of all n_m movement fields m. Because $\mathbf{a}$ is a weight vector it holds $\sum_{k=1}^{n_m} a_k = 1$ (Sect. 3.3 Internal Perspective, Sect. 3.4, Eq. (3.4)).

$\mathcal{A}$ **Motor Activation Space.** A set of possible motor activations $\mathbf{a} \in \mathcal{A}$ issued by an agent through a motor topology M. Due to the constraint $\sum_{k=1}^{n_m} a_k = 1$ on motor activations $\mathbf{a}$, the space $\mathcal{A}$ is a n_m-simplex (Sect. 3.5).

χ **State Transition Function.** Returns the resulting state $x + 1$ when taking an action q in a given state x. (Sect. 3.5, Fig. 3.4).

Prediction

P **General Stimulus Prediction Model.** Notation for a general stimulus prediction model without specified representation.

p **General Stimulus Prediction Function.** A prediction function to predict future stimuli $\mathbf{o}_{x+1}$ from actions q and current stimuli $\mathbf{o}_x$. Refers to a prediction function for a specific action q when written as p_q (Sect. 3.6).

$\mathbf{P}$ **Linear Prediction Operator.** A linear prediction operator which predicts a future stimulus $\mathbf{o}_{x+1}$ from a current stimulus $\mathbf{o}_x$ for a specific action q (Sect. 4.2).

Behavior

B **Behavioral Policy.** A given rule how the agent selects motor activations $\mathbf{a}$ given sensory stimuli $\mathbf{o}_x$.

C **Sensorimotor Experience Sampling Function.** A function which samples actions q and experienced signals i over the lifetime of the agent. The function returns an unordered set of triplets $\{(i_0, i_1, q)\} = C(B, E)$ with visual signals i_0 and i_1 recorded before and after taking an action q (Sect. 3.7, Eq. (3.10)).

$\mathcal{B}$ **Overall Sensorimotor Experience.** A characteristic of the agent's behavior defined as $\mathcal{B} = C(B, E)$. The sensorimotor experience $\mathcal{B}$ serves as input to the problem formulated in this work (Sect. 3.7, Eq. (3.10)).

Special Operators

ρ **Dimension Reduction Operator.** An operator which defines how fields s or m transform input to a single output value (Sect. 3.4).

ϵ **Dimension Expansion Operator.** An operator which defines how fields s or m can project their activation value back to the original input domain (Sect. 3.4).

Sensor–Motor Analogies

When considering a sensorimotor systems described by the above given dictionary, a number of inherent analogies and dualities exist between sensory and motor subsystems. These are briefly summarized in what follows. The comparison lists for each symbol its counterpart. To emphasize the relationship

between analogous or dual properties, a slightly more abstract terminology is used. (For a more compact overview, analogies between discretized representations are omitted.)

i, q Signals emitted by (q), or available to (i) the agent. These signals are independent of a subsequent layer of sensorimotor structures S and M.

$\mathcal{I}, \mathcal{Q}$ Function spaces describing the set of possible peripheral signals.

d_s, d_m The dimension of the domain of peripheral signals. On the sensor side, d_s denotes the dimensionality of the physical structure recording the signal i. On the motor side, d_m denotes the degrees of freedom of the considered agent.

y, w A location in the domain of a peripheral signal. A point in a topologically organized sensor / motor domain.

s, m A description of the morphology of sensorimotor primitives. Both, s and m describe the influence area of a sensory or motor primitive.

n_s, n_m The number of sensory and motor primitives. For a well balanced sensorimotor system, it can be conjectured that n_s and n_m should be of the same order of magnitude for similar dimensions d_s and d_m.

S, M A description of the morphology of sensorimotor structures. Both, S and M describe the topology of a sensory or motor structure.

$\mathcal{S}, \mathcal{M}$ Function spaces describing the set of possible sensorimotor structures.

$\mathbf{o}, \mathbf{a}$ Signals recorded or emitted through a layer of sensorimotor structures S and M. These signals are internal to the agent and their effect, or meaning, is defined only in conjunction with topologies S and M.

$\mathcal{O}, \mathcal{A}$ Sets of possible signals recorded or emitted by the agent.

ρ, ϵ Operators which define how signals traverse sensorimotor layers. The two operators are dual in the sense that ρ defines an operation on afferent signals and ϵ defines an operation on efferent signals.

3.3 An Adaptive Visual Sensorimotor System

This section discusses in depth the principal aspects and components of the considered sensorimotor system. In particular, it also provides a perspective on the relationship of sensorimotor structures S, M and P with respect to biological prototypes. A schematic overview is given in Fig. 3.1.

Embodiment. The agent's body physically implements sensory and motor systems in areas Φ_s and Φ_m. These areas are assumed to be given and represent the physical space where sensory and motor signals are recorded and encoded. In a biological system, Φ_s corresponds for example to the surface of a retina recording a projection of the environment. Similarly on the motor side, Φ_m corresponds to a layer or volume of neural tissue where each location represents a particular motor signal. Hence, considering a d_s-dimensional sensor, the sensor area Φ_s is a topographic map of $\mathbb{R}^{d_s}$ and represents the domain on which visual signals $i : \Phi_s \to \mathbb{R}$ are defined. Analogously, for an agent with d_m degrees of freedom, the motor area Φ_m is a topographic map of $\mathbb{R}^{d_m}$ and represents the domain on which motor signals $q : \Phi_m \to \mathbb{R}$ are defined.[1] Commutative diagrams in Fig. 3.2 illustrate the relationships between a physical area, its parametrization, and a signal encoded on top of it. In the remainder of this work, however, the morphology of Φ_s and Φ_m can be abstracted. Of principal importance is the existence of visual signals $i \in \mathcal{I}$ and motor signals $q \in \mathcal{Q}$ contained in the agent's sensor space $\mathcal{I}$ and motor space $\mathcal{Q}$. Note that the fact that Φ_s and Φ_m are not explicitly considered, does not mean that their structure has no influence on the organized sensorimotor system, rather, their physical shape influences how signals i and q are recorded, and these signals in turn influence structures processing i and q. However, the assumption that Φ_s and Φ_m are given, implies that part of the agent's morphology is predefined. On the sensor side, this concerns physical structures used to form an image of the environment on the sensor. For a camera-type eye, this is for example a lens system and the shape of the retinal surface. On the motor side, actions are encoded in Φ_m which essentially means it is assumed that the kinematics of the considered agent is given.

Adaptive Sensorimotor Structures. The agent records visual signals i through a layer of n_s light receptive elements $S = [s_1, s_2, \ldots, s_{n_s}]$. Likewise, it encodes motor signals q using a layer of n_m motor primitves $M = [m_1, m_2, \ldots, m_{n_m}]$. These layers define how visual signals i are reduced to compact visual stimuli $\mathbf{o}$ and how the activation of a limited number of motor primitives $\mathbf{a}$ composes

---

[1]In a physical implementation the motor area Φ_m is at most a volume. Thus, if $d_m > 3$, a mapping of $\mathbb{R}^{d_m}$ to Φ_m is required. Such mappings of higher dimensional spaces to volumes or layers of neural tissue is commonly found in biological systems. For a model see e.g. [124].

Sensory System

Motor System

behavioral policy B

stimulus $\mathbf{o} = [o_1, o_2, \ldots, o_{n_s}]^\top$

activation $\mathbf{a} = [a_1, a_2, \ldots, a_{n_m}]^\top$

prediction model P

topology $S = \{s_1, s_2, \ldots, s_{n_s}\}$

topology $M = \{m_1, m_2, \ldots, m_{n_m}\}$

area Φ_s

area Φ_m

signal $i : \Phi_s \to \mathbb{R}$

signal $q : \Phi_m \to \mathbb{R}$

environment E

sensorimotor layer

higher-level

lower-level

signal complexity

decreasing

increasing
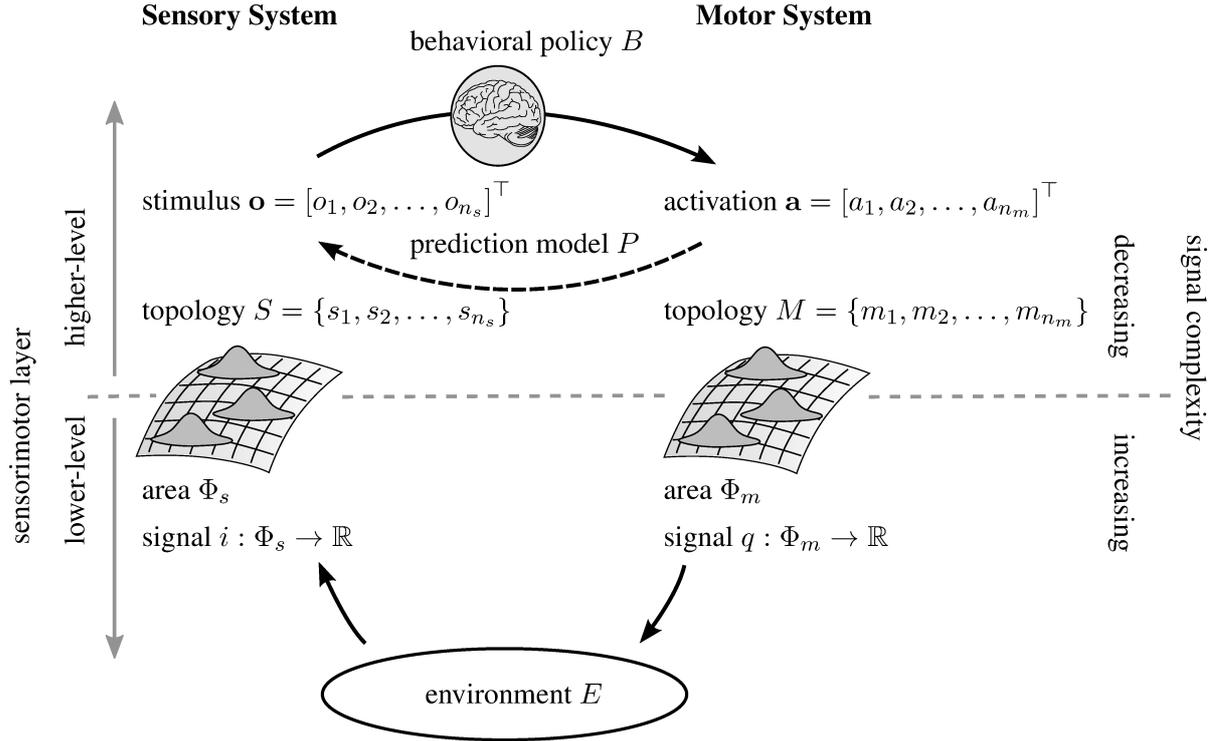
Figure 3.1: Schematic overview of the considered sensorimotor system with the sensory pathway on the left and the motor pathway on the right. The figure illustrates lower-level sensorimotor layers as grids and higher-level sensorimotor layers as Gaussian shaped receptive fields.

and action q. The topological arrangement and physical shape of these n_s visual receptors and n_m motor primitives is initially completely unknown. Their structure represents the principal variable of the considered problem. A mathematical definition of S and M is presented in Sect. 3.4, and a more in depth discussion on visual receptors and motor primitives and their biological prototypes follows below in this section. A forward sensorimotor map P, later used to drive the organization of S and M, is implicitly induced for any S and M. Prediction models suitable to implement such a forward sensorimotor map are revised in Sect. 3.6. In the context of a bigger visual perceptual system, it shall be noted that sensory and motor structures S and M, can be considered to form a single layer in a stack of hierarchically organized sensorimotor layers. In biological systems, a typical example of such a layered and hierarchical structure is the optic tectum or superior colliculus reviewed in Sect. 2. In the case where several sensorimotor layers are considered, an intermediate layer receives input and projects output from and to neighboring layers instead of receiving input directly from Φ_s and encoding output directly on Φ_m. In this sense, the two sets of variables S and M describe the topology of the connection between two different sensorimotor layers at two different levels of abstraction. In this work, usually the intuition is given that the higher dimensional layer represents the periphery of

$$\Phi_s \xrightarrow{\;i\in\mathcal{I}\;} \mathbb{R} \qquad\qquad\qquad \Phi_m \xrightarrow{\;q\in\mathcal{Q}\;} \mathbb{R}$$

$$\phi_s \uparrow \qquad\qquad \phi_m \uparrow$$

$$\mathbb{R}^{d_s} \qquad\qquad\qquad\qquad \mathbb{R}^{d_m}$$

(a) Sensor Space   (b) Motor Space

Figure 3.2: Sensor and motor spaces $\mathcal{I}$ and $\mathcal{Q}$ are defined on physical areas Φ_s and Φ_m in the given agent. Visual signals $i \in \mathcal{I}$ and motor signals $q \in \mathcal{Q}$, are functions defined on domains Φ_s and Φ_m. The given functions ϕ_s and ϕ_m describe the parametrization of Φ_s and Φ_m in $\mathbb{R}^{d_s}$ and $\mathbb{R}^{d_m}$. For a computational approach on how to represent higher dimensional spaces in lower dimensional cortical maps, see e.g. [124]. Note that a $\mathbb{R}^{d_m}$ and Φ_m are connected in one direction only, since, if $d_m > 3$, a surjective function results for an agent embodied in 3-dimensions.

the agent's sensorimotor system, however, this is not a necessary assumption. The considered layers could be located at an arbitrary depth along the sensorimotor pathway of the processing system.

From an Internal Perspective. From an internal perspective, the agent records with its current sensor topology S a visual stimulus $\mathbf{o}$ which appears to the agent as a display of n_s orderless pixels. Similarly, motor primitives can be thought of as n_m strings which the agent can pull by specifying a motor activation $\mathbf{a}$, and which in consequence can change pixel values in an unknown way. This means the morphological configuration of the agent is completely unknown to the agent itself and no information whatsoever is available about the topological organization S of the sensor. Likewise, the arrangement of the agent's motor primitives M is unknown, and their effects can only be observed via changes in the observed stimulus. Hence, initially also the forward sensorimotor map P is unknown for the given agent. Furthermore, in the sense of a purely reactive sensorimotor loop, the agent has no information about past events nor does it know about its state within the world. It is assumed, though, that the agent possesses a given behavioral policy B specifying how motor primitives are activated based on recorded visual stimuli.

From an External Perspective. From an external perspective, the agent is not stateless but is found at each time step in a particular state $x \in \mathcal{X}$ depending on the agent's configuration and its absolute location and orientation in its environment. Sect. 3.5 provides the details on a view of the agent's sensorimotor system from an external perspective.

Visual Receptors S. A visual receptor is specified by a visual *receptive field* function which linearly integrates luminance from the sensory area it covers. This area can be the sensor surface onto which an image of the environment is projected. In this case the receptive field integrates luminance directly

41

at the periphery of the agent. If the visual receptor belongs to a layer further away from the periphery of the visual system, then the input area of the receptive field function is a previous layer of receptors. Accordingly, a visual receptor is defined as a real-valued function on a given d_s-dimensional domain where d_s typically equals 2. The receptive field functions for all n_s visual receptors are described by the variable S. An example of visual receptors at the very periphery of a biological visual system are the photoreceptor cells located in the retina of camera-type eyes. A next layer of visual receptors integrating input from the photoreceptor cell layer is the ganglion cell layer. Other layers of visual receptors follow in the superior colliculus and the visual cortex. In essence, visual receptive fields can appear anywhere along the visual pathway, no matter at which hierarchical level, even as distant from the periphery of the visual system as the receptors in the frontal eye field (FEF) discussed in Sect. 2. The definition of a visual receptor solely requires an input domain and a receptive field function.

Motor Primitives M. In biological systems, the equivalent of a sensory receptive field is the influence area of an efferent neuron on the domain of a considered motor space. Analogously to a sensory receptive field function, this influence area can be described by a function, typically of Gaussian-like shape, which attributes a "preference" to certain locations in the domain of the motor space, see e.g. [100, 38, 116, 64]. Classic studies such as the one by Lee et al. used the term *movement fields* to address such influence areas of neurons in the motor layers e.g. of the superior colliculus. This term is adopted in this work. A movement field represents the dual of a visual receptive field. It is defined as a real-valued function on a d_m-dimensional domain of a given motor space which either covers a subsequent layer of motor primitives or lies at the periphery of the considered agent where it maps to a particular motor signal. The movement field functions for all motor primitives n_m are described by the variable M, analogous to the variable S describing visual receptive fields. In biological systems, recordings in the motor cortex and the motor layers of the superior colliculus have shown that motor signals are typically encoded as the activation of not only one but a set of neurons. Such an encoding of motor signals is called *population coding*. A popular model originally introduced in [38] proposed that a specific target location in the motor space is related to the activity of a population of neurons via a weighted linear combination of active locations. This model is commonly referred to as the *Population Vector* method, where the population vector denotes the target location of the resulting motor command. Despite the popularity of this model, a number of questions remain unsolved [66, 75, 108, 110]. In [108, 109] and [3], alternative models were proposed where movement fields are modeled as probability distributions which can be combined in a multiplicative way. In this sense, a population code represents a probability density function over the underlying motor space.

However, to the best knowledge of the author, no final consensus has been reached so far regarding how a specific motor command is selected or read out from a population of active neurons. The present work assumes that for an active motor primitive, a location in the underlying motor space or motor layer is selected stochastically according to a distribution defined by the movement field function of the respective motor primitive.

Stimulus Prediction Model P. The stimulus prediction model, or forward sensorimotor map, is a mechanism which provides a prediction of a future visual stimulus based on the current visual stimulus and a motor signal. In a biological system such a stimulus prediction mechanism corresponds to a corollary discharge circuit as reviewed in Chapter 2. The artificial agent considered in this work implements a corollary discharge circuit via a stimulus prediction model parametrized by the variable P. This predictive circuit takes input signals from both, sensory and motor areas. In biological systems, the signals originating in the motor area are elicited by corollary discharge neurons which record motor activity according to their receptive fields. In this work, it is assumed that corollary discharge is recorded by the same elements which trigger motor signals, i.e. the motor primitives described in the previous paragraph. This means, motor primitives work in two directions. In one direction motor signals are triggered, in the other direction motor activity is recorded and provided as input to the forward sensorimotor map. In the latter case, the movement field of a motor primitive linearly integrates motor activity from the underlying motor space or motor layer analogous to the receptive field function of visual receptors.

3.4   Observation and Action Model

Stimuli. To observe the world, the agent's sensor works in two steps. First, the given sensor space $\mathcal{I}$ records light projected from the environment. How the environment E is projected onto the sensor is defined by a function $\iota : \mathcal{X} \to \mathcal{I}$, where each element $i \in \mathcal{I}$ is a function $i : \Phi_s \to \mathbb{R}$ returning the projected intensity at each point of the light receptive area Φ_s of the sensor. When the agent records such a visual signal in a particular state $x \in \mathcal{X}$ it is written as i_x. In a second step, the n_s receptive elements spatially integrate projected intensity over the light receptive area of the sensor according to their receptive field functions $s_j : \Phi_s \to \mathbb{R}^+$ which integrate intensity $i_x(y)$ like

$$o_j(i_x) = \int s_j(y) i_x(y) dy, \tag{3.1}$$

where $o_j(i_x)$ is the value observed by the j-th receptor. In this sense the following notation is used henceforth to refer to visual stimuli and sensor topologies:

$$\mathbf{o}_x = \begin{bmatrix} o_1(i_x) \\ o_2(i_x) \\ \vdots \\ o_{n_s}(i_x) \end{bmatrix}, \qquad S = \{s_1, s_2, \ldots, s_{n_s}\}, \;\; s_j > 0. \qquad (3.2)$$

Note that it is assumed here that luminance cannot be subtracted, and hence receptive field functions $s_j(y)$ are defined on a positive range $\mathbb{R}^+$.[2] The transformation of i through a sensor topology S, is implemented via an integration operation, which means the operator ρ as listed in Sect. 3.2 takes in this case the form of an integral. To describe the complete structure of the involved sensor, the two steps of the described observation model – the projection ι and the sensor topology S – have to be considered. The first part of the observation model, the transformation ι is considered to be given and is defined by the physical laws governing the agent's world and the shape of its sensor area Φ_s as described in Sect. 3.3.

Actions. On the motor side, the equivalent to the sensor space is the action space $\mathcal{Q}$ spanned by the given d_m degrees of freedom of the agent. Similar to receptive field functions, each motor primitive is defined by a movement field function m_k. Instead of integrating input like receptive fields s_j, the function $m_k : \Phi_m \to \mathbb{R}^+$ defines a probability for choosing a particular output location in the action space, or a particular primitive in an underlying layer of motor primitives. Dual to the recording of stimuli, the execution of an action works in two steps. First, the agent specifies an activation $\mathbf{a}$ for all n_m motor primitives composing a probability density function according to which a particular motor signal $q : \Phi_m \to \mathbb{R}$ is generated like

$$q = \epsilon \left(\sum_k a_k m_k \right), \qquad (3.3)$$

where the operator ϵ defines how an action q is sampled from a probability density $\sum_k a_k m_k$. The way the operator ϵ is implemented is related to the discussion in the paragraph on motor primitives in Sect. 3.3. One possibility which is later adopted, is that ϵ samples a location according to the given probability density function and a motor signal q is defined as the Dirac delta function at that location

---

[2]Although, in biological systems, more complex receptive fields are common too, this work does consider positive receptive field functions only.

in the sense of a spike of activity. In a second step, the agent executes the motor signal denoted by q which changes the agent's state according to the state transition function χ. Similarly to the function image projection function ι on the sensor side, χ depends on the physical laws governing the agent's world, the agent's kinematics and its current state x. The notation used to describe activation of motor primitives and the topology of the motor system is

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n_m} \end{bmatrix}, \qquad M = \{m_1, m_2, \ldots, m_{n_m}\}, \;\; m_k > 0 \qquad (3.4)$$

Like s_j, the field functions m_k are defined on a positive range $\mathbb{R}^+$. Being probability distributions, this is a natural constraint.

Prediction. Regarding a stimulus prediction operator P for the described observation and action model, it is assumed in this work that q changes i_x in a *predictable* way to $f_q(i_x)$, where f is a functor modifying i_x. The stimulus observed by the agent after taking an action q is therefore

$$\mathbf{o}_j\left(f_q(i_x)\right) = \int s_j(y) f_q(i_x)(y)\, dy, \qquad \mathbf{o}_{x+1} = \begin{bmatrix} o_1(i_{x+1}) \\ o_2(i_{x+1}) \\ \vdots \\ o_{n_s}(i_{x+1}) \end{bmatrix}, \qquad (3.5)$$

where for convenience the definition $i_{x+1} = f_q(i_x)$ is introduced. The request for a predictable change in $\mathcal{I}$ poses a constraint on the action model. This constraint and how it relates to an agent acting in a 3-dimensional world is discussed in Sect. 3.5. The action model in Eq.3.5 also imposes that f be linear. However, since f is actually an operator acting from a function space to a function space, this is not as limiting as it might seem at first glance. Consider the example $f(i)(y) = f_s(y)i(f_p(y))$ where $f_s(y) : \mathbb{R}^2 \to \mathbb{R}$ and $f_p(y) : \mathbb{R}^2 \to \mathbb{R}^2$ are any linear or nonlinear functions. The first one (f_s), modifies for example the intensity of the given image i, and the second one (f_p), warps i in an arbitrary way. The corresponding operator f is linear as can be quickly checked for all y:

$$\begin{aligned} (\alpha i_1 + \beta i_2)(y) &= f_s(y)(\alpha i_1 + \beta i_2)(f_p(y)) \\ &= \alpha f_s(y)\, i_1(f_p(y)) + \beta f_s(y)\, i_2(f_p(y)) \end{aligned}$$

recorded stimulus **o**	motor action **a**	corollary discharge **ã**
visual receptive fields s	motor movement fields m	motor movement fields m
recorded luminance i	selected motor action q	selected motor action q
sensor surface $\mathcal{I}$	motor space $\mathcal{Q}$	motor space $\mathcal{Q}$
(a) Visual Stimulus	(b) Motor Activation	(c) Corollary Discharge
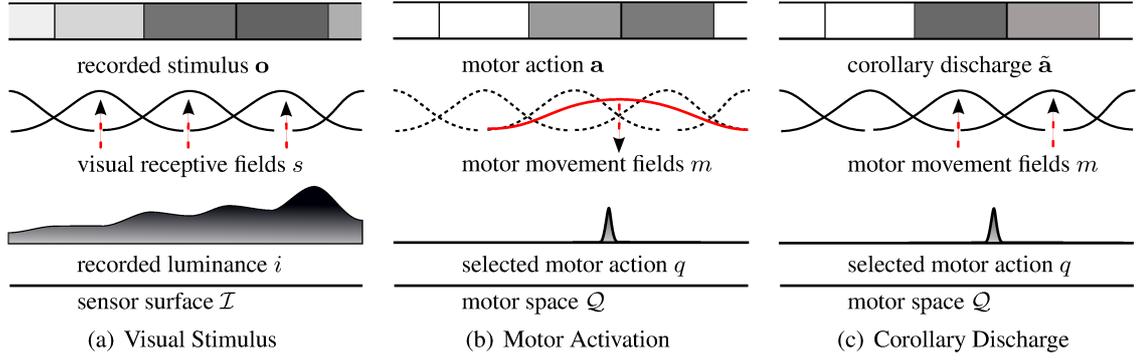
Figure 3.3: Graphical representation of efferent and afferent stimulus transitions implemented by Eq. (3.1), Eq. (3.3), and Eq. (3.6). (a) Visual receptive fields s recording a stimulus **a** from a visual signal i. (b) A particular activation **a** of a number of movement fields m generates an action q. (c) Movement fields m report corollary discharge **ã** from a given action q.

$$= \alpha f(i_1)(y) + \beta f(i_2)(y),$$

which can be summarized as, the warping of a linear combination of images is identical to the linear combination of individually warped images.

Corollary Discharge. Motor primitives can not only trigger motor signals but can also record motor activity. Biological prototypes of such "motor receptive fields" have been reviewed in Chap. 2. There, the graphical model proposed in Fig. 2.3 introduced corollary discharge neurons which record activity from an underlying motor layer. Here, the functionality of efferent motor primitives and afferent corollary discharge neurons are both implemented by the movement fields m. The generation of motor signals q in an efferent direction has been described in Eq. (3.3). In an afferent direction, a movement field m_k estimates its original activation a from a triggered motor action q like

$$\tilde{a}_k(q) = \int m_k(w)q(w)dw. \tag{3.6}$$

Note that this relationship implements the dimension reduction operator ρ as an integral, analogous to Eq. 3.1, and the resulting value does not exactly equal the original activation which generated q.

46

Compare also Fig. 3.3. With Eq. (3.6), the complete corollary discharge signal can be written as

$$\tilde{\mathbf{a}} = \begin{bmatrix} \tilde{a}_1(q) \\ \tilde{a}_2(q) \\ \vdots \\ \tilde{a}_{n_m}(q) \end{bmatrix}. \tag{3.7}$$

3.5   From an External Perspective

In general, an action q induces a change in state x represented by the state transition function $\chi_q :$ $\mathcal{X} \rightarrow \mathcal{X}$. In Sect. 3.4 it is stated that functions χ must induce a predictable change of i_x. This poses a constraint on χ, requiring that it induces a transformation $f_q : \mathcal{I} \rightarrow \mathcal{I}$ in sensor space. Thus, the projection on the sensor surface after taking an action f_q must be perfectly reconstructable solely from the previous projection. If the set of actions is constrained to have this property, then for all agent purposes an action can be fully described as the function f_q instead of considering the full agent state model acted on by functions χ_q. This requirement is too strict to satisfy exactly for most general applications. However, one particular exception arising in many biological systems are eye movement actions in camera-type eyes. In this case the surface onto which the world is projected is a sphere and the eye movement rotates the projection on the sphere. A counter-example for actions q which do not induce a χ_q with a camera-type eye are actions which lead to motion parallax. In this case the resulting projection i_{x+1} in $\mathcal{I}$ contains information (from previously hidden locations in the environment) which cannot be predicted from i_x. Also notice that the existence of an action χ_q does not guarantee that the observed stimuli $\mathbf{o} \in \mathcal{O}$ are predictable in the observation space $\mathcal{O}$. For this to happen, the action must be such that the integrating receptors are transformed nicely one to another before and after the action is taken, corresponding to a permutation of the observed stimulus. This is of particular importance in this work as the sensor topology S is considered a variable and for a given action q, predictability in $\mathcal{O}$ is dependent on S, or vice versa. The relationships between $\mathcal{X}, \mathcal{I}$, and $\mathcal{O}$ are depicted in Fig. 3.4.

3.6   Stimulus Prediction Models

A forward sensorimotor map capable of predicting a future visual stimulus from a current stimulus and a selected motor signal corresponds to a prediction model P. Following the central idea outlined in Sect. 1.4, the characteristics of such a stimulus prediction model will be of central concern for the organization of sensorimotor structures S and M. This section sketches the basic form of a suitable

Figure 3.4: Relationships and constraints between the state space $\mathcal{X}$, the sensor space $\mathcal{I}$, and the space of sensor stimuli $\mathcal{O}$ for a given action q. Functions χ_q, f_q and p_q denote action transitions with χ_q defined by q, the agent's body and the environment. The function ι describes how a visual signal $i \in \mathcal{I}$ is recorded as a projection of the environment depending on the agent's state $x \in \mathcal{X}$. This transformation is implicitly given by the agent's body. The function o defines how a stimulus $\mathbf{o} \in \mathcal{O}$ is recorded from a visual signal $i \in \mathcal{I}$ via a sensor topology S. This transformation is defined in Eq. 3.1

stimulus prediction model.

In general, given an action functor p_q and a stimulus $\mathbf{o}_x$, the stimulus $\mathbf{o}_{x+1}$ can be approximated with functions $p_q : \mathbb{R}^{n_s} \to \mathbb{R}^{n_s}$. Applying p_q to the initially observed sensor values $\mathbf{o}_x$, the sensor values obtained after applying the action q can be approximated like:

$$\mathbf{o}_{x+1} \approx p_q\left(\mathbf{o}_x\right). \tag{3.8}$$

Considering a motor signal q a parameter of p, a functor for all possible actions q can be written as

$$\mathbf{o}_{s+1} \approx p\left(q, \mathbf{o}_x\right). \tag{3.9}$$

With respect to the questions formulated later in Sect. 3.10, the following can be anticipated. If it is requested that the prediction operator p does work well specifically for a sensor topology S and a motor layout M, then it can be expected that the complexity of p depends on S and M, where complexity refers to the number of parameters required to describe the operator.

3.7 Input to the Problem

The input to the formulated problem is solely related to sensorimotor activity experienced by the considered agent. This activity is generated by the agent's behavioral policy which at each time step t selects an action q depending on the currently experienced stimulus. The behavioral policy B is assumed to be part of the given system and can be thought of as a simple brain of the agent, compare also Fig. 3.1. Thus, for an agent following a policy B, each time step a sensorimotor experience $e_t = (i_0, i_1, q)$ results, where i_0 and i_1 denote afferent signals recorded before and after the efferent signal q is elicited. Considering a function C which can sample sensorimotor experiences $e_t = C_t(B, E)$, a

characteristic of the sensorimotor activity of an agent living in an environment E can be obtained as a set of triplets $\{(i_0, i_1, q)\}$ like

$$\mathcal{B} = C_{\text{lifetime}}(B, E). \tag{3.10}$$

This set of sensorimotor experiences $\mathcal{B}$ is henceforth referred to as the *sensorimotor experience* of the agent. In line with observations discussed in Sect. 1.3, an important hypothesis of this work is that the characteristic of an agent's sensorimotor interaction described by $\mathcal{B}$ is the principal driving force for the mutual development of sensorimotor structures S and M. Thus, $\mathcal{B}$ represents the input to the problem considered in this work.

3.8 A General Cost Function

A common line of thinking in biology proposes that evolutionary adaptation implicitly optimizes some underlying criterion which is related to the fitness of an organism, see also Sect. 1.1 and in particular [83]. From an abstract perspective, it can be argued that similarly any autonomous and adaptive artificial system should optimize a certain overall cost function in order to temporally maximize its resource-efficiency, task completion rate, or in general its functional subsistence. It is therefore assumed that the agent considered in this work develops so as to optimize an underlying cost function c_{agent}. Clearly, such a function c_{agent} strongly depends on the agent's body and behavior, and with it on the structure of its sensorimotor apparatus (S, M). Here, it is proposed that a developmental process for the considered artificial agent should implicitly strive to optimize a loosely defined optimization problem

$$\min_{(S,M,B)} c_{\text{agent}}(S, M, B, E), \tag{3.11}$$

which can always be separated into

$$\min_{B} \left[\min_{(S,M)} c_{\text{agent}}(S, M; B, E) \right]. \tag{3.12}$$

Note that in this form, the full problem can be locally solved by iteratively optimizing first variables S and M while keeping B constant and then optimizing B while keeping S and M constant. When considering a given and fixed behavior B, sensorimotor structures (S, M) can be optimized independently by addressing the inner problem in (3.12), compare also Fig. 3.5.

With the hypothesis that the agent's behavior B and environment E enter the problem as sensorimotor experience $\mathcal{B}$ as defined in Sect. 3.7, the inner optimization problem given in (3.12) can be
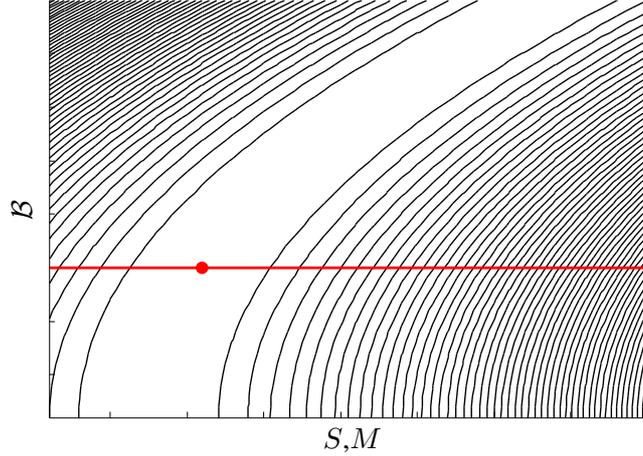
Figure 3.5: An abstract illustration of an optimization for sensorimotor structures (S, M) according to a hypothetical cost function c_{sm} as proposed in Eq. (3.13). For a fixed behavior and environment reflected by $\mathcal{B}$, the optimal sensorimotor structure (S, M) is found along the red line. The plot illustrates dimensions of (S, M) collapsed on the x-axis, and dimensions of $\mathcal{B}$ collapsed on the y-axis.

rewritten as

$$(S^*, M^*) = \operatorname{argmin}_{(S,M)} \quad [c_{sm}(S, M; \mathcal{B})] \ . \tag{3.13}$$

$$\text{s.t.} \quad \mathcal{B} = C(B, E)$$

In this equation the predictor operator P is implicitly present because the sensorimotor structure (S, M) automatically induces a forward sensorimotor map. Thus, concerning the further elaboration of Eq. (3.13), it is clear that in order to incorporate the ideas outlined in Chapter 1, the cost function c_{sm} must be related to the accuracy and simplicity of the induced prediction model P.

3.9 Assumptions

In summary, the problem considers an artificial agent which complies with the following assumptions:

1. The agent lives in a given, static world.

2. The agent can record visual signals $i : \Phi_s \to \mathbb{R}$, where Φ_s is a given d_s-dimensional input domain defined by the physical structure of the agent's sensor.

3. The agent can trigger motor signals $q : \Phi_m \to \mathbb{R}$, where Φ_m is a given d_m-dimensional input domain defined by the kinematics of the agent and its degrees of freedom.

4. The agent executes a given and fixed action selection policy B.

5. The agent executes actions which induce predictable transformations of visual signals i. This means future visual signals i_1 are perfectly predictable solely from previous signals i_0 and executed actions q, irrespectively of the agent's state x.

6. The sensor topology S of the agent consists of a fixed number of n_s light sensitive receptors.

7. The motor topology M of the agent consists of a fixed number of n_m motor primtives.

8. The agent's sensory and motor primitives work according to the specification given in Sect. 3.4.

3.10 Questions Addressed

For a sensorimotor system as described in this chapter, a given sensorimotor structure (S, M) implicitly induces a forward sensorimotor map P. As outlined in Sect. 1.4, this work proposes that the quality of the considered sensorimotor system is related to the accuracy and simplicity of P. These requirements are reflected by a cost function c_{sm} introduced in Sect. 3.8. Together with the remaining definitions specified in this chapter, the principal questions addressed in this work are:

1. Given a sensor topology S and sensorimotor experiences $\mathcal{B}$, what motor layout M optimizes c_{sm}?

2. Given a motor layout M and sensorimotor experiences $\mathcal{B}$, what sensor topology S optimizes c_{sm}?

3. Given the sensorimotor experiences $\mathcal{B}$, does a process which optimizes 1.) and 2.) concurrently converge? What sensorimotor topology (S, M) results?

Chapter 4

Solution

Contents

This chapter proposes a solution to the principal questions posed in Sect. 3.10. The first section introduces a discretization of the addressed problem such as to represent it on a computer. In Sect. 4.2 a concrete model for visual stimulus prediction is instantiated and Sect. 4.3 reviews a number of elements which hold promise with respect to the formulation of a cost function c_{sm} for the evolution of sensor and motor topologies S and M. Eventually, Sect. 4.5– 4.7 follow the approach formulated in Sect. 1.4, addressing questions (1), (2) and (3) from Sect. 3.10 in Sections 4.5, 4.6, and 4.7. In Sect. 4.8 the different characteristics of the proposed cost functions are discussed, and in Sect. 4.9 methods to solve the resulting optimization problems are described.

4.1 Discretization

In order to represent the problem described in Chapter 3 on a computer, continuous signals and functions have to be discretized. This section describes the chosen discretizations.

Grid Discretization. A general discretization of functions defined on continuous domains Φ_s and Φ_m is obtained by discretizing sensor and motor areas Φ_s and Φ_m in a grid-like manner. Thus, defining the resolution of discretized sensor and motor areas as N_s and N_m, visual and motor signals i and q can be represented as a real-valued vectors $\mathbf{i}$ and $\mathbf{q}$ of size N_s and N_m respectively. To represent motor signals q which are of the form of a Dirac delta function, a vector with a single non-zero entry denoting the location of the peak of the function is used. To represent receptive fields and movement fields in a discretized form, functions s and m are discretized accordingly as real-valued vectors $\mathbf{s}$ and $\mathbf{m}$. Sensor and motor topologies S and M, as introduced in Eq. (3.2) and Eq. (3.4), can thus be represented as matrices $\mathbf{S}$ and $\mathbf{M}$ of size $N_s \times n_s$, respectively $n_m \times N_m$.[1] With this notation, the observation of stimuli $\mathbf{o}$ and triggered motor signals q as described in Eq. (3.1) and Eq. (3.3) can be written as

$$\mathbf{o}_x = \mathbf{S} \cdot \mathbf{i}_x, \qquad\qquad \mathbf{q} = \epsilon \left(\mathbf{M} \cdot \mathbf{a} \right), \qquad\qquad (4.1)$$

where $\mathbf{S}$ describes with each row a receptive field, and $\mathbf{M}$ describes with each column a movement field. The estimation of an activation $\tilde{\mathbf{a}}$ having generated q as described by Eq. (3.7) becomes now

$$\tilde{\mathbf{a}} = \mathbf{M}^\top \cdot \mathbf{q}. \qquad\qquad (4.2)$$

When batches of signals $\mathbf{i}_x$ are recorded, signals $\mathbf{i}_x$ and stimuli $\mathbf{o}_x$ can be arranged in columns of matrices $\mathbf{O}$ and $\mathbf{I}$, which allows for the following notation when computing visual stimuli $\mathbf{o}$ from discretized visual signals $\mathbf{i}_x$:

$$\mathbf{O} = \mathbf{S} \cdot \mathbf{I}. \qquad\qquad (4.3)$$

Gaussian Model. A simpler way to discretize functions describing the sensor and motor topologies S and M, is to use a particular model to describe visual receptive fields or motor movement fields s and m. A well accepted approach is to use multivariate Gaussians to describe receptive field functions,

---

[1]The format of $\mathbf{M}$ was chosen to be transposed with respect to the format of $\mathbf{S}$. In this way, a natural application of $\mathbf{S}$ and $\mathbf{M}$ to $\mathbf{i}$ and $\mathbf{q}$ results in Eq. (4.1).

both for biological plausibility as well as for their amenable mathematical properties [92]. This approach was adopted to represent M in an approach to organize the motor layer of a corollary discharge circuit as described in the first part of Sect. 4.5. With a Gaussian model, a motor movement field m_k can be represented via two variables $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ describing m_k like

$$m_k(w) = e^{-\frac{1}{2}(w-\boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(w-\boldsymbol{\mu}_k)}, \tag{4.4}$$

where $\boldsymbol{\mu}_k$ is a real-valued vector of length d_m specifying the location of m_k on Φ_m, and the covariance matrix $\boldsymbol{\Sigma}_k$ of size $d_m \times d_m$ describes the shape and size of the receptive field.

A representation for the forward sensorimotor map p is developed in the next section.

4.2 Stimulus Prediction

In Sect. 3.6 in the previous chapter, a forward sensorimotor map for the considered visual system was introduced as a visual stimulus predictor P, written $p_q(\mathbf{o})$ for a specific action q and stimuli $\mathbf{o}$, or $p(q, \mathbf{o})$ for any action q and stimuli $\mathbf{o}$. It was anticipated that the complexity of such a prediction operator is related to the sensorimotor structure described by S and M. Consequently, if it is desired to organize S and M, such as to induce a simpler sensorimotor map, it is in the operator p where this request has to be accommodated. This section first proposes a concrete form for p, and subsequently outlines how to infer or enforce simplicity in p.

Regarding the formulation of the prediction operator, the following is observed. Considering the static environment E and a spatially rigid sensor layout S, the class of functions from which a stimulus predictor p should be chosen can be restricted. In Appendix B.3, an argument is provided which motivates a reduction of these functions to the linear function set. The argument relies on assumption (5) listed in Sect. 3.9, which requires that the actions executed by the agent lead to perfectly predictable changes of a visual signal i on the sensor area Φ_s. Taking into account linear prediction functions, Eq. (3.9) can be rewritten as

$$\mathbf{o}_{x+1} \approx \mathbf{P}(q)\,\mathbf{o}_x, \tag{4.5}$$

where $\mathbf{P}(q)$ is the matrix representation of a linear prediction function $p(q)$.

To incorporate the tendency towards simpler prediction operators, it is now left to decide on the complexity of the prediction model. A common and natural approach to select simpler models is to evaluate the number of parameters required by the model. For a linear predictor, this requirement can

be translated by forcing the predictor to be sparse. In this sense, equation Eq. 4.5 is revised as:

$$\mathbf{o}_{x+1} \approx \mathbf{P}(q)\,\mathbf{o}_x, \quad \mathbf{P}(q) \text{ sparse.} \tag{4.6}$$

This equation is still ill defined since the notion of sparsity is vague and nothing has been said about the prediction error. It is the author's belief that these cannot be canonically defined, so several alternatives can be proposed as a means of mixing or balancing the importance of sparsity and allowed error:

1. Fix sparsity and minimize some norm of the prediction error. For example one can say that for each location q in the function $\mathbf{P}(q)$, a matrix is found where each row of the matrix $\mathbf{P}^q$ must have a single non-zero entry (sparsity) and under this set the norm error is minimized. Other sparsity sets can be chosen such as $\mathbf{P}^q$ having at most k non-zero entries or that $\mathbf{P}^q$ must be permutation matrices.

2. Minimize the prediction error and infer sparsity. A strategy which first obtains the minimum norm error solution for every location q in $\mathbf{P}(q)$ with a rule which grounds entries of this solution to zero and subsequently deduces sparsity.

3. Simultaneously minimize both prediction error and sparsity. For example the well known LASSO algorithm allows for a single parameter to weight the importance of sparsity versus norm error [127].

4. Empirically, it is found in this work that, requesting $\mathbf{P} \geq \mathbf{0}$ is often sufficient to drive $\mathbf{P}$ towards a sparse solution. This conclusion is motivated by results presented in Sect. 5.3 and Sect. 5.4.

Any of these methods will obtain prediction matrices $\mathbf{P}^q$ as well as associated prediction errors E^q from several samples of the sensor values before and after executing a specific action q from randomly chosen states x. In Sect. 5.2 an empirical study is presented which investigates the sparsity of prediction operators resulting from the coupling of different sensor topologies and motor actions q. Of course, solutions $\mathbf{P}^q$ for particular locations q do not define a function $\mathbf{P}(q)$ for the entire motor space. An approach how $\mathbf{P}(q)$ can be approximated over the entire motor space is part of Sect. 4.5.

4.3 Elements of the Cost Function

When addressing the questions posed in Sect. 3.10, the principal difficulty is to find a formulation for a cost function c_{sm} which reflects the properties sketched in Sect. 1.4 and detailed in Sect. 4.2. Such

a cost function must not only be sensitive to the accuracy of a sensorimotor map induced between S and M, but also to the map's simplicity or sparsity. How to assess or induce a sparse and at the same time accurate prediction model has been previously outlined in Sect 4.2. This section first describes two possible hierarchical levels along the sensorimotor pathway on which a costfunction c_{sm} can be implemented. Subsequently, it is shown that a particular and recurrently appearing correlation term is invariant on both levels and might play an important role with respect to a measure for sensorimotor structures adapted to a specific temporal behavior.

Considering the sensorimotor pathway depicted in Fig. 3.1, the accuracy and sparsity of the linear predictor $\mathbf{P}$ can be evaluated at two different levels. Either a comparison between predicted signals and effectively experienced signals is done at a level of visual stimuli $\mathbf{o}$ and motor activations $\mathbf{a}$, or the comparison is done at the level of visual and motor signals $\mathbf{i}$ and $\mathbf{q}$. The level of stimuli $\mathbf{o}$ and activations $\mathbf{a}$ encodes a more compact representation further away from the agent's periphery. It is referred to as the *higher-level* sensorimotor layer. The level of signals $\mathbf{i}$ and $\mathbf{q}$ records the raw signals closer to the agent's periphery. It is referred to as the *lower-level* sensorimotor layer.

Higher-Level Sensorimotor Layer.   At the level of stimuli $\mathbf{o}$ and motor activations $\mathbf{a}$, the quality of a linear prediction operator $\mathbf{P}^q$ can be directly assessed via

$$E^q = \sum_q \left\| \mathbf{P}^q_{\mathbf{S},\mathbf{M}} \mathbf{S} \mathbf{i}^q_0 - \mathbf{S} \mathbf{i}^q_1 \right\|^2. \tag{4.7}$$

where, with the notation introduced in Sect. 4.1, $\mathbf{i}_0$ and $\mathbf{i}_1$ refer to visual signals recorded before and after taking action q. Importantly, this measure can be written using only stimuli $\mathbf{o}^q = \mathbf{S}\mathbf{i}^q$. Signals $\mathbf{i}^q$ do not appear isolated. The notation $\mathbf{P}^q_{\mathbf{S},\mathbf{M}}$ indicates that the linear predictor $\mathbf{P}$ depends on the current sensorimotor structure $\mathbf{S}$ and $\mathbf{M}$, and also on the executed action q.

Lower-Level Sensorimotor Layer.   Different from the error measure in Eq. (4.7) which operates on stimuli $\mathbf{o}^q = \mathbf{S}\mathbf{i}^q$ in observation space $\mathcal{O}$, the comparison of predicted and experienced signals $\mathbf{i}^q$ at the sensorimotor level closer to the agent's periphery has to be done in sensor space $\mathcal{I}$. To enable such a comparison of stimuli $\mathbf{S}\mathbf{i}$ at the lower-level sensorimotor layer of signals $\mathbf{i}$, a reconstruction $S^+(\mathbf{S}\mathbf{i})$ of an original signal $\mathbf{i}$ can be used. Recalling that an orthogonal projection from the domain of signals $i \in \mathcal{I}$ onto the subspace $\mathbf{S}$ is achieved by the operator $\mathbf{S}^\top \left(\mathbf{S}\mathbf{S}^\top \right)^{-1} \mathbf{S}$, an optimal implementation for S^+ is therefore $\mathbf{S}^\top \left(\mathbf{S}\mathbf{S}^\top \right)^{-1}$. However, for situations where the inner product $\mathbf{S}\mathbf{S}^\top$ is expected to be close to the identity, this reconstruction can be approximated by the adjoint operator, which in this

case corresponds to $\mathbf{S}^\top$. In what follows, it is motivated that in the context of the tackled problem, the solutions for $\mathbf{S}$ can be expected to allow for the use of the adjoint operator for the purpose of signal reconstruction. First, note that sensor topologies with positive, non-overlapping receptive fields $\mathbf{s}$ naturally fulfill the constraint $\mathbf{SS}^\top = \mathbb{D}$, where $\mathbb{D}$ is a diagonal matrix. To assume that $\mathbf{SS}^\top$ stays close to $\mathbb{D}$ appears plausible considering the following. On the one hand, receptive fields obey $\mathbf{s} \geq \mathbf{0}$ as stated in Eq. 3.2, and on the other hand, although a constraint for non-overlapping visual fields has not been introduced, an excessive overlap of receptive fields would decrease the accuracy of reconstructed signals and can therefore be expected to appear only to a very limited extent. In a second argument, it remains to be explained why the adjoint operator $\mathbf{S}^\top$ can be applied in situations where $\mathbf{SS}^\top \approx \mathbb{D}$ as opposed to $\mathbf{SS}^\top \approx \mathbb{I}$. To this end, it is important to note that the reconstruction operator directly interacts with the prediction operator as can be seen when writing the accuracy of a predictor at the sensorimotor level closer to the periphery as

$$E^q = \sum_q \left\| \mathbf{S}^\top \mathbf{P}^q_{\mathbf{S},\mathbf{M}} \mathbf{S} \mathbf{i}^q_0 - \mathbf{i}^q_1 \right\|^2 . \tag{4.8}$$

In this formulation the predictor $\mathbf{P}^q_{\mathbf{S},\mathbf{M}}$ can absorb scaling factors contained in the diagonal of $\left(\mathbf{SS}^\top \right)^{-1}$. Furthermore, matrix inversion is avoided, which – from a biological perspective – is appealing since matrix inversion is an operation unlikely to be implemented by neural tissue. Different from Eq. (4.7), visual signals $\mathbf{i}$ appear here explicitly as experienced future sensor signals $\mathbf{i}_1$, suggesting that the agent requires access to these signals. However, in the next paragraph it is shown that an optimization problem with a cost function according to Eq. (4.8) does not require access to signals $\mathbf{i}$ explicitly.

The viability of using Eq. (4.8) in the context of an optimization problem will be verified in Chap. 5. At this point it can be anticipated that with respect to the design of a cost function c_{sm}, Eq. (4.8) differs from Eq. (4.7) in that it does not only include a request for an accurate prediction of visual stimuli $\mathbf{Si}$, but it is also sensible to the error between available signals $\mathbf{i}$ and stimuli $\mathbf{Si}$ recorded by the agent. Note that, when working with Eq. (4.8), the structure of $\mathbf{S}$ is exposed, meaning the agent has to *know* about $\mathbf{S}$. Interestingly though, to optimize $\mathbf{S}$ and $\mathbf{M}$ the agent does *not* require access to signals $\mathbf{i}$, even though $\mathbf{i}_1$ appears explicitly in Eq. (4.8). Why becomes clear in the next paragraph.

Related to Spatiotemporal Correlation. In this paragraph, the structure of Eq. (4.7) and Eq. (4.8) is compared. By expanding both formulations, it is found that an intrinsic term related to spatiotemporal correlation of visual stimuli is invariant at both of the addressed sensorimotor levels.

Considering in a first step the error measure at the higher-level sensorimotor layer, the summand

on the right side of Eq. (4.7) can be expanded into three terms as

$$\operatorname{tr}\left\{\left(\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}^{q}-\mathbf{Si}_{1}^{q}\right)^{\top}\left(\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}^{q}-\mathbf{Si}_{1}^{q}\right)\right\}$$

$$=\operatorname{tr}\left\{\mathbf{i}_{0}^{q\top}\mathbf{S}^{\top}\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}{}^{\top}\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}^{q}\right\}-2\operatorname{tr}\left\{\mathbf{i}_{0}^{q\top}\mathbf{S}^{\top}\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}{}^{\top}\mathbf{Si}_{1}^{q}\right\}+\operatorname{tr}\left\{\mathbf{i}_{1}^{q\top}\mathbf{S}^{\top}\mathbf{Si}_{1}^{q}\right\}$$

$$=\underbrace{\operatorname{tr}\left\{\left(\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}^{q}\right)^{\top}\left(\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}^{q}\right)\right\}}_{\substack{\text{Term 1: spatial correlation}\\ \text{(predicted stimuli)}}}-2\underbrace{\operatorname{tr}\left\{\left(\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}^{q}\right)^{\top}(\mathbf{Si}_{1}^{q})\right\}}_{\substack{\text{Term 2: spatiotemporal correlation}\\ \text{(predicted stimuli, future stimuli)}}}+\underbrace{\operatorname{tr}\left\{(\mathbf{Si}_{1}^{q})^{\top}(\mathbf{Si}_{1}^{q})\right\}}_{\substack{\text{Term 3: spatial correlation}\\ \text{(future stimuli)}}}\quad. \quad (4.9)$$

In this form it can be seen that each term relates to a correlation between different pairs of stimuli. Searching for a term which relates to the agent's temporal behavior, it is found that a temporal relationship is established by the second term which correlates a prediction of future stimuli $\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}$ with actually experienced future stimuli $\mathbf{Si}_{1}$. Intuitively it is clear, that this term, as opposed to the first and the third term, which merely correlate either predicted or future stimuli among themselves, is an important component in Eq. (4.7) and addresses the relationship of the spatial structure of the considered sensorimotor system with its behavioral characteristics.

Expanding Eq. (4.8) in the same manner as done above for Eq. (4.7), it is found that the second term remains identical:

$$\operatorname{tr}\left\{\left(\mathbf{S}^{\top}\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}^{q}-\mathbf{i}_{1}^{q}\right)^{\top}\left(\mathbf{S}^{\top}\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}^{q}-\mathbf{i}_{1}^{q}\right)\right\}$$

$$=\operatorname{tr}\left\{\mathbf{i}_{0}^{q\top}\mathbf{S}^{\top}\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}{}^{\top}\mathbf{S}\mathbf{S}^{\top}\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}^{q}\right\}-2\operatorname{tr}\left\{\mathbf{i}_{0}^{q\top}\mathbf{S}^{\top}\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}{}^{\top}\mathbf{Si}_{1}^{q}\right\}+\operatorname{tr}\left\{\mathbf{i}_{1}^{q\top}\mathbf{i}_{1}^{q}\right\}$$

$$=\underbrace{\operatorname{tr}\left\{\left(\mathbf{S}^{\top}\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}^{q}\right)^{\top}\left(\mathbf{S}^{\top}\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}^{q}\right)\right\}}_{\substack{\text{Term 1: spatial correlation}\\ \text{(predicted signals)}}}-2\underbrace{\operatorname{tr}\left\{\left(\mathbf{P}_{\mathbf{S},\mathbf{M}}^{q}\mathbf{Si}_{0}^{q}\right)^{\top}(\mathbf{Si}_{1}^{q})\right\}}_{\substack{\text{Term 2: spatiotemporal correlation}\\ \text{(predicted stimuli, future stimuli)}}}+\underbrace{\operatorname{tr}\left\{\mathbf{i}_{1}^{q\top}\mathbf{i}_{1}^{q}\right\}}_{\substack{\text{Term 3: spatial correlation}\\ \text{(future signals)}}}\quad.$$

$$(4.10)$$

The first and third term spatially correlate predicted signals respectively future stimuli at the close peripheral level. The expanded notation also shows that peripheral signals $\mathbf{i}$ enter in the first and second term always as visual stimuli $\mathbf{Si}$ seen through the sensor topology $\mathbf{S}$. Explicitly, visual signals $\mathbf{i}$ appear only in the third term which is independent of $\mathbf{S}$ and $\mathbf{M}$. For an optimization problem concerned with $\mathbf{S}$ and $\mathbf{M}$, this means the third term is constant and visual signals $\mathbf{i}$ need *not* to be available to an agent dealing with such an optimization. Note that since Eq. (4.7) and Eq. (4.8) are formulated as the sum of samples generated by motor signals $\mathbf{q}$, the same independence from $\mathbf{q}$ cannot be assumed. However, access to $\mathbf{q}$ can be assumed to be plausible via proprioceptive feedback.

The conclusion drawn in this paragraph is that correlation between predicted stimuli and actual future stimuli appears naturally in Eq. (4.7) *and* Eq. (4.8) as

$$\left(\mathbf{P}^q_{\mathbf{S},\mathbf{M}} \mathbf{Si}^q_0 \right)^\top \mathbf{Si}^q_1, \tag{4.11}$$

and hence, is a component of principal interest with respect to a cost function addressing the problem formulated in Chap. 3.

4.4 A Measure of Visual Sensorimotor Coupling

Before effectively dealing with the optimization of sensorimotor topologies $\mathbf{S}$ and $\mathbf{M}$ in Sect. 4.5 and Sect. 4.6, it is first investigated in this section how a given sensor topology S affects the sparsity of predictors associated to motor actions q. This work has been published in [102].

To quantify the relationship between an action q and a sensor topology S in terms of sparse prediction, a performance score for the given sensor topology is introduced. For a given action q, this score is a function $\tau^q(\mathbf{P}^q, E^q)$ depending on the prediction matrix $\mathbf{P}^q$ and the prediction error E^q. A given sensor topology S is considered to go well with a given action q if it induces a high performance score. The choice of this performance score is usually tied with the choice of the sparsity / prediction optimization algorithm, see also the different possibilities outlined in 4.2.

For the experiments described next, approach (2) from Sect. 4.2 was used, where first the prediction error is minimized while entries of $\mathbf{P}^q$ are constrained to be greater or equal to zero. To provide the performance score, the Gini index is employed [51]. This index is a well known sparsity measure which complies with a number of desirable properties. Among them the following three: i) changes in smaller coefficients affect the measure more than changes in larger coefficients, ii) the measure is independent of the number of coefficients considered, and iii) the measure is 0 for the least sparse case and 1 for the most sparse case. Thus, in short:

$$\mathbf{P}^q = \underset{}{\mathrm{argmin}} \quad \sum_a \left\| \mathbf{P}^q \, \mathbf{o}^a_x - \mathbf{o}^a_{x+1} \right\|^2 \tag{4.12}$$
$$\text{s.t.} \quad \mathbf{P}^q \geq \mathbf{0}$$

$$\tau^q(\mathbf{P}^q) = \mathrm{Gini}\left(\mathbf{P}^q\right), \tag{4.13}$$

where $\mathbf{P}^q$ is defined as an average over a number of samples a for an action q. This combination was found to provide good invariance to sampling noise and overall results consistent with what was

to be expected. Note that as privileged observers we, the reader and the authors, have access to the underlying sensor topology hence can judge what is to be expected, but this information is not given to the algorithm. Also, the constraint $\mathbf{P}^q \geq \mathbf{0}$ feels natural in a biological context and the chosen approach does not require any further parameter as it is the case when fixing sparsity (number of non-zero entries in $\mathbf{P}^q$), or when implementing a regularized version of the least squares solution such as the LASSO method (regularization parameter).

Results for the score τ^q are presented in Sect. 5.2. In Fig. 5.6 also a comparison of measures τ^q versus prediction errors E^q is presented. Interestingly, the obtained results indicate that for positive linear predictors $\mathbf{P}^q \geq \mathbf{0}$, the local maxima of the sparsity measure τ coincide with the local minima of the mean squared error E. These empirical results are a hint that the request for sparse predictors $\mathbf{P}^q \geq \mathbf{0}$ can be simply addressed through a minimization of E^q.

4.5   Organizing Motor Topologies

This section relates to question (1) in Sect. 3.10 for which a given sensor topology S is considered and a layout of motor primitives optimal according to a (still to be defined) cost function is sought. The solution is organized in two parts. First, the organization of motor topologies for a given sensor topology $\mathbf{S}$ is addressed with motor movement fields following the Gaussian model described in Sect. 4.1 and a cost function in the sense of Eq. (4.7) at the higher-level sensorimotor layer. This work has been published in [104]. In a second part, the constraint that S must be composed of multivariate Gaussian movement fields is relaxed and the more general grid discretization introduced in Sect. 4.1 is used in conjunction with a cost function according to Eq. (4.8) at the lower-level sensorimotor layer, see also [103].

Higher-Level Sensorimotor Layer.   This approach is a direct implementation of the model derived in Chap. 2 shown in Fig. 2.3. In this model, motor movement fields correspond to corollary discharge neurons labeled (c), and work according to the description given in Chap. 3, illustrated in Fig. 3.3(c). The cost function follows Eq. (4.7) and aims at reducing the prediction error directly at the level of recorded visual stimuli $\mathbf{o}$ by adapting the weights of feed forward connections (b) and the layout and shape of corollary discharge neurons (c) as depicted in Fig. 2.3.

Starting with the graphical model introduced in Chap. 2 and the assumptions from Chap. 3, a prediction model can directly be written as described next: feed forward connections shown in gray in Fig. 2.3 are interpreted as manipulators controlling the discharge rate of receptor-receptor connections

(yellow); and when grouping the connections of a single corollary discharge neuron (CDN) together, the feed forward connection weights of the j-th CDN can be written in the form of the previously introduced linear predictor $\mathbf{P}_k$. Thus, with respect to the model shown in Fig. 2.3, the entry in $\mathbf{P}_k(j, l)$ specifies how much the observation $o_{j,x}$ of receptor j contributes to the predicted observation $o_{l,x+1}$ of receptor l. Combining contributions of different corollary discharge neurons according to the CDN layer (c) of Fig. 2.3, the eventually predicted sensor stimulus is composed as a linear combination

$$\mathbf{Si}_1^q = \underbrace{\left(\sum_k^{n_m} m_k(\mathbf{q}) \mathbf{P}_k \right)}_{\mathbf{P}^q} \mathbf{Si}_0^q, \tag{4.14}$$

where $m_k(\mathbf{q})$ denotes the activation of a particular CDN depending on the action $\mathbf{q}$ coded by the underlying layer. Hence, matrices $\mathbf{P}_k$ and their activation functions m_k together define the prediction function $\mathbf{P}(\mathbf{q})$ defined over the entire action space. Because the function m_k models the receptive field of a CDN k on the underlying population of motor neurons, the previous assumption is followed and each m_k is implemented as a multivariate Gaussian according to Eq. (4.4). Measurements motivating a Gaussian model for movement fields in the SC-MD-FEF pathway can be found in [118]. Results comparing the approximated Gaussian model to the actual required shape for movement fields according to the model discussed here are presented in Sect. 5.3 and illustrated in Fig. 5.9.

In conclusion, the free parameters of the model introduced in this approach are $\mathbf{P}_k$, $\mathbf{\Sigma}_k$, and $\boldsymbol{\mu}_k$, compare also Eq. (4.4). These parameters define the plasticity of the modeled corollary discharge circuit. $\mathbf{P}_k$ defines a prediction operator for each movement field or corollary discharge neuron m_k, $\mathbf{\Sigma}_k$ and $\boldsymbol{\mu}_k$ code for topological plasticity in the CDN layer allowing for changes in position and shape of each movement field m_k with respect to the underlying motor space $\mathcal{Q}$.

To learn the free parameters $\mathbf{P}_k$, $\mathbf{\Sigma}_k$, and $\boldsymbol{\mu}_k$, the agent collects a set of sensorimotor experiences $\{(\mathbf{i}_0, \mathbf{i}_1, \mathbf{q})\}$ as described in Sect. 3.7. The adaptation of movement fields m_k is then chosen to follow a minimization on the prediction error like

$$(\mathbf{\Sigma}^*, \boldsymbol{\mu}^*, \mathbf{P}^*) = \operatorname{argmin}_{(\mathbf{\Sigma}, \boldsymbol{\mu}, \mathbf{P})} \sum_q \left\| \left(\sum_k^{n_m} m_k(\mathbf{q}^q) \mathbf{P}_k \right) \mathbf{Si}_0^q - \mathbf{Si}_1^q \right\|^2 \tag{4.15}$$

$$\text{s.t.} \quad \mathbf{P}_k \geq \mathbf{0}$$

As in Eq. (4.7), signals $\mathbf{i}$ do not appear isolated in this formulation, but are only "seen through" the sensor topology $\mathbf{S}$. How a solution can be found to this optimization problem is described in Sect. 4.9.

Lower-Level Sensorimotor Layer. In a second approach, an implementation of a cost function based on Eq. (4.8) is used to find a solution for a motor topology represented by the matrix $\mathbf{M}$. In this approach, movement fields are modelled according to the more general model where fields m are discretized in a grid-like manner as described in the first part of Sect. 4.1. The reason why a similar implementation using a cost function based on Eq. (4.7) is set aside is explained in Sect. 4.7. The approach presented here is an evolution of the previous one, in the sense that it will later allow for an extension which addresses question (3) posed in Sect. 3.10.

Like in the previous approach, a predictor $\mathbf{P}^q$ for a particular action q is considered to be a linear combination of predictors $\mathbf{P}_k$ as described by Eq. (4.14). Different from the previous approach, however, the constraint that movement fields m_k must follow a Gaussian model as described in Eq. (4.4) is relaxed. Movement fields are now discretized in a general grid-like manner as described in the first part of Sect. 4.1. With this discretization of a motor topology M, where location and shape of movement fields m_k are encoded in columns $\mathbf{m}_k$ of a matrix $\mathbf{M}$, Eq. (4.14) becomes

$$\mathbf{Si}_1^q = \underbrace{\left(\sum_k^{n_m} \left(\mathbf{m}_k^\top \mathbf{q} \right) \mathbf{P}_k \right)}_{\mathbf{P}^q} \mathbf{Si}_0^q, \tag{4.16}$$

Then, using Eq. (4.8) as the basic structure to formulate an optimization problem, a cost function for the organization of $\mathbf{M}$ can be written as

$$(\mathbf{M}^*, \mathbf{P}^*) = \operatorname{argmin}_{(\mathbf{M}, \mathbf{P})} \quad \sum_q \left\| \mathbf{S}^\top \left(\sum_k^{n_m} \left(\mathbf{m}_k^\top \mathbf{q}^q \right) \mathbf{P}_k \right) \mathbf{Si}_0^q - \mathbf{i}_1^q \right\|^2 . \tag{4.17}$$
$$\text{s.t.} \quad \mathbf{M} \geq \mathbf{0}, \quad \mathbf{P}_k \geq \mathbf{0}$$

The savvy reader will notice that the apparent ambiguity which arises by the interaction between $\mathbf{P}_k$ and $\mathbf{M}$ nearly disappears with the positivity constraints. In Sect. 4.9, two different methods are described on how to address this problem.

4.6 Organizing Sensor Topologies

This section relates to question (2) in Sect. 3.10 and is organized in two parts, analogously to the previous section. First, the organization of sensor topologies $\mathbf{S}$ for a given motor topology $\mathbf{M}$ is addressed using a cost function at the higher-level sensorimotor layer as introduced in Eq. (4.7). This part focuses in particular on a cost function based on correlation as introduced in Eq. (4.11). Results have been published in [105]. In a second part, the organization of sensor topologies $\mathbf{S}$ is addressed

(a) Coherent Receptive Fields   (b) Scattered Receptive Fields
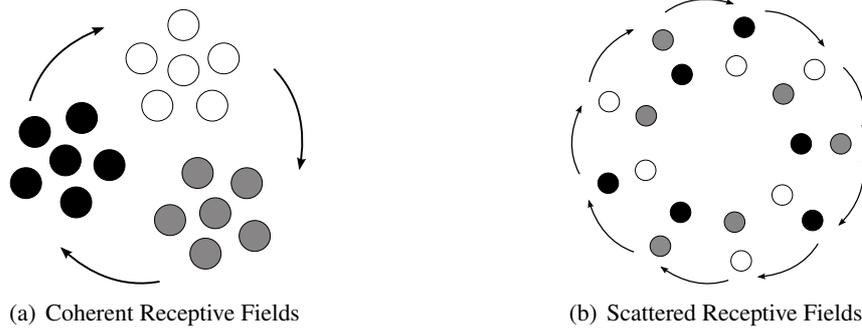
Figure 4.1: Spatially coherent and scattered receptive fields. The illustration shows two different example sensor layout. Each layout is composed of three receptive fields with receptive field functions non-zero at six discrete spatial locations on the sensor area. Both sensor topologies are perfectly predictable under certain rotational actions but are composed (a) of spatially *coherent* receptive fields, and (b) of spatially *scattered* receptive fields. While both layouts comply with the request for self-similarity under certain transformations, layout (b) is undesirable as a visual sensor since it is unable to extract spatial details from a visual signal projected onto the sensor area.

at the lower-level sensorimotor layer via a cost function similar to the one proposed in Eq. 4.17 in the previous section. This approach has been described in [103].

In Sect. 1.3, it was argued that an underlying principle for the organization of sensory structures is related to self-similarity. In particular, the last part of Sect. 1.3 reviewed work described in [21] where a measure for self-similarity was used to organize an abstract sensor layout based on experienced sensor displacements. Here, this potential principle for the organization of sensor topologies is resumed and incorporated into an optimization problem as advertised in step (3) of the approach described in Sect. 1.4. In this sense, a cost function has to be found which induces the following two principal properties in sensor topologies:

1. Spatially *coherent* visual receptive fields develop. For an illustration of coherent receptive fields versus scattered receptive fields, see Fig. 4.1

2. The topological layout of visual receptive fields reflects stimulus translations induced by the agent's behavior according to the self-similarity criterion discussed in Sect. 1.3.

How the first property is stimulated is investigated with two different methods in the two parts of this section. With a cost function formulated at the higher-level sensorimotor layer, spatially coherent receptive fields are formed driven by maximizing the correlation of light stimuli integrated by each single visual receptive field. With a cost function formulated at the lower-level sensorimotor layer, the formation of spatially coherent visual receptive fields is naturally induced by minimizing the reconstruction error of recorded stimuli **o** with respect to original signals **i**. The implementation of

the second property is addressed via the maximization of spatiotemporal correlation as introduced in Eq. (4.11). This component of the cost function naturally appears at the lower-level and higher-level sensorimotor layer.

Higher-Level Sensorimotor Layer. Considering the formulation of an optimization problem following a cost function constructed from Eq. (4.7), the following is foreseeable: the minimization of the stimulus prediction error does not lead to the desired development of spatially coherent receptive fields. For example, a sensor layout consisting of rotationally symmetric but spatially completely incoherent receptive fields can be imagined, and still, the activation of its receptors is perfectly predictable under a number of rotational sensor displacements, compare also Fig. 4.1. Thus, the capability to predict a future stimulus and the tendency towards spatially coherent receptive fields are not directly related. This has been confirmed empirically; via Eq. (4.7), property (1), spatially coherent receptive fields, cannot be induced. However, the spatiotemporal correlation term implicitly contained in Eq. (4.7) can be expected to be sensitive to the spatial coherence of visual receptive fields, because, given the predominantly low spatial frequency of natural images, visual receptive fields $\mathbf{s}$ which are scattered – i.e. cover a wide spatial area – record essentially blurred gray values, whereas more compact receptive fields are able to provide sharper stimuli from the underlying signal. According to these thoughts, the cost function at the higher-level sensorimotor layer is developed around spatiotemporal correlation as introduced in Eq. (4.11). In this context, it is clear that the search space of the problem has to be constrained, since a correlation measure is per se unbounded. For this purpose, a constraint set $\mathcal{R}$ for $\mathbf{S}$ is chosen as $\mathcal{R} = \{\mathbf{S} : \mathbf{S} \geq \mathbf{0},\ \mathbf{S}^\top \mathbf{1} = \mathbf{1}\}$, such as to guarantee that the visual receptive fields occupy the whole sensor area and luminance cannot be subtracted. In the remainder of this paragraph, an optimization problem based on spatiotemporal cross-correlation is formulated incorporating properties (1) and (2) outlined in the beginning of this section. For the sake of clarity, the two properties are addressed step-by-step.

First, the development of spatially coherent receptive fields is addressed by considering an immobile agent with a reduced set of sensorimotor experiences $\mathcal{B} = \{(\mathbf{I}, \mathbf{q})\}$. In this case, a reasonable sensor topology $\mathbf{S}$ is considered to be one which leads to high correlation within a batch $\mathbf{SI}$ of recorded stimuli. The rational behind this is that bigger differences between receptive field activations recorded in a spatial neighborhood indicate that the agent is able to pick-up more information from images contained in $\mathbf{I}$. A first version for a cost function c_{sm} is thus proposed like

$$c_{sm}(\mathbf{S}) = -\sum_x \left(\hat{\mathbf{S}} \mathbf{i}_x \right)^\top \left(\hat{\mathbf{S}} \mathbf{i}_x \right), \quad \hat{\mathbf{S}} = \frac{\mathbf{S}}{\sqrt{\mathbf{S11}^\top}}, \quad (4.18)$$

65

where correlation is normalized with respect to the size of a receptive field such that different sized receptive fields are comparable. The notation of $\hat{\mathbf{S}}$ implies that the division and square root operators are applied element wise.

For a mobile agent, the organization of sensor topologies related to motor activity is addressed with an active agent and a full set of sensorimotor experiences $\mathcal{B} = \{(\mathbf{I}_0, \mathbf{I}_1, \mathbf{q})\}$. To establish a temporal relationship between receptive fields, the previous cost function is adapted to compute correlation between pre- and post-action stimuli. The reader is reminder that it is a priori unknown how to temporally relate receptive fields and how stimuli change under an action $\mathbf{q}$. However, this problem is naturally solved by the predictor $\mathbf{P}^q$ which describes a mapping of receptors for a given action, allowing for a comparison of stimuli at different points in time. Thus, Eq. 4.18 is revised like

$$c_{sm}(\mathbf{S}) = -\sum_q \sum_x \left(\hat{\mathbf{S}}\mathbf{i}_{x,1}^q\right)^\top \left(\mathbf{P}_{\mathrm{pls}(\mathbf{S})}^q \hat{\mathbf{S}}\mathbf{i}_{x,0}^q\right), \quad \hat{\mathbf{S}} = \frac{\mathbf{S}}{\sqrt{\mathbf{S}\mathbf{1}\mathbf{1}^\top}}, \tag{4.19}$$

where, in accordance with the results obtained in Sect. 4.4, it is proposed that the prediction operator $\mathbf{P}_{\mathrm{pls}}^q$ is the solution to a positive least squares problem computed for each action q from a batch of samples $(\mathbf{I}_0^q, \mathbf{I}_1^q)$ as

$$\mathbf{P}_{\mathrm{pls}}^q = \mathrm{argmin}_{(\mathbf{P}^q)} \quad \|\mathbf{P}^q \mathbf{I}_0^q - \mathbf{I}_1^q\|^2 . \tag{4.20}$$
$$\mathrm{s.t.} \quad \mathbf{P}^q \geq \mathbf{0}$$

The request for a self-similar sensor layout – or in other words a simple prediction model – is intrinsically present in the cost function given in Eq. (4.19) because correlation can be expected to be higher if translated receptive fields match well with the location and size of receptive fields at the previous time step (on average). Thus, the Euclidean distance measure used in [21] to assess self-similarity is replaced in Eq. 4.19 with a correlation measure between predicted and future stimuli $\mathbf{P}^q\mathbf{S}\mathbf{i}_0$ and $\mathbf{S}\mathbf{i}_1$.

Eventually an optional cost on the growth of receptive fields can be added to Eq. (4.19) in the form of $C(\mathbf{S}) = \omega\|\mathbf{S}\|_2^2$. This term provides control over the smoothness of the receptive field boundaries. For $\omega = 0$ solutions with hard receptive field boundaries are obtained. For $\omega > 0$ solutions with increasingly overlapping receptive fields result.

Collecting the introduced components, the final optimization problem at the distant sensorimotor layer is written as

$$\mathbf{S}^* = \mathrm{argmin}_{(\mathbf{S})} \quad -\sum_q \sum_x \left(\hat{\mathbf{S}}\mathbf{i}_{x,1}^q\right)^\top \left(\mathbf{P}_{\mathrm{pls}(\mathbf{S})}^q \hat{\mathbf{S}}\mathbf{i}_{x,0}^q\right) + C(\mathbf{S}) , \quad \hat{\mathbf{S}} = \frac{\mathbf{S}}{\sqrt{\mathbf{S}\mathbf{1}\mathbf{1}^\top}}. \tag{4.21}$$
$$\mathrm{s.t.} \quad \mathbf{S}^\top \mathbf{1} = \mathbf{1}, \ \mathbf{S} \geq \mathbf{0}$$

Solutions for this optimization problem can be found with methods described in Sect. 4.9.

Lower-Level Sensorimotor Layer. The formulation of an optimization problem based on Eq. 4.8 is straight forward. Using sensorimotor experiences $\mathcal{B} = \{(\mathbf{I}_0, \mathbf{I}_1, \mathbf{q})\}$ and prediction operators $\mathbf{P}^q_{\text{pls}(\mathbf{S})}$, it can be written as

$$\mathbf{S}^* = \text{argmin}_{(\mathbf{S})} \quad \sum_q \left\| \mathbf{S}^\top \mathbf{P}^q_{\text{pls}(\mathbf{S})} \mathbf{S} \mathbf{I}^q_0 - \mathbf{I}^q_1 \right\|^2 . \tag{4.22}$$

$$\text{s.t.} \quad \mathbf{S} \geq \mathbf{0}$$

Interestingly, this formulation implicitly addresses both properties requested for the organization of sensor topologies through a single mechanism: the development of spatially coherent receptive fields (property 1) is directly induced by penalizing inaccurate reconstructions of visual signals $\mathbf{I}_1$ from predicted stimuli $\mathbf{P}^q_{\text{pls}(\mathbf{S})} \mathbf{S} \mathbf{I}_0$ via the adjoint operator $\mathbf{S}^\top$, and at the same time, the organization of sensor topologies related to motor activity (property 2) is addressed via the comparison of predicted visual signals with future signals. Thus, different from the solution at the higher-level sensorimotor layer (where the clustering of receptive fields is induced through correlation), a solution at the lower-level sensorimotor layer can be formulated as a prediction error (because the formation of spatially coherent receptive fields is induced through the reconstruction of signals $\mathbf{I}^q_0$).

Why (4.22) also encodes a tendency towards self-similar sensor layouts dependent on the agent's motor activity becomes clear when recalling the relationship between sparsity and accuracy for positive linear predictors $\mathbf{P}^q_{\text{pls}}$ as investigated in Sect. 4.4. According to the hypothesis stated there, and the empirical results presented in Sect. 5.2, it can be expected that the mean squared error in Eq. (4.22) reaches a minimum for sensor topologies which not only feature compact receptive fields, but also induce particularly sparse prediction operators $\mathbf{P}^q_{\text{pls}}$. This in turn means, the cost function in Eq. (4.22) is minimized by sensor layouts which increase (on average) the self-similarity property of stimulus translations experienced under the considered set of motor actions.

Solutions for 4.22 can be found using the method described in Sect. 4.9. A major advantage of the problem formulated in 4.22 is that its cost function naturally constrains $\mathbf{S}$. No normalization or additional constraints apart from $\mathbf{S} \geq \mathbf{0}$ are required.

On a final note, different from solutions proposed in the previous and the next section, in (4.21) and (4.22) the prediction operator $\mathbf{P}^q$ is not an optimization variable. Instead, it is a function dependent on the current sensor topology $\mathbf{S}$ and the sensorimotor experiences recorded for actions q. Thus it is assumed in this section, that for each action q a batch $\mathbf{I}^q$ of visual signals $\mathbf{i}$ is available, and a predictor $\mathbf{P}^q$ can be written as a function of a triplet $(\mathbf{I}_0, \mathbf{I}_1, q)$, as in Eq. (4.20). The assumption

that sensorimotor experiences consist of batches of sensory recordings $(\mathbf{I}_0^q, \mathbf{I}_1^q)$ per action q (instead of single recordings $(\mathbf{i}_0^q, \mathbf{i}_1^q)$ per action) will be relaxed again in the next section.

4.7   Organizing Sensor and Motor Topologies Conjointly

This section addresses the third and final question formulated in Sect. 3.10 with the goal of showing that sensor and motor topologies $\mathbf{S}$ and $\mathbf{M}$ can be developed conjointly with a unified approach. To do so, solutions from sections 4.5 and 4.6 are revised, and a cost function is proposed which enables the organization of motor and sensor topologies in a single problem. This approach is part of the work described in [103].

Initial attempts to formulate a unified problem at the higher-level sensorimotor layer by fusing (4.15) and (4.21) were not successful. The reasons can be summarized as follows. On the one hand, the cost function of problem (4.15) is formulated as a pure error measure on stimulus prediction and is therefore not suitable for organizing sensor topologies $\mathbf{S}$ as pointed out in the first part of Sect. 4.6. On the other hand, despite the initial appeal of the simple spatiotemporal cross-correlation measure, attempts to integrate $\mathbf{M}$ in a cost function according to (4.21) were discontinued due to the difficulties of constraining the optimized arguments and the necessity for non-trivial normalizations.

Considering problem (4.17) and (4.22), a formulation to address the concurrent organization of sensor and motor topologies in one single optimization problem seems feasible. Both cost functions are based on the same formulation and can either evolve $\mathbf{S}$ or $\mathbf{M}$. Thus, an optimization problem on the same basis incorporating the organization of sensor *and* motor topologies concurrently can be written as

$$(\mathbf{S}^*, \mathbf{M}^*, \mathbf{P}^*) = \text{argmin}_{(\mathbf{S},\mathbf{M},\mathbf{P})} \quad \sum_q \left\| \mathbf{S}^\top \left(\sum_k^{n_m} \left(\mathbf{m}_k^\top \mathbf{q}^q \right) \mathbf{P}_k \right) \mathbf{S} \mathbf{i}_0^q - \mathbf{i}_1^q \right\|^2 . \qquad (4.23)$$
$$\text{s.t.} \quad \mathbf{S} \geq \mathbf{0}, \ \mathbf{M} \geq \mathbf{0}, \ \mathbf{P}_k \geq \mathbf{0}$$

As in Eq. (4.17), stimulus prediction is implemented in this formulation as a combination of n_m linear predictors $\mathbf{P}^k$ each associated to a motor movement field $\mathbf{m}_k$. Thus, different from problem (4.22), predictors are not a direct function of $\mathbf{S}$ and $\mathbf{q}$, but are learned conjointly with $\mathbf{S}$ and $\mathbf{M}$. Also, sensorimotor experiences enter the problem as a set of triplets $\mathcal{B} = \{(\mathbf{i}_0, \mathbf{i}_1, \mathbf{q})\}$ as proposed in Sect. 3.7, such that for each action q a single sensory recording $(\mathbf{i}_0, \mathbf{i}_1)$ is sufficient. Finally, it shall be recalled here that to optimize (4.23) the agent does *not* require access to individual visual signals $\mathbf{i}$. Signals $\mathbf{i}$ appear in (4.23) only as visual stimuli $\mathbf{S}\mathbf{i}$ in relation to $\mathbf{S}$, $\mathbf{M}$ and $\mathbf{P}$, see also Sect. 4.3. On the motor

side, the agent *does* require access to motor actions $\mathbf{q}$, although, this can be considered plausible since actions $\mathbf{q}$ are self-initiated motor commands. Methods to find solutions for Eq. (4.23) are described in the next section.

4.8   Comparison of Optimization Criteria

This section summarizes two thoughts comparing cost functions formulated according to Eq. (4.8) at the level of signals $\mathbf{i}$ and $\mathbf{q}$, and cost functions formulated according to Eq. (4.7) at the level of stimuli $\mathbf{o}$ and activations $\mathbf{a}$.

The first comment relates to the desired property of spatially coherent receptive fields. It has been argued in Sect. 4.6, that a problem based on Eq. (4.7) is not suitable for the organization of visual sensor topologies since it has no means to induce spatially coherent receptive fields. However, comparing the basis for the finally proposed solution (Sect. 4.7) in its expanded form Eq. (4.10) to Eq. (4.9), it can be seen that the only difference between the two formulations consists of the additional expression $\mathbf{SS}^\top$ in the middle of the first term in Eq. (4.8). With this observation, it becomes apparent that it is precisely this expression which favors non-overlapping receptive fields and induces spatially coherent sensory primitives in $\mathbf{S}$. The reason for this to happen can be explained in two steps: Firstly, non-overlapping receptive fields lead to a diagonal matrix $\mathbf{SS}^\top = \mathbb{D}$, previously discussed in Sect. 4.3, and secondly, diagonal matrices are favored by the considered minimization problem since they lead to small values for the first term in Eq. (4.10).

The second comment concerns the internal perspective of the considered agent. In cost functions based on Eq. (4.8), visual signals $\mathbf{i}$ appear in isolation, implying the agent requires access to the physical layer where signals $\mathbf{i}$ are recorded. However, considering Eq. (4.10), it can be seen that in an optimization problem, the agent does not require access to signals $\mathbf{i}$ because those signals appear in isolation only in the third term which is independent of the optimized variables $\mathbf{S}$ and $\mathbf{M}$. Everywhere else signals $\mathbf{i}$ appear as $\mathbf{Si}$ "seen through" the sensor topology $\mathbf{S}$.

4.9   Optimization Methods

The optimization problems proposed in this work can be addressed by a number of methods. For example, to find a (locally optimal) solution, different gradient descent methods are applicable and readily available in both batch and online versions. All results presented in Chapt. 5 were obtained with gradient descent methods via a batch approach which takes a full set of sensorimotor experiences

$\mathcal{B} = \{(\mathbf{i}_0, \mathbf{i}_1, \mathbf{q})\}$ as input.[2] While it is no problem to find solutions for $\mathbf{S}^*$, $\mathbf{M}^*$ and $\mathbf{P}^*$ with an online method, convergence is much slower and the batch approach was therefore preferred for practical reasons. However, it should be noted that under different circumstances an online implementation might be preferable, e.g. for a purely biologically inspired implementation in a robot with stronger memory constraints and a longer exploration phase where triplets $\{(\mathbf{i}_0, \mathbf{i}_1, \mathbf{q})\}$ are sequentially becoming available as experienced.

To find $\mathbf{S}^*$, $\mathbf{M}^*$ and $\mathbf{P}^*$ in optimization problems (4.17), (4.21), (4.22) and (4.23), the cost function is iteratively minimized using a projected gradient descent method [1]. With this method, constraints on $\mathbf{S}$, $\mathbf{M}$ or $\mathbf{P}$ are respected by projecting the gradient of the respective problem onto the constraint set. In problems (4.21) and (4.22) where prediction operators are not considered variables but directly depend on the sensor topology $\mathbf{S}$ and actions $\mathbf{q}$, a linear predictor $\mathbf{P}^q_{\mathrm{pls}} \geq \mathbf{0}$ is learned for each action $\mathbf{q}$ such that it best satisfies $\mathbf{SI}^q_1 = \mathbf{P}^q_{\mathrm{pls}} \mathbf{SI}^q_0$ in a positive least squares sense using the optimization method known from [9]. In problems (4.17) and (4.23), where prediction operators $\mathbf{P}^q \geq \mathbf{0}$ are arguments of the optimization, the derivative of the cost function with respect to $\mathbf{P}^q$ is required. Note that, even though $\mathbf{P}^q \geq \mathbf{0}$ cannot be obtained as a closed form solution, a partial derivative with respect to $\mathbf{P}^q$ can still be found in closed form by applying the implicit function theorem to the Karush-Kuhn-Tucker optimality conditions of the positive least squares optimization problem [12]. The stepsize of the gradient descent was adapted in all cases according to the Armijo rule [4].

Slightly different from the rest of the optimization problems, Eq. (4.15) was solved using the Levenberg-Marquardt algorithm which was found to have nice convergence properties for this problem while being relatively simple to implement, see for example [95]. The constraint $\mathbf{P}^k \geq \mathbf{0}$ was implemented in this case by adding an exponential penalty function to the optimization.

To conclude, it is important to note that all the proposed cost functions are expected to be non-convex, which means nothing prevents a gradient descent from converging to a locally optimal solution. From a biological perspective, locally optimal solutions can be accepted. In some cases, the privileged perspective of the observer allows for the detection of undesirable solutions. In these cases, additional empirical tests are conducted to verify for example the validity of a suspected global minimum. Such checks are described when discussing the results in Chap. 5.

In addition to locally optimal solutions, the existence of non-isolated minima is expected, due to symmetries of $\mathbf{S}^*$ and $\mathbf{M}^*$ on sensor and motor areas Φ_s and Φ_m. However, this is of no concern since different (e.g. rotational) symmetric sensor or motor layouts are equivalent in a topological sense and actually desired to support self-similar stimulus translations.

---

[2]For problems (4.21) and (4.22) input is considered to be available as $\mathcal{B} = \{(\mathbf{I}_0, \mathbf{I}_1, \mathbf{q})\}$.

Chapter 5

Results

Contents

This chapter presents results obtained with solutions proposed in Chap. 4. In a first section, a concrete instance of the problem described in Chap. 3 is introduced. This setup was used to obtain all presented results. The remaining sections discuss results corresponding to the problems formulated in the previous chapter. Sect. 4.5– 4.7 are directly associated to the three principal questions posed in Sect. 3.10. Each section provides an individual discussion with respect to the presented results.

5.1 A Concrete Instance of the Problem

This section introduces the particular setup used to obtain the results described in subsequent sections. The considered configuration is an instance of the model described in Chapter 3 where an agent with four degrees of freedom observes a planar environment textured by a very high resolution image e (2448×2448 pixels) depicting a real world scene, see Fig. 5.3. The agent's sensor area Φ_s, is a bounded and flat area assumed to be parallel to the environment. From an external perspective, the agent is found at each time step in a state x, which is defined by the agent's absolute position, orientation, and current distance from e. The agent records in each state x a signal vector $\mathbf{i}_x$ with a sensor area
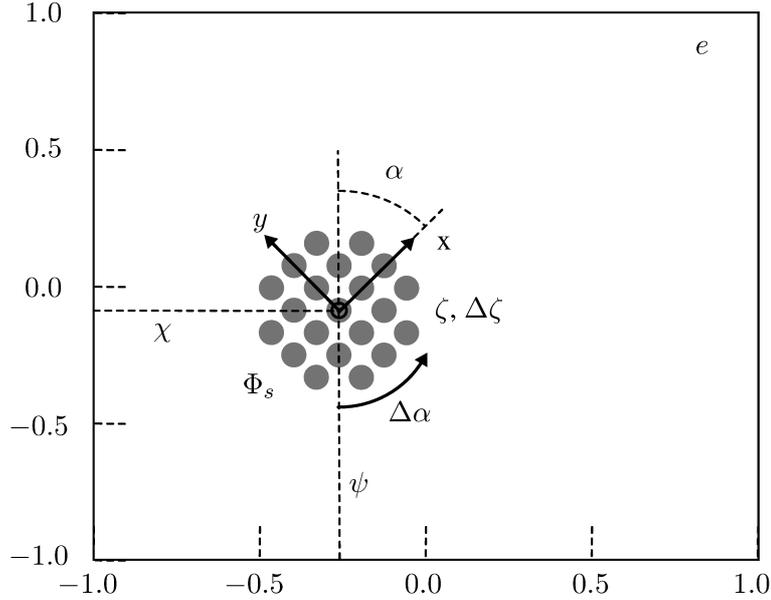
Figure 5.1: A particular instance of the problem described in Chapter 3. An agent with a planar sensor area Φ_s observes a flat environment textured with a high resolution image e. At each time step, the agent is found in a state defined by the agent's position (χ, ψ), its orientation α, and its distance ζ from e. The agent can change its state by taking an action from a 4-dimensional motor space spanned by the following types of actions: translations in two directions (x, y), rotations $(\Delta\alpha)$, and changes in distance to e, denoted $\Delta\zeta$, which dilate the stimulus. The agent is illustrated with a circular sensor surface Φ_s discretized as a regular grid. The environment extends over the indicated range $[-1, +1]$.

discretized in a grid-like manner as described in Sect. 4.1. From signals $\mathbf{i}_x$, stimuli $\mathbf{o}_x$ are observed via a sensor topology $\mathbf{S}$ which is discretized in the same grid-like fashion. The agent can modify the current observation $\mathbf{o}_x$ by executing actions from a 4-dimensional motor space. The four types of actions are: translations (x, y), rotations $(\Delta\alpha)$, and dilations of visual stimuli induced by a change in distance with respect to e denoted as $\Delta\zeta$. This setup is illustrated in Fig. 5.1. The sensor area is illustrated as a grid of points where each point denotes a discrete sampling location. An illustration of a discretization of sensor and motor areas is shown in Fig. 5.2. This discretization with 481 locations on the sensor area and 225 locations on motor areas was used in particular for the results discussed in Sect. 5.5. An illustration of sensorimotor experience triplets $(\mathbf{i}_0, \mathbf{i}_1, \mathbf{q})$ used as input for all addressed problems are shown in Tab. 5.2. Each row shows a sample triplet. Visual signals $\mathbf{i}$ are discretized into 2877 locations on a disk. Discretized actions are visualized as reshaped vectors $\mathbf{q}$ with a single non-zero entry as introduced in Sect. 4.1. The different states x of the agent in which these samples were taken are depicted in Fig. 5.3. The discretization of the sensor area as shown in Fig. 5.3 and Tab. 5.2, was used in particular for results discussed in Sect. 5.4.
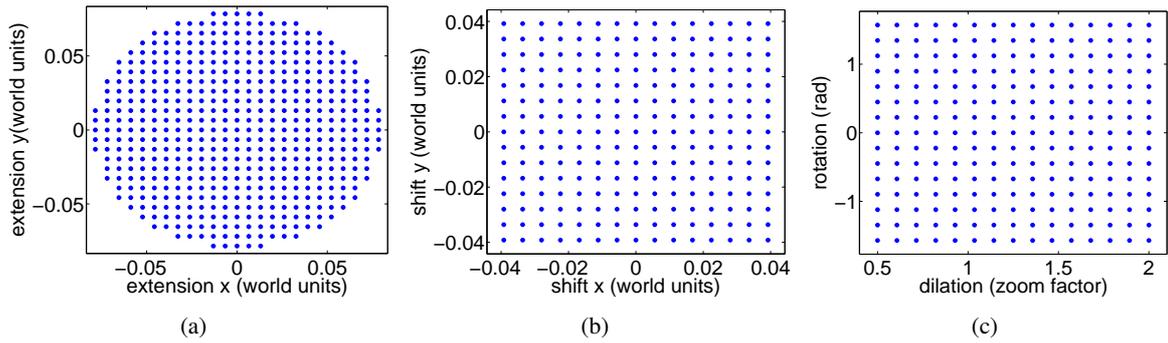
Figure 5.2: (a) Discretization of a sensor space defined on a disk (481 discrete locations); (b) discretization of a motor space for a behavior inducing horizontal and vertical translation actions (225 discrete locations); (c) discretization of a motor space for a behavior inducing dilation and rotation actions. Sensor area and translation distances are specified in environment coordinates ranging from -1 to 1 in x- and y-direction (225 discrete locations).

Coding of Stimulus Dilation Actions. Note that slightly different from translation and rotation actions, the action $\Delta\zeta$ which dilates stimuli recorded by the agent is encoded like $\zeta = \log(z)$, where z can be seen as a zoom factor, while ζ can be interpreted as the distance of the visual sensor to the observed scene. Thus, an action $\Delta\zeta$ means moving the sensor along the vertical direction changing the distance to the observed image e. With this choice, a situation independent, and at the same time energetically plausible representation of an action dilating the stimulus is obtained. Situation independence is achieved in the sense that an action composition like $+\Delta\zeta_a - \Delta\zeta_a$ leaves a sensor stimulus $\mathbf{i}_x$ invariant. This means the agent's current state x does not influence the effect of an action $\Delta\zeta_a$. Energetically plausible refers to the fact that in a physical setup, $\Delta\zeta$ might directly relate to voltage or current applied to an actuator moving the agent for example towards or away from an observed scene. Thus, encoding dilation as $\zeta = \log(z)$ appears reasonable, as moving away $(-\Delta\zeta_a)$ or towards $(+\Delta\zeta_a)$ a scene requires the same amount of energy, which would not be reflected by the zoom factor z.

Note on the Action Space from a Biological Perspective. From a biological point of view the chosen action space includes degrees of freedom which are not directly implemented by the oculomotor system e.g. of primates and other animals with binocular vision. This concerns in particular the two action dimensions which lead to rotation and dilation of visual stimuli. However, despite not being directly supported by anthropomorphic oculomotor systems, actions which induce rotation and dilation of visual stimuli do play a very important role, e.g. during locomotion, or in visual sensorimotor systems of airborne insects. They also typically appear in sensorimotor interaction patterns of binocular vision systems engaged in object manipulation. Therefore, it is considered important in this work to

Figure 5.3: The environment e of the setup considered in this chapter. The illustration also shows three example situations where the agent is taking actions from the considered 4-dimensional action space. Top left: rotation by $45°$. Top right: translation of 15 discrete steps in x-direction, and 6 discrete steps in y-direction. Bottom: dilation with $z = 1.5$. The taken samples and their associated actions $\mathbf{q}$ are shown in Tab. 5.2. The underlying image has a resolution of 2448×2448 pixels and was taken in Guerrero / México, courtesy of Fam. Urióstegui-Arellano.

Action Type	$\mathbf{i}_0$	$\mathbf{i}_1$	$\mathbf{q}$
Translation			
Rotation			
Dilation			
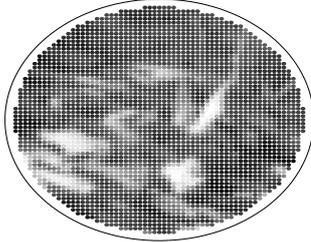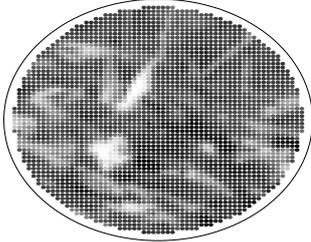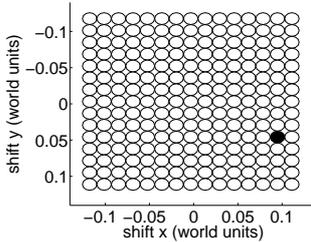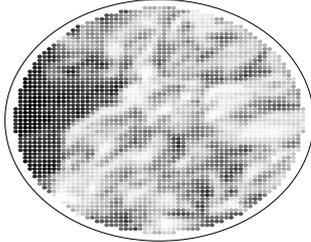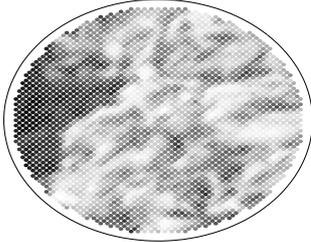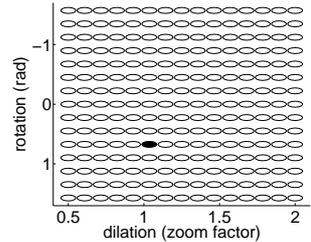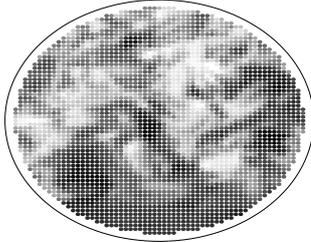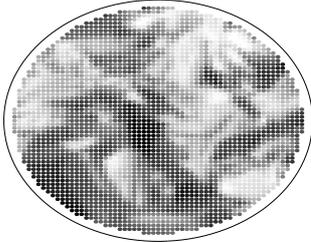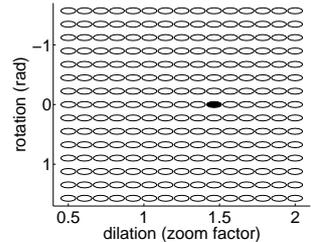
Table 5.2: Example sensorimotor experience triplets $(\mathbf{i}_0, \mathbf{i}_1, \mathbf{q})$ as obtained from agent states depicted in Fig. 5.3. Each visual signal $\mathbf{i}$ is sampled at 2877 discrete locations in the environment. For each discrete location a gray-scale value was obtained by filtering the underlying image with a Gaussian filter with standard deviation corresponding to the radius of depicted signal pixels. Discretized actions are visualized as reshaped vectors $\mathbf{q}$ with a single non-zero entry as introduced in Sect. 4.1.

include actions inducing rotation and dilation of recorded stimuli.

Given Sensor Topologies. For experiments conducted with given sensor topologies, such as the ones presented in Sect. 5.3, two different sensor layouts have been investigated. A regular grid-like configuration, and a non-uniform, fovea-inspired layout with a logarithmic parametrization, also commonly used to describe growth spirals found in nature, see also Sect. 1.3.

The grid configuration was chosen because of its relevance with respect to basically all artificial image sensors, e.g. CCD sensors available off-the-shelf. The logarithmic distribution was chosen to analyse stimulus change patterns for foveating visual systems. The two layouts are shown in Fig. 5.4. The foveal layout shown in Fig. 5.4(b) was generated according to a logarithmic spiral. Retinotopic layouts found in living organisms with binocular vision follow closely such a density distribution up
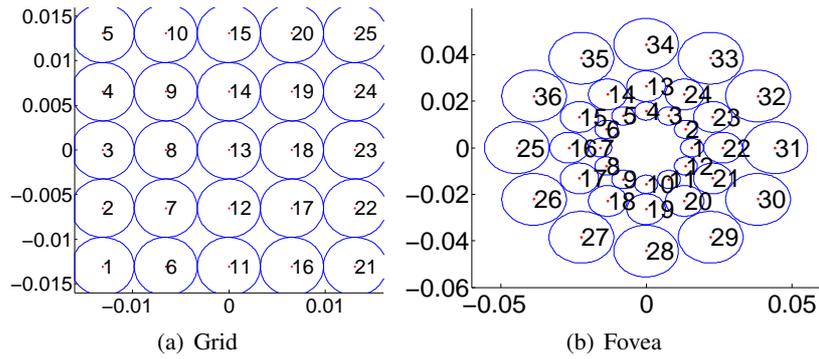
Figure 5.4: Two sensor topologies considered for experiments with fixed sensor topologies. Each visual receptive field is implemented with a Gaussian model. Circles are drawn with the radii of the standard deviation of each receptive field function. Axis units refer to the size of the 2-dimensional environment e as depicted in Fig. 5.1. a) Uniform grid-like 5×5 layout with 25 receptive fields. b) Foveal layout parametrized in polar coordinates (ρ, ϕ) according to a growth spiral with 12 branches each with 3 receptive fields following $\rho = 0.0063e^{0.33\phi}$.

to a small area in the very center which deviates from this law. In [117] an approximation of this deviation is formulated, however this area is not considered in the layout shown in Fig. 5.4(b).

5.2 A Measure of Visual Sensorimotor Coupling

This section presents results regarding the evaluation of a coupling between sensor topologies and motor actions in the sense of a sparsity measure as introduced in Eq. (4.13). Measures were obtained for the two sensor layouts shown in Fig. 5.4 with respect to the 4-dimensional action space as introduced in Sect. 5.1. To allow for a visualization of the results in three dimensions, both sensor layouts were measured for two 2-dimensional subspaces of the full 4-dimensional action space, one subspace covering translations in x- and y-direction, the other including rotations and dilations. For each subspace, measurements τ^q were obtained for $10\,000$ randomly chosen motor actions $\mathbf{q}$. Each measurement was obtained by sampling a particular action $1.4 \cdot n_s$ times from randomly chosen states x in environment e. Fig. 5.5 visualizes the measurements for all combinations of the two sensor layouts and action subspaces. The results are discussed in detail below.

In an empirical investigation, the results obtained for the sparsity measures on linear predictors $\mathbf{P}^q$ are compared to results for the mean squared error E^q of linear predictors $\mathbf{P}^q$. By visual inspection evidence is found that the local maxima of the sparsity measure coincide with the local minima of the mean squared error. Results for E^q are plotted in Fig. 5.6.

(a) Grid layout x- /y-translation

(b) Grid layout rotation / dilation

(c) Foveal layout x- /y-translation

(d) Foveal layout rotation / dilation
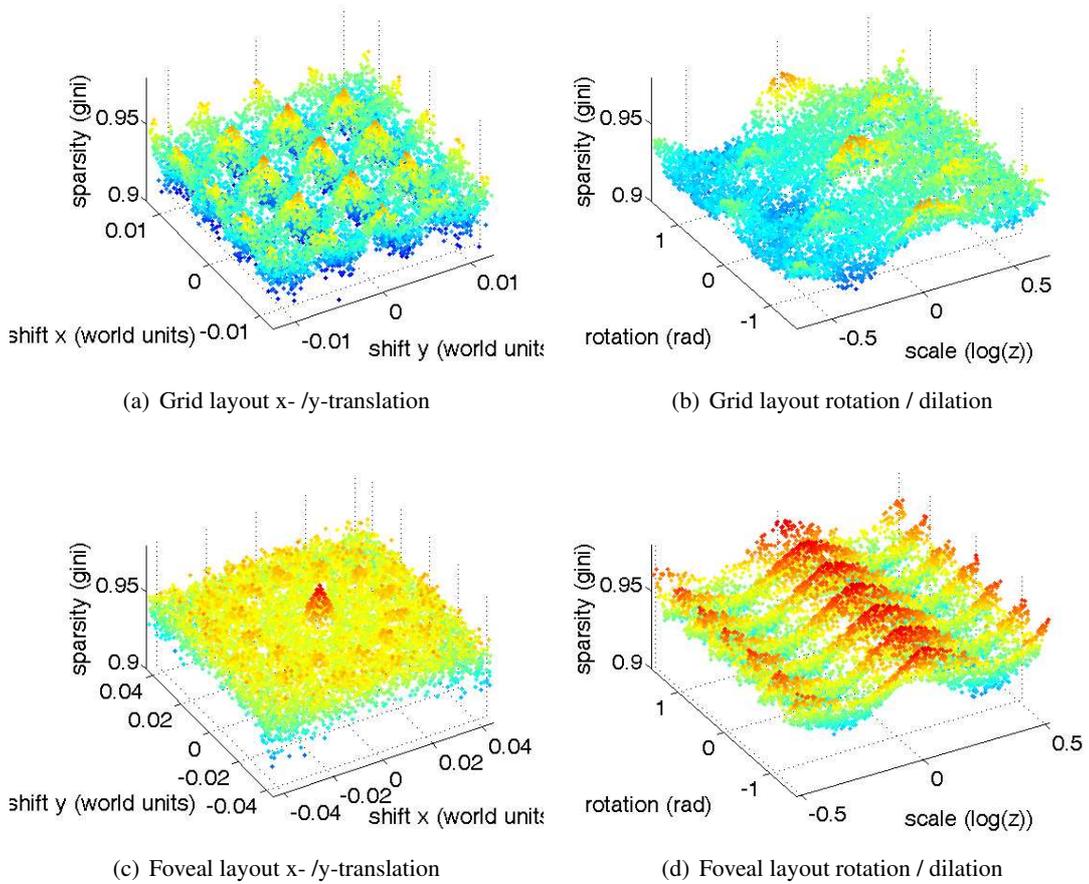
Figure 5.5: Sensorimotor coupling of two visual sensor topologies and two 2-dimensional action spaces. sensor topologies. The sensor layouts "grid" and "fovea" are shown in Fig. 5.4. The two action spaces cover i) translation actions in x- and y-directions, and ii) rotation and dilation actions as introduced in Sect. 5.1. In each plot, the z-coordinate shows the sparsity measure τ^q introduced in Eq. (4.13) for 10 000 randomly chosen actions $\mathbf{q}$.

Uniform Grid Sensor Layout. Measurements for horizontal and vertical translation actions of the uniform grid sensor layout are shown in Fig. 5.5(a). The plot shows clear peaks when a translation equals a combination of horizontal and vertical receptor distances. The peaks corresponding to larger action steps are slightly smaller because larger displacement actions lead to a bigger number of peripheral and unpredictable receptors, provoking noise in the prediction operator which is responsible for less sparse solutions. This is a desirable effect, since actions with a bigger number of unpredictable receptors should qualify less well in the sense of self-similar stimulus translations. The results for the grid layout under rotation and dilation are shown in Fig. 5.5(b). Excluding the peak for the zero-action, only two significant peaks are visible, located at zoom level $z = 1$ and $90°$ and $-90°$ rotation. This makes sense, as these are the actions which achieve a perfect permutation of receptors. In between,

77

the scores are significantly lower with certain angles achieving a slightly better score than others.

In summary, while this sensor achieves good measures for x- and y-translation actions with step-sizes related to receptor distances, the topology does not qualify well for zoom and rotation actions. Thus, with this layout stimulus prediction for x- and y-translations can be computed using a reduced amount of parameters and operations, whereas prediction for actions inducing stimulus rotations and dilations requires more resources.

Foveal Sensor Layout. Measurements for horizontal and vertical translation actions of the foveal sensor layout are shown in Fig. 5.5(c). Disregarding the peak for the zero-action, this plot does not show clear peaks except for a ring of local optima corresponding to shifts of the center location to receptor positions on the second ring; although, these shifts are assigned a significantly lower score than the peaks in Fig. 5.5(a)). Much in contrast, the measurements for the rotation and scale action space shown in Fig. 5.5(d) show strongly expressed peaks at zoom levels $z = 1$, $z = 0.60$ and $z = 1.68$. The number of peaks in angular direction corresponds to the number of receptors on the semicircle between $-90°$ and $90°$ in the foveal sensor layout shown in Fig. 5.4(b). The three rows of peaks in the direction of stimulus dilation actions correspond to a mapping of the three receptor rings of the foveal layout.

In conclusion, a foveal sensor favors simplified linear prediction operators for rotational stimulus transformations but sacrifices on the other hand sparsity under translational displacements.

Discussion. A clear shift in the characteristics of τ-scores is observed when moving from the grid sensor layout to the foveal layout. While the grid layout achieves higher scores under translation, the foveal sensor requires a more descriptive prediction function for the same actions. This coincides with the fact that animals with a foveating visual system usually compensate image translations by appropriate sensor movements, e.g. saccadic gaze shifts or smooth pursuit of foveated objects. Hence, the handling of image translations is "outsourced" to active sensor repositioning allowing for reduction of image transformations for which the sensor itself is badly optimized. Contrariwise, the grid-like sensor does not qualify well for rotation and scaling while the foveal sensor significantly facilitates the description of stimulus changes for these actions. Again, this matches well with observations in living organisms. Foveating visual systems can be assumed to frequently experience image rotation and scaling because foveal sensors are typically used to engage in object-oriented behavior ranging from prey-catching to in-hand manipulation; activities which typically involve self-induced actions resulting in approximation, adjustment or repositioning of an object. Assuming the oculomotor system

78

(a) Grid layout, τ sparsity measurements

(b) Grid layout, E error measurements

(c) Foveal layout, τ sparsity measurements

(d) Foveal layout, E error measurements
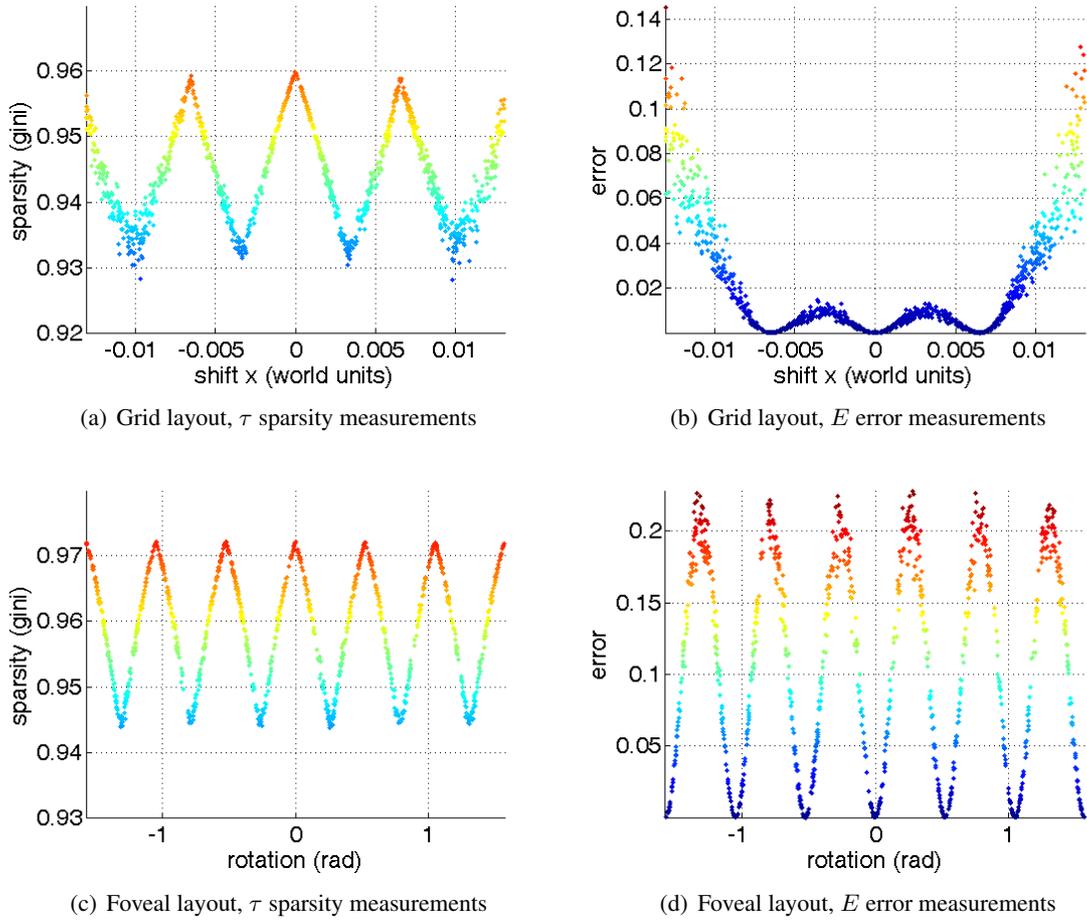
Figure 5.6: A comparison of sparsity scores τ^q and mean squared errors E^q for both sensor layouts and two selected cut-sections of the investigated action spaces.

compensates for horizontal and vertical translations of the target object, these behaviors mostly induce rotation and scaling of the object's projection on the observing sensor – transformations for which the foveal sensor layout is well adapted in the sense that they can be compensated by a computationally inexpensive post-processing step. Giving consideration to these observations, it seems reasonable to assume that the morphology of a sensor has strong ties to the agent's behavior (and vice versa).

Relationship Between Sparsity Measure and Mean Squared Error. At the end of Sect. 4.4, it was conjectured that for positive linear prediction operators $\mathbf{P}^q \geq \mathbf{0}$, the most accurate prediction operators might be achievable for actions $\mathbf{q}$ for which the operator $\mathbf{P}^q$ is also particularly sparse. For the sensor layouts and motor spaces considered in this section, this expectation was empirically confirmed. Fig. 5.6 illustrates the results; for visualization purposes, the figure shows for each sensor layout a particularly relevant dimension of the full action space (left side), and compares the same

cut-section to previously obtained sparsity measures (right side). While these results cannot replace an analytical proof, they are a strong hint that the minimization of the prediction error in problems proposed in Chap. 4, not only leads to more accurate predictors, but also discovers locations in the action space where prediction can be achieved with particularly sparse matrices $\mathbf{P}^q$. This mechanism can be seen at work in the next section where it is used to organize the movement fields of the motor layer.

5.3   Organizing Motor Topologies

This section presents results for the optimization of motor topologies according to problem (4.15). Because problem (4.17) is part of the unified approach formulated in problem (4.23), motor topologies obtained on the basis of problem (4.17) are presented in Sect. 5.5.

The results described in this section are organized in two groups. For each sensor layout shown in Fig. 5.4, a 2-dimensional sub-space of the action space described in Sect. 5.1 was considered. For the grid layout, it was chosen to analyze the adaptation process of movement fields under translational actions. The foveal layout is used to train a motor layer for rotation and zoom actions. Since in problem (4.15), the motor topology M is composed of motor movement fields m^k modeled as multivariate Gaussians according to Eq. (4.4), the shape of movement fields m_k is described by covariance matrices $\mathbf{\Sigma}_k$ and their location is encoded by the mean $\boldsymbol{\mu}_k$. The results presented in this section were obtained with an implementation where $\mathbf{\Sigma}_k$ was reduced to a diagonal matrix thereby constraining movement fields m^k to be axis-aligned Gaussians. This reduction in dimensionality is based on empirical evidence collected as a part of the result validation discussed at the end of this section. To find a solution for all variables $\mathbf{\Sigma}^*$, $\boldsymbol{\mu}^*$ and $\mathbf{P}^*$ of problem (4.15), a set of sensorimotor experiences $\mathcal{B} = \{\mathbf{i}_0, \mathbf{i}_1, \mathbf{q}\}$ is collected and the Levenberg-Marquardt optimization algorithm is used, see also the optimization procedures described in Sect. 4.9.

The obtained results are illustrated in Fig. 5.7 and Fig. 5.8. Plots on the left depict initial and final configuration of movement fields on the motor area showing position and size of the movement fields. Prediction matrices $\mathbf{P}^k$ are shown on the right for a number of selected movement fields. For each experiment the positions $\boldsymbol{\mu}_k$ were initialized according to a uniform random distribution. The shapes of the movement fields $\mathbf{\Sigma}_k$ were all initially set to a fixed value. The prediction matrices are initialized to zero. It is important to note that with a randomized initialization, nothing prevents the adaptation process from converging to a locally optimal solution. Eventually, one has to be aware of boundary effects when inspecting the results of the organization of motor movement fields. This is due to the

fact that one inevitably has to rely on a finite range for action sampling. Thus, some disturbances are expected to be observed for motor primitives located at the sampling border.

Grid Layout. When solving problem (4.15) using the grid sensor layout under translational actions, it is found that motor movement fields converge to a configuration where they are distributed in the action space on a regular grid. This is visible in Fig. 5.7(c), in particular for non-boundary movement fields. Notably, the positions of motor movement fields on the motor area coincide with the spacing of receptors in the sensor layout. It is noted that the prediction matrix of the motor primitive with index 14 converged to zero which has the same effect on the final cost function as if this motor primitive had been removed. From the privileged perspective of the reader, it can be seen that motor primitive 14 should have been placed somewhere between motor primitives 16 and 4 to improve the present solution. In the shown case, the algorithm converged to a local optimal solution where motor primitive 14 has no contribution and the area between motor primitives 16 and 4 is covered by slightly more outstretched neighbor movement fields. This slightly increases the prediction error in this neighborhood of the action space but has no severe impact on the prediction ability.

In Fig. 5.7(d), the learned prediction matrices of nine selected motor primitives are shown. Interestingly, the entries for unpredictable receptors 1 to 5 converged to be non-zero in the diagonal. This is due to the fact that for natural images with low spatial frequency and small sensor translation distances, the future activation of a receptor with unpredictable input is best described by values of previously close receptors. Despite the fact that matrices $\mathbf{P}^k$ are expected to be sparse in general due to spatiotemporal relationships between visual receptors, it is found that the converged configuration features prediction matrices with an exceptionally small number of non-zero entries.

To address questions regarding the global optimal solution, a number of optimizations were run where each run started from a different randomly initialized configuration. Measuring for each converged configuration the total prediction error over all sampled actions, it was possible to confirm with a high degree of certainty that the globally optimal configuration is the one where all 25 CDNs are arranged on a regular 5×5 grid. No other configuration with a smaller overall prediction error could be found for the setup addressed here.

Foveal Layout. When solving problem (4.15) using the foveal sensor layout under rotational actions and stimulus dilations, it is found that motor movement fields converges to a configuration where they are regularly distributed on concentric circles. This configuration can be seen in Fig. 5.8(c). The movement fields are aligned in three columns which correspond to three different dilation levels and

(a) Initial configuration of the motor movement fields (iter. 0).

(b) Prediction matrices after 1 iteration.

(c) Final configuration of the motor movement fields after 1 000 iteration. Note, CDN 14 was suppressed during the optimization, see also explanations in the text.

(d) Prediction matrices after 1 000 iterations. The numbering of prediction matrices corresponds to the numbering of movement fields on the left.

Figure 5.7: **Translations.** Initial and final configurations of an optimization problem as formulated in Eq. (4.15). The problem is iterated with the Levenberg–Marquardt algorithm and optimizes 25 motor movement fields located on a motor layer Φ_m covering translational actions. Left: Representation of the motor area Φ_m for translations where each point represents a shift relative to the sensor's original position. In grey, 5 000 sampled displacements used to train the model are shown. In black, the movement fields (visualized as ellipses) of each motor primitive (Σ, μ). Right: Prediction matrices $\mathbf{P}^k$ of nine selected motor movement fields. Each matrix is shown as a table where entries are color-coded in grayscale (black $= 0$, white $= 1$). Note, without any specific assumptions, motor movement fields converged to locations in the motor area which correspond to translational actions which match exact shifts of visual receptive fields, and which allow for particularly sparse prediction operators $\mathbf{P}^k$).

span 180° along the rotational dimension. As for the grid layout, the organization of movement fields on the motor area happens to reflect the spatial layout of the visual receptive fields in the sensor area.

In Fig. 5.8(d) the learned prediction matrices of nine selected motor primitives are shown. As for the grid layout, visual receptors which cannot be predicted accurately (e.g. the ones at the periphery of the sensor) happen to be approximated by their own previous value (diagonal matrix entries). Also, it is observed again that all $\mathbf{P}^k$ are exceptionally sparse.

Unlike the solution in Fig. 5.7(c) which converged to a local optima, the results presented for the foveal setup represent what can be suspected to be the globally optimal solution. This assumption is supported by the fact that no other solution could be found which led to a smaller overall prediction error. Thus, it is likely that the positioning of the movement fields shown in Fig. 5.8(c) corresponds to the globally optimal one.

Result Validation. In this paragraph the hypothesis is validated that for the discussed motor spaces and sensor topologies, multivariate Gaussians are a suitable model of motor movement fields. For this purpose, it is useful to first inspect the underlying function $\mathbf{P}(\mathbf{q})$. This is difficult since $\mathbf{P}(\mathbf{q})$ defines for every action $\mathbf{q}$ a matrix $\mathbf{P}^q$ of size $n_s \times n_s$. Nonetheless, to get an impression of what the trained model is actually approximating, a particular entry of this matrix was visualized for a number of random actions. In Fig. 5.9(a) the selected entry is plotted using prediction matrices $\mathbf{P}^q$ learned by linear regression from multiple samples for each action $\mathbf{q}$. For a comparison, Fig. 5.9(b) shows the same matrix entry obtained from our model $\sum_k m_k(\mathbf{q})\mathbf{P}^k$ using the parameters learned for the foveal layout. Comparing the two plots, two things become apparent: first of all, the values plotted in Fig. 5.9(a) resemble closely a multivariate Gaussian and are therefore well approximated by the linear interpolation shown in Fig. 5.9(b); secondly, even though a selected prediction matrix entry for the non-uniform sensor layout and the rotation-zoom action space was sampled, the resulting distribution resembles an axis-aligned Gaussian. The second observation justifies the previous decision to restrict $\boldsymbol{\Sigma}_k$ to be diagonal matrices. Thus, in conclusion, the Gaussian model approximates well the underlying function for the sensorimotor system considered here. However, it shall be noted that the applicability of this model also depends on the receptive field functions of the given sensor (which are Gaussian in this case). Also note that the constraint that movement fields must be multivariate Gaussians has been relaxed in problems (4.17) and (4.23). Corresponding results where arbitrary movement field functions develop are shown in Sect. 5.5.
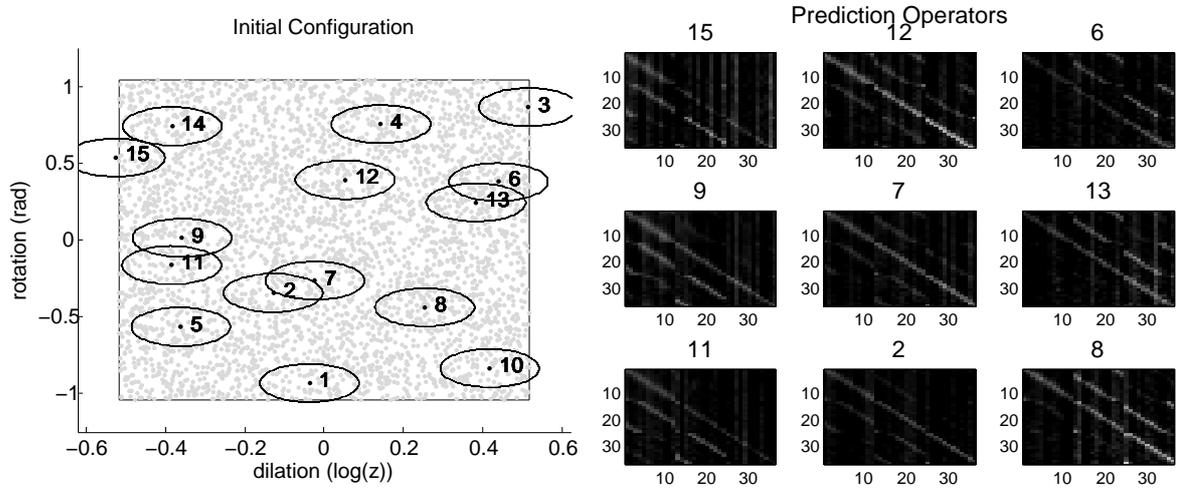
(a) Initial configuration of the motor movement fields (iter. 0).

(b) Prediction matrices after 1 iteration.
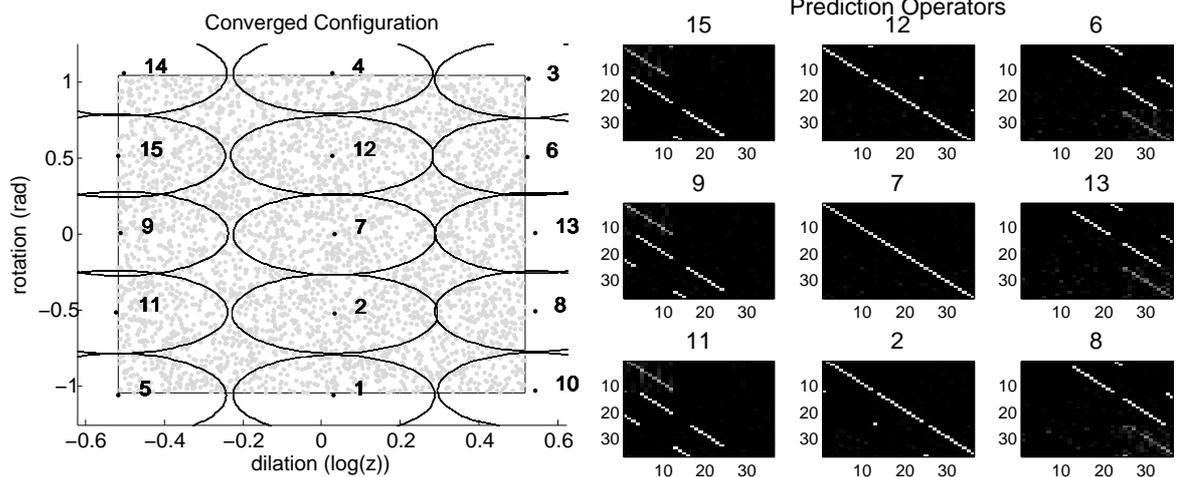
(c) Final configuration of the motor movement fields after 600 iterations.

(d) Prediction matrices after 600 iterations. The numbering of prediction matrices corresponds to the numbering of movement fields on the left.

Figure 5.8: **Rotation and Dilation.** Initial and final configurations of an optimization problem as formulated in Eq. 4.15. The problem is iterated with the Levenberg–Marquardt algorithm and optimizes 15 motor movement fields located on a motor layer Φ_m covering rotational and dilating actions. Left: Representation of the motor area Φ_m for rotations and dilations where each point represents a rotation and change in distance z relative to the sensor's original position. In grey, the sampled displacements used to train the model (5 000). In black, the movement fields (visualized as ellipses) of each motor primitive (Σ, μ). Right: Prediction matrices $\mathbf{P}^k$ of nine selected motor movement fields. Each matrix is shown as a table where entries are color-coded in grayscale (black $= 0$, white $= 1$). Note, without any specific assumptions, motor movement fields converged to locations in the motor area which correspond to rotational actions and dilations which match exact shifts of visual receptive fields, and which allow for particularly sparse prediction operators $\mathbf{P}^k$.

(a) Sampled   (b) Interpolated
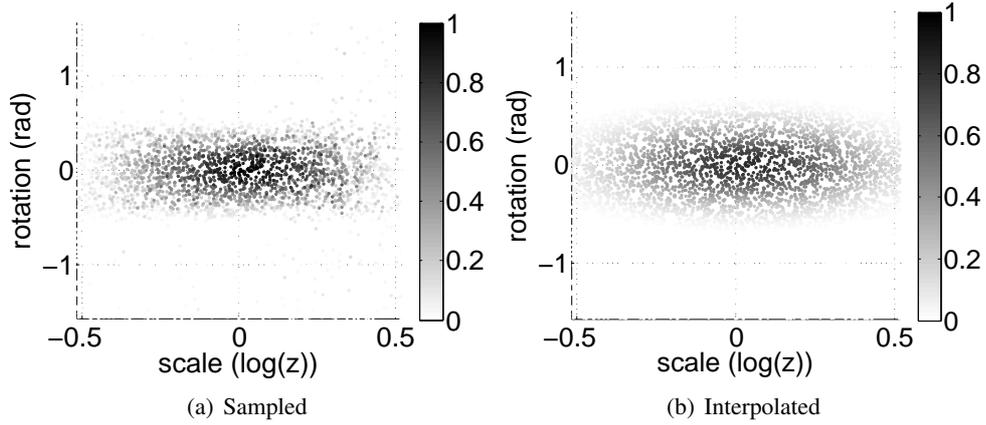
Figure 5.9: Activation of the receptor-receptor connection $(18, 18)$ in the foveal layout plotted over the rotation-zoom motor area. The left plot shows $\mathbf{P}_{(18,18)}$, where for each location $\mathbf{q}$ a $\mathbf{P}^q$ was explicitly computed by linear regression. The right plot shows $\mathbf{P}_{(18,18)}$ approximated by $\sum_k m_k(\mathbf{q})\mathbf{P}_{k(18,18)}$ with parameters learned as shown in Fig. 5.8(c) and 5.8(d).Other entries in $\mathbf{P}$ than $(18, 18)$ show similar activation distributions centered at different locations.

Conclusion. On the basis of the obtained results, it is observed that optimal motor layouts for problem (4.15) feature movement fields which have particularly sparse prediction matrices associated. Hence, the optimization discovers and takes advantage of locations in the action space where visual stimulus prediction can be done with an especially simple prediction model. Even though it is expected that the prediction operators be sparse in general, due to spatio-temporal constraints between visual receptors, the number of non-zero entries in linear prediction operators in the found solutions is sparser than expected. Since problem (4.15) can directly be interpreted as an implementation of the model of a corollary discharge circuit as deduced and depicted in Fig. 2.3, the results described in this section also influence the physical implementation of such a circuit. The fact that the obtained solutions tend to feature particularly sparse prediction matrices means the feed forward network of a corollary discharge circuit can be built with a particularly small number of neural connections. Thus, a majority of the feed forward connections shown at layer (b) in Fig. 2.3 can be removed. Furthermore, in a computational sense, the calculation of predicted stimuli requires a smaller number of operations due to the reduced number of connections.

5.4   Organizing Sensor Topologies

This section presents results for the optimization of sensor topologies according to problem (4.21). Because problem (4.22) is part of the unified approach formulated in problem (4.23), sensor topologies

(a) Converged Matrix $\mathbf{S}^*$ (b) Topology of $\mathbf{S}^*$ on the Sensor Surface
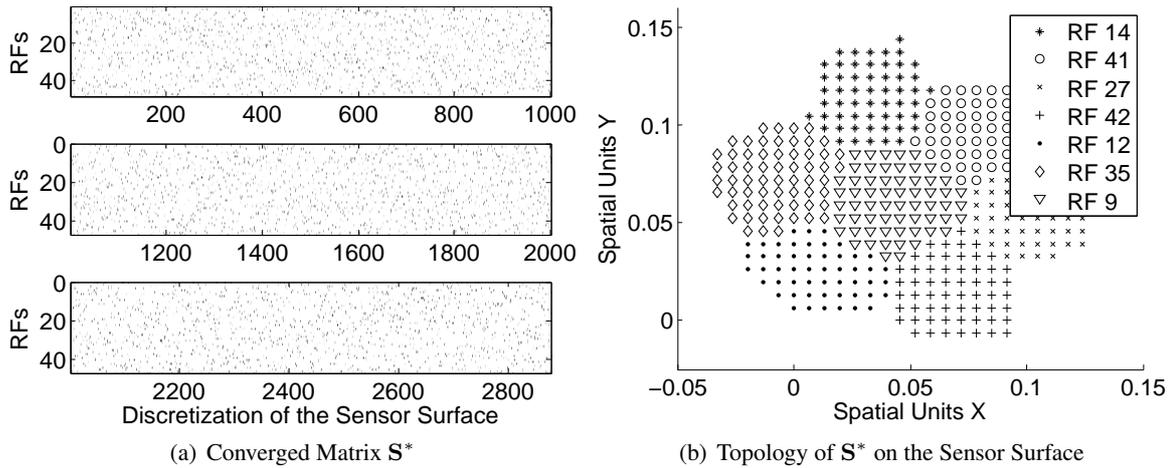
Figure 5.10: Emergent clustering of receptive fields. Left: A converged but topologically orderless matrix $\mathbf{S}$ as seen by the algorithm; each entry specifies the contribution of a location on the sensor surface to a receptive field; the sensor area is discretized into 2877 pixels (x-axis), and the matrix $\mathbf{S}$ codes for 48 receptive fields. Right: The sensor area and the coverage of 7 selected receptive fields at spatial locations where their contribution is predominant; this view reveals the implicitly present topological clustering in $\mathbf{S}$.

obtained on the basis of problem (4.23) are presented in Sect. 5.5.

To demonstrate the correlation principle introduced in Eq.(4.18), first the results for an immobile agent are shown. This example – although for now discarding any meaningful behavior – shows a crucial capability of the proposed method, namely the development of spatially coherent visual receptive fields. Figure 5.10 highlights the discovery of topological order from the orderless sampling of underlying visual signals. The figure takes the privileged perspective of an external observer. From this perspective, the spatial locations where the sensor surface was sampled are known and as such it is possible to plot the topological ordering of receptive fields on the sensor area as shown in Fig. 5.10(b). In this 2-dimensional visualization it was chosen to show at each discrete sensor area location the predominant receptor. The clustering property of the receptive field elements is clearly demonstrated. Since in this case no action is taken, this clustering is a sole consequence of the interaction between the correlation based cost function of problem (4.21) and the low spatial frequency of the observed environment (which is characteristic for natural images). Note that the agent does not have access to the sampling locations of the sensor surface and is thus unaware of the final topological ordering. The proposed algorithm operates solely on matrix $\mathbf{S}$ which is absent of any topological meaning even in the final converged state. For illustration purposes, the matrix $\mathbf{S}$ is shown in Fig. 5.10(a).

For active agents two different behaviors as shown in Fig. 5.11(a) and Fig. 5.11(b) are considered. The first consists of a uniform action probability distribution of 2-dimensional translations over the

(a) Translation   (b) Rotation and Dilation
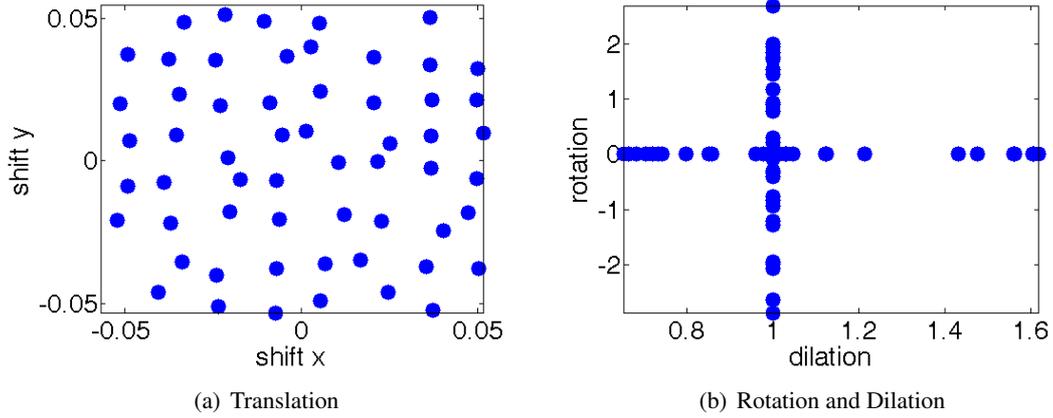
Figure 5.11: Two different behaviors represented as action distributions. Left: uniform 2-dimensional translations in a given range covering 10 times the distance between discrete sampling locations on the sensor surface in each direction. Shift units are normalized with respect to the environment. Right: independent zoom and rotation actions distributed uniformly on each axis. Rotations are given in radians and dilations are given as a scale factor. Both operate with respect to the center of the sensor surface. Zoom actions range from 0.6 to 1.66 and rotations cover $-\pi$ to π.

sensor area in a given range. This scenario relates to translational unbiased oculomotor control causing random stimulus displacements. The second behavior is composed of independent zoom and rotation actions distributed uniformly on each axis. This mimics the behavior of an object manipulating agent where the oculomotor system stabilizes the sensor on target, mechanically compensating for image translations but not image rotations or scaling. These setups demonstrate that the agent's behavior induces different topologies of receptive fields on the sensor surface. In Fig. 5.12 the converged layouts for the two considered action distributions are shown. The nature of the two converged topologies exhibits macroscopic differences: in the translation only case a tendency for hexagonal tiling structures over the entire sensor surface can be identified (apart from boundary effects), whereas in the rotation and zoom case the receptors organize radially in clear circular rings. Unlike in Fig. 5.10(b), the 3-dimensional perspective shows the smooth overlapping between receptive field elements. Such overlap is obtained for $\omega > 0$ in the C-term of Eq. (4.21).

To better comprehend the resulting sensor layouts, the reader is referred back to the work of Clippingdale and Wilson [21] reviewed in the last part of Sect. 1.3, where the fitness of a layout relates directly to the distance between predicted and original point locations. In the problem considered here, just as in [21], a perfect sensor layout is one where receptors exactly map one onto another for every considered action resulting in $\mathbf{P}_{\mathrm{pls}}^{q}$ matrices where each row contains exactly one non-zero entry. Any deviation from this case leads to an increase in prediction error and lowers correlation. This fact allows

87
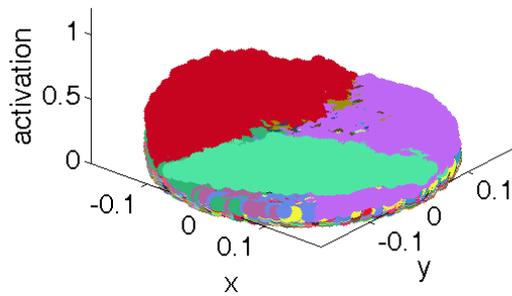
(a) Translation (Iteration 1)
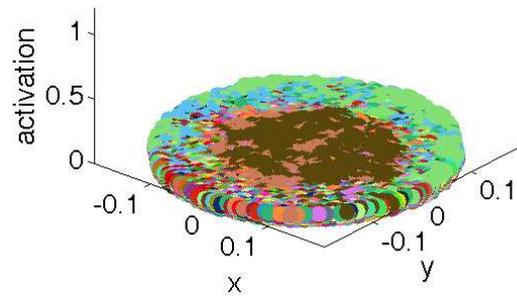
(b) Rotation and Dilation (Iteration 1)
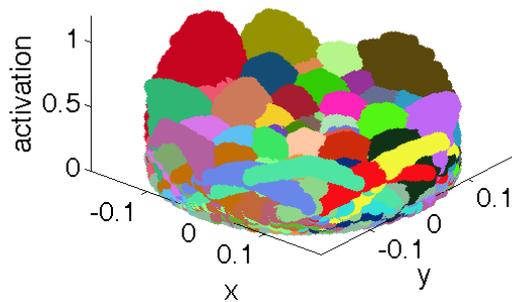
(c) Translation (Iteration 7)
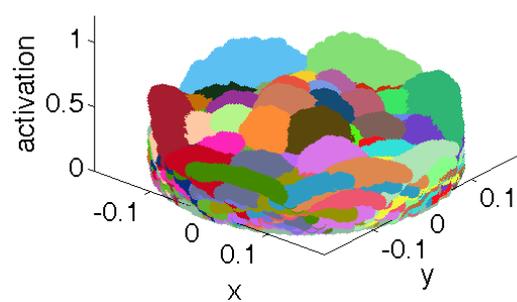
(d) Rotation and Dilation (Iteration 7)
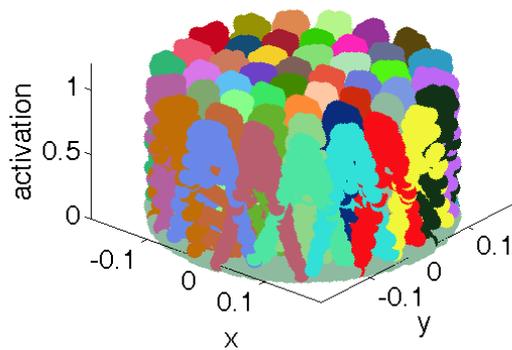
(e) Translation (Iteration 55)
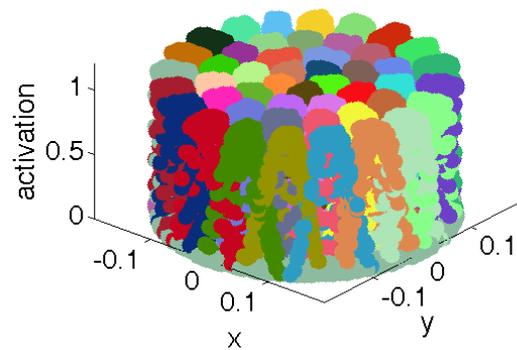
(f) Rotation and Dilation (Iteration 40)

Figure 5.12: Evolution of sensor topologies optimized for behaviors visualized in Fig. 5.11(a) and Fig. 5.11(b). Left: translations. Right: dilation and rotation. Converged configurations are shown in the last row, figures (e) and (f).

for the replacement of the Euclidean distance as used by Clippingdale and Wilson by one based solely on correlation between sensory readings disregarding any knowledge about the sensor topology.

5.5   Organizing Sensor and Motor Topologies Conjointly

This section presents results for a joint optimization of sensor and motor topologies according to problem (4.7). On the basis of the 4-dimensional action space introduced in Sect. 5.1, two different sets of sensorimotor experiences, $\mathcal{B}_1$ and $\mathcal{B}_2$, are considered. These two sets of sensorimotor experiences are used to co-develop two sets of sensor and motor topologies $\mathbf{S}_1^*$, $\mathbf{S}_2^*$ and $\mathbf{M}_1^*$, $\mathbf{M}_2^*$. In a first setup, $\mathcal{B}_1$ is recorded using sensor translation actions sampled from a 2-dimensional motor space as shown in Fig. 5.2(b). Triplets $(\mathbf{i}_0, \mathbf{i}_1, \mathbf{q})$ in $\mathcal{B}_1$ are sampled choosing actions $\mathbf{q}$ with uniform probability from the available discrete actions. This scenario relates to translational unbiased oculomotor control causing random stimulus displacements. The second behavior is composed of mixed zoom and rotation actions where $\mathcal{B}_2$ samples combined sensor rotations and stimulus dilations from a 2-dimensional motor space as shown in Fig. 5.2(c). As for $\mathcal{B}_1$, triplets $(\mathbf{i}_0, \mathbf{i}_1, \mathbf{q})$ were sampled with uniform probability from the available discrete actions. Behavior $\mathcal{B}_2$ mimics, for example, an object manipulating agent where the oculomotor system stabilizes the sensor on target, mechanically compensating for image translations but not image rotations or scaling. Note, different from the sensorimotor experience recorded according to the actions shown in Fig. 5.11(b), the set of sensorimotor experiences $\mathcal{B}_2$ considered in this setup is composed of combined dilation and rotation actions sampling the entire motor space shown in Fig. 5.2(c). The resulting sensor and motor topologies $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{M}_1$, $\mathbf{M}_2$ are shown in Fig. 5.13 and Fig. 5.14. The results demonstrate that different behaviors $\mathcal{B}_1$ and $\mathcal{B}_2$ induce sensorimotor structures of different macroscopic nature. Note that even though the proposed algorithm is unaware of the topological order present in recorded stimuli $\mathbf{i}$, visual receptors cluster as smoothly overlapping receptive fields and motor primitives appear as spatially coherent Gaussian-like areas.

Development of Spatially Coherent Motor Movement Fields   Previous results on the organization of motor topologies presented in Sect. 5.3 confirmed the evolution of a topological arrangement of movement fields according to given sensorimotor experience. However, movement fields m_k shown in Fig. 5.7 and Fig. 5.8 are modeled according to multivariate Gaussians. Thus, while on the sensor side, the results on the organization of visual receptive fields presented in Sect. 5.4 showed that it is possible to develop compact visual receptive fields considering a more general discretization, the problem of forming spatially coherent movement fields using a grid-like discretization has been omit-

ted so far on the motor side. The results discussed in this section demonstrate now, that when solving problem (4.23) using a grid-like discretization of the motor topology, spatially coherent movement fields evolve concurrently with compact visual receptive fields. The reason for this tendency towards spatially coherent movement fields can be found in the request for an accurate prediction of future stimuli $\mathbf{i}_1$ in problem (4.23): only if motor movement fields $\mathbf{m}_k$ specialize on spatially limited areas, accurate linear combinations of prediction operators $\mathbf{P}^q$ for particular actions $\mathbf{q}^q$ are possible. At the same time, the optimization problem forces movement fields to cover the entire motor area Φ_m which experiences samples $\mathbf{q}$ in order to be able to assemble predictors $\mathbf{P}^q$ for every $\mathbf{q}^q$.

Conclusion. The unified approach for the co-development of visual sensor and motor structures is based on two main hypotheses. The first (Sect. 3.7 and Sect. 3.8) proposes that sensorimotor structures can develop according to a general cost function c_{sm} where the agent's behavior and environment are decoupled and enter the problem as the agent's overall sensorimotor experience $\mathcal{B}$ as in

$$(S^*, M^*) = \operatorname{argmin} \quad [c_{sm}(S, M; \mathcal{B})] \quad .$$
$$\text{s.t.} \quad \mathcal{B} = \{(\mathbf{i}_0, \mathbf{i}_1, \mathbf{q})\}$$

The second hypothesis proposes that sensor and motor topologies $\mathbf{S}$ and $\mathbf{M}$ evolve such as to optimize i) the reconstruction of higher dimensional signals, and ii) stimulus predictability. Per se, it is not clear if the introduced hypotheses are justifiable. However, the proposed framework is capable of reproducing some characteristics of in-nature observed sensorimotor structures, and captures inherent principles present in phylogenetic and or ontogenetic development of biological systems. Therefore, even though the true evolutionary cost function is unknown, it might be claimed that the made assumptions could hold, and that the proposed framework with its simple underlying principles has explanatory power not found in other computational models.

(a) $\mathbf{S}_1$, iteration 1

(b) $\mathbf{M}_1$, iteration 1

(c) $\mathbf{S}_1$, iteration 80

(d) $\mathbf{M}_1$, iteration 80

(e) $\mathbf{S}_1$, iteration 400

(f) $\mathbf{M}_1$, iteration 400
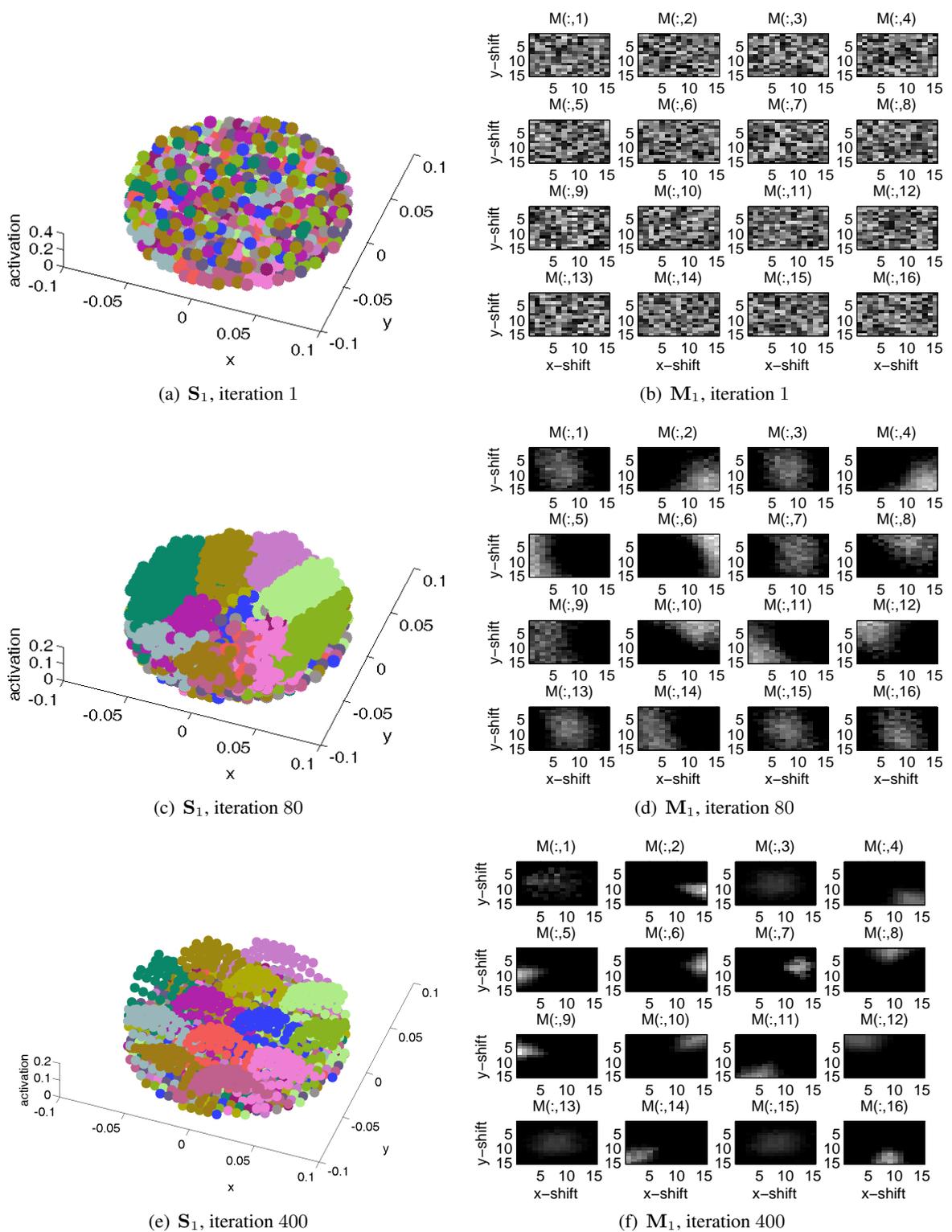
Figure 5.13: Sensor and motor topologies obtained for translation actions uniformly sampled from a motor space as shown in Fig. 5.2(b). Left: the evolution of $\mathbf{S}$. Each color denotes a different visual receptive field, and each dot shows the activation of that field at the respective location on the sensor area. Right: the evolution of $\mathbf{M}$. Note, some motor fields happen to overlap, and therefore appear less pronounced as their contribution is combined in problem (4.23) according to Eq. (4.16).

(a) $\mathbf{S}_2$, iteration 1

(b) $\mathbf{M}_2$, iteration 1

(c) $\mathbf{S}_2$, iteration 80

(d) $\mathbf{M}_2$, iteration 80

(e) $\mathbf{S}_2$, iteration 400

(f) $\mathbf{M}_2$, iteration 400
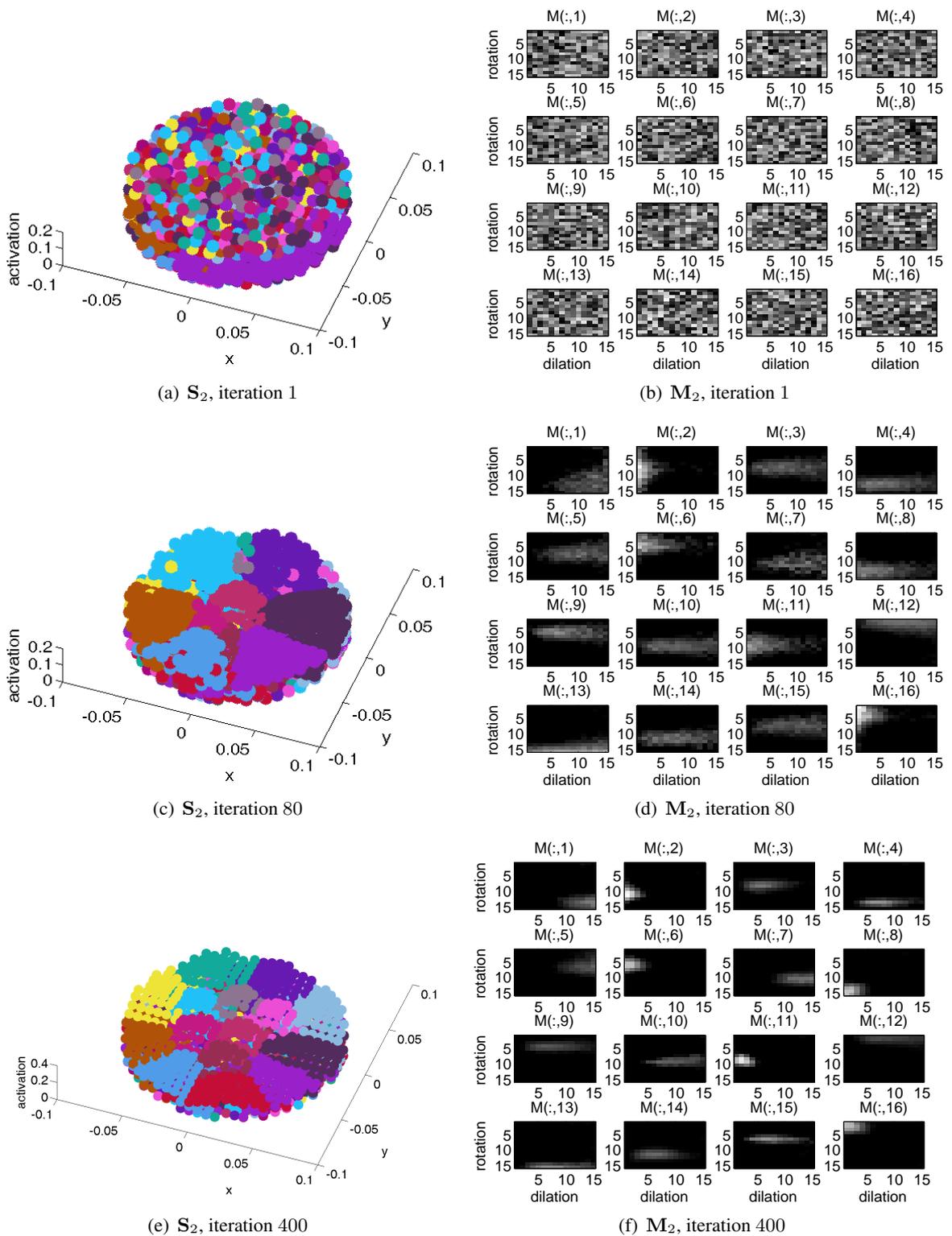
Figure 5.14: Sensor and motor topologies obtained for rotation and dilation actions uniformly sampled from a motor space as shown in Fig. 5.2(c). Left: the evolution of $\mathbf{S}$. Each color denotes a different visual receptive field, and each dot shows the activation of that field at the respective location on the sensor area. Right: the evolution of $\mathbf{M}$. In this case, elongated elliptic fields develop reflecting the higher axial resolution of sensor $\mathbf{S}_2^*$ compared to its radial resolution.

Chapter 6

Discussion & Outlook

Contents

Efficient autonomous robotic systems require a body which is highly adapted to the system's particular task and environment. In biology, a common line of thinking proposes that adaptation implicitly optimizes some underlying criterion which is related to the overall fitness of the organism. However, applying such optimization methods to the design of entire artificial systems is not straight forward. The criteria underlying adaptive processes in biological systems is in general unknown, of considerable complexity, or impossible to evaluate within a reasonable timescale.

Addressing the development of a computational method for the automated design of behavior-dependent visual sensorimotor structures, this thesis proposed that it is possible to isolate a simple and at the same time computationally tractable criterion encoding principal characteristics of visual sensorimotor layouts observed in living organisms.

The two main hypotheses evaluated throughout the work are, i) an animal's lifelong sensorimotor experience is the principal driving force for the development of sensorimotor layouts, and ii) the criterion optimized throughout the development of visual sensorimotor layers is related to simple stimulus prediction models, which for visual sensors is shown to be related to sparse linear predictors and self-similarity in stimulus transformations. These hypotheses were tested for different evaluation criteria applied to both, the organization of sensor layouts, as well as the organization of motor topologies.

Exploiting a number of analogies between sensory and motor systems, the thesis eventually proposed a solution which addresses the organization of sensor and motor layouts in a unified manner.

93

The developed method can serve in two ways: Firstly, it can be used to synthesize sensorimotor layouts for artificial visual systems and secondly, it represents a model capable of explaining morphological aspects of biological visual sensorimotor systems.

6.1 Discussion

This section comments on a number of points addressed throughout this thesis. Some points resumed here have been addressed conclusively, others are still open for discussion or relate to limitations in the proposed approach.

Violating the Assumption of Perfectly Predictable Stimuli. An important assumption associated to the problem formulated in Chap. 3 is the perfect predictability of future visual signals i for executed motor actions q, see also point (5) in Sect. 3.9. However, this assumption can be restrictive, and in fact not all presented results have been obtained in experiments which adhere to this assumption. In particular actions which translate the sensor violate the request for perfect predictability since after taking the action, the boundary of the sensor records information which was not available before taking the action. For such actions, the fraction of unpredictable sensor area increases with the step size of the action. A similar effect appears in 3-dimensional environments for agent states x and actions q which lead to motion parallax in the projected signals i. This has been previously discussed in Sect. 3.5. Although situations leading to motion parallax do not arise with a problem instance as introduced in Sect. 5.1, the experience with unpredictable stimuli at the periphery of investigated sensor areas, has shown that the proposed solution is robust with respect to unpredictable stimuli. Visual signals which are (partly) unpredictable with the considered linear model merely increase the level of noise present when learning prediction operators $\mathbf{P}_k$ or $\mathbf{P}_{\text{pls}}^q$. For unpredictable peripheral visual receptors (e.g. under translation actions), prediction operators have been observed to converge to configurations where those receptors estimate their future activation value equal to the currently experienced signal. Considering the predominantly low spatial frequency of natural images, this makes sense as a best guess for unpredictable receptors and actions of limited step size. According to these observations, it can be expected that the proposed solution will also be robust with respect to partly unpredictable stimuli due to motion parallax.

Motor movement fields. In Sect. 3.4 which describes the observation and action model, it has been defined by Eq. (3.3) that from a given motor activation $\mathbf{a}$, a low-level action q is sampled via motor

movement fields m_k as $q = \epsilon \left[\sum_k \mathbf{a}_k m_k \right]$ where ϵ is a sampling function and actions $\mathbf{a}$ obey $\sum_k \mathbf{a}_k = 1$. In the opposite direction, movement fields m_k estimate motor activation $\tilde{\mathbf{a}}$ in the sense of corollary discharge as described in Eq. (3.6). While the latter is perfectly dual to the integration of visual stimuli on the sensor side as described by Eq. (3.1), a similar analogy between the generation of low-level actions q from activations $\mathbf{a}$ and a process on the sensor side is not equally clear. However, it could be argued that on the sensor side, the dual operation to generating low-level actions q is the reconstruction of visual signals i. In this case, a duality between a projection operator S^+, as discussed in Sect. 4.3, and the generation of actions q from motor activation $\mathbf{a}$ as defined by Eq. (3.3) is expected. Along these lines, it could be argued further that Eq. (3.3) should consider an inverse operator M^+ which generates some combined probability density function imposed on the higher dimensional motor space of actions q. This probability density function would not exactly correspond to the linear combination of movement field functions m_k as proposed by Eq. (3.3) which can be considered a simplified approximation of the true probability density function similar to the adjoint operator $\mathbf{S}^\top$ approximating the optimal reconstruction $\mathbf{S}^\top \left(\mathbf{SS}^\top \right)$. However, more work is required to formally justify an approximation for an operator M^+ dual to S^+.

Limitations. Two important topics which recurrently surfaced while working on this thesis, but proved to be hard to tackle, are i) agents, environments, and actions which induce state-dependent changes in sensory stimuli, and ii) the encoding of a task in the considered agent which features a temporally variable sensorimotor system. These two points are still unsolved. A summary is provided in what follows. For related directions in future work, see also Sect. 6.2.

- **Encoding of the Agent's Task:** Assumption (4) in Sect. 3.9 states that the considered agent follows a given action selection policy B. Given this policy B, the input to problems addressed in this work is a set of sensorimotor experiences $\mathcal{B}$ sampled in a given environment E according to Eq. (3.10). With assumption (4), the behavior of the agent enters the problems addressed through the given action selection policy B. This is in accordance with the hypothesis that sensorimotor structures develop according to the agent's overall sensorimotor experience as formulated in Sect. 3.7. However, for robotic applications it is in general difficult to specify B. Typically it is desirable to implicitly induce a policy B by defining a task at a more abstract level and using e.g. a reward driven learning strategy. Furthermore, considering the addressed problem at a larger scope, it can be questioned if it is suitable to encode behavior at the level of an action selection policy for an agent with a temporally variable morphology. A more abstract encoding of the agent's task is thus considered an open problem which has to be addressed in

future work.

- **State-dependent Stimulus Changes:** Assumption (5) in Sect. 3.9 states that future visual stimuli i_1 must be predictable solely from previous stimuli i_0 and executed actions q irrespectively of the agent's state x. This assumption is crucial in this work and allows for prediction functions $p(q, \mathbf{o}_x)$ independent of x as introduced in Eq. (3.9). However, in most robotic setups, changes in recorded stimuli are often dependent on the system's state. In these cases, state-independence needs to be dropped from assumption (5) in Sect. 3.9 and prediction functions become dependent on x. A possible approach in this direction is outlined in the next section under the topic termed *state dependent sensorimotor contingencies*.

6.2   Future Work

This section proposes a number of topics which are beyond the scope of this thesis but are considered branches of future work well worth exploring.

Other Sensory Modalities.   This work explored the organization of *visual* sensorimotor structures. However, nothing constrains the proposed concept to visual perception. Other sensory modalities could be considered without changes to the principal concept just by replacing the observation and action model. For example, an implementation for an auditory sensorimotor system can be imagined where sensory stimuli and motor actions are recorded and emitted in a frequency domain. In this case, receptive elements and motor primitives would specialize on certain frequency bands depending on the interaction of the agent's body and its environment with respect to acoustic properties. Picturing a frequency space mapped onto a spiral inspired by the cochlea found in the mammalian auditory system, Fig. 6.1 sketches the idea of an auditory sensor by illustrating a number of auditory receptive fields placed on the spiral shaped sensor area.

Hierarchical Organization of Sensorimotor Structures.   In the present work, usually the intuition is given that visual signals $\mathbf{i}$ are directly recorded at the periphery of the agent, and motor actions $\mathbf{q}$ are at the lowest level of topologically coded motor signals. However, this does not necessarily need to be the case. In an approach where several layers of sensorimotor structures $\mathbf{S}$ and $\mathbf{M}$ are considered, signals $\mathbf{i}$ and $\mathbf{q}$ could be located at an arbitrary depth along the sensorimotor pathway. In this case, the idea is that sensory and motor primitives in sensorimotor layers further from the agent's periphery would code for more and more abstract stimuli and motor actions. Such structures with layered levels
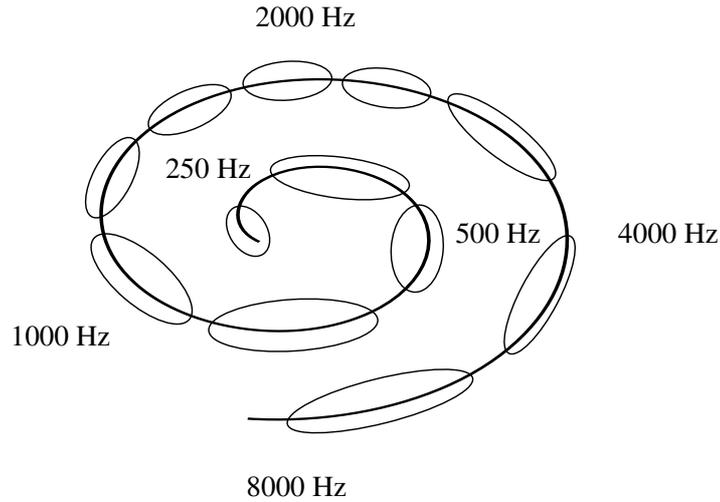
Figure 6.1: The concept for the organization of sensorimotor structures as proposed by this thesis could be applied to other sensory modalities. The illustration sketches twelve auditory receptive fields covering a given frequency range with increased resolution around 2000Hz.

of abstraction have been extensively studied in nervous systems of living organisms, see e.g. [52]. A hierarchical approach is also imaginable to support state-dependent sensorimotor contingencies as proposed for investigation in the next paragraph.

State Dependent Sensorimotor Contingencies. Sensorimotor contingencies are considered to be laws governing the sensory changes induced by motor actions [81]. In other words, a contingency is a descriptor for the changes observed when an action is taken in a certain situation. Sensorimotor theory as promoted in [81] suggests that an agent equipped with a number of learned sensorimotor contingencies can inquire about its state by probing which sensorimotor contingencies are obeyed in its current situation. In this work, a prediction operator $\mathbf{P}^q$ can be considered a sensorimotor contingency for the action q being true irrespective of the state x the agent is in. This independence of predictors $\mathbf{P}^q$ from the agent's state is deliberate in this work. In this sense, descriptors $\mathbf{P}^q$ are general *state-independent* sensorimotor contingencies of the considered agent. *State-dependent* sensorimotor contingencies naturally appear if state-dependent prediction functions $p^q(\mathbf{o}_x, x)$ are introduced, e.g. when attempting to overcome the limitation imposed by assumption (4) from Sect. 3.9. Therefore, relaxing the assumption for state-independent stimulus changes and working towards an implementation of a perceptual system based on evaluating state-dependent sensorimotor contingencies go naturally together. In a concrete implementation this means, depending on the state of the agent, changes in sensory stimuli are described by a different operator p^q for a given action q. A first attempt in this direction might try to deal with 3-dimensional depth perception where different states x relate to the

distance of the agent to an observed scene. For a biological prototype employing such a strategy of active depth perception, see also the Praying Mantis described in Appendix A.

Task Encoding. As outlined in the previous section, a particular difficulty encountered in this work is the encoding of a task which would allow for a more flexible definition of the agent's behavior. Ultimately, this problem was postponed by considering a given action selection policy according to which sensorimotor experience is sampled. However, in future work, it is desirable to induce the agent's behavior in a more abstract way, e.g. via a given task. In the literature, a wide variety of techniques and methods to evolve task-specific behaviors have been described but, their application to the problem at hand is in general difficult for a number of reasons. On the one hand, one might try to employ a type of reinforcement learning method which directly rewards particular situations encountered by the agent. Apart from typically encountered problems such as hard-to-find desired states and therefore slow learning processes, another major difficulty arises for agents with variable morphologies: it is challenging to define desired states. In particular when trying to formulate rewards from an agent intrinsic perspective, a changing body complicates the situation considerably. Trying to steer clear of these problems, one might try to induce behavior via a more abstract approach which simply rewards "survival" or a similar criterion related to the functional subsistence of the agent. Considering biological evolution, such an approach bears great potential, both for a flexible development of the agent's body as well as for its behavior [107]. However, in the end it is unlikely that this approach eradicates arising difficulties. A principal problem is that, requiring rewards to be based on the evaluation of an agent's entire lifespan poses big problems with respect to computational and temporal resources. Furthermore, assigning rewards at a more abstract level naturally results in a reduced ability to define a specific task. Although, it is the author's belief, that when the adaptivity of an agent's body is increased, the rigidity of the agent's task definition must be relaxed simultaneously.

Appendix A

Visual Sensorimotor Systems in Nature

During the course of this work, a collection of animals with specifically adapted visual sensorimotor systems for visual perception have been reviewed. These are listed in this chapter along with a short description of their visual sensorimotor system. Of course, the present list can represent only a very limited selection from the many different visual systems found in living organisms. It is restricted to some examples which appeared to be of particular interest at different stages of this work. Thus, the list does not aspire to be conclusive in any way nor is it presented in any particular order. For a more in depth introduction on the evolution of visual perception, see e.g. [115] or [82]. Here, the intent is to provide the reader with a few samples contributing to a broad perspective on adaptation in visual sensorimotor systems. In accordance with the arguments in favor of "simple brains" as presented at the beginning of this thesis, this perspective essentially reveals that less complex animals tend to develop more specifically adapted visual sensorimotor systems whereas animals with a larger nervous system tend to possess a visual sensorimotor apparatus which supports more general usage strategies.

Trilobites. Fossils of trilobite eyes represent today the oldest preserved visual system [20, 126]. They date back to the early Cambrian period and have been found in many different species of the highly successful class of Trilobita which existed over more than 250 million years. Similar to the compound eyes of modern arthropods, their visual organs were composed of individual *eyelets* ranging in number from a few dozen up to several thousand. Each of these eyelets captured light through a small lens made of calcite. These calcite structures have been preserved in fossil records and have enabled paleontologists to reconstruct the shape of trilobite eyes. Two examples are shown in Fig. 1.6. As can be seen there, a big diversity in eye morphology existed between different trilobite species. Some of them show lens topologies where equally sized eyelets are arranged according to a hexagonal,

or less frequantly square packing. Other trilobite eyes have more complex but still highly regular sensor topologies, e.g. eyes where receptors are placed in patterns of intersecting logarithmic spirals as described in Sect. 1.3.

How trilobites moved their visual sensors with respect to the environment is mostly unknown. Fossils basically reveal that most trilobite eyes were fixed with respect to the front part of the animal's body, except for some species which developed eye stalks. In terms of visual sensorimotor properties relevant in the context of the present work, the configuration of sensory organs and motor apparatus can be considered similar to the one found in modern insects. In both classes of animals the visual sensor is composed of photoreceptors arranged on a convex surface and is moved together with the body of the animal. Thus, it can be expected that the characteristics of raw stimulus changes, induced for example through locomotion in free swimming or floating trilobites, is in principle comparable to the optic flow patterns processed by the visual system of flying insects. Interestingly, however, recent evidence suggests that some trilobite eyes might have implemented a system of muscles which was able to move the receptor areas inside each eyelet with respect to the fixed lenses [112]. But, if such a system might have allowed the animal to execute some kind of saccading behavior or image stabilization is currently unknown (B. Schoenemann, personal communication, July 2012). Today, the closest extant relative of trilobites is the horseshoe crab (*Limulus*). Its visual sensor consists of about 1000 ommatidia and is particularly well suited for neurophysiological studies. For a review on the extensive work done by Hartline and Ratliff in this area, see e.g. [7].

Nematodes. Several nematode species (roundworms) exhibit *positive phototaxis* which is the ability to discover a light source and direct their body movement towards it. To achieve this, the worms use a simple but effective sensorimotor strategy. They continuously oscillate the anterior tip of their body in order to sense a gradient in light intensity. With this strategy the worms manage to navigate through difficult terrain covered by grass or other low vegetation using only a couple of photoreceptors. In [16] for example, Burr et al. describe this behavior for the gravid female nematode *Mermis nigrescens* which attempts to move towards a bright area to place its eggs at locations where grasshoppers feed. For a detailed description of the visual sensor of *Mermis nigrescens* see also [70].

From the perspective taken in this work, it can be argued that the worm augments the limited capabilities of its sensory system by combining it with appropriate motor actions such that it can perceive a phenomenon which otherwise cannot be assessed by solely analyzing sensory feedback from an immobile position. By moving its sensor, the worm can detect a light gradient with a very limited sensor and minimal signal processing. In fact, the visual sensor of *Mermis nigrescens* consists

of only two photoreceptors and its nervous system is composed of 302 neurons.

The Copepod *Copilia*. Copepods are group of small crustaceans which live in sea and freshwater habitats and are about one to two millimeters long. Some are planctonic, and thus float passively in the sea, while others are parasitic and attach themselves to other animals. Still others are predators of their smaller relatives. To the later group belongs *Copilia*, a tiny but ferocious predator with a sophisticated visual sensorimotor system. Similar to the strategy employed by the nematodes discussed in the previous paragraph, it employs an active signal recording strategy using two very limited visual sensors which it moves in fast scanning movements horizontally across its field of view. Fig. A.1 illustrates Copilia's visual sensorimotor system. Each of the two sensors consists only of seven visual receptors. Both of them are placed inside the animals body and recording light which is guided through two fixed lenses located at the front of the animal. A single muscle pulls the two sensors rhythmically together and generates the horizontal scanning movement. In this way the animal achieves to add "an additional dimension" to the otherwise point-like vision of it sensors. By scanning along a horizontal line the animal can detect and eat plankton which passes in front of it when migrating in a vertical direction. The eye of Copilia has been described by several authors, among them [34, 42, 29, 114].
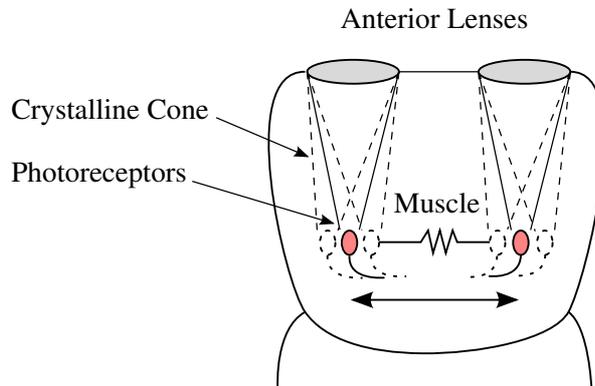
Figure A.1: Sketch of Copilia's visual sensorimotor system seen from above. Ellipses marked in red denote light receptive areas inside the animal. Each is composed of seven photoreceptors. The two anterior lenses collecting light from the environment are shown in gray. The photoreceptors of the animal's light receptive structures are embedded in a crystalline cone which is suspended in a system of ligaments and muscles. By contracting those muscles, the animal can move its photoreceptors generating a horizontal scanning movement with a "saw-tooth" temporal profile.

Similar to *Copilia*, the copepod *Labidocera* has two two eyes composed of only ten visual receptors each and executes a scanning movement covering about 35° of the dorsal visual field. In addition to these rhythmic scanning movements the eyes of *Labidocera* also execute a kind of stabilization movement coupled with the tail of the animal which might serve to compensate for movements to

which the animal is exposed while floating in the open water. See also [61].

The Seasnail *Oxygyrus*.   The *Oxygyrus* is a heteropod mollusc, a sea snail which also uses a scanning strategy for visual perception. However, different from the copepod *Copilia* which scans along a horizontal line with a point-like sensor, *Oxygyrus* added an additional dimension to its field of view by scanning in two-dimensions with a one-dimensional sensor. Its eye consists of a horizontal band of about 410 by 6 receptors which in some species expands along a straight line and in others is formed in a horseshoe shape. To find food, the snail moves this band of receptors continuously up and down [60]. Fig. A.2 illustrates this mechanism.
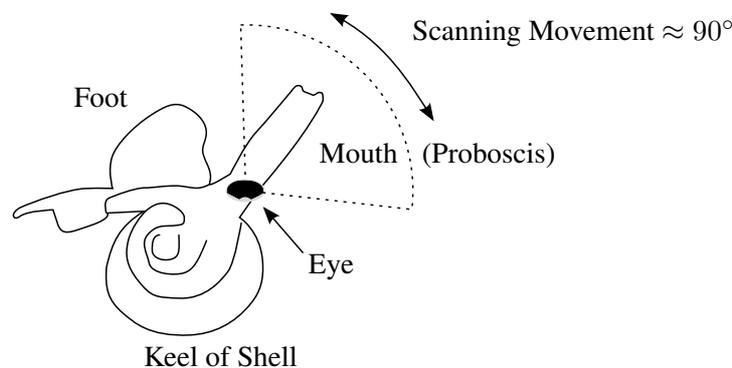
Figure A.2: Sketch of *Oxygyrus* in its normal orientation (swimming upside down). The elongated part in the upper right corner of the image is the animal's mouth. The black area indicates its left eye.

The Mantis Shrimp (*Stomatopoda*).   Mantis Shrimps or Stomatopods (not really shrimps) are crustaceans with a complex compound eyes which feature a horizontal band of color sensitive. They can move their eyes independently and with three different degrees of freedom: horizontal, vertical, and rotation. Their eye movements are attributed to three different types of movements, i) target tracking, ii) scanning, and iii) stabilization via a optokinetic nystagmus. The second type of movements, scanning is most often executed in a direction approximately perpendicular to the color sensitive band of receptors. Thus, although mantis shrimps have a two-dimensional receptor surface, their color vision is limited to one-dimension but is extended to two-dimensions via specific sensorimotor patterns, see [57].

The Fiddler Crab (*Uca*).   The Fiddler crab has two compound eyes, each of them located on a eyestalk. Contrary to typical vision systems, the center of the eyes points at the sky while the periphery of the eyes observes the crab's direct surroundings. Not surprisingly, the receptor density distribution of

these eyes is very different compared to eyes featuring e.g. a central fovea. In fact exactly conversely, resolution is highest around the eye's edges and lowest in the eye's center. This is useful for the crab since it uses the periphery of its visual sensor to detect other crabs and the center to observe the sky where it is highly sensitive to movements in order to detect airborne predators [138].

The Praying Mantis (*Mantodea*). The compound eyes of insects are fixed to the animal's exoskeleton. Thus, animals with compound eyes are not able to estimate distance using convergent eye movements or lens accommodation as animals with binocular camera-type visual systems do. Instead, some species make use of a behavior referred to as *peering movements* to estimate depth. By translating their head and body from side to side, these animals slightly change their perspective with respect to a target object and estimate its distance by observing the effects of induced motion parallax. Such movements can be particularly well observed in some locust species like the *gregarious locusts* or the praying mantis. These species execute clearly visible pendulum movements of head and body before jumping or attacking a target. Peering movements have also been observed in other insects like crickets, the fruit fly *Rhagioletis cerasi* or the wasp *Mellinus arvensis*. For a review on the topic see [56].

Jumping Spiders (*Salticidae*). Instead of building webs, jumping spiders hunt their prey by attacking it with an accurate jump. For this purpose they heavily rely on their visual capabilities. In total they feature eight camera-type eyes. While the lateral and much simpler eyes are mainly used for motion detection, the much larger anterior eyes evolved a very particular architecture which is highly specialized to the purpose of the particular visual task faced by jumping spiders. The lenses of these eyes is fixed in the spider's carapace, but the retina inside the eye is mobile and can move horizontally and vertically in a large range. Furthermore, it is also able to rotate about $+/-30°$. In addition to this advanced motor capabilities, the topology of the spider's retinae is equally peculiar. Each eye has its receptors arranged in a boomerang-like layout where the density of the receptors is highest in the narrow region of the "knee" of the boomerang. Also, the retina is composed of several (partially transparent) layers of receptors, which have been found to play an important role in depth perception. Eventually, when the recording area of the two eyes are put together, the field of view of the spider is revealed to have an x-like shape with the two bent regions of the retinae facing each other. During normal behavior, the animal moves this "template" with continuous scanning movements in a coordinated manner. It has been suggested, that in this way the spider classifies other spiders by analyzing their legs distinguishing if they are prey, predators, or mating partners. Thus, using a built-in "line detector", jumping spiders deduce the kind of opponent they face with an active sensorimotor

strategy [30, 58, 59].

Mammals with a *Visual Streak*.　In a number of mammals, for example sheep, pigs, horses, or the red kangaroo (Macropus rufus) a horizontally elongated region of high ganglion cell density, a *visual streak*, can be observed [44, 49, 50]. Such distributions could account for the fact that horizontal image translations are more frequently experienced by these animals, because as opposed to predators, they have very limited binocular vision and their behavior is less "object oriented". For these species it is more important to observe the horizon, a behavior which induces horizontal image shifts. Compare also Fig. 1.7.

The African Elephant (*Loxodonta Africana*).　Like the mammals described above, elephants also feature a ganglion cell distribution which forms a horizontal visual streak. However, in addition they also have an area of high receptor concentration in the upper temporal area of their retina. It is assumed that this part of the retina is pointed at the trunk of the animal which is used for complex object manipulation tightly coupled to visual feedback. See also [123, 85].

Appendix B

Derivations

Contents

For a gradient descent approach to problems proposed in Chap. 4, partial derivatives of the different cost functions must be computed. Sect. B.1 provides the derivatives required to address the final problem proposed in Sect. 4.7. Additionally, Sect. B.2 provides a derivation of the derivative of the positive least squares prediction operator $\mathbf{P}^q_{\text{pls}}$ with respect to a changing sensor topology $\mathbf{S}$. This derivative is not required in the final problem proposed in Eq. (4.23), but is required to solve some of the proposed problems. It is also of importance with respect to the relationship between sparsity and mean square error as discussed at the end of Sect. 5.2. Sect. B.3 provides an argument for the applicability of linear functions for visual stimulus prediction. Sect. B.4 discusses the reason why linear predictors $\mathbf{P}^q_{\text{pls}}$ converge to locations in the motor area which allow them to be particularly sparse.

B.1   Derivatives for the Proposed Sensorimotor Optimization Problem

This section provides the partial derivatives for problem (4.23) with respect to $\mathbf{S}$, $\mathbf{M}$, and $\mathbf{P}$. To facilitate derivation, a slight change in notation is introduced first. The linear combination of prediction operators can be written using a matrix multiplication, if all n_m predictors $\mathbf{P}^k$ are represented in

vectorized form as the columns of a matrix $\bar{\mathbf{P}}$ like

$$c_{sm}^{(4.23)}(\mathbf{S}, \mathbf{M}, \bar{\mathbf{P}}) = \sum_q \left\| \mathbf{S}^\top \left[\bar{\mathbf{P}} \mathbf{M}^\top \mathbf{q}^q \right]_\square \mathbf{S} \mathbf{i}_0^q - \mathbf{i}_1^q \right\|^2. \tag{B.1}$$

Note that, when using this notation, a prediction operator $\mathbf{P}^q$ is first computed in vectorized form and then reshaped into the usual matrix form. The reshape from a vector of length n_s^2 to the final matrix $\mathbf{P}^q$ of size $n_s \times n_s$ is denoted using square brackets. For any matrix $\mathbf{X}$, this operator is defined as $[\text{vec}(\mathbf{X})]_\square = \mathbf{X}$. For a full specification, a reshape operation requires the declaration of the dimension of the resulting matrix, however, since these dimensions are usually known in the given context, they can be omitted.

To provide the reader with a compact notation, the final result will be written in a hierarchical manner. Derivatives for $\mathbf{S}$, $\mathbf{M}$, and $\mathbf{P}$ are assembled by consecutively applying the chain rule. For this purpose, Eq. (B.1) is first decomposed into subfunctions as in

$$c_{sm}^{(4.23)}(\mathbf{S}, \mathbf{M}, \bar{\mathbf{P}}) = \sum_q \left(\mathbf{S}^\top \left[\bar{\mathbf{P}} \mathbf{M}^\top \mathbf{q}^q \right]_\square \mathbf{S} \mathbf{i}_0^q - \mathbf{i}_1^q \right)^\top \underbrace{ \left(\underbrace{\mathbf{S}^\top}_{Y} \underbrace{ \underbrace{\left[\underbrace{\bar{\mathbf{P}} \mathbf{M}^\top \mathbf{q}^q}_{H} \right]_\square \underbrace{\mathbf{S} \mathbf{i}_0^q}_{I} - \mathbf{i}_1^q}_{Z} \right) }_{X} }_{C}. \tag{B.2}$$

The chain rule to compose the final result from derivatives of subfunctions introduced above is

$$\frac{\partial Z\left(Y\left(\mathbf{X}\right)\right)}{\partial \text{vec}\left(\mathbf{X}\right)^\top} = \frac{\partial Z\left(Y\right)}{\partial \text{vec}\left(Y\right)^\top} \cdot \frac{\partial Y\left(\mathbf{X}\right)}{\partial \text{vec}\left(\mathbf{X}\right)^\top}. \tag{B.3}$$

Additionally, the following relationship is repeatedly applied to simplify intermediate results

$$\text{vec}\left(\mathbf{ABC}\right)^\top = \text{vec}\left(\mathbf{B}\right)^\top \left(\mathbf{C} \otimes \mathbf{A}^\top \right). \tag{B.4}$$

When applying the chain rule in the present context, the first multiplicand is always a vector, and thus, this relationship given in Eq. (B.4) can always be used to eliminate Kroenecker products which appear when deriving subfunctions of the cost function. In some cases the definition of the commutation matrix K_{mn} defined as

$$\text{vec}\left(\mathbf{X}^\top\right) = K_{mn} \text{vec}\left(\mathbf{X}\right) \tag{B.5}$$

is used for matrices $\mathbf{X^{n \times m}}$. Note that, in the case where $m = 1$ or $n = 1$, the commutation matrix equals the identity. To facilitate replication, each step of the derivation procedure provides page number and index of applied rules according to [68].

Partial Derivative for S

$$\frac{\partial C}{\partial \text{vec}\,(X)^\top} = 2X^\top \qquad\qquad \text{p. 178 (8)} \qquad\qquad \text{(B.6)}$$

$$\frac{\partial C}{\partial \text{vec}\,(Z)^\top} = \frac{\partial C}{\partial \text{vec}\,(X)^\top}\,Y \qquad\qquad \text{p. 183 (5)} \qquad\qquad \text{(B.7)}$$

$$\frac{\partial C}{\partial \text{vec}\,(Y)^\top} = \text{vec}\left(\left(\left(\frac{\partial C}{\partial \text{vec}\,(X)^\top}\right)^\top Z^\top\right)^\top\right) \qquad \text{p. 183 (5) and (B.4)} \qquad \text{(B.8)}$$

$$\frac{\partial C}{\partial \text{vec}\,(I)^\top} = \frac{\partial C}{\partial \text{vec}\,(Z)^\top}\,[H]_\square \qquad\qquad \text{p. 183 (3) and (B.4)} \qquad \text{(B.9)}$$

$$\frac{\partial C}{\partial \text{vec}\,(\mathbf{S})^\top} = \frac{\partial C}{\partial \text{vec}\,(Y)^\top} + \text{vec}\left(\left(\left(\frac{\partial C}{\partial \text{vec}\,(I)^\top}\right)^\top \mathbf{i}_0^{q\top}\right)^\top\right) \qquad \text{p. 183 (1),(3) and (B.5)} \qquad \text{(B.10)}$$

$$\frac{\partial c_{sm}^{(4.23)}(\mathbf{S}, \mathbf{M}, \bar{\mathbf{P}})}{\partial \text{vec}\,(\mathbf{S})^\top} = \sum_q \left(\frac{\partial C}{\partial \text{vec}\,(Y)^\top} + \text{vec}\left(\left(\left(\frac{\partial C}{\partial \text{vec}\,(I)^\top}\right)^\top \mathbf{i}_0^{q\top}\right)^\top\right)\right) \qquad \text{(B.11)}$$

Partial Derivative for M

$$\frac{\partial C}{\partial \text{vec}\,(X)^\top} = 2X^\top \qquad\qquad \text{p. 178 (8)} \qquad\qquad \text{(B.12)}$$

$$\frac{\partial C}{\partial \text{vec}\,(Z)^\top} = \frac{\partial C}{\partial \text{vec}\,(X)^\top}\,Y \qquad\qquad \text{p. 183 (5)} \qquad\qquad \text{(B.13)}$$

$$\frac{\partial C}{\partial \text{vec}\,([H]_\square)^\top} = \text{vec}\left(\left(\left(\frac{\partial C}{\partial \text{vec}\,(Z)^\top}\right)^\top I^\top\right)^\top\right) \qquad \text{p. 183 (3) and (B.4)} \qquad \text{(B.14)}$$

$$\frac{\partial C}{\partial \text{vec}\,(\mathbf{M})^\top} = \text{vec}\left(\mathbf{q}^q\,\frac{\partial Z}{\partial \text{vec}\,(H)^\top}\,\bar{\mathbf{P}}\right)^\top \qquad \text{p. 183 (4) and (B.4), (B.5)} \qquad \text{(B.15)}$$

$$\text{(B.16)}$$

$$\frac{\partial c_{sm}^{(4.23)}(\mathbf{S}, \mathbf{M}, \bar{\mathbf{P}})}{\partial \text{vec}\,(\mathbf{M})^\top} = \sum_q \left(\frac{\partial C}{\partial \text{vec}\,(\mathbf{M})^\top}\right) \qquad\qquad \text{(B.17)}$$

Partial Derivative for P

$$\frac{\partial C}{\partial \text{vec}\,(X)^\top} = 2X^\top \qquad\qquad \text{p. 178 (8)} \qquad\qquad \text{(B.18)}$$

$$\frac{\partial C}{\partial \text{vec}\,(Z)^\top} = \frac{\partial C}{\partial \text{vec}\,(X)^\top}\,Y \qquad\qquad \text{p. 183 (5)} \qquad\qquad \text{(B.19)}$$

$$\frac{\partial C}{\partial \text{vec}\,([H]_\square)^\top} = \text{vec}\left(\left(\left(\frac{\partial C}{\partial \text{vec}\,(Z)^\top}\right)^\top I^\top\right)^\top\right) \qquad \text{p. 183 (3) and (B.4)} \qquad \text{(B.20)}$$

$$\frac{\partial H}{\partial \text{vec}\,(\bar{\mathbf{P}})^\top} = \text{vec}\left(\left(\left(\frac{\partial C}{\text{vec}\,([H]_\square)^\top}\right)^\top \mathbf{q}^{q\top}\mathbf{M}\right)^\top\right) \qquad \text{p. 183 (4) and (B.4), (B.5)} \qquad \text{(B.21)}$$

$$\frac{\partial c_{sm}^{(4.23)}(\mathbf{S},\mathbf{M},\bar{\mathbf{P}})}{\partial \text{vec}\,(\bar{\mathbf{P}})^\top} = \sum_q \left(\frac{\partial H}{\partial \text{vec}\,(\bar{\mathbf{P}})^\top}\right) \qquad\qquad \text{(B.22)}$$

B.2 Derivatives for the Linear Positive Least Squares Predictor

Even though the solution for a positive least squares solution $\mathbf{P}_{\text{pls}}^q$ cannot be computed analytically, its derivative with respect to $\mathbf{S}$ can still be found in closed form by applying the implicit function theorem to the Karush-Kuhn-Tucker optimality conditions of the positive least squares optimization problem [12].

Karush-Kuhn-Tucker Optimality Conditions. The prediction operator $\mathbf{P}_{\text{pls}}^q$ can be written as a system of implicit functions (equations and conditions) $g(\mathbf{P}_{\text{pls}}^q; \mathbf{\Gamma}, \mathbf{S}) = 0$ using the Karush-Kuhn-Tucker Optimality Conditions (KKT), where in the case of the given optimization problem, $\mathbf{P}_{\text{pls}}^q$ and $\mathbf{\Gamma}$ are both functions of $\mathbf{S}$.

$$\text{tr}^2\left\{\mathbf{P}_{\text{pls}}^q\mathbf{Si}_0 - \mathbf{Si}_1\right\} = \text{tr}\left\{\left(\mathbf{P}_{\text{pls}}^q\mathbf{Si}_0 - \mathbf{Si}_1\right)\left(\mathbf{P}_{\text{pls}}^q\mathbf{Si}_0 - \mathbf{Si}_1\right)^\top\right\}$$

$$= \text{tr}\left\{\mathbf{P}_{\text{pls}}^q\mathbf{Si}_0\mathbf{i}_0^\top\mathbf{S}^\top\mathbf{P}_{\text{pls}}^{q\top} - \mathbf{Si}_1\mathbf{i}_0^\top\mathbf{S}^\top\mathbf{P}_{\text{pls}}^{q\top} - \mathbf{P}_{\text{pls}}^q\mathbf{Si}_0\mathbf{i}_1^\top\mathbf{S}^\top + \mathbf{Si}_1\mathbf{i}_1^\top\mathbf{S}^\top\right\}$$

$$= \text{tr}\left\{\mathbf{P}_{\text{pls}}^q\mathbf{Si}_0\mathbf{i}_0^\top\mathbf{S}^\top P^\top\right\} - 2\cdot\text{tr}\left\{\mathbf{P}_{\text{pls}}^q\mathbf{Si}_0\mathbf{i}_1^\top\mathbf{S}^\top\right\} + \text{tr}\left\{\mathbf{Si}_1\mathbf{i}_1^\top\mathbf{S}^\top\right\} \quad \text{(B.23)}$$

$$\mathcal{L} = \text{tr}\left\{\mathbf{P}_{\text{pls}}^q\mathbf{Si}_0\mathbf{i}_0^\top\mathbf{S}^\top P^\top\right\} - 2\cdot\text{tr}\left\{\mathbf{P}_{\text{pls}}^q\mathbf{Si}_0\mathbf{i}_1^\top\mathbf{S}^\top\right\} + \text{tr}\left\{\mathbf{Si}_1\mathbf{i}_1^\top\mathbf{S}^\top\right\} + \text{tr}\left\{\mathbf{\Gamma}^\top\mathbf{P}_{\text{pls}}^q\right\} \quad \text{(B.24)}$$

With [68], using p. 178 (11) for the first summand and p. 177 (2) for the second and third summand, the KKT can be written as:

$$\frac{\partial \operatorname{vec}(\mathcal{L})}{\partial \operatorname{vec}\left(\mathbf{P}_{\text{pls}}^{q}\right)^{\top}} = 2 \cdot \mathbf{P}_{\text{pls}}^{q}\left(\mathbf{S}\mathbf{i}_{0}\mathbf{i}_{0}^{\top}\mathbf{S}^{\top}\right) - 2 \cdot \mathbf{S}\mathbf{i}_{1}\mathbf{i}_{0}^{\top}\mathbf{S}^{\top} + \mathbf{\Gamma} = 0$$

$$\mathbf{P}_{\text{pls}}^{q} \geq 0$$

$$\mathbf{\Gamma} \leq 0$$

$$\mathbf{P}_{\text{pls}}^{q} \circ \mathbf{\Gamma} = 0 \tag{B.25}$$

where "∘" is used to denote the Hadamard product.

Obtaining $\mathbf{\Gamma}$ from the KKT conditions is easy for the addressed problem, it simply is:

$$\mathbf{\Gamma} = 2 \cdot \left(\mathbf{S}\mathbf{i}_{1}\mathbf{i}_{0}^{\top}\mathbf{S}^{\top} - \mathbf{P}_{\text{pls}}^{q}\mathbf{S}\mathbf{i}_{0}\mathbf{i}_{0}^{\top}\mathbf{S}^{\top}\right). \tag{B.26}$$

Applying the Implicit Function Theorem. According to the implicit function theorem the derivative of the two sets of functions $\mathbf{P}_{\text{pls}}^{q}$ and $\mathbf{\Gamma}$ with respect to the sensor topology $\mathbf{S}$ can be written using the Jacobian D_g of a function g as:

$$D_g(\mathbf{P}_{\text{pls}}^{q}; \mathbf{\Gamma}, \mathbf{S}) = \left[\underbrace{\mathbf{X}}_{\mathbf{S}} \quad \underbrace{\mathbf{Y}}_{\mathbf{P}_{\text{pls}}^{q}, \mathbf{\Gamma}}\right]$$

$$\frac{\partial \operatorname{vec}\left(\mathbf{P}_{\text{pls}}^{q}, \mathbf{\Gamma}\right)}{\partial \operatorname{vec}(\mathbf{S})^{\top}} = \mathbf{Y}^{-1}\mathbf{X} = \left[\begin{array}{c} \text{result from } \mathbf{P}_{\text{pls}}^{q} \text{ equations} \\ \hline \text{result from } \mathbf{\Gamma} \text{ equations} \end{array}\right]. \tag{B.27}$$

With the notation

$$g(\mathbf{P}_{\text{pls}}^{q}; \mathbf{\Gamma}, \mathbf{S}) = \left[\begin{array}{c} \underbrace{2 \cdot \mathbf{P}_{\text{pls}}^{q}\left(\mathbf{S}\mathbf{i}_{0}\mathbf{i}_{0}^{\top}\mathbf{S}^{\top}\right) - 2 \cdot \mathbf{S}\mathbf{i}_{1}\mathbf{i}_{0}^{\top}\mathbf{S}^{\top} + \mathbf{\Gamma}}_{g_1} \\ \underbrace{\mathbf{P}_{\text{pls}}^{q} \circ \mathbf{\Gamma}}_{g_2} \end{array}\right], \tag{B.28}$$

the Jacobian D_g writes

$$D_g = \begin{pmatrix} \frac{\partial \operatorname{vec}(g_1)}{\partial \operatorname{vec}(\mathbf{S})^\top} & \frac{\partial \operatorname{vec}(g_1)}{\partial \operatorname{vec}\left(\mathbf{P}^q_{\mathrm{pls}}\right)^\top} & \frac{\partial \operatorname{vec}(g_1)}{\partial \operatorname{vec}(\boldsymbol{\Gamma})^\top} \\ \frac{\partial \operatorname{vec}(g_2)}{\partial \operatorname{vec}(\mathbf{S})^\top} & \frac{\partial \operatorname{vec}(g_2)}{\partial \operatorname{vec}\left(\mathbf{P}^q_{\mathrm{pls}}\right)^\top} & \frac{\partial \operatorname{vec}(g_2)}{\partial \operatorname{vec}(\boldsymbol{\Gamma})^\top} \end{pmatrix}. \tag{B.29}$$
$$\underbrace{\phantom{\frac{\partial \operatorname{vec}(g_1)}{\partial \operatorname{vec}(\mathbf{S})^\top}}}_{\mathbf{X}} \quad \underbrace{\phantom{\frac{\partial \operatorname{vec}(g_1)}{\partial \operatorname{vec}(\boldsymbol{\Gamma})^\top}}}_{\mathbf{Y}}$$

To derive $g1$ [68] p. 190 (6) can be applied

$$\frac{\partial \operatorname{vec}(g_1)}{\partial \operatorname{vec}(\mathbf{S})^\top} = 2 \cdot \left(\mathbf{S} \mathbf{i}_0 \mathbf{i}_0^\top \otimes \mathbf{P}^q_{\mathrm{pls}} \right)$$
$$+ 2 \cdot \left(\mathbb{I}_m \otimes \mathbf{P}^q_{\mathrm{pls}} \mathbf{S} \mathbf{i}_0 \mathbf{i}_0^\top \right) \mathbf{K}_{mn}$$
$$- 2 \cdot \left(\mathbf{S} \mathbf{i}_0 \mathbf{i}_1^\top \otimes \mathbb{I}_m \right)$$
$$- 2 \cdot \left(\mathbb{I}_m \otimes \mathbf{S} \mathbf{i}_1 \mathbf{i}_0^\top \right) \mathbf{K}_{mn} \tag{B.30}$$

which with [68] p. 183 (3) yields

$$\frac{\partial \operatorname{vec}(g_1)}{\partial \operatorname{vec}\left(\mathbf{P}^q_{\mathrm{pls}}\right)^\top} = \underbrace{2 \left(\mathbf{S} \mathbf{i}_0 \mathbf{i}_0^\top \mathbf{S}^\top \right)}_{\mathbf{E}} \otimes \underbrace{\mathbb{I}_m}_{\mathbf{F}} \tag{B.31}$$

$$\frac{\partial \operatorname{vec}(g_1)}{\partial \operatorname{vec}(\boldsymbol{\Gamma})^\top} = \mathbb{I}. \tag{B.32}$$

To derive $g2$, [68] p. 185 (16) can be applied like

$$\frac{\partial \operatorname{vec}(g_2)}{\partial \operatorname{vec}(\mathbf{S})^\top} = 0$$

$$\frac{\partial \operatorname{vec}(g_2)}{\partial \operatorname{vec}\left(\mathbf{P}^q_{\mathrm{pls}}\right)^\top} = \underbrace{\operatorname{diag}(\operatorname{vec}(\boldsymbol{\Gamma}))}_{\mathbf{d_1}} \tag{B.33}$$

$$\frac{\partial \operatorname{vec}(g_2)}{\partial \operatorname{vec}(\boldsymbol{\Gamma})^\top} = \underbrace{\operatorname{diag}\left(\operatorname{vec}\left(\mathbf{P}^q_{\mathrm{pls}}\right)\right)}_{\mathbf{d_2}}. \tag{B.34}$$

Note, in this particular situation, the sub-matrices $\frac{\partial \operatorname{vec}(g_2)}{\partial \operatorname{vec}\left(\mathbf{P}^q_{\mathrm{pls}}\right)^\top}$ and $\frac{\partial \operatorname{vec}(g_2)}{\partial \operatorname{vec}(\boldsymbol{\Gamma})^\top}$ of $\mathbf{X}$ are not invertible as they carry zeros in the diagonal due to $\mathbf{P}^q_{\mathrm{pls}}$ being sparse. But on a closer look, it can be seen that it is sufficient to invert $\frac{\partial \operatorname{vec}(g_1)}{\partial \operatorname{vec}(\mathbf{S})^\top}$ to compute $\frac{\partial \operatorname{vec}\left(\mathbf{P}^q_{\mathrm{pls}}\right)}{\partial \operatorname{vec}(\mathbf{S})^\top}$ with Eq. (B.27). This is true because i) only the upper part of $\mathbf{Y}^\top \mathbf{X}$ is required, and ii) the lower part of $\mathbf{X}$ is $\mathbf{0}$, and therefore when multiplying $\mathbf{Y}^\top \mathbf{X}$ the top right corner of $\mathbf{Y}$ is multiplied by $\mathbf{0}$.

To invert the upper left block of $\mathbf{Y}$ blockwise inversion can be used. If a matrix is devided in four blocks $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ its inverted form can be written as

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1} & \cdots \\ \cdots & \cdots \end{bmatrix} \tag{B.35}$$

Taking into account Eqs. (B.31), (B.32), (B.33), and (B.34), the required part of $\mathbf{Y}^{-1}$ is

$$\mathbf{Y}^{-1} = \begin{bmatrix} \underbrace{\left(\mathbf{E}^{-1} \otimes \mathbf{F}^{-1}\right) + \left(\mathbf{E}^{-1} \otimes \mathbf{F}^{-1}\right) \cdot \left[\mathbf{d_2} - \mathbf{d_1}\left(\mathbf{E}^{-1} \otimes \mathbf{F}^{-1}\right)\right]^{-1} \cdot \mathbf{d_1} \cdot \left(\mathbf{E}^{-1} \otimes \mathbf{F}^{-1}\right)}_{\mathbf{Y}_{\text{tl}}^{-1}} & \cdots \\ \cdots & \cdots \end{bmatrix}, \tag{B.36}$$

and due to i) and ii) from above:

$$\frac{\partial \operatorname{vec}\left(\mathbf{P}_{\text{pls}}^q\right)}{\partial \operatorname{vec}\left(\mathbf{S}\right)^\top} = \mathbf{Y}_{\text{tl}}^{-1} \frac{\partial \operatorname{vec}\left(g_1\right)}{\partial \operatorname{vec}\left(\mathbf{S}\right)^\top}. \tag{B.37}$$

B.3 Proof of Linearity for Visual Stimulus Prediction

Here it is demonstrated that under the assumptions summarized at the end of Chap. 3, the class of functions from which a stimulus predictor p^q should be chosen can be restricted to the linear function set. The argument relies on assumption (5) listed in Sect. 3.9, which requires that the actions executed by the agent lead to perfectly predictable changes of a visual signal i on the sensor area Φ_s.

First it is noted that the observation function Eq. (3.1) is linear in i. Then, with perfectly predictable observations, equation Eq. (3.8) is perfectly satisfied, meaning that each receptor value satisfies

$$\mathbf{o}_{x+1} = p^q(\mathbf{o}_x) \iff \begin{bmatrix} o_1(q(i)) \\ o_2(q(i)) \\ \vdots \\ o_{n_s}(q(i)) \end{bmatrix} = p^q\left(\begin{bmatrix} o_1(i) \\ o_2(i) \\ \vdots \\ o_{n_s}(i) \end{bmatrix}\right). \tag{B.38}$$

Since q and o_j are linear, given any two images i_1 and i_2, and any two scale factors, α and β, the

previous satisfies

$$\begin{bmatrix} o_1(q(\alpha i_1 + \beta i_2)) \\ o_2(q(\alpha i_1 + \beta i_2)) \\ \vdots \\ o_{n_s}(q(\alpha i_1 + \beta i_2)) \end{bmatrix} = p^q \left(\begin{bmatrix} o_1(\alpha i_1 + \beta i_2) \\ o_2(\alpha i_1 + \beta i_2) \\ \vdots \\ o_{n_s}(\alpha i_1 + \beta i_2) \end{bmatrix} \right)$$

$$\alpha \begin{bmatrix} o_1(q(i_1)) \\ o_2(q(i_1)) \\ \vdots \\ o_{n_s}(q(i_1)) \end{bmatrix} + \beta \begin{bmatrix} o_1(q(i_2)) \\ o_2(q(i_2)) \\ \vdots \\ o_{n_s}(q(i_2)) \end{bmatrix} = p^q \left(\alpha \begin{bmatrix} o_1(i_1) \\ o_2(i_1) \\ \vdots \\ o_{n_s}(i_1) \end{bmatrix} + \beta \begin{bmatrix} o_1(i_2) \\ o_2(i_2) \\ \vdots \\ o_{n_s}(i_2) \end{bmatrix} \right),$$

which, when equation (B.38) is replaced on the left hand side

$$\alpha p^q \left(\underbrace{\begin{bmatrix} o_1(i_1) \\ o_2(i_1) \\ \vdots \\ o_{n_s}(i_1) \end{bmatrix}}_{\mathbf{x}} \right) + \beta p^q \left(\underbrace{\begin{bmatrix} o_1(i_2) \\ o_2(i_2) \\ \vdots \\ o_{n_s}(i_2) \end{bmatrix}}_{\mathbf{y}} \right) = p^q \left(\alpha \underbrace{\begin{bmatrix} o_1(i_1) \\ o_2(i_1) \\ \vdots \\ o_{n_s}(i_1) \end{bmatrix}}_{\mathbf{x}} + \beta \underbrace{\begin{bmatrix} o_1(i_2) \\ o_2(i_2) \\ \vdots \\ o_{n_s}(i_2) \end{bmatrix}}_{\mathbf{y}} \right)$$

$$\alpha p^q(\mathbf{x}) + \beta p^q(\mathbf{y}) = p^q \left(\alpha \mathbf{x} + \beta \mathbf{y} \right),$$

proves linearity of p^q whenever the action is perfectly predictable. $\square$

B.4   Sparsity of Prediction Operators.

It has been argued in Sect. 5.3 that when solving problem (4.15), particularly sparse matrices $\mathbf{P}_k$ result. This paragraph provides a hint for why this is the case.

In problem (4.15), but also in problem (4.23), n_m prediction matrices $\mathbf{P}_k$ are learned, such that for linear combinations $\mathbf{P}^q = \sum_k^{n_m} m_k(\mathbf{q})\mathbf{P}_k$ the most accurate predictor results for every action $\mathbf{q}$ executed by the agent. Considering the constraint $\mathbf{P}_k \geq \mathbf{0}$ and thinking of predictors $\mathbf{P}_k$ in a vectorized form, all $\text{vec}(\mathbf{P}_k)$ are constrained to the positive orthant. In this case, linear combinations can only assemble predictors $\mathbf{P}^q$ which lie within the area spanned by all $\text{vec}(\mathbf{P}_k)$. Thus, to cover a bigger area in the positive orthant in order to assemble a bigger variety of predictors, matrices $\mathbf{P}_k$ must be chosen such that they are closer to the boundaries of the positive orthant, which in turn means
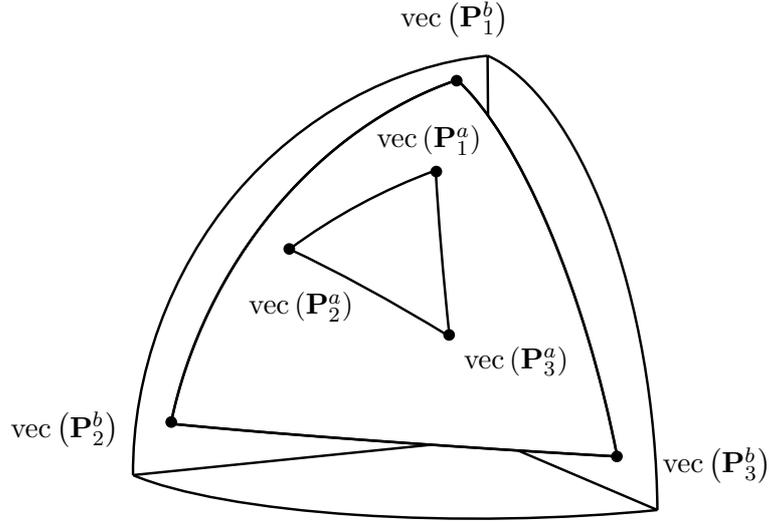
Figure B.1: Visualization of a 3-dimensional positive orthant with two sets of prediction operators a and b, shown in vectorized form $\mathrm{vec}\left(\mathbf{P}_k\right)$. The first set a spans a smaller set of possible linear combinations $\mathbf{P}^q = \sum_k^{n_m} m_k(\mathbf{q})\mathbf{P}_k$. The second set b spans a bigger set of possible linear combinations $\mathbf{P}^q$. The illustration shows that the set b, which can account for a bigger variety of prediction operators $\mathbf{P}^q$, is composed of $\mathbf{P}_k$ which are closer to the corners of the positive orthant, and thus sparser.

choosing sparser $\mathbf{P}_k$. This argument further supports the empirical finding that requesting $\mathbf{P}_k \geq \mathbf{0}$ is sufficient to drive predictors $\mathbf{P}_k$ towards particularly sparse solutions, see also point (4) in Sect. 4.2.

Fig. B.1 provides an illustration in 3-dimensional space of a positive orthant and two sets of predictors $a = \{\mathrm{vec}\left(\mathbf{P}_1^a\right), \mathrm{vec}\left(\mathbf{P}_2^a\right), \mathrm{vec}\left(\mathbf{P}_2^a\right)\}$ and $b = \{\mathrm{vec}\left(\mathbf{P}_1^b\right), \mathrm{vec}\left(\mathbf{P}_2^b\right), \mathrm{vec}\left(\mathbf{P}_3^b\right)\}$, where the first set a spans a smaller set of possible linear combinations, and the second set b spans a bigger set of possible linear combinations. The nodes of set b are closer to the corners of the positive orthant and thus sparser. [1]

---

[1] Note that, for example, for the results presented in Fig. 5.7, the space visualized in Fig. B.1 is 25^2-dimensional, since prediction matrices for the presented experiment are of size 25×25 for a sensor with 25 receptors.

Bibliography

[1] P.-A. Absil, R. Mahoney, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[2] E. Ahissar and D. Kleinfeld. Closed-loop neuronal computations: focus on vibrissa somatosensation in rat. *Cerebral Cortex*, 13:53–62, 2003.

[3] C. H. Anderson. Unifying perspectives on neuronal codes and processing. In *In Proc. Int. Conf. on Artificial Neural Networks (ICANN)*, 1996.

[4] L. Armijo. Minimization of functions. having continuous partial derivatives. *Pacific Journal of Mathematics*, 16:1–3, 1966.

[5] J. J. Atick and A. N. Redlich. What does the retina know about natural scenes? *Neural Computation*, 4:196–210, 1992.

[6] H. B. Barlow. Critical limiting factors in the design of the eye and the visual cortex. *Proceedings of the Royal Society of London. Series B, Biological Sciences.*, 212:1–24, 1981.

[7] Robert. B. Barlow. Vision in horseshoe crabs. In John T. Tanacredi, Mark L. Botton, and David Smith, editors, *Biology and Conservation of Horseshoe Crabs*, pages 223–235. Springer, 2009.

[8] M. Barnsley. *Fractals Everywhere, 2nd edition*. Morgan Kaufmann, 1993.

[9] J. Barzilai and J. Borwein. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8:141–148, 1988.

[10] A. P. Batista, C. A. Buneo, L. H. Snyder, and R. A. Andersen. Reach plans in eye-centered coordinates. *Science*, 285:257–260, 1999.

[11] C. C. Bell, V. Han, and N. B. Sawtell. Cerebellum-like structures and their implications for cerebellar function. *Ann. Rev. Neuroscience*, 31:1–24, 2008.

[12] Dimitri P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Athena Scientific, 1996.

[13] Belinda Y. Betsch, Wolfgang Einhäuser, Konrad P. Körding, and Peter König. The world from a cat's perspective – statistics of natural videos. *Biological Cybernetics*, 90:41–50, 2004.

[14] Valentino Braitenberg. *Vehicles: Experiments in synthetic psychology*. MIT Press, 1984.

[15] Alfred Brückstein, Robert J. Holt, Igor Katsman, and Ehud Rivlin. Head movements for depth perception: Praying mantis versus pigeon. *Autonomous Robots*, 18:21–42, 2005.

[16] A. H. Jay Burr, Colette P. F. Babinszki, and Amanda J. Ward. Components of phototaxis of the nematode *Mermis nigrescens*. *Journal of Comparative Physiology*, 167:245–255, 1990.

[17] M. C. Bushnell, M. E. Goldberg, and D. L. Robinson. Behavioral enhancement of visual responses in monkey cerebral cortex: I. modulation in posterior parietal cortex related to selective visual attention. *J. Neurophysiology*, 46:755–772, 1981.

[18] C. Cherniak. Neural component placement. *Trens Neurosci.*, 18:522–527, 1995.

[19] D. B. Chklovskii. Synaptic connectivity and neuronal morphology: viewpoint two sides of the same coin. *Neuron*, 43:609–617, 2004.

[20] Euan Clarkson, Riccardo Levi-Setti, and Gabor Horváth. The eyes of trilobites: The oldest preserved visual system. *Arthropod Structure & Development*, 35:247–259, 2006.

[21] S. M. Clippingdale and R. Wilson. Self-similar neural networks based on a kohonen learning rule. *Neural Networks*, 9(5):747–763, 1996.

[22] C. L. Colby, J. R. Duhamel, and M. E. Goldberg. Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *J. Neurophysiology*, 76:2841–2852, 1996.

[23] C. L. Colby and M. E. Goldberg. Space and attention in parietal cortex. *Ann. Rev. Neuroscience*, 22:319–349, 1999.

[24] L. Craighero, L. Fadiga, G. Rizzolatti, and C. Umiltà. Action for perception: A motor-visual attentional effect. *Exp. Psych.: Human Perception and Performance*, 25:1673–1692, 1999.

[25] T. B. Crapse and M. A. Sommer. Corollary discharge across the animal kingdom. *Nat. Rev. Neuroscience*, 9:587–600, 2008.

[26] T. B. Crapse and M. A. Sommer. Corollary discharge circuits in the primate brain. *Curr. Opin. Neurobiology*, 18:552–557, 2008.

[27] J. D. Crawford, W. P. Medendorp, and J. J. Marotta. Spatial transformations for eye-hand coordination. *J. Neurophysiology*, 92:10–19, 2004.

[28] F. Delcomyn. Corollary discharge to cockroach giant interneurons. *Nature*, 269:160–162, 1977.

[29] A. C. Downing. Optical scanning in the lateral eyes of the copepod copilia. *Perception*, 1:247–261, 1972.

[30] Oskar Drees. Untersuchungen über die angeborenen verhaltensweisen bei springspinnen *(Salticidae). Zeitschrift für Tierpsychologie*, 9:169–207, 1952.

[31] J. R. Duhamel, C. L. Colby, and M. E. Goldberg. The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255:90–92, 1992.

[32] D. Edwards, W. Heitler, and F. Krasne. Fifty years of a command neuron: the neurobiology of escape behavior in the crayfish. *Trens Neurosci.*, 22:153–161, 1999.

[33] J.-P. Ewert and H.-W. Borchers. Reaktionscharakteristik von neuronen aus dem tectum opticum und subtectum der erdkröte *Bufo Bufo* (l.). *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 71:165–189, 1971.

[34] Sigmund Exner. *Die Physiologie der facettirten Augen von Krebsen und Insecten*. Franz Deuticke, 1891.

[35] Dario Floreano, Toshifumi Kato, Davide Marocco, and Eric Sauser. Coevolution of active vision and feature selection. *Biol. Cybern.*, 90:218–228, 2004.

[36] N. Franceschini, J. M. Pichon, and C. Blanes. From insect vision to robot vision. *Phil. Trans. Biological Sciences*, 337:283–294, 1992.

[37] N. Franceschini, F. Ruffier, and J. Serres. A bio-inspired flying robot sheds light on insect piloting abilities. *Curr. Biol.*, 17:329–335, 2007.

[38] A. P. Georgopoulos, R. Caminiti, J. K. Kalaska, and J. T. Massey. Spatial coding of movement: A hypothesis concerning the coding of movement direction by motor cortical populations. *Exp. Brain Res.*, 7:328–336, 1983.

[39] B. Girard and A. Berthoz. From brainstem to cortex: Computational models of saccade generation circuitry. *Prog. Neurobiol.*, 77:215–251, 2005.

[40] M. E. Goldberg and M. C. Bushnell. Behavioral enhancement of visual responses in monkey cerebral cortex: Ii. modulation in frontal eye fields specifically related to saccades. *J. Neurophysiology*, 46:773–787, 1981.

[41] A. C. Graham. *The Book of Lieh-tzǔ: A Classic of Tao*. Columbia University Press, 1960.

[42] R. L. Gregory, H. E. Ross, and N. Moray. The curious eye of *Copilia. Nature*, 201, 1964.

[43] William C. Hall and Adonis K. Moschovakis. *The Superior Colliculus*. Taylor & Francis Inc., 2003.

[44] Rudolf Hebel. Distribution of retinal ganglion cells in five mammalian species (pig, sheep, ox, horse, dog). *Anat. Embryol.*, 150:45–51, 1976.

[45] Toshihiko Hosoya, Stephen A. Baccus, and Markus Meister. Dynamic predictive coding by the retina. *Nature*, 436:71–77, 2005.

[46] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Physiology*, 160:106–154, 1962.

[47] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Physiology*, 195:215–243, 1968.

[48] D. H. Hubel, T. N. Wiesel, and S. LeVay. Plasticity of ocular dominance columns in monkey striate cortex. *Phil. Trans. R. Soc. Lond. B*, 278:377–409, 1977.

[49] A. Hughes. A comparison of retinal ganglion cell topography in the plains and tree kangaroo. *Physiology*, 244:61–63, 1975.

[50] A. Hughes. The topography of vision in mammals of contrasting life style: comparative optics and retinal organisation. In F. Crescitelli, editor, *Handbook Of Sensory Physiology Vol. 7(5)*, pages 613–756. Springer, 1977.

[51] N. Hurley and S. Rickard. Comparing measures of sparsity. *Transactions on Information Theory*, 811, 2008.

[52] Saitoh Kazuya, Ariane Ménard, and Sten Grillner. Tectal control of locomotion, steering, and eye movements in lamprey. *J Neurophysiol.*, 97:3093–3108, 2007.

[53] Andrew J. King, Mary E. Hutchings, David R. Moore, and Colin Blakemore. Developmental plasticity in the visual and auditory representations in the mammalian superior colliculus. *Nature*, 332:73–76, 1988.

[54] M. Kohler and S. Moya-Sola. Reduction of brain and sense organs in the fossil insular bovid myotragus. *Brain Behav. Evol.*, 63:125–140, 2004.

[55] T.o Kohonen, M. R. Schroeder, and T. S. Huang, editors. *Self-Organizing Maps*. Springer-Verlag New York, Inc., 3rd edition, 2001.

[56] Karl Kral and Michael Poteser. Motion parallax as a source of distance information in locusts and mantids. *Journal of Insect Behavior*, 10:145 – 163, 1996.

[57] M. F. Land, J. N. Marshall, D. Brownless, and T. W. Cronin. The eye-movements of the mantis shrimp *Ondontodactylus scyllarus* (curstacea: Stomatopoda). *Journal of Computational Physiology A*, 167:155 – 166, 1990.

[58] Michael F. Land. Movements of the retinae of jumping spiders (salticidae: Dendryphantinae) in response to visual stimuli. *Journal of Experimental Biology*, 51:471–493, 1969.

[59] Michael F. Land. Structure of the retinae of the principal eyes of jumping spiders (salticidae: Dendryphantinae) in relation to visual optics. *Journal of Experimental Biology*, 51:443–470, 1969.

[60] Michael F. Land. Scanning eye movements in a heteropod mollusc. *Journal of Experimental Biology*, 96:427–430, 1982.

[61] Michael F. Land. The functions of eye and body movements in *Labidocera* and other copepods. *Journal of Experimental Biology*, 140:381–391, 1988.

[62] Cecilia Laschi, Gioel Asuni, Eugenio Guglielmelli, Giancarlo Teti, Roland Johansson, Hitoshi Konosu, Zbigniew Wasik, Maria Chiara Carrozza, and Paolo Dario. A bio-inspired predictive sensory-motor coordination scheme for robot reaching and preshaping. *Auton. Robots*, 25:85–101, 2008.

[63] S. B. Laughlin, R. R. D. van Steveninck, and J. C. Anderson. The metabolic cost of neural information. *Nat. Neurosci.*, 1:36–41, 1998.

[64] C. Lee, W. H. Rohrer, and D. L. Sparks. Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature*, 332:357–360, 1988.

[65] L. Lichtensteiger and P. Eggenberger. Evolving the morphology of a compound eye on a robot. In *Proc. 3rd Europ. Worksh. on Adv. Mobile Robots*, pages 127–134, 1999.

[66] G. E. Loeb, I. E. Brown, and S. H. Scott. Directional motor control. *TINS*, 19:137–138, 1996.

[67] Max Lungarella and Olaf Sporns. Mapping information flow in sensorimotor networks. *PLoS Computational Biology*, 2:1301–1312, 2006.

[68] Helmut Lütkepohl. *Handbook of Matrices*. John Wiley & Sons, 1997.

[69] R. C. Miall and D. M. Wolpert. Forward models for physiological motor control. *Neural Networks*, 9:1265–1279, 1996.

[70] A. K. Mohamed, C. Burr, and A. H. Burr. Unique two-photoreceptor scanning eye of the nematode mermis nigrescens. *The Biological Bulletin*, 212:206–221, 2007.

[71] Adonis K. Moschovakis. The superior colliculus and eye movmement control. *Curr. Opin. Neurobiology*, 6:811–816, 1996.

[72] C. F. Moss and S. R. Sinha. Neurobiology of echolocation in bats. *Curr. Opin. Neurobiol.*, 13:751–758, 2003.

[73] T. D. Mrsic-Flogel, S. B. Hofer, C. Creutzfeldt, I. Cloëz-Tayarani, J. P. Changeux, T. Bonhoeffer, and M. Hübener. Altered map of visual space in the superior colliculus of mice lacking early retinal waves. *J. Neuroscience*, 25:6921–=6928, 2005.

[74] A. Murthy, S. Ray, S. M. Shorter, E. G. Priddy, J. D. Schall, and K. G Thompson. Frontal eye field contributions to rapid corrective saccades. *J. Neurophysiology*, 97:1457–1469, 2007.

[75] F. A. Mussa-Ivaldi. Do neurons in the motor cortex encode movement direction? an alternative hypothesis. *Neurosci. Lett.*, 91:106–111, 1988.

[76] K. Nakamura and C. L. Colby. Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proc. Natl. Acad. Sci. USA*, 99:4026–4031, 2002.

[77] J. E. Niven, M. Vähäsöyrinki, M. Juusola, and A. S. French. Interactions between light-induced currents, voltage-gated currents, and input signal properties in *Drosophila* photoreceptors. *J. Neurophysiol.*, 91:2696–2706, 2004.

[78] Jeremy E. Niven and Simon B. Laughlin. Review, energy limitation as a selective pressure on the evolution of sensory systems. *Experimental Biology*, 211:1792–1804, 2008.

[79] B. A. Olshausen and D. J. Field. Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.*, 14:481–487, 2004.

[80] Lars Olsson, Chrystopher L. Nehaniv, and Daniel Polani. From unknown sensors and actuators to actions grounded in sensorimotor perceptions. *Connection Science*, 18, 2006.

[81] J. K. O'Regan and Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5):939–1031, 2001.

[82] Andrew Parker. *In The Blink Of An Eye: How Vision Sparked The Big Bang Of Evolution.* Basic Books, 2004.

[83] G. A. Parker and J. Maynard Smith. Optimality theory in evolutionary biology. *Nature*, 348:27–33, 1990.

[84] Chandana Paul. Morphological computation, a basis for the analysis of morphology and control requirements. *Robotics and Autonomous Systems*, 54:619–630, 2006.

[85] John D. Pettigrew, Adhil Bhagwandin, Mark Haagensen, and Paul R. Manger. Visual acuity and heterogeneities of retinal ganglion cell densities and the tapetum lucidum of the african elephant *(Loxodonta africana)*. *Brain, behavior and Evolution*, 75:251–261, 2010.

[86] Rolf Pfeifer and J. Bongard. *How the Body Shapes the Way We Think.* MIT Press, 2006.

[87] Rolf Pfeifer, Fumiya Iida, and Gabriel Gómez. Morphological computation for adaptive behavior and cognition. *Int. Congress Series*, 1291:22–29, 2006.

[88] Rolf Pfeifer and Christian Scheier. *Understanding Intelligence.* MIT Press, 1999.

[89] David Pierce and Benjamin Kuipers. Map learning with uninterpreted sensors and effectors. *Artificial Intelligence*, 92:169–229, 1997.

[90] Raphael Pinaud, Liisa A. Tremere, and Peter Weerd, editors. *Plasticity in the Visual System: From Genes to Circuits.* Springer US, 2006.

[91] Xaq Pitkow and Markus Meister. Decorrelation and efficient coding by retinal ganglion cells. *Nature Neuroscience*, 15:628–635, 2012.

[92] A. Pouget and L. Snyder. Computational approaches to sensorimotor transformations. *Nature Neuroscience*, 3:1192–1198, 2000.

[93] J. F. Poulet and B. Hedwig. A corollary discharge mechanism modulates central auditory processing in singing crickets. *J. Neurophysiology*, 89:1528–1540, 2003.

[94] J. F. Poulet and B. Hedwig. New insights into corollary discharges mediated by identified neural pathways. *Trends in Neuroscience*, 30:14–21, 2007.

[95] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007.

[96] P. Prusinkiewicz and A. Lindenmayer. *The Algorithmic Beauty of Plants*. Springer Verlag New York, 1990.

[97] C. Quaia, L. M. Optican, and M. E. Goldberg. The maintenance of spatial accuracy by the perisaccadic remapping of visual receptive fields. *Neural Networks*, 11:1229–1240, 1998.

[98] Michael B. Reiser and Michael H. Dickinson. A test bed for insect-inspired robotic control. *Phil. Trans. R. Soc. Lond. A*, 361:2267–2285, 2003.

[99] G. Rizzolatti and L. Craighero. Spatial attention: Mechanisms and theories. *In Sabourin et al., Adv. in Psych. Sc.: Biological and Cognitive Aspects*, 1998.

[100] D. A. Robinson. Eye movements evoked by collicular stimulation in the alert monkey. *Vision Research*, 12:1795–1808, 1972.

[101] Jonas Ruesch and Alexandre Bernardino. Evolving predictive visual motion detectors. In *Proc. Int. Conf. on Development and Learning (ICDL), Shanghai, China*. IEEE, 2009.

[102] Jonas Ruesch, Ricardo Ferreira, and Alexandre Bernardino. A measure of good motor actions for active visual perception. In *Proc. Int. Conf. on Development and Learning (ICDL), Frankfurt, Germany*. IEEE, 2011.

[103] Jonas Ruesch, Ricardo Ferreira, and Alexandre Bernardino. An approach toward self-organization of artificial visual sensorimotor structures. In *Proc. Int. Conf. on Biologically Inspired Cognitive Architectures, Palermo, Italy*. Springer, 2012.

[104] Jonas Ruesch, Ricardo Ferreira, and Alexandre Bernardino. Predicting visual stimuli from self-induced actions: an adaptive model of a corollary discharge circuit. *IEEE Transactions on Autonomous Mental Development*, 4:290 – 304, 2012.

[105] Jonas Ruesch, Ricardo Ferreira, and Alexandre Bernardino. Self-organization of visual sensor topologies based on spatio-temporal cross-correlation. In *Proc. 12th Int. Conf. on Simulation and Adaptive Behavior (SAB), Odense, Denmark*. Springer, 2012.

[106] Jonas Ruesch, Manuel Lopes, Alexandre Bernardino, Jonas Hörnstein, José Santos-Victor, and Rolf Pfeifer. Multimodal saliency-based bottom-up attention a framework for the humanoid robot icub. In *Proc. Int. Conf. on Robotics and Automation (ICRA), Pasadena, USA*. IEEE, 2008.

[107] Kepa Ruiz-Mirazo, Jon Umerez, and Alvaro Morena. Enabling conditions for "open-ended evolution". *Biology and Philosophy*, 23:67–85, 2008.

[108] Terence D. Sanger. Probability density estimation for the interpretation of neural population codes. *J. Neurophysiology*, 76:2790–2793, 1996.

[109] Terence D. Sanger. Probability density methods for smooth function, approximation and learning in populations of tuned spiking neurons. *Neural Computation*, 10:1567–1586, 1998.

[110] Terence D. Sanger. Neural population codes. *Current Opinion in Neurobiology*, 13:238–249, 2003.

[111] Peter H. Schiller. The role of the monkey superior colliculus in eye movement and vision. *Invest Ophthalmol.*, 11:451–460, 1972.

[112] Brigitte Schoenemann. Aus trilobiten-sicht. *Fossilien*, pages 148–156, 2011.

[113] Jens Schouenborg. Learning in sensorimotor circuits. *Curr. Opin. Neurobiology*, 14(6):693–697, 2004.

[114] I. R. Schwab. Flatlanders. *British Journal of Ophthalmology*, 88:988, 2004.

[115] Ivan R. Schwab. *Evolution's Witness, How Eyes Evolved*. Oxford University Press, 2011.

[116] A. B. Schwartz, R. E. Kettner, and A. P. Georgopoulos. Primate motor cortex and free arm movements to visual targets in three-dimensional space i. relation between single cell discharge and direction of movement. *J. Neuroscience*, 8:2913–2927, 1988.

[117] E. L. Schwartz. Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. *Vision Research*, 20(1):645–669, 1980.

[118] M. A. Sommer and R. H. Wurtz. What the brain stem tells the frontal cortex. i. oculomotor signals sent from superior colliculus to frontal eye field via mediodorsal thalamus. *J. Neurophysilogy*, 91:1381–1402, 2004.

[119] M. A. Sommer and R. H. Wurtz. Influence of the thalamus on spatial visual processing in frontal cortex. *Nature*, 444:374–377, 2006.

[120] M. A. Sommer and R. H. Wurtz. Brain circuits for the internal monitoring of movements. *Ann. Rev. Neuroscience*, 31:317–338, 2008.

[121] R. Sperry. Neural basis of the spontaneous optokinetic response produced by visual inversion. *J Comp Physiol Psychol*, 43:482–489, 1950.

[122] M. V. Srinivasan, S. B. Laughlin, and A. Dubs. Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences.*, 216:427–459, 1982.

[123] Jonathan Stone and Paul Halasz. Topography of the retina in the elephant loxodonta africana. *Brain, Behavior and Evolution*, 34:84–95, 1989.

[124] Nicholas V. Swindale. How different feature spaces may be represented in cortical maps. *Network: Computation in Neural Systems*, 15, 2004.

[125] S. J. Tan, W. Amos, and S. B. Laughlin. Captivity selects for smaller eyes. *Curr. Biol.*, 15:R540–R542, 2005.

[126] A. T. Thomas. Developmental palaeobiology of trilobite eyes and it evolutionary significance. *Earth Science Reviews*, 71:77–93, 2005.

[127] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Royal Statistical Society. Series B*, 58:267–288, 1996.

[128] A. S. Tolias, T. Moore, S. M. Smirnakis, E. J. Tehovnik, A. G. Siapas, and P. H. Schiller. Eye movements modulate visual receptive fields of v4 neurons. *Neuron*, 29:757–767, 2001.

[129] M. M. Umeno and M. E. Goldberg. Spatial processing in the monkey frontal eye field. i. predicitive visual responses. *J. Neurophysiology*, 78:1373–1383, 1997.

[130] M. M. Umeno and M. E. Goldberg. Spatial processing in the monkey frontal eye field. ii. memory responses. *J. Neurophysiology*, 86:2344–2352, 2001.

[131] J. H. van Hateren. Real and optimal neural images in early vision. *Nature*, 360:68–70, 1992.

[132] S. Vaziri, J. Diedrichsen, and R. Shadmehr. Why does the brain predict sensory consequences of oculomotor commands? optimal integration of the predicted and the actual sensory feedback. *J. Neuroscience*, 26:4188–4197, 2006.

[133] W. E. Vinje and J. L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287:1273–1276, 2000.

[134] E. von Holst and H. Mittelstaedt. Das reafferenzprinzip (wechselwirkung zwischen zentralnervensystem und peripherie). *Naturwissenschaften*, 37:464–476, 1950.

[135] Barbara Webb. Neural mechanisms for prediction: Do insects have forward models? *Trends in Neuroscience*, 27:278–282, 2004.

[136] Rüdiger Wehner. 'matched filters'- neural models of the external world. *Comp. Physiol. A*, 161:511–531, 1987.

[137] Daniel M. Wolpert, J. Diedrichsen, and J. R. Flanagan. Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12:739–751, 2011.

[138] Jochen Zeil and J. M. Hemmi. The visual ecology of fiddler crabs. *Journal of Computational Physiology A*, 192:1–25, 2006.