Predicting Sensory Performance

Maps, Optimality, Feature Extraction, and Learning

Andreas B. Sichert

Cover illustration: Photograph © courtesy
of Guido Westhoff
Autor: Andreas B. Sichert
Title: Predicting Sensory Performance
Subtitle: Maps, Optimality, Feature Extraction, and
Learning

Funding has been provided by the Bernstein Center
for Computational Neuroscience (BCCN) – Munich,
the CoTeSys Exzellenz Cluster, and the Sloan-Swartz
Foundation.

Physik Department
Technische Universität München

Predicting Sensory Performance

Maps, Optimality, Feature Extraction, and Learning

Andreas B. Sichert

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Th. Hugel
Prüfer der Dissertation:	1. Univ.-Prof. Dr. J.L. van Hemmen
	2. Univ.-Prof. Dr. R. Netz

Die Dissertation wurde am 14.07.2010 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 20.09.2010 angenommen.

Preface

Wow, you are *reading* my PhD thesis! I mean, you are reading – which is an almost unbelievably complex process. It is, however, a good example of what makes neuroscience so fascinating. Let's take a closer look at the above example of *reading* to understand what neuroscience is all about.

At first glance, reading starts with some tiny black dots on a white sheet of paper which form geometrical structures that we denote as letters. We can *perceive* them because our eyes and the subsequent neuronal processing of the visual system are able to translate pictures in the external world into neuronal information that forms a corresponding representation within our brain. The visual system, however, does not settle information processing by presenting clouds of black dots. At this point, information processing starts to get really sophisticated. Cognitive neuronal circuits group the neuronal representation of these clouds of dots to mental *objects* – the letters. The individual letters are *related* to each other to form words that represent real objects, attributes, activities etc. To finally gain a reasonable *meaning*, the brain uses the underlying grammar that is the specific methodology of the particular language to build sentences out of individual words.

For anyone who is not already impressed let me point out two further issues. First, letters, words, and sentences are only meaningful because we relate these clouds of dots and the corresponding abstract mental objects to past sensory experience. We can interpret the word "barking" because our visual system relates the black dots on the sheet of paper at hand to the sound of a dog heard some time ago. This is due to the fact that different modalities are able to interact faithfully, for instance, by means of combining the correct memories of visual and auditory sensory events.

Second, the answer to the question of where such intelligent neuronal circuits come from is elusive. May evolution be an answer, viz., are these neuronal structures encoded in our genes? No, not in the first place. Otherwise there would be genes for Bavarian, German, English, and other languages. The ability to read, to interpret abstract methodology and symbols, is *learned*. Moreover, there is no teacher that gives babies lectures in their mother tongue. So, our brain teaches itself, that is it learns purely from sensory experience to build itself in a way to be best suited for processing sensory experience.

Comparing these abilities – our abilities – with those of man-made and so called "cognitive" systems is impressive. In contrast to nature we have, however, just started to invent such systems. From my personal perspective that is what makes the field of neuroscience so fascinating: Discovering the nervous system, a hell of a machinery!

The tools for discovering the nervous system are manifold. It turns out that a broad interaction of different disciplines is necessary to enlighten this complex topic. By means of this thesis we look at the nervous system through the glasses of theoretical neuroscience. Theoretical neuroscience is situated at the intersection of biology, medicine, and physics employing mathematical tools. In short, this dissertation focuses on "Predicting Sensory Performance" since sensory experience is the major driving force for the whole system – including its development. For that purpose, we will derive mathematical descriptions of sensory systems, feature extraction and learning, and apply them to specific settings, viz., different senses.

Chapter 1

The first chapter gives an introduction to fundamental concepts of sensory processing such as that of a neuronal map which are needed to understand the rest of the work. Besides, the references given in the compact introduction should provide the reader with a good starting point for own studies.

Chapter 2

In order to exploit the information provided by sensory systems, animals reconstruct the spatio-temporal environment from their sensory systems. In this chapter we derive a universal framework that allows to calculate the specific layout of the involved neuronal network by means of a general mathematical principle, viz., stochastic optimality. That is, given a known physical signal transmission and rudimental knowledge of the detection process, this approach allows to estimate the possible performance and to predict neuronal properties of biological sensory systems.

Chapter 3

In this chapter we apply our framework derived in Chap. 2 to the auditory system. That is, coming from the general mathematical principle of optimality we deduce an archetype model and corresponding temporal receptive fields for auditory signal enhancement and echo suppression. We compare the performance of both the neuronal and the mathematical implementation of the model to state-of-the-art algorithms and biological solutions.

Chapter 4

The lateral-line system is a unique mechano-sensory facility of aquatic animals that enables them not only to localize prey, predators, obstacles, and conspecifics but also to recognize hydrodynamic objects as we demonstrate in this chapter. We present both an analytic approach that shows the effect of submerged moving object shape and an explicit model explaining how aquatic animals can distinguish different shapes. Furthermore, we predict natural limitations of this sensory capability.

Chapter 5

Two groups of snakes possess an infrared detection system to perceive a heat-image of their environment. This chapter explores the different stages of the infrared pathway such as transmission, detection, and neuronal en- and decoding in order to provide fundamental insights to this sensory system. The thorough physical description of the involved phenomena as well as the incorporation of anatomical data into an explicit model extend the present picture of this sensory system.

Chapter 6

For neuronal processing of sensory information uni- and multimodal maps play a key role. The precise interaction of such maps is not present at birth but has to be learned during the development of an animal by means of sensory experience. This chapter presents the concept of integrated Multimodal Teaching that combines supervised spike-timing-dependent plasticity and multimodal integration to inherently develop a faithful interaction of different sensory modalities.

Thank you

Many people have supported me during my time at the Physik Department of the Technische Universität München. In the first place, I thank Leo van Hemmen, my thesis supervisor. He has introduced me to the topic of theoretical neuroscience, viz., biophysics. His dedication to scientific research is very impressive, and through many inspiring discussions I have learned a lot about science.

Writing this thesis would not have been possible without the pleasant interaction and cooperation with my colleagues at the "T35" chair and a few people beyond. In the context of this thesis, I may say thank you to Moritz Bürck, Stefanie Bürck, Roland Utz, Andreas Vollmayr, and Christine Voßen who have proofread parts of the thesis and have given helpful comments.

As a whole, I was given the opportunity to make many valuable experiences during the time of my PhD work; in addition, I was able to join the Bayerische EliteAkademie and to visit Harvard University. I have also met exceptional people and have made new friends. For all of this I am very grateful.

Finally, I thank my friends, my parents, my brother, and Babsi, who have always been there to support me.

Munich, June 2010

Contents

Die Reise von tausend Meilen
beginnt zu deinen Füßen.

Laotse

1. Processing Sensory Information

This chapter provides a synopsis of the wide area of sensory information processing. Looking through the glasses of this work, we apply Occam's razor by means of introducing and focussing on the essential concepts used in this thesis enabling us to derive predictions on sensory performance in the subsequent chapters. It starts with the question of what a stimulus in external world is, discusses the conceptual steps in the sensory pathway, and presents a brief overview of neuronal representation by means of maps. Finally, we focus our attention on map formation and introduce synaptic plasticity, a tool for understanding development of a common multi-sensory space that enables beings to act faithfully in the external world.

1.1 Steps of sensory processing

Sensory processing can be divided into three *conceptual* steps as depicted in Fig. 1.1 $(A-C)$. First, a stimulus is physically mapped onto detector organs. Second, the detector organs respond to the stimulus in a specific way and translate the physical signal into neuronal information, viz., spikes of the corresponding primary sensory neurons; for introductions to neurons, neuronal computation see e.g. [164, 267], and for the different levels of modeling neuroscience e.g. [119]. Third, the information contained within the spike-trains of the sensory system is processed neuronally so as to provide a useful representation of the original stimulus to, for instance, initiate motor commands.

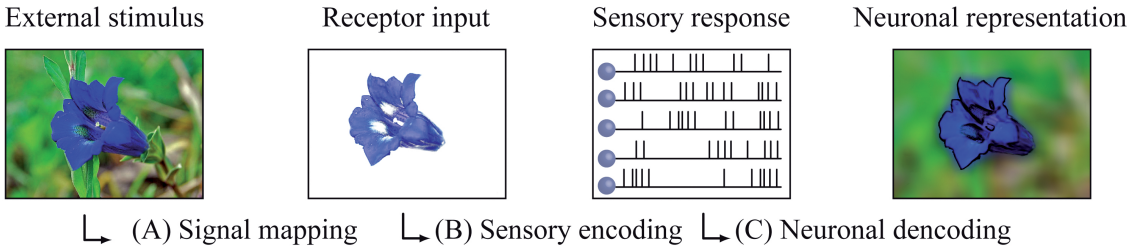

Figure 1.1: *Three conceptual steps $(A-C)$ of sensory processing. Every stimulus in the external world emits signals, in our example light. (A) First, the signals are mapped onto the sensory organs where they affect the physical state of specialized (protein) structures within the (photo) receptors. (B) Second, the sensory neurons respond only to specific stimuli, for instance blue light, by means of generating corresponding spikes. (C) Third, the neuronal system decodes the imbedded data within the spike-trains to reveal the essential information in a useful way.*

1.1.1 Signal mapping

Initially, any object in the external world reveals its presence by generating different signals. A running animal may, for instance, generate a sound and a changing visual image as it moves, as well as vibrations and an infrared profile. The generation and transmission process follows specific physical laws. This fact, as simple as it may look at first glance, provides a powerful tool for understanding sensory processing: the *mathematical* description of sensory stimuli as physical observable quantities.

In concrete terms, given any signal that varies as a function of spatial position $\mathbf{x}$ and time t, it is possible to calculate the time-dependent input $\mathbf{i}^{\mathbf{x}}(t)$ to the receptor organs. That is, the physical mapping of signals $\mathbf{s}^{\mathbf{x}}(t)$ onto sensory detectors can be described by a set of *transfer functions* h,

$$h : \mathbf{s}^{\mathbf{x}}(t) \mapsto \mathbf{i}^{\mathbf{x}}(t) \ . \tag{1.1}$$

Moreover, real objects in the external world typically emit a bundle of signals through several channels based on different physical pathways at the same time. That is, they permanently leave a *multi-sensory* fingerprint that can be detected and processed by sensory systems to form a coherent representation of the external world in space-time. This issue will lead us to *multimodal integration* and later on to the question, how this precise interplay of our senses evolves. We leave the issue of multimodal processing for the time being and now resume discussing the fundamental steps of sensory processing within one modality.

1.1.2 Sensory encoding

In the second step of sensory processing, the signal that is mapped onto the sensory detectors is translated into neuronal information – spikes. This process can also be described by a set of transfer functions k. In the beginning of the process, the external signal affects the physical state of the detector. Coming back to the example in Fig. 1.1, light in the visual range – simplified – changes the state of the wavelength-sensitive photo receptors in the retina. By means of this physical process, spike generation is induced. The resulting spikes $\mathbf{t}_i(t)$ consequently encode the intensity of the corresponding specific wavelength [263, 267, 281] at a particular detector position i and time t,

$$k : \mathbf{i}^{\mathbf{x}}(t) \mapsto \mathbf{t}_i(t) \ . \tag{1.2}$$

Two aspects of sensory encoding are of further interest in the context of this thesis. The first issue is that sensory encoding includes at least two steps of signal filtering. The physical change of state within the receptor is specific to a particular stimulus characteristic and hence filters the signal. In our example, the photo receptor responds only to blue light; for review of the visual signal cascade see [263]. So, the wavelength-specificity selects out light of all other colors, cf. Fig. 1.1.

In addition, the process of spike generation also filters the signal. We now take a closer look at this important issue. First, translating sensory signals into neuronal information compares to an analog-digital-conversion process because a spike code is a binary code constituted of zeros (no spikes at this point in time) and ones (spike at this particular time). Second, in the process of spike generation at the level of one single neuron, biophysical dynamics of cell compartments limit the sampling rate, e.g. refraction time limits the spike frequency [267, 164]. This results in a restrained sampling accuracy, that is, the analog-digital-conversion process principally reduces the information content. Third, the neuronal transmission process also alters the imbedded information. Synapses, for instance, typically feature complex non-linear dynamics that contribute to an iterative sequence of fragmentation, selection, and amplification of information, and thus filter the signal [89]. Fourth, there are different kinds of neuronal coding that additionally may modify the original information. Neurons can respond to spike rates ("rate coding") but in other cases neurons respond

to firing onset, change of spike rates, skipping, bursting, or complex temporal pattern – just to list a few [18, 37, 247]. We see, generating and transmitting neuronal information presents definitely a filtering process. There is no doubt, however, that the neuronal filtering processes need not generally be bugs. They may, on the contrary, be necessary features reducing the arriving physical signals to the essential ones and thus making external sensory information streams computationally feasible.

The second important aspect of sensory encoding in this thesis is *topography*. Some sensory organs are organized in such a way that they are able to receive and conserve topologically ordered information. In our example, the retina shows a topological organization of the detectors and lenses assuring a well defined topological relation between positions in the external world $\mathbf{x}$ and the detector positions i. Translating spatial information from the external world into neuronal information, however, induces a transfer from continuous topologies to discrete topologies. This spatial analog-digital-conversion process at least causes uncertainties in the sensory pathway.

Topography, moreover, is not only a spatial characteristic. It is also important in the temporal domain. Within the auditory system it is crucial to know the temporal sequence of events in order to, for instance, process interaural time differences. Within the limitation of a neuronal system our brain is even able to translate neuronally temporal data into spatial information to gain a useful representation of external space. The Jeffress' model is a famous example of such a translation process [132].

1.1.3 Neuronal decoding

In the next stage of sensory processing, the neuronal system has to decode the spike-trains $\mathbf{t}_i(t)$. They may come from different sensory neurons i which may additionally have a spatial or temporal relation. In mathematical terms, the neuronal system *principally* has to solve the *neuronal inverse* problem. That is, the system has to invert the composition of transfer functions h (signal generating and signal mapping) and k (sensory detecting and neuronal encoding) to reveal the original signal $\mathbf{s}$. In general, it is hardly possible to find an exact solution. The system, however, can try to find an *optimal* inverse function l to reconstruct $\hat{\mathbf{s}}$ the original signal from the neuronal data

$$l : \mathbf{t}_i(t) \mapsto \hat{\mathbf{s}}^{\mathbf{x}}(t) \ . \tag{1.3}$$

Additionally, in many cases the sensory system even settles for computing the essential stimulus information instead of reconstructing the complete stimulus. Energy costs and the limitations of neuronal processing capabilities may be reasons for that.

The limitations each sensory system is faced with lead to another global issue in sensory processing. A sensory system can only pose and answer the following question: Given uncertainties, limitations, and a sensory response – what is the most *probable* signal that could have caused this sensory response? For illustration, the finite number of detectors is not able to sample continuous signals or space without uncertainties. The limiting temporal accuracy of neuronal computation as well as the neuronal noise level additionally increase processing impreciseness. Consequently, and no matter what causes these system restrictions, the question itself reduces naturally to a question of *conditional* probability. So, it is not surprising that everywhere in the course of sensory processing we find neuronal structures behaving like estimators [7, 59, 61, 67, 69, 113, 120, 183, 220, 237, 238, 239] that, for instance, can optimally solve the neuronal inverse problem [75, 76, 252]. It remains, however, to reveal how these estimators work on the level of neuronal circuitry and to quantify their performance.

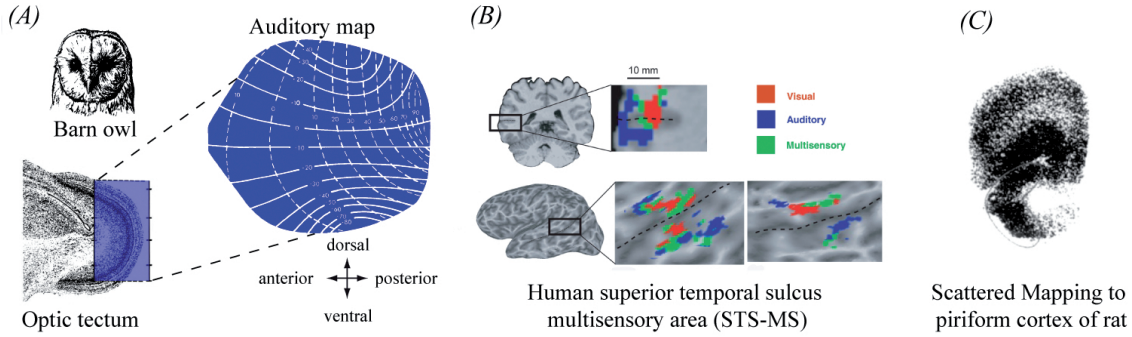

Figure 1.2: *Three categories of maps based on their topography; adapted from [11, 154, 208]. Nelson and Bower [208] suggest three map groups, i.e. continuous maps (A), patchy maps (B), and scattered maps (C). (A) Map corresponds to a continuous mapping of auditory space in the optic tectum of the barn owl showing a global topography [154]. (B) The patchy mapping of visual, auditory, and multisensory sensitive areas in the human superior temporal sculus multisensory area shows a topography only within small and distinct areas of the whole organization [11]. (C) In spite of the spatial organization in the latter two examples, the scattered mapping of olfactory input to the piriform cortex of the rat lacks any spatial organization. Depending on the computational task, the different structures lead to different computational advantages as briefly addressed in the main text.*

In summary, the physics of external signal transduction dictate the sensory data structure. The aim of all the resulting neuronal effort is to achieve a reasonable and useful representation of the original stimulus information, viz., the signal. The central nervous system (CNS) can represent the computed information in different ways. In this thesis the population code and, more importantly, the topologically ordered neuronal map play an essential role. The map accuracy defined as to how precise a map can represent an external stimulus is a good measure for the performance of the corresponding sensory system. So we will consider maps in the next paragraph in more detail.

1.2 Neuronal maps

A major role in sensory processing is reserved for neuronal *maps*[1] [116, 161]. The concept of a map suggests that there is a lawful relation between the surface of the neuronal substrate and a relevant aspect of the external space. To cut a long story short, a neuronal map is typically an array of neurons in which neighboring neurons respond to "similar" sensory stimuli. This simplification holds true at least for spatial sensory maps on which we focus in the course of this thesis.

As an example, visual input in the mammalian brain is processed through multiple cortical layers that are organized according to the topography of the retinal input cells ("retinotopic organization") [281, 143]. In this way, neighboring neurons respond to visual input from neighboring points in space, and thus form a spatial map. Spatial maps have been discovered in various sensory systems in many groups of vertebrates [10, 40, 62, 109, 123, 149, 151, 192, 259, 264, 295] and even in some non-vertebrates [130].

[1] In 1879, Helmholtz [114] remarked "Dass durch das Entlangführen des tastenden Fingers an den Objecten die Reihenfolge kennen gelernt wird, in der sich ihre Eindrücke darbieten, dass diese Reihenfolge sich als unabhängig davon erweist, ob man mit diesem oder jenem Finger tastet, dass sie ferner nicht eine einläufig bestimmte Reihe ist, deren Elementen man immer wieder vor- oder rückwärts in derselben Ordnung durchlaufen muss, um von einem zum anderen zu kommen, also keine linienförmige Reihe, sondern ein flächenhaftes Nebeneinander, oder nach Riemann's Terminologie, eine Mannigfaltigkeit zweiter Ordnung, das alles ist leicht einzusehen." That is, Helmholtz already recognized that a neuronal representation of a two-dimensional surface constitutes a two-dimensional manifold in the brain.

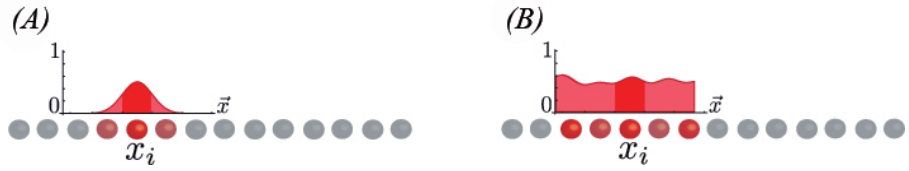

Figure 1.3: *The firing profile of a spatial map encodes the likelihood of finding an object at a certain position in sensory space. Focusing on the firing rate of only a* single *neuron such as neuron i with preferred direction x_i and ignoring its neighbors prevents a faithful perception of reality. A faithful perception can only be achieved if the activity of the whole neuronal map (just compare the "neighborhood" of i with the rest) is taken into account. Consequently, even though the nominal value of the firing rate of neuron i is identical in case A and B, the represented physical reality differs significantly in both cases.*

Neuronal maps are not limited to spatial representations. Prominent examples are found in the auditory system. There are maps representing frequency, interaural time difference, interaural amplitude difference, and even amplitude modulation frequencies [84, 187, 214, 225, 245, 265]. So, we see that maps are neither restricted to particular sensory modalities nor particular domains such as space.

A typical way of categorizing in physics is to search for characterizing *length scales*. Looking at maps, we can find two types of length scales. First and obviously we can consider the dimensions of the neuronal array. Following the presentation of Dehaene and Cohen [58], one may hence distinguish maps by their size into macromaps (a few centimeters), mesomaps (up to centimeters), and micromaps (millimeters). Second, beside the length scale of map dimensions we can consider the scale up to which a common topologic map organization can be identified. Nelson and Bower [208] therefore suggest three categories, continuous map, patchy map, and scattered map; see Fig. 1.2. These categories presumably arise mainly from computational efficiency reasons as we will see in the next section. Strictly speaking, the latter, the scattered map, is not a map as it shows no topography.

Intrinsically tied to the concept of a neuronal map is that of *receptive fields*. A receptive field of a sensory neuron is defined as the region of space in which the presence of a stimulus alters the activity of the neuron. Since receptive fields of sensory maps are easily accessible experimentally they are well studied, and the corresponding data shed light on the underlying map properties.

1.2.1 Map properties

As we have seen just above, maps are used in many areas in the brain. The compelling question [116] therefore is: What is the functional advantage of neuronal maps? That is, why choose a map structure for neuronal processing? Here, we focus on neuronal maps with a global topography, viz., continuous maps; cf. Fig. 1.2. From the theoretical perspective, the fundamental argument is that, in contrast to *arbitrary* population coding or clustered neuronal architectures [51], neuronal maps ensure a topographic neuronal organization.

This conceptual advantage of maps arises from combining basic topological information with actual stimulus-evoked responses. That is, in case of sensory maps, the preferred direction of the sensory organs or receptors is linked with their actual activity. In other words, merging topography and activity allows for interpreting the firing profile of map neurons with respect to their neighbors, viz., context-dependent. For example, as illustrated in Fig. 1.3, although the activity of neuron i with preferred direction x_i is the same in (A) and (B), in the context of its neighboring neurons its information content differs drastically from case (A) to (B).

Moreover, the map organization underlying the neuronal processing allows for an efficient

representation of continuously varying input signals as *stochastic quantities*. For instance, it allows the likelihood interpretation of a firing pattern on a spatial map, that is, the *likelihood* to find a sensory object at a certain position [60, 116, 131, 227, 248]. In the context of maps, we cannot retrieve anything from the firing profile Fig. 1.3 (B) but a noisy background whereas Fig. 1.3 (A) presents quite nicely the probability of finding a corresponding sensory object. That is, the latter correlates with a positional estimator for sensory objects. While sensory maps encoding spatial or temporal estimators are the most evident ones, maps, especially maps in higher brain areas, are not limited to spatial or temporal topology. They are found to encode various stimuli underlying various topologies (review by Knudsen et al. [161]). So, getting to the point, the real power of any kind of map is to provide a *systematic* representation of complex stimuli.

In the context of this thesis, the likelihood interpretation of the firing pattern evoked by a sensory object presented above is crucial. Beyond this issue, maps provide other computational advantages. One, as we have seen previously, relates to the degrees of topologic order. Nelson and Bower [208] suggested that different categories (see Fig. 1.2) of topologic order are due to minimizing computational effort similar to parallel computers. The general hypothesis is that map organization reflects map usage and input structure in order to achieve a minimal communication overhead and an optimal load balancing. The former and the latter are found to be realized neuronally by the three kinds of maps mentioned above, namely, continuous map, patchy map, and scattered map; cf. Fig. 1.2. Continuous maps, for instance, are found to be used mainly for local operations in the problem space such as feature extraction and spatial filtering. We have to bear in mind that computational tasks are performed mainly but not only by means of synaptic connections. Chemical processes are also used for information processing, for instance, to inhibit a complete area of neurons representing similar features of the stimulus. Thus the concept of a map does not only minimize neuronal wiring effort but enables reasonable chemical control. "Load balancing" is achieved by local map distortions that reflect differences in sensory densities. For instance, there is a much higher density of somatosensory receptors in the tongue compared to the skin in our back. The increased sensory sensibility of the tongue is reflected by distortions of the corresponding sensory map where more neurons encode the same physical space than in our back. The patchy maps seem to additionally include non-local operations to analyze local sensory data within a more global context. But this is still matter of ongoing debate. The scattered maps such as found in the olfactory system reflect the non-topographic nature of the input. Here, the system identifies complex stimuli and answers the question of what rather than that of where.

1.2.2 Collection of computational maps

There is an even more convincing explanation for the omnipresence of maps in the brain. The real power of computational maps can be appreciated only when the *interplay* of several maps is considered; for review see [161]. In the context of this thesis we define *computational* as an arbitrary transformation in the representation of information. To give a famous example of how information is processed in a sequence of computational maps we consider the auditory system; cf. Fig. 1.7. Frequency-depending time differences are computed within a map in the central nucleus of the inferior colliculus (ICC). In a next step, this map is translated into a spatial map in the external nucleus of the inferior colliculus (ICX) where the information is further projected into the optic tectum (for review see [154, 155, 161]). Although the various early sensory cues in the auditory system are hardly related to space, in the optic tectum (OT) the auditory map represents space and even connects smoothly to spatial maps of other modalities.

The OT and its well-studied mammalian homologue, the superior colliculus (SC) [261], is a beautiful example of the connection between different maps. All sensory systems that provide information in a map-like form project to the SC/OT [261]; cf. Fig. 1.4. Within the SC/OT, multisensory as well as predominantly monosensory layers are found and are shown to process spatial

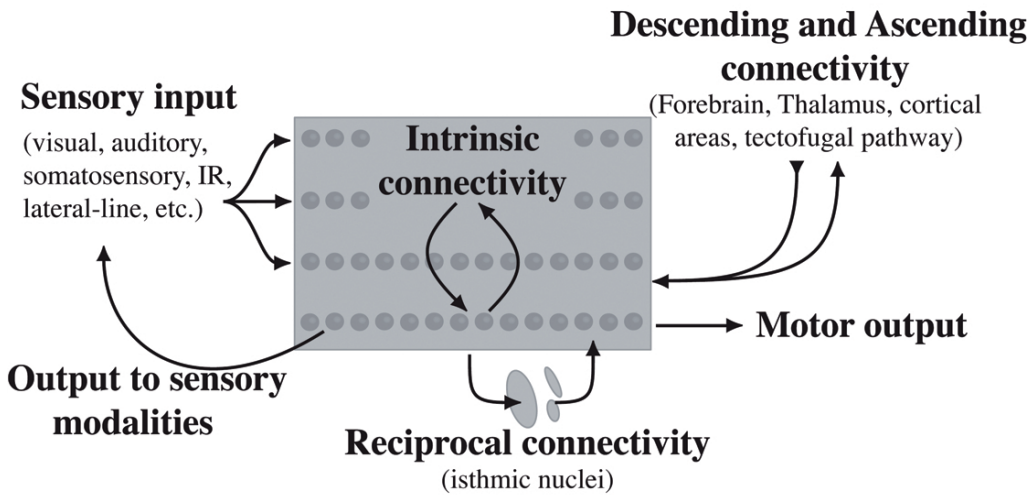

Figure 1.4: *The superior colliculus (SC). The superior colliculus and its non-mammalian homologue, the optic tectum (OT), are well studied mid-brain structures to which all sensory systems project that provide information in a map-like form [261]. Within the SC/OT, spatial sensory information of different monosensory maps is computed and combined into multisensory maps. The combined information is used for delivering unambiguous motor commands, for refining sensory modalities, and for further processing in higher brain areas [166, 167, 179, 216, 260]. Because all maps in the SC/OT are mutually aligned, we can speak of a collection of maps that represent an aligned sensory space. The figure is adapted from [179].*

information. In multisensory maps, most neurons respond to more than one sensory modality. In doing so, the neuronal response to more than one sensory modality can be complex. In case of spatial and temporal coincidence we can observe super-additivity, cross-modal enhancement [134, 224, 261], or inverse effectiveness [224] whereas in case of non-coincidence cross-modal suppression can be found [140]. Multimodal maps are widespread in the brain [139].

Both the multisensory maps and the monosensory layers are found to be aligned in the SC/OT to gain a unified multisensory representation of sensory space [147, 261]. The combined sensory information can then be used to generate directional motor responses [166, 167, 179, 216, 260]. This is only possible since, within the SC, there are not only sensory but also motor maps, i.e., motor neurons organized in a map-like structure. In motor maps, systematic variations of a movement are represented topographically across the neuronal array. Direct evidence for this hypothesis has recently been found in eye-tracking experiments [104].

In a collection of aligned neuronal maps, moreover, external *objects* can be *identified* by their position encoded by means of the firing pattern in different sensory maps. This does not mean that an object is adequately described if only the position is known. It rather implies that position serves as appropriate and – in contrast to higher computational levels [65, 98] – necessary information for defining a sensory object; cf. Fig. 1.5. When combining different sensory systems, the spatial information is needed to bind information associated with the same position in the monosensory maps into one single multimodal percept for further processing.

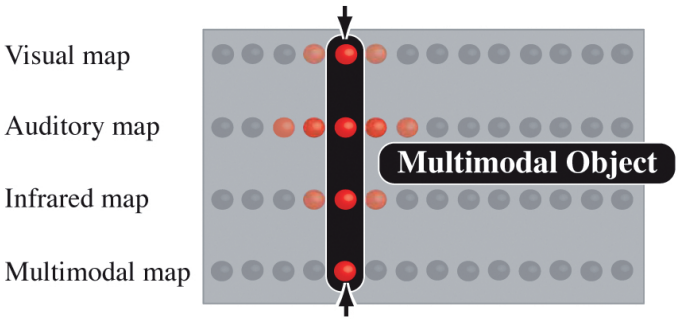

Figure 1.5: *Object formation by spatial information binding. Within an* atlas *there is a simple way to define a sensory object by means of grouping spatial information, viz., spatial congruent activity pattern (red column). Up to now, such a mechanism is speculative. However, it can be relevant, for instance, to select a multimodal target (e.g. running and barking dog) within a dynamic environment. There are lines of evidence supporting the existence of such a mechanism. A prominent example are the neuronal "AND" and "OR" circuits within the OT that link infrared and visual map [210]. These circuits are supposed to identify external objects such as prey and enhance their context-dependend saliency.*

Bearing in mind the advantages of neuronal maps that are due to topography we next face the question of how a neuronal map can develop. Later in this thesis, we additionally study how a collection of maps can evolve.

1.2.3 Map development

One might argue that neuronal maps simply exist because their neuronal architecture *follows* the sensory surface of their input modality. From this point of view, for instance, the visual layers are retinotopically organized *because* they receive their input from the retina. Similarly, a frequency map just reflects the tonotopic organization of the cochlea.

Interestingly, this argument does not hold for every sensory map as it already becomes apparent in the existence of auditory maps for interaural time and amplitude differences. Regarding examples in the spatial domain, let us consider the sensory units of the frog lateral-line system [295] or the snake infrared system [109]. These sensory units receive a complex superposition of input from several different spatial locations. Computing a map from such a complex input [75, 76, 95, 252] is certainly *not* straightforward. So, how can maps evolve?

The most obvious idea is to imagine an explicit genetic mechanism encoding map topography, viz., connectivity. In this case, however, we have to genetically encode a large number of possible connections. Our brain consists of about 10^{12} neurons and 10^{15} synaptic connections. That is, we have to encode a number of possible connections comparable to the number of atoms in the universe [267]. Of course, our $2 - 3 \cdot 10^4$ human genes cannot achieve this *directly*. Consequently, *intrinsic* mechanisms must determine map development.

In the last years, neurobiologists formulated four main hypotheses as to how topographic-correct synaptic map connections arise; reviews e.g. [34, 181, 190, 271]. First, following the *address label hypothesis*, synaptic structures possess a series of positional labels telling ingrowing axons or pre-synaptic axon terminals how to arrange. Second, neighboring axons *conserve pre-synaptic topography* during the growth process. Therefore, the post-synaptic structures are topographically arranged. Third, both the ingrowing axons and the target structures develop in a topographically

arranged time sequence. The *temporally matching* entails a spatial map topography. Fourth, pre-synaptic axon terminals arrange themselves topologically by means of electrical *activity*. Although evidence for each of these hypotheses can be found [271], several mechanisms seem to cooperate to ensure a proper map topography. To give an example, electrical activity-based adaption is supposed to fine-tune map connectivity after some imprecise prewiring which can be based on other mechanisms such as chemical labeling [13, 146].

Precise map topography is crucial for sensory performance. In the next section, we therefore shed light on *activity-based*, viz., spike-dependent map formation by considering local mechanisms and global structures responsible for the plasticity of map topography.

1.3 Learning in sensory systems

In the context of sensory systems, learning basically means a change in the input-output relation. That is, sensory activity (input) leads to a different map activity (output). Consequently, learning takes place by means of changing the connection strengths between (input and output) neurons. In 1949, Hebb already translated this general insight of activity-based plasticity into neuronal terms by formulating his famous rule[2] [112]. In short, if a post-synaptic neuron excites a spike shortly after one of its pre-synaptic neurons has fired, this particular connection is strengthened. In other words, Hebb's learning rule describes that temporal correlations of neuronal activity – if causal – are learned. To cut a long story short, neuronal activity relates typically to events in the external world and thus Hebb's rule impressively enables us to understand very generally how to learn correlation in the external world by means of a *local* neuronal mechanism.

In the context of this thesis "Predicting Sensory Performance" two questions directly connected with Hebb's rule arise. First, can we understand aspects of sensory performance by means of studying *local* mechanisms of synaptic plasticity? Second, what provokes learning, i.e, the adaption of sensory abilities? More precise, which *global* and intrinsic mechanisms assure proper guiding input for local synaptic plasticity to evolve a complex brain structure, for instance, such as the aligned collection of maps found in the SC?

1.3.1 Local mechanisms

In the last decades, neurobiological evidence has been found to support Hebb's rule. Various mechanisms underlying long-term potentiation (LTP) and long-term depression (LTD) of synaptic transmission have been studied in great detail even if not yet fully understood; reviews e.g. [13, 34, 128, 177, 185].

In addition, to gain further insight into this topic, Hebb's essence, synaptic weight development due to correlated activity of corresponding pre- and post-synaptic neurons, has also been formulated in mathematical terms; for detailed review see e.g. [115, 145]. The development of synaptic connection strength J_{ij} between a pre-synaptic neuron i and a post-synaptic neuron j can be divided into three parts as illustrated in Fig. 1.6. First, the arrival of a pre-synaptic spike at time t_m^i induces the weight J_{ij} to change by an amount ηw^{pre}. Second, if the considered post-synaptic neuron j excites a post-synaptic spike at t_n^j, all connections are changed by the amount ηw^{post}. The quantity $\eta > 0$ is a small parameter and the weight changes w^{pre} and w^{post} can either be positive or negative. In Fig. 1.6 we assumed $w^{\text{pre}} > 0$ and $w^{\text{post}} < 0$. The third part is due to the spike-timing of pre and post spikes. Given the time differences $t_m^i - t_n^j$ between pre- and post-synaptic spikes, J_{ij} is changed by the amount of $\eta W(t_m^i - t_n^j)$, where the learning window $W(t_m^i - t_n^j)$ denotes a

[2] "When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased"; [112] p. 62

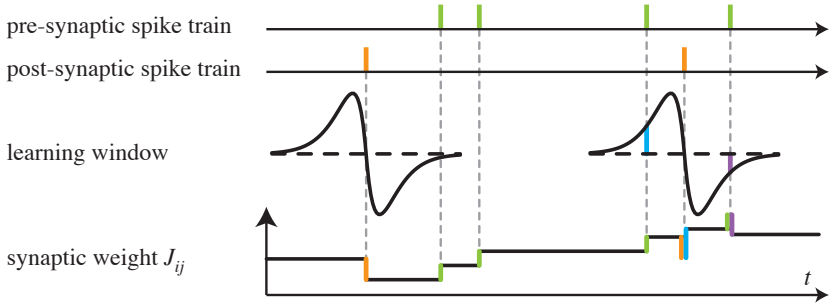

Figure 1.6: *The synaptic learning process [115, 145, 146]. The diagram illustrates the learning process due to the three parts of STDP. In this case, every spike of the pre-synaptic spiketrain (top row) strengthens the synaptic weight J_{ij} by w^{pre} (green). All post-synaptic spikes (second upper row) induce a negative weight change w^{post} (orange). The contribution of the third part of Spike-timing-dependent plasticity (STDP), the correlation term, depends on the spike timing and the learning window (middle row). In the left part, the pair of pre-post spikes is temporally too far separated as compared to the dimension of the learning window. The remaining two other possibilities are depicted in the right part. If a pre-synaptic spike precedes a post-synaptic spike, the synapse strength increases (blue). In case of opposite temporal order, the synapse is weakened (purple).*

real valued function. We have depicted a typical form of a learning window in the third upper row of Fig. 1.6. Finally, we can combine the three parts and derive a Hebb-type learning rule based on spike-timing for a time interval $[t, t + \Delta t]$

$$\Delta J_{ij} = \eta \left[\sum_{t_m^i} w^{\mathrm{pre}} + \sum_{t_n^j} w^{\mathrm{post}} + \sum_{t_m^i,\, t_n^j} W(t_m^i - t_n^j) \right] . \tag{1.4}$$

The above equation can also be formulated in integral form. We therefore define the pre- and post-synaptic spike trains $\mathbf{t}^{\mathrm{pre;post}}(t) = \sum_n \delta(t - t_n^{i;j})$ and compute

$$\Delta J_{ij} \quad = \quad \eta \int_t^{t+\delta t} dt' \left[w^{\mathrm{pre}} \mathbf{t}'^{\mathrm{pre}}_i(t) + w^{\mathrm{post}} \mathbf{t}_j^{\mathrm{post}}(t') + \int_t^{t+\delta t} dt'' W(t' - t'') \mathbf{t}_i^{\mathrm{pre}}(t') \mathbf{t}_j^{\mathrm{post}}(t'') \right] . \tag{1.5}$$

Both, equation (1.4) and (1.5), represent a mathematical Hebb-type learning rule since the equations correlate pre- and post-synaptic activity. In particular, the learning rule is based on the precise timing sequence of pre- and post-synaptic spikes. Such learning schemes are therefore denoted as Spike-timing-dependent plasticity (STDP); for detailed review see e.g. [1, 15, 55, 78, 115, 145].

Kempter, Gerstner, and van Hemmen demonstrated [145] that learning schemes based on pre-synaptic ν_i^{pre} and post-synaptic ν_i^{post} mean firing rates such as

$$\frac{d}{dt} J_{ij} = a_0 + a_1 \nu_i^{\mathrm{pre}} + a_2 \nu_j^{\mathrm{post}} + a_3 \nu_i^{\mathrm{pre}} \nu_j^{\mathrm{post}} + O[(\nu_i^{\mathrm{pre}})^2] + O[(\nu_i^{\mathrm{post}})^2] \tag{1.6}$$

are not capable to catch the neuronal dynamics on a plausible temporal level. To give an example, mean rate descriptions require that all changes of neuronal activity are slow at a time scale of the learning window W. That is not necessarily the case. The Hebbian learning window in cortical areas, for instance, is about 100 ms as compared to the frequency of 40 Hz corresponding to a time scale of 25 ms on which neuronal activity varies.

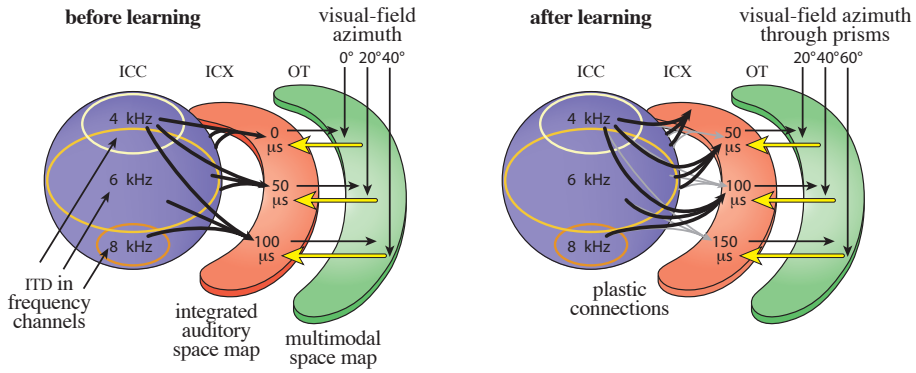

Figure 1.7: *Site for plasticity and origin of guiding input for map adaption in the auditory pathway; adapted from [78, 156]. The auditory system processes the initial sensory cues, for instance, interaural time differences (ITD) to gain a spatial representation of sound source locations in the optic tectum (OT) that is aligned to multimodal maps in the OT. In short, sensory cues, i.e., frequency-specific ITD maps, are present in the central nucleus of the inferior colliculus (ICC) and are projected to the external nucleus of the inferior colliculus (ICX). From there, spatial-auditory information is mapped to the optic tectum. Experiments providing abnormal visual input demonstrate two important findings. First, there is a specific site for plasticity. The synaptic connections between ICC and ICX change in such a way that the ICX map acquires OT topography. Second, audition adapts to input from the OT. In result, vision seems to guide auditory map development [159, 160].*

This short example illustrates not only the advantages of STDP as compared to rate-based learning schemes but also reflects an issue already mentioned above. The way of neuronal encoding, for instance rate coding versus spike patterns that are phase-locked to the input signals, may reduce and refine the imbedded information and thus correspond to a neuronal filter in the process of neuronal encoding; see Sec. 1.1.2.

In this section we have quantitatively seen that pre-synaptic input that drives post-synaptic activity can induce plasticity of corresponding connection strengths. To understand and predict the consequences of local synaptic events to sensory performance we have to consider the global context wherein the local events take place.

1.3.2 Global structures

In this section, we introduce some, but not many important thoughts about learning beyond local synaptic events. *Local* learning rules can describe how the local connection strength changes according to the corresponding pre- and post-synaptic activity. However, without knowing what "corresponding" means in the context of a specific neuronal setup of a sensory system, local learning rules cannot tell us how the sensory system develops as a whole. Therefore, we have to ask at least three questions that will guide us through the subsequent lines. First, *what* drives post-synaptic activity that results in synaptic plasticity? Second, *where* in the structure of pre- and post-synaptic connections does plasticity appear? Third, *how* does the pre-synaptic input manipulate post-synaptic activity to result in reasonable global connections? Hereunto, we take an exemplary look at the auditory pathway.

We have seen above that pre-sypatic input driving post-synaptic activity can lead to changes of the corresponding connection strength. The first question beyond local synaptic events is: *What* drives the post-synaptic activity? That is, which input projects to the post-synaptic neuron. To

give a concrete example, as we can see in Fig. 1.7 the external nucleus of the inferior colliculus (ICX) receives sensory input from the central nucleus of the inferior colliculus (ICC) *and* the optic tectum (OT). Consequently, the activity of neurons in the spatial map in ICX is driven by two input origins. Very closely connected to the first question is another one: What constitutes the input from OT? As this is topic of Chap. 6, we leave it for the moment.

To predict how the auditory system behaves, we have to consider the next question of *where* plasticity appears, i.e., which connections are changed by means of activity-based learning? We know that maps in the OT are mutually aligned and that this is a result of sensory experience during growth. The answer to the *where* question can hence tell us, for instance, which input guides alignment and which connections adapt. In our particular example, we see that only the ICC–ICX connections show plasticity and adapt to input from OT. The latter input is therefore denoted as *guiding*; cf. Fig. 1.7. For a detailed description of the ICC-ICX connections, consider Knudsen et al.'s paper [156].

Having clarified the what and where, "only" the *how* remains. In case of our example, Friedel and van Hemmen [78] have mathematically analyzed different kinds of guiding input and their practical consequences to the learning process. They derived, for instance, that only inhibitory input from the OT, called *selective disinhibition*, can reorganize map topography, viz., the ICC–ICX mapping can be reorganized so as to compensate even a spatial 180° shift between guiding (OT) and sensory (ICC) input. This corresponds to famous prism glass experiments done by Knudsen and others in the barn owl [156, 159] – performed similarly in other animals, such as hamster [200], cat [278, 280], clawed frog [44], ferret [148, 150], and snake [96]. In contrast, excitatory input leads to a less precise, less stable, and slower learning process which results additionally in an ambiguous mapping [78]. However, in other scenarios excitatory guiding input faithfully describes neurobiological reality [146, 174].

In summary, deriving mathematical descriptions of local synaptic events, that is, making them quantitatively predictive, is important but not enough to gain insight into the development of brain structures and their function as a whole. We have to consider the global environment in which the synaptic events appear to predict what they cause. In the context of this thesis, such an environment is constituted by means of neuronal maps that belong to sensory systems. We provide a new model for intrinsic map development in the last chapter of this work. In the next chapter, we will focus our attention on a general framework of *optimal map formation* to understand how sensory information can be processed by means of maps and how the neuronal connectivity pattern can be predicted.

Man sollte alles so einfach wie möglich sehen - aber auch nicht einfacher.

Albert Einstein

2. Optimal Map Formation

In the struggle for survival in a complex and dynamic environment, nature has developed a multitude of sophisticated sensory systems. In order to exploit the information provided by these sensory systems, higher vertebrates reconstruct the spatio-temporal environment from their sensory modalities. As we introduced in the last chapter, most animals are able to compute a neuronal representation of the outside world in form of a neuronal map. Here we present a universal framework that allows to calculate the specific layout of the involved neuronal network by means of a general mathematical principle, viz., stochastic optimality*. That is, given a known physical signal transmission and rudimental knowledge of the detection process, our approach allows to estimate the possible performance and to predict neuronal properties of biological sensory systems.*

2.1 Starting point

A mouse hears a rustling in the grass, sees some leaves moving and – escapes from the predator. Thus, the perception of the outside world by sensory systems and the consequent translation of their response into a reliable neuronal representation that allows, for instance, directional motor commands is an essential concept for surviving.

In the last chapter, we introduced the advantageous concept of a neuronal map for representing the external world (see Sec. 1.2), for instance, the auditory space in the example just mentioned in the above paragraph. Moreover, we subdivided the sensory information processing into three fundamental steps. These are *signal mapping*, *sensory encoding*, and *neuronal decoding*; see Sec. 1.1 and the example in Fig. 1.1. In this chapter, we will now translate these considerations into a mathematical framework of *optimal map formation*.

2.2 Mathematical model

Following the idea of mapping, encoding, and decoding, we want to understand how to extract the stimulus at the best from the sensory responses. Mathematically speaking we have to derive the *inverse transfer function* l (1.3) that can perform an optimal reconstruction of a particular stimulus given the sensory responses and computational limitations. This inverse transfer function can then be translated into a neuronal connectivity pattern (see Sec. 1.1).

The derivation we provide below is based on three reasonable simplifications. First, we assume that all sensory maps are purely *monosensory*. Although it has been questioned whether there are maps without influence from other sensory systems [262], our assumption can be justified by the finding that many spatial maps are clearly dominated by a single sensory modality [280].

Second, we simplify the internal processes within a specific sensory receptor by identifying the receptor input with its response. That is, we set the transfer function k (1.2) to be the identity function, and hence h (1.1) connects directly the stimulus characteristics, i.e., the sensory input connects to the corresponding sensory responses on a one-to-one basis. Although this may not be given for a realistic setting, the framework of optimal map formation is without loss of generality because we can incorporate the filtering properties of the second transfer function k by redefining the first transfer function h. In return, we get a set of equations that are more understandable and comparable with common techniques; see e.g. Chap. 5.

The third important assumption is that of a *linear* relation between the stimulus and the receptor response of the sensory system. That means that the detector responses change proportionally to the signal strength. For example, the multipoles of a submerged moving object, i.e. the signal strengths, translate linearly to the detectable water velocities at the lateral-line organs [251, 75]. Nonlinear relations between stimulus and detector responses, e.g., a logarithmic response [169, 213, 172, 136, 53], can in principle be treated with our model as well; see Appendix A.2 for details.

2.2.1 Definition of the problem

As we have introduced, an object generates a signal $s^{\mathbf{x}}(t)$ varying in time t and position $\mathbf{x}$ in the external world. The signal may be, for instance, the time-dependent sound pressure at a particular location or may denote the presence of edges or movement at a particular position within the visual field.

The signal induces a response $r_i(t)$ in a set of N sensory detectors. Depending on the problem at hand a single detector i with $0 \leq i \leq N$ can be a complete sensory organ, such as the left ear, or part of a detector array such as a specific interval of best frequrncies in the cochlea. In principle, the detector combines information from past signals within the whole sensory space. The response is therefore described by

$$r_i(t) = \int_{\text{all space}} \mathrm{d}\mathbf{x} \int_{-\infty}^{t} \mathrm{d}\tau \; s^{\mathbf{x}}(\tau) h_i^{\mathbf{x}}(t-\tau) \tag{2.1}$$

where the *transfer function* $h_i^{\mathbf{x}}(t)$ incorporates the physics of signal transmission and detection. The transfer function can be different for each detector i. Auditory transfer functions, for example, incorporate the position of sound source and ear with respect to the head midline and therefore differ between right and left ear. In general, we can safely assume that $h_i^{\mathbf{x}}(t) = 0$ for large values of $|\mathbf{x}|$ and t. This reflects our intuition that events occuring far away or long ago will not influence the state of a sensor. We will need this property later on. Moreover, since any detector can only react to temporal-causal, i.e., past signals we set $h_i^{\mathbf{x}}(t) = 0$ for $t < 0$. We can then rewrite the response function (2.1) with adapted integration limits as a convolution with respect to time,

$$r_i(t) = \int \mathrm{d}\mathbf{x} \int_{-\infty}^{\infty} \mathrm{d}\tau \; s^{\mathbf{x}}(\tau) h_i^{\mathbf{x}}(t-\tau) =: \int \mathrm{d}\mathbf{x} \; (s^{\mathbf{x}} \star h_i^{\mathbf{x}})(t) \; . \tag{2.2}$$

The above equation describes the response of an ideal system. In biological systems the quality of the detector response is limited by at least three factors.

First, information may get lost during the transfer from the outside object to the inside sensory system. Second, noise influences all steps in the detection and reconstruction process [71]. Finally, limitations of the neuronal hardware, for instance, the limited dynamic range of receptors, constrain possible solutions; see Sec. 2.4 for details.

Within our mathematical model we incorporate these three restrictive factors by introducing additional noise terms. Accordingly, a term describing background noise $\xi^{\mathbf{x}}(t)$ must be added to

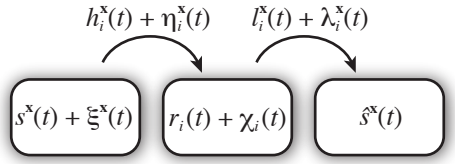

Figure 2.1: Physical mapping: *signal $s^{\mathbf{x}}(t)$ with background noise $\xi^{\mathbf{x}}(t)$ is mapped onto a noisy receptor response $r_i(t) + \chi_i(t)$ through the noisy transfer function $h_i^{\mathbf{x}}(t) + \eta_i^{\mathbf{x}}(t)$.* Optimal map formation: *the (possibly noisy) inverse transfer function $l_i^{\mathbf{x}}(t) + \lambda_i^{\mathbf{x}}(t)$ gives an estimate $\hat{s}^{\mathbf{x}}(t)$ of the signal.*

Signal	$s^{\mathbf{x}}(t) + \xi^{\mathbf{x}}(t)$
Transfer function	$h_i^{\mathbf{x}}(t) + \eta_i^{\mathbf{x}}(t)$
Receptor response	$r_i(t) + \chi_i(t)$
Inverse transfer function	$l_i^{\mathbf{x}}(t) + \lambda_i^{\mathbf{x}}(t)$
Estimated signal	$\hat{s}^{\mathbf{x}}(t)$

Table 2.1: *Functions and error terms describing detection and processing of sensory information.*

the signal. Furthermore, we assume that transfer function and sensory response are hampered by additional noise terms $\eta_i^{\mathbf{x}}(t)$ and $\chi_i(t)$, respectively. Consequently, (2.2) is modified so as to read

$$r_i(t) = \int \mathrm{d}\mathbf{x}\ [(s^{\mathbf{x}} + \xi^{\mathbf{x}})(\star\, h_i^{\mathbf{x}} + \eta_i^{\mathbf{x}})](t) + \chi_i(t)\ . \tag{2.3}$$

To reconstruct the estimated signal from the detector responses $r_i(t)$, the above transformation must be "inverted" in some appropriate way. We therefore calculate the time-dependent inverse transfer functions $l_i^{\mathbf{x}}(t)$ between detector i and the map at position $\mathbf{x}$. When applying $l_i^{\mathbf{x}}(t)$ to the receptor responses at i, we obtain accordingly to (1.3) the estimate

$$\hat{s}^{\mathbf{x}}(t) = \sum_i [r_i(\star\, l_i^{\mathbf{x}} + \lambda_i^{\mathbf{x}})](t) \tag{2.4}$$

of the original signal $s^{\mathbf{x}}(t)$. Here the hat on $\hat{s}^{\mathbf{x}}(t)$ denotes a reconstruction and the term $\lambda_i^{\mathbf{x}}(t)$ represents the noise due to the concrete realization of the theoretical inverse transfer function. We note that in contrast to elsewhere [217, 229] the present model is non-iterative. This will result in a purely feedforward network structure when it comes to a neuronal realization in Sec. 2.4.

Figure 2.1 illustrates the whole mathematical procedure of sensory information processing as described in the steps of Fig. 1.1. All the relevant terms are summarized in Table 2.1. In the next section, we will indicate how to calculate inverse transfer functions $l_i^{\mathbf{x}}(t)$ that enable optimal signal reconstruction.

2.2.2 Optimal reconstruction

We want to tune our sensory system to *optimally* reconstruct not only one specific situation but the *typical* environment. In other words, biologically relevant signals belong to a class of signals that we denote as "typical". Consequently, a specific sensory signal is a concrete realization of a class of typical, biologically relevant signals. That is, it is a stochastic quantity. We therefore minimize the *expectation value* of the squared difference between signal and reconstruction.

This is possible because all quantities and functions (see Fig. 2.1) involved in both, the process of physical mapping and the neuronal process of optimal map formation (Sec. 2.4) are self-averaging as we demonstrate in [35]. The mathematical definition of self-averaging allows for a description in terms of expectation values.

To derive the inverse transfer functions $l_i^{\mathbf{x}}(t)$ that enable optimal signal reconstruction for a class of typical signals, we can next minimize the expectation value of the squared error between estimated and real signal

$$E\{\mathbf{l}^{\mathbf{x}}(t),t\} := \left\langle \int_{t-T}^{t} \mathrm{d}t' \int \mathrm{d}\mathbf{x} \;\; [s^{\mathbf{x}}(t') - \hat{s}^{\mathbf{x}}(t')]^2 \right\rangle = \int_{t-T}^{t} \mathrm{d}t' \int \mathrm{d}\mathbf{x} \;\; \left\langle [s^{\mathbf{x}}(t') - \hat{s}^{\mathbf{x}}(t')]^2 \right\rangle \; . \tag{2.5}$$

Here the brackets $\langle . \rangle$ denote the expectation value with respect to the different types of noise, and T is a typical processing time.

To be mathematically precise, an expectation value is an integral on a probability space with respect to a probability measure p. For arbitrary functions f and g, if $\langle |f-g|^2 \rangle = 0$ then $f = g$ *with respect to* p or, physically, looking at the world through p's glasses: what p finds important pops up clearly whereas what p finds "irrelevant" has hardly any weight. The latter need not correspond to what we "think" ourselves; see van der Waerden [276].

A quadratic form of the error term has been proven to be a reasonable and practical choice in many physical optimizing problems; see [193]. In case of independent Gaussian error terms, the formulation via a quadratic error is under certain conditions identical to results obtained by means of *maximum-likelihood* estimates [135, 144]; see Sec. 2.3.2.

Mathematically, the error (2.5) is a functional assigning to every set of inverse transfer functions one specific value. Minimization of functionals in the above integral form is a central and well-studied aspect of the calculus of variations [45, 85, 138, 32]. For the present situation, the first variation with respect to every inverse transfer function $l_j(\mathbf{x},t')$ is to vanish. That is,

$$\frac{\partial \left\langle [s^{\mathbf{x}}(t') - \hat{s}^{\mathbf{x}}(t')]^2 \right\rangle}{\partial \, l_j^{\mathbf{x}}(t')} = 0 \qquad \text{for every } j. \tag{2.6}$$

In order to solve (2.6), we have to substitute (2.4) for the estimate $\hat{s}^{\mathbf{x}}(t)$ and replace $r_i(t)$ by its description (2.3). Expanding the square, we encounter expectation values of products consisting of varying combinations of noise and signal terms. Here we assume that all noise terms as well as the signal itself are stochastically independent of each other so that the expectation of a product of independent term factorizes; for instance,

$$\left\langle s^{\mathbf{x}}(t) \eta_i^{\mathbf{x}'}(t') \right\rangle = \left\langle s^{\mathbf{x}}(t) \right\rangle \left\langle \eta_i^{\mathbf{x}'}(t') \right\rangle \; . \tag{2.7}$$

For a product consisting of the same kind of term we need to consider the definition of the autocorrelation of a quantity $f^{\mathbf{x}}(t)$ as given by

$$\left\langle f^{\mathbf{x}}(t) f^{\mathbf{x}'}(t') \right\rangle = \delta(\mathbf{x}-\mathbf{x}')\delta(t-t')(\mu_f^2 + \sigma_f^2) \tag{2.8}$$

with μ_f the mean and σ_f the variance of the quantity $f^{\mathbf{x}}(t)$. That is, we assume in a first step that the values for different spatio-temporal positions are completely uncorrelated, a kind of worst-case analysis.

Since the means of all noise terms μ_f vanish we get the following correlation terms

$$\left\langle \xi^{\mathbf{x}}(t) \xi^{\mathbf{x}'}(t') \right\rangle = \delta(\mathbf{x}-\mathbf{x}')\delta(t-t')\sigma_\xi^2 \; , \tag{2.9a}$$

$$\langle \chi_i(t) \chi_j(t') \rangle = \delta_{ij}\delta(t-t')\sigma_\chi^2 \; , \tag{2.9b}$$

$$\left\langle \eta_i^{\mathbf{x}}(t) \eta_j^{\mathbf{x}'}(t') \right\rangle = \delta_{ij}\delta(\mathbf{x}-\mathbf{x}')\delta(t-t')\sigma_\eta^2 \tag{2.9c}$$

$$\text{with} \quad |\mathbf{x}| < x^{\max} \quad \text{and} \quad 0 < t < t^{\max} \; . \tag{2.9d}$$

Through the final equation we take into account that the noise $\eta_i^{\mathbf{x}}(t)$ vanishes for large values of t and $|\mathbf{x}|$, in the same way as for the transfer function $h_i^{\mathbf{x}}(t)$.

The autocorrelation (2.8) of the signal $s^{\mathbf{x}}(t)$ itself depends on the problem at hand. Either the detectors of the sensory system measure absolute signal strengths (μ_S), e.g., vision, or modulations of a mean value of the signal (deviation σ_S), e.g., audition. In any case, one has to choose the corresponding biologically relevant term and put the other equal to zero. In the following, we choose the expectation value μ_S^2 of the signal as the appropriate quantity and therefore take σ_S^2 to be zero,

$$\left\langle s^{\mathbf{x}}(t) s^{\mathbf{x}'}(t') \right\rangle = \delta(\mathbf{x}-\mathbf{x}')\delta(t-t')\ \mu_s^2\ . \tag{2.10}$$

While (2.9) incorporates reasonable assumptions for all noise terms, the correlation (2.10) for the signal is a strong hypothesis. Signals are namely characterized by spatio-temporal continuity; e.g., objects and their corresponding signals usually do not disappear from one point in time to the next. A Gaussian correlation term

$$\left\langle s^{\mathbf{x}}(t) s^{\mathbf{x}'}(t') \right\rangle = A \exp\left(-|\mathbf{x}-\mathbf{x}'|^2/(2\sigma_x^2)\right) \exp\left(-|t-t'|^2/(2\sigma_t^2)\right)\ , \tag{2.11}$$

for instance, can take into account correlations between neighboring points in space and time. Here σ_x and σ_t are typical spatial and temporal correlation scales. The application of such a Gaussian correlation, however, does not greatly alter the further derivation but only smoothens the final estimated signal. For reasons of clarity, we will therefore stick to the relation (2.10); see Appendix A.4 and [35] for details.

Returning to (2.6) we have to solve it and in so doing apply the correlations (2.9) and (2.10) so as to arrive at

$$l_j^{\mathbf{x}}(t)\left[\sigma_\chi^2 + (\mu_s^2+\sigma_\xi^2)\int_{\substack{|\mathbf{y}|<y^{\max}\\ 0<\tau<t^{\max}}} \mathrm{d}\mathbf{y}\mathrm{d}\tau\ \sigma_\eta^2\right] + (\mu_s^2+\sigma_\xi^2)\sum_i \int \mathrm{d}\mathbf{y}\ \left[(h_i^{\mathbf{y}} \star l_i^{\mathbf{x}}) \circ h_j^{\mathbf{y}}\right](-t) = \mu_s^2 h_j^{\mathbf{x}}(-t) \tag{2.12}$$

for details see Appendix A.3. The open circle $\circ$ denotes the autocorrelation integral

$$(a \circ b)(t) := \int_{-\infty}^{\infty} \mathrm{d}\tau\ a(\tau) b(t+\tau)\ . \tag{2.13}$$

In order to simplify (2.12), we define two new noise measures,

$$\tau^2 := \frac{\sigma_\xi^2}{\mu_s^2} \tag{2.14}$$

and

$$\sigma^2 := \frac{\sigma_\chi^2}{\mu_s^2} + \int_{\substack{|\mathbf{y}|<y^{\max}\\ 0<\tau<t^{\max}}} \mathrm{d}\mathbf{y}\mathrm{d}\tau\ \frac{\sigma_\eta^2(\mu_s^2+\sigma_\xi^2)}{\mu_s^2}\ . \tag{2.15}$$

The parameter τ represents an inverse signal-to-noise ratio. Therefore, it is often reasonable to assume a small value of τ. The parameter σ, on the other hand, describes the overall measurement noise by relating *detection* and *transmission* noise, σ_χ and σ_η, to the signal mean amplitude μ_s. A priori, its value cannot be assumed to be small and has to be adjusted according to the situation at hand.

In order to further simplify (2.12) we switch to Fourier space, where convolution (2.2) and correlation (2.13) become ordinary multiplications combined with complex conjugations. Using (2.14) and (2.15) and denoting Fourier transforms by capital letters and the complex conjugation by an overline, (2.12) simplifies to

$$\sum_i L_i^{\mathbf{x}}\left[\sigma^2\delta_{ij} + (1+\tau^2)\int \mathrm{d}\mathbf{y}\ H_i^{\mathbf{y}}\overline{H_j^{\mathbf{y}}}\right] = \overline{H_j^{\mathbf{x}}}\ . \tag{2.16}$$

Equation (2.16) is the main result of our derivation. In principle, it allows us to calculate the inverse transfer functions $L_i^{\mathbf{x}}$ for optimal signal reconstruction. A calculation of the second variation – see (A.16) in the Appendix for details – then confirms that the inverse transformation we have found indeed minimizes the error. For convenience we will introduce an alternative notation in the next section.

2.2.3 Matrix notation

To rewrite (2.16) in a more practical notation we introduce "matrices" $\mathcal{H}$ and $\mathcal{L}$ by putting

$$\mathcal{H}_{[ix]} = H_i^{\mathbf{x}} \ , \qquad \mathcal{L}_{[xi]} = L_i^{\mathbf{x}} \ . \tag{2.17}$$

The notations illustrate that transfer functions and inverse transfer functions are linear transformations from a continuous space (the outside world) into a discrete space (the neuronal map) and vice versa. Therefore, $\mathcal{H}$ and $\mathcal{L}$ are only formally matrices with a spatial coordinate $\mathbf{x}$ varying in $\mathbb{R}$. The matrix multiplication involving the spatial coordinate must consequently be understood as an integration. A discretization of space, as is usual in numerics, would lead to a true matrix formulation.

In addition, we introduce the covariance matrix $\mathcal{C}(\mathbf{R})$ of the receptor response $\mathbf{R}$ as described, e.g., in [135, 144]. In our case we find

$$\mathcal{C}(\mathbf{R}) \quad := \quad \left\langle (\mathbf{R} - \langle \mathbf{R} \rangle)\overline{(\mathbf{R} - \langle \mathbf{R} \rangle)}^T \right\rangle = \mu_s^2 \left(\sigma^2 \mathbb{1} + \tau^2 \overline{\mathcal{H}} \mathcal{H}^T \right) \tag{2.18}$$

where the superscript T denotes the matrix transpose and $\mathbb{1}$ the identity matrix. Equation (2.16) now simplifies to

$$\mathcal{M}\mathcal{L}^T = \overline{\mathcal{H}} \ \text{ with } \mathcal{M} := \mu_s^{-2}\mathcal{C} + \overline{\mathcal{H}}\mathcal{H}^T \ . \tag{2.19}$$

Given $\mathcal{M}$ as an invertible matrix, denoted as the 'model matrix', the solution for $\mathcal{L}$ turns out to be

$$\mathcal{L} = \left(\mathcal{M}^{-1}\overline{\mathcal{H}}\right)^T = \overline{\mathcal{H}}^T \left(\mu_s^{-2}\mathcal{C} + \mathcal{H}\overline{\mathcal{H}}^T \right)^{-1} \ . \tag{2.20}$$

This equation gives a unique solution for the optimal reconstruction for any given set of transfer functions and noise constants (σ, τ). Using (2.4) in matrix form we find

$$\hat{\mathbf{S}} = \mathcal{L}\mathbf{R} \tag{2.21}$$

as estimated signal from the measured response vector $\mathbf{R}$.

2.3 Relation to common methods

The challenge of signal reconstruction has a long tradition, and, accordingly, one may ask how the above formalism relates to methods that have been established in this field. In the following, we will discuss the relation of our model to methods based on the pseudo-inverse and to the maximum-likelihood approach.

2.3.1 Pseudo-inverse.

If the noise terms can be neglected, the interpretation of (2.19) is straightforward. In this case the covariance matrix $\mathcal{C}$ vanishes and the resulting equation leads to

$$\mathcal{L} = \overline{\mathcal{H}}^T \cdot \left(\mathcal{H} \cdot \overline{\mathcal{H}}^T \right)^{-1} \ . \tag{2.22}$$

The reconstruction functions $\mathcal{L}$ that we have just found fulfill the properties of the Moore-Penrose *pseudo-inverse*[1] of $\mathcal{H}$ [12]. At hindsight, this makes sense since the pseudo-inverse generates an

[1] The explicit expression for $\mathcal{L}$ in (2.22) only holds if $(\mathcal{H}\cdot\mathcal{H}^T)$ is invertible.

approximate inverse matrix that minimizes the quadratic error; see Appendix A.5. An exact inversion may not be possible for a matrix $\mathcal{H}$ that is, e.g., rectangular instead of square or of incomplete rank.

But even in the more general situation of non-vanishing noise terms, we can observe strong relations between our framework and methods based on the pseudo-inverse. The point is that, for the calculation of the pseudo-inverse, a regularization has to be introduced to suppress noise terms, typically high-frequency variations. The so-called *Tikhonov-Miller regularization* adds a positive term $\alpha\mathbb{1}$ to make it more stable [193, 268, 269, 228]. The regularized equation then reads

$$\left(\alpha\mathbb{1} + \overline{\mathcal{H}} \cdot \mathcal{H}^T\right) \cdot \mathcal{L}^T = \overline{\mathcal{H}} \; . \tag{2.23}$$

Comparing this equation with (2.19) we see that α corresponds exactly to our term σ^2 if $\tau = 0$. Hence, in this special case our general approach is identical to methods using the Tikhonov-Miller regularization.

2.3.2 Maximum-likelihood approach.

The *maximum-likelihood* analysis [135, 144] is a common tool in the interpretation of measurement data. Within the maximum-likelihood scheme one computes the stimulus that is the most likely one *given* a set of detector responses R. Experiments have shown that optimal or near-optimal stimulus combinations can indeed *describe* several phenomena of sensory processing [67, 126, 165, 2, 113, 202]. A method of optimal stimulus combination such as the maximum-likelihood approach is therefore highly relevant to neuronal information processing and ought to be included into our model; see Chap. 1.

The maximum-likelihood approach tries to find the most probable input signal S given the detector responses R, a known transfer function $\mathcal{H}$, and *no* a priori knowledge about the signal ($\sigma_s = \infty$). We now assume a linear relation

$$\mathsf{R} = \mathcal{H}\mathsf{S} + \boldsymbol{\chi} \tag{2.24}$$

with $\boldsymbol{\chi}$ representing the noise. We assume the noise to follow a Gaussian distribution with zero mean and the standard deviation σ_χ.

The method minimizes the noise χ. That is, based on the fundamental definitions of Bayesian statistics, it maximizes the conditional probability density function

$$p(\mathsf{R}|\mathsf{S}) \propto \exp\left[-\frac{1}{2\sigma_\chi^2}(\mathsf{R} - \mathcal{H}\mathsf{S})^T(\overline{\mathsf{R} - \mathcal{H}\mathsf{S}})\right] \tag{2.25}$$

with respect to the signal S. This leads to a linear system of equations

$$\mathsf{S} = \underbrace{\left(\overline{\mathcal{H}}^T\mathcal{H}\right)^{-1}\overline{\mathcal{H}}^T}_{=:\mathcal{L}_{ML}}\mathsf{R} \; . \tag{2.26}$$

Using the above assumptions for our model, viz., $\sigma_s = \infty$, $\eta = 0$, and $\xi = 0$, (2.19) reduces to

$$\left(\overline{\mathcal{H}}\mathcal{H}^T\right)\mathcal{L}^T \;\; = \;\; \overline{\mathcal{H}} \; . \tag{2.27}$$

To test whether the two filters are equal, we insert $\mathcal{L}_{ML}$ into (2.27). Application of the transposition rules shows that with the assumptions we used $\mathcal{L} = \mathcal{L}_{\mathcal{ML}}$, the two strategies are identical; for details we refer to elsewhere [135, 144, 242, 236]. We will use this method (2.25) to quantify the localization capability of the lateral-line system, for instance, of fish in the next chapter.

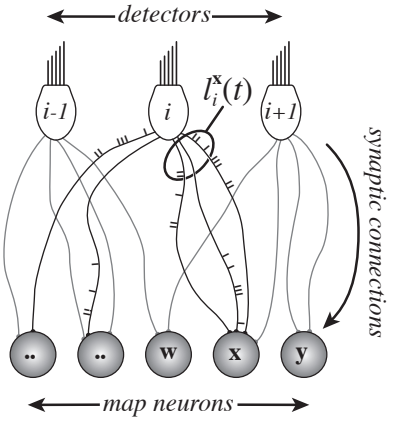

Figure 2.2: *Neuronal realization of unimodal map formation. Each sensor (here hair cells labeled i) connects to several map neurons. The map neurons (encoding the location $\mathbf{x}$) may receive (multiple) connections from each sensor. Each connection has a well-defined strength and temporal delay t. In this way, the transformation $l_i^{\mathbf{x}}(t)$ can be reliably represented in a neuronal network [76].*

2.4 Neuronal realization of the model

In this section we relate the general mathematical algorithm of optimal stimulus reconstruction to a concrete neuronal context. We therefore have to verify first whether the assumptions we made in Sec. 2.2.2 are fulfilled in neuronal processing. That is, we need to check whether the neuronal quantities and functions of optimal map formation are *self*-averaging. To this end we note on the one hand that firing of neurons is correlated with neuronal input and that neuronal noise can be described by a stochastic process, for instance, a Gaussian one; we will see in a minute why. Our framework can cope with any distribution of neuronal noise as long as the mean is zero. On the other hand the optimal inverse transfer functions $l_i^{\mathbf{x}}(t)$ are *learned* synaptic connections between the maps associated with different modalities and hence reflect properties of the underlying learning process. Effective learning is slow because it needs many independent repetitions. Accordingly, time scales for learning and individual realizations of an external signal can be separated. In other words, learning is a self-averaging process where only *averaged* quantities enter by the very nature of the process; see [145]. As mentioned before, quantities and functions within the physical mapping process are self-averaging as well; please see also [35]. In conclusion, the conditions needed to exploit the mathematical framework as derived in Sec. 2.2 are fulfilled.

Consequently, we can now translate the inverse transfer functions $l_i^{\mathbf{x}}(t)$ into neuronal hardware. In such an architecture, the actual processing is performed by the synaptic connections between neurons and detectors. Spatial processing is governed by the topographic structure of the network; that is, which detector is connected to which neuron. Temporal processing on the other hand is determined by the distribution of delays within the set of connections. Figure 2.2 shows an example of such a neuronal setup.

In the above derivation we have already taken into account the discrete character of detectors and the ensuing map through a discrete number of inverse transfer functions. Furthermore, the discrete, "spiky" character of response and reconstruction by the neuronal realization is already taken care of by the noise terms χ_i and $\lambda_i^{\mathbf{x}}$. That is, we are left with the temporal discretization of the inverse transfer functions $l_i^{\mathbf{x}}(t)$. This discretization is realized by a sampling procedure where a number of dendrites with appropriate delays is chosen to represent the complete $l_i^{\mathbf{x}}(t)$. It has indeed been shown that a limited number of synaptic connections suffices to sample the time course of $l_i^{\mathbf{x}}(t)$ [76]. Even more so, the map-neuron response is robust with respect to the sampling method of the temporal delays [178] as well.

Consequently, as illustrated in Fig. 2.2, our unified framework can be implemented by means of a simple feedforward network of excitatory and inhibitory connections in order to form a unimodal map from arbitrary input [76, 252, 178]. It does not, however, explain how such a connectivity pattern is established in a real biological system. Here the correct synaptic connections have to be *learned*. It has been shown [74, 78] that a teacher such as the visual system can generate correct synaptic strengths so that a map can indeed develop in other modalities by means of (supervised) STDP; for details see Sec. 1.2.3, Sec. 2.5.3, and Chap. 6. Thanks to the present method we can compare the learned connectivity pattern with the optimal one as given by (2.16) and (2.20).

A meaningful comparison of the mathematically optimal network architecture with an actual biological setup, though, may not be straightforward. In real biological systems, error reduction as in (2.5) to its minimum – that is, realizing the optimal connectivity – may not be possible because of neuronal limitations. The limited neuronal accuracy that results can be included into our framework by reducing the error only below a certain error threshold, which may even vary in space. For instance, the sampling arrays of animal eyes are non-uniform, with different parts of the visual field being sampled with different spatial and spectral resolution [125, 257, 293]. Such a focus on specific spatio-temporal domains can mathematically be realized by introducing a positive weighting function into the integral of (2.5). Accordingly, when reducing the global error below a certain threshold, the areas within the focus of the weight function have to reach a higher level of optimization, i.e., of resolution, than the rest.

As indicated in Sec. 1.2, the concept of receptive fields is included in our formalism. It may be good to remember, though, that there are two mapping functions ($\mathcal{LH}$) and $\mathcal{L}$ projecting directly onto the map. Since the rows of these mapping functions contain the information from which areas a specific map neuron receives input, the rows describe the receptive fields. We will use the concept of receptive fields when we come to auditory signal processing in Chap. 3. There, we take a closer look at the neuronal implementation by means of a concrete setting and exploit the connection between optimal map formation and receptive fields.

Taken together, the formalism of optimal map formation is capable to deliver an optimal neuronal connectivity pattern, just as illustrated in Fig. 2.2, and hereby directly gives a forecast of how the receptive fields are shaped. In the next section, we provide a very simple but concrete example in the context of the visual system to illustrate the capabilities of the present framework.

Spatial example: visual processing

Within the visual system each sensory neuron is basically tuned to a particular spatial position. In mathematical terms, every retinal neuron i receives input from a spatial position $\mathbf{x}_i$, its preferred position, and neighboring positions within a region determined by resolution ρ. The transfer function corresponding to such a sensory system is

$$h_i^{\mathbf{x}}(t) = \exp\left(-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\rho^2}\right)\delta(t)\ , \tag{2.28}$$

and its Fourier transform reads

$$H_i^{\mathbf{x}} = \exp\left(-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\rho^2}\right)\ . \tag{2.29}$$

Within our exemplary setup we assume that the signal position $\mathbf{x} = (u, v)$ encodes positions $u, v \in [-1/2, 1/2]$. As a reminder, we have rescaled positions so as to make them dimensionless and fit in the square $[-1/2, 1/2]^2$. From the above ansatz (2.28) and (2.16) we calculate the matrix

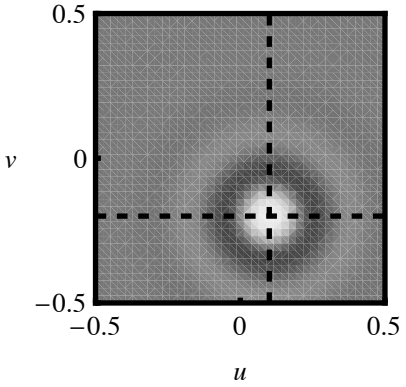

Figure 2.3: *Spatial receptive field. Connection strengths to a map neuron encoding the position* $(u, v) = (0.1, -0.2)$*. The sensory neurons are distributed on a* 40×40 *grid with preferred positions* $u, v \in [-1/2, 1/2]$ *and a tuning curve width* $\rho = 0.9$*. We chose* $\sigma = 1$ *and* $\tau = 0$*. A clear center-surround receptive field emerges. Receptor neurons that have a preferred position matching that of the map neuron have excitatory connections (white spot). Receptor neurons having a slightly off-set position* inhibit *the map neuron (dark circle). Neurons with preferred positions far away from the map neuron have connection strength zero (gray).*

components

$$M_{ij} = \sigma^2 \delta_{ij} + (1+\tau^2) \exp\left(-\frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{4\rho^2}\right) \times \left[\operatorname{erf}\left(\frac{u_i + u_j - 1}{2\rho}\right) - \operatorname{erf}\left(\frac{u_i + u_j + 1}{2\rho}\right)\right] \times \left[\operatorname{erf}\left(\frac{v_i + v_j - 1}{2\rho}\right) - \operatorname{erf}\left(\frac{v_i + v_j + 1}{2\rho}\right)\right] \tag{2.30}$$

where $\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(-y^2\right) \mathrm{d}y$ is the error function. To find the connection strengths $l_i^{\mathbf{x}}$, we numerically calculate the model matrix $\mathcal{M}$ for a discretized space and parameters $\sigma = 1$ and $\tau = 0$. With the matrix $\mathcal{M}$ we then determine the connection strengths $\mathbf{L}$. By an inverse Fourier transformation we can numerically obtain $l_i^{\mathbf{x}}$ for each map position $\mathbf{x}$ as shown in Fig. 2.3. Here the connections from all receptors to a map neuron i, i.e., its receptive field, are plotted for an arbitrary preferred position $\mathbf{x}_i = (0.1,\ -0.2)$. Clearly, the receptors encoding the preferred position have strong projections to the map neuron (bright spot in Fig. 2.3) but, interestingly, the receptors that encode slightly differing locations contribute negatively (dark circle in Fig. 2.3).

Such a center-surround profile is called "Mexican hat" and is, e.g., realized by lateral inhibition, a well-known phenomenon first described by Mach [184] in the visual system in 1866. Up to now this mechanism, studied in the mammalian visual system [281, 143], has been discovered as well in, for instance, insect vision [130], snake infrared vision [255, 252], electric field detection in electric fish [250], and surface wave detection in the back swimmer [203].

In contrast to many models such as pop-out [153] or saliency detection [129, 175] a center-surround receptive field is a natural consequence of our model and thus explains lateral inhibition as *optimal* for map-formation purposes. In the following chapter, we visit the auditory system and apply our concept of optimal map formation to elaborately derive an archetype model for signal enhancement and echo suppression. In the Appendix A.1, we additionally provide an easy step-by-step "recipe" so as to find the optimal connectivity in an arbitrary biological setup.

Next, we leave the framework of optimal map formation for the time being until we consider the sensory performance of the auditory system (Chap. 3), the lateral-line system (Chap. 4), and the infrared system of snakes (Chap. 5) and revisit general aspects of cross-map computation that follow from combining mono-sensory maps.

2.5 Multimodal interaction

In the last section, we gained a deeper understanding of monosensory maps and their computational power. That is, we can revisit a topic of Sec. 1.2, *Cross-map computation*, and take a closer look at this important issue. On the level of maps, we generally distinguish two categories of multimodal interaction, viz., cross-map computation: *integration* and *pooling* of information.

2.5.1 Integration

Congruent spatial information from different sensory systems can be integrated into a single merged and, hence, multimodal map. Such an integrated map, as compared to unimodal information processing, features increased information reliability and saliency as well as an improved sensitivity in both space and time [166, 167, 104, 237]. For example, if visual and auditory sensory systems both register a signal, e.g., "brown ahead" and "barking ahead", it is very probable that the signal corresponds to an actual object rather than to a sensory artefact. At the same time, the integrated signal will be stronger and allows for faster reactions (e.g., "escape!"). In some cases an integrated signal is even *optimal* [102, 202].

More general neuronal models describing multimodal integration and based on statistical methods have been presented elsewhere [59, 60, 61]. Concrete theoretical models of multimodal integration within the SC have been developed as well [7, 220, 238, 239, 186, 274].

2.5.2 Pooling

Not only can the monosensory maps be merged into a more reliable multisensory map, but the diverse information, thus, signal characteristics within the monosensory maps can be accessed simultaneously as well; remember Fig. 1.5. This simultaneous accessing is only possible since all monosensory maps are aligned and consequently space-time can serve to link the different modalities. Consequently, an object at one specific position can be identified and characterized in order to select motor responses in a complex environment. For example, a rattle snake may detect spatial coherent activity in its visual and/or infrared map. Only if the encoded object is visible *and* warm it will be identified as a living prey object. If it is visible and *not* warm the snake will discard the information. Experimental evidence for such a pooling of information is provided by neuronal AND and OR processing steps for the combination of visual and infrared map [210, 211]. These prominent examples of pooling in the SC/OT could enable target selection and thus ensure appropriate motor commands in a complex environment.

Despite increased reliability of an integrated map, its individual input streams cannot be distinguished anymore. That is, the information about which monosensory map has determined the position is lost. Within the above example the multimodal map may indicate a multimodal event ahead, but the triggering modality, that is, visual, auditory, or yet another modality remains unresolved.

In summary, integration of information allows for a reliable spatial determination of an object, the key task of *object formation*. On the other hand, pooling of information assures an access to the details of an object necessary for *object identification*. Switching between integration and pooling corresponds to a switch between parallel and serial data processing to best fit different tasks.

To enable efficient multimodal interaction such as integration and pooling, alignment of the different mono- and multisensory maps is of crucial importance. Only then can a multimodal stimulus at a specific spatial location be identified. An alignment of sensory maps, however, is not present at birth and must be *learned* [116, 262, 156], as discussed in the next section.

2.5.3 Development

We next come back to the question of how sensory maps can be aligned, i.e., how a common multimodal space can evolve that we have briefly introduced in Sec. 1.2 and Sec. 1.3. An obvious solution to such an alignment process would be the existence of one dominant modality as reference for all other modalities [158, 156]. This reference map would then automatically lead to modifications of all other maps.

And indeed, experimental and physiological studies have shown that, in many animals, destruction or disturbance of the visual pathway leads to disorganized and abnormal sensory maps in non-visual modalities. These findings have been obtained in hamster [200], cat [280, 278], clawed frog [44], ferret [150], barn owl [159], and in snakes [96]. Psychophysical experiments with congenitally blind and normally sighted humans have shown that visual input early in life is necessary for multimodal interaction to occur [122, 230, 233]. Consequently, vision seems to serve as a "teacher" for non-visual modalities.

A plausible argument supporting the idea of vision as teacher input is the intrinsic topographic order of the retina. It is known that layers of neurons can self-organize into topographic maps, provided that initially a small set of correctly organized neurons exists [284]. For a review the reader is referred to [271]. This together with the subsequent development of layers in the visual cortex (for mice, see [133]) may allow the intrinsic topography of the retina to step-by-step dictate the organization and alignment of higher visual and, potentially, also multimodal maps.

The general mechanism facilitating the precision of such an alignment of maps is spike-timing-dependent plasticity (STDP) [16, 17, 55, 87, 115, 145, 188, 254, 294]. An example where the alignment has been studied in detail, both experimentally and theoretically, is audio-visual integration within the OT of the barn owl. Here experiments [127, 158] have shown that the auditory map follows systematic changes within the visual input. Although the precise nature of this teaching signal has not been clarified experimentally, *selective neuronal disinhibition*, or gating, seems to play a key role [103, 285]. Although excitatory and inhibitory teaching input can theoretically account for proper map alignment [78, 56] only *inhibitory* teaching input is able to re-align already established map connections [78].

In summary, the above studies support the idea of vision as teacher modality to align other monosensory maps, but there are contradicting findings as well. We can summarize these findings into two major points. First, vision is not needed at all as teacher input for the learning process of sensory maps. Second, vision shows plasticity as it is influenced by other modalities and as it improves during development. Concerning the first point, it has been shown both theoretically and experimentally that, although imprecise, a map of *azimuthal* sound location can be learned *without* any visual input [146, 158] though admittedly on a genetically determined substrate. In addition, non-visual modalities can influence each other as well, e.g., audition can influence haptics [31]. Moreover, somatosensory receptive fields already shrink in a postnatal phase when only auditory, but no visual neurons are present [279, 277]. For the second aspect, behavioral and psychophysical studies show that visual perception can even be influenced by other modalities such as haptics [68] or audition [249, 77, 258]. More importantly, vision itself can improve, respectively sharpen, as found in the visual system of young cats [279, 277].

Altogether the experimental and theoretical findings we have presented above question the current picture of vision-guided map alignment [148]. Wallace and Stein [279] have pointed out that the development of different modalities starts in parallel and in temporal coincidence with the appearance of multimodal integration. They hereby suggest a common mechanism driving both map development and multimodal integration. In Chap. 6, we will stick to this idea, namely that a combination of all available sensory input intrinsically guides its alignment by developing the concept of *integrated multimodal teaching* that addresses the issues mentioned above.

In the next chapters, we turn to more sophisticated settings than in Sec. 2.4 and gain new insights into the capabilities, characteristics, and limitations of theses senses.

Es hört doch jeder nur, was er versteht.
Johann Wolfgang von Goethe

3. Auditory processing: Exemplifying optimality

In this chapter we apply our model of optimal map formation to the auditory system. That is, we deduce an archetype model and corresponding temporal receptive fields for auditory signal processing, viz., handling echoes. Based on our model, we predict the ability of both mathematical and neuronal implementations to suppress echoes and extracts original signals in various echo-rich scenarios, even in the absence of exact information on the specific echo form. In addition, we compare state-of-the-art technical algorithms, biological solutions and our archetype model derived from the general framework. Originating from this particular setting, the comparison hints to limitations of our general framework and promising sites for future work.

3.1 Introduction

In the present chapter we apply the framework of optimal map formation we previously derived in Chap. 2 to a concrete setting in the *temporal* domain. In order to do so, we take the auditory pathway as a reference for an archetype model of signal enhancement and echo suppression that yield optimal temporal receptive fields. To get started, we initially review the essential facts and specific questions of interest according to this setting.

In natural environments, auditory signals are followed by echoes, or reflections, that degrade the signal. In the non-biological field, this setting is called *reverberation*. Not surprisingly, biological systems processing acoustic signals have developed neuronal mechanisms of echo suppression [20, 221]. Humans, for example, do not consciously perceive echoes arriving less than about 20 ms after the original signal, depending on the signal character [20]. We therefore expect delays and suppression, in neuronal terms, inhibition, to shape the receptive fields of our model. For a quantitative understanding, however, more specific questions need to be asked. What is the shape of the receptive field our model predicts? How does this shape connect to biological reality? Can different functions be attributed to different regions of the receptive field? Can we identify universal characteristics shared by different receptive fields?

We have also chosen the context of auditory signal processing because this is a site where technical solutions, viz., algorithms that address some of the above questions to a certain extent have already been realized. This promises a fruitful comparison of the models and biology. Although artificial algorithms are quite sophisticated we remark that "[...] for acoustic echo cancellation filters, simple and robust algorithms may outperform more sophisticated solutions [...]" [105]. So, there are artificial algorithms with good performance that provide enough simplicity to be comparable to the archetype model at hand. In doing so, we apply our present model to a neuronal setting and contrast our model and its collection of receptive fields to what we find in nature and technics. Finally, similarities and differences of the derived archetype model may point out in which way algorithms can be improved by means of bio-analogous cognitive architectures.

3.2 Archetype model

In this section we briefly recapitulate the essential steps of our framework of optimal map formation in the concrete setting of auditory signal processing. That is, we derive an archetype model of signal enhancement and echo suppression based only on the mathematical principle of optimality.

3.2.1 Optimality: An architectural principle

The model derivation is based on the idea of signal reconstruction by optimal receptive fields that inverse the corruption process of echoes and noise. By deriving the necessary neuronal connection strengths we form the basis for a smooth integration of our archetype model into the existing knowledge on auditory processing such as related physiological and psychophysical results.

For the starting point, the model minimizes the expectation value of the quadratic error $E := \langle (s_t - \hat{s}_t)(s^t - \hat{s}^t) \rangle$ between the original auditory signal, $\mathbf{s} = (\dots, s_\mu, \dots)$, and its *optimal* reconstruction $\hat{\mathbf{s}} = (\dots, \hat{s}_\mu, \dots)$. The reconstructed signal $\hat{\mathbf{s}}$ minimizes E and is thus optimal in the least-square sense. This is the definition of optimality we have already introduced previously.

The optimal reconstruction $\hat{\mathbf{s}}$ is computed by the tensor l. Its coefficients suffice the set of equations $(\partial E / \partial {l_\mu}^\nu = 0)$. The resulting reconstruction algorithm computes $\hat{\mathbf{s}}$ only by means of the sensory response $\mathbf{r} = (\dots, r_\mu, \dots)$ of the auditory system, $\hat{s}_t = {l_t}^\tau r_\tau$. The coefficients ${l_t}^\tau$ of the matrix l are describing how (in neuronal terms: with which connection strength) a value of the incoming signal at the time $t - \tau$ is mapped onto the output signal at time t. Neuronally this can be realized through a delay line with delay time τ and a synapse of suitable strength. So, l contains the set of optimal temporal receptive fields that compensate for the corruption of the original signal within the sensory response.

The sensory response itself will be, in principle, a convolution of the original signal and the transfer function h that in this setting we denote as echo function, viz., $r_t = {h_t}^\tau s_\tau$. In the technical field the echo function h is more commonly specified as "room impulse response" (RIR). It depends on the physical surrounding, e.g., the "room". In analogy, the scenario biologically denoted as echo suppression is technically referred to as de-reverberation.

As in the general case (see Chap. 2), any *auditory* sensory system also has to cope with uncertainties no matter whether they come from measurement errors, variances within the assumed physical transmission process, or simply through the fact that space and time are continuous quantities that sensors and especially neurons cannot continuously represent. Artificial auditory sensors, for instance, typically average over a specific amount of time and give rise to measurement errors. The temporal dynamics of neurons are even inherently limited by characteristics such as refractoriness. In consequence, the input-output relation within any system will be corrupted by errors or noise. We thus rewrite the simple convolution $r_t = {h_t}^\tau s_\tau$ by adding the noise term ς accounting for measurement and transmission failures to our sensory response. In addition, our signal of interest may, and in general, will be disturbed by complementary signals we pool and refer to as background noise we model by adding a random variable α to the signal. We hence arrive at

$$r_t = {h_t}^\tau (s_\tau + \alpha_\tau) + \varsigma_t \ . \tag{3.1}$$

Before calculating the receptive fields l we need to discuss if the assumptions regarding the expectation values we have made in deriving the general framework (2.16) are valid in the present context of auditory signal processing. Of course, the input signal s_t is deterministic and we would not expect any problem. However, we do not know *ex ante* the exact values of s_t. We can overcome this problem by looking at "biologically relevant" signals. Such signals belong to a class of signals that we denote as "typical". Consequently a specific sensory signal is a concrete realization of a class of typical signals. That is, the input signal s_t is a stochastic quantity with a defined mean μ_S and we use an expectation value to compute E in the same way as in the general framework.

Furthermore, we also take all appearing cross-correlations to be zero, that is now the signal value at one specific point in time does not tell us anything about the signal value in the next time frame. The same holds true for both types of noise. We take the noise to be independent at different points in time, each with standard deviation $\sigma_{\varsigma/\alpha}$. Please note that the allocative function for the noise need not to be Gaussian. We, as expected, can therefore use the same assumptions as in case of (2.16). That is, only the autocorrelations remain and with $\delta_{\mu\nu}$ as Kronecker delta are given by

$$\langle s_\mu s_\nu \rangle = \mu_s^2 \delta_{\mu\nu},\ \langle \varsigma_\mu \varsigma_\nu \rangle = \sigma_\varsigma^2 \delta_{\mu\nu},\ \langle \alpha_\mu \alpha_\nu \rangle = \sigma_\alpha^2 \delta_{\mu\nu}\ . \tag{3.2}$$

We finally obtain a linear equation for the coefficients of the receptive fields l depending only on the transfer function h and the dimensionless parameters $\sigma := \sigma_\varsigma/\mu_S$ and $\eta := \sigma_\alpha/\mu_S$,

$$l^\mu{}_\gamma \left[h_\nu{}^\delta h^\gamma{}_\delta \left(1+\eta^2\right) + \sigma^2 \delta^\gamma_\nu \right] = h_\nu{}^\mu \tag{3.3}$$

where $h_\nu{}^\delta$ is the transpose of the matrix $h^\gamma{}_\delta$. The parameters σ and η correspond to inverse signal-to-noise ratios of σ – the mean signal strength to the variance of the detector messurement errors – and of η – the mean signal strength to the variance of the acoustic noise. The receptive fields l now match the echo function h and allow calculating the reconstruction $\hat{\mathbf{s}}$ of the original auditory signal $\mathbf{s}$. Since we have neither specified the original signal nor used any information such as input correlations (3.2), our algorithm can reconstruct *any* arbitrary signal. To get an optimal reconstruction performance in the case of noise, we simply need to adjust the two model parameters σ and η. If we have access to the noise levels $\sigma_{\varsigma/\alpha}$ and the typical value of the original input strength μ_S, the definition of σ and η gives us a good estimate of these values. In case prior knowledge of intrinsic correlations of the original signal, we can incorporate this into μ_S that hence transforms into a matrix with off-diadonal elements and leave the above *minimal ansatz*.

We have now derived an archetype model from our framework of optimal map formation in a concrete *temporal* setting. In the next subsequent section we will analyze the performance and the robustness of the model by applying it to exemplary echo functions.

3.2.2 Receptive fields: A computational building block

By analyzing the mathematical model derived in the previous section we can observe the intrinsic capabilities and characteristics of our approach. We deal with three exemplary but explicit temporal transfer functions, namely, the echo functions $h^{\mathrm{d;r;e}}$; cf. Fig. 3.1. In doing so, we take a delta function in the time domain as input signal. This function corresponds to a click. It appears slightly smeared out and is followed by an echo as defined by the echo function in the sensory response. The simplest environment featuring an echo would be a single solid wall in a free space. Here, the echo would be a simple, weakened repetition of the original signal. Hence for the first echo function we assume a single, discrete reflection of a click and label the echo "d" for "discrete". The second echo function, "r" for "realistic", mimics a typical room impulse response (RIR) where we have not one but many walls and where the signal is followed by a short silence and many subsequent reflections [170]. We take these reflections to have a maximum at about 20 ms after the click. For the third echo function we assume the closed space to be even more restricted and hence do not suppose any gap between signal and reflections. Here we take an exponential decay of the reflections and therefore mark it with "e". So, roughly following some exemplary RIRs [105], we cover the range of possible complex echo functions by three simplified scenarios, i.e., explicitly assuming the reflecting boundaries at very far (case "d"), normal (case "r"), and close (case "e") distance. In tensor notation ($r_t = h_t{}^\tau s_\tau$), they are

$$h^{\mathrm{d}}_{\mu\nu} = \exp\left[-\frac{(\mu-\nu)^2}{2\kappa^2}\right] + \exp\left[-\frac{(\mu-\nu+\beta_{\mathrm{d}})^2}{2\kappa^2}\right]\ , \tag{3.4}$$

$$h^{\mathrm{r}}_{\mu\nu} = \exp\left[-(\mu-\nu)^2/2\kappa^2\right] + \mathcal{A}(\nu,\mu)\ , \tag{3.5}$$

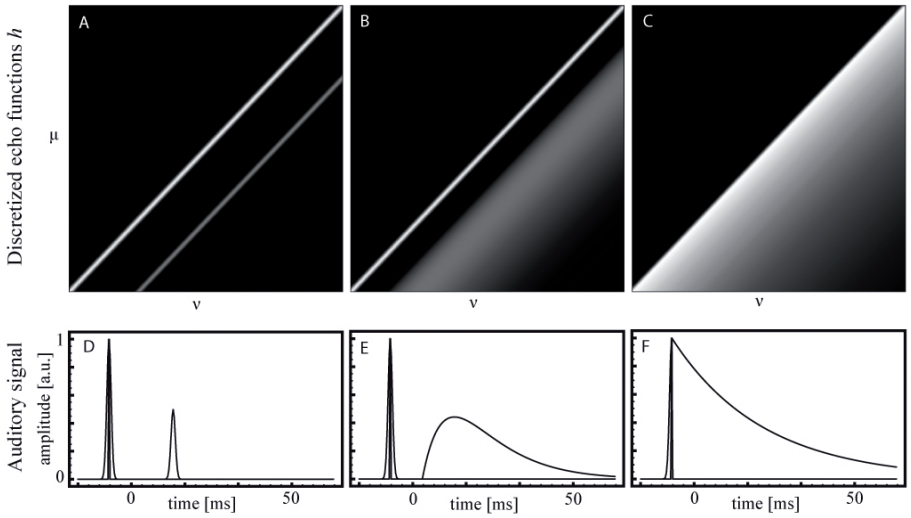

Figure 3.1: *Three variations of echo functions h with the original signal* $\mathbf{s}$ *and the resulting sensory response* $\mathbf{r}$*; settings "d", "r", and "e" from left to right. The upper graphs* $(A - C)$ *depict the discretized echo functions* $h^{\mathrm{d,r,e}}$ *(black* $= 0$*, white* $= 1$*) in matrix notation as defined by (3.4) – (3.6). The lower graphs* $(D - F)$ *show the original signal* $\mathbf{s}$ *(a delta function, filled line) and resulting sensory responses* $\mathbf{r}^{\mathrm{d,r,e}}$ *(solid line) in arbitrary units. Throughout this paper we use three exemplary echo functions, viz., discrete (left), realistic (middle), and exponential (right) echo, as explained in Section 3.2.2.*

$$h^{\mathrm{e}}_{\mu\nu} = \begin{cases} \exp\left[-(\mu-\nu)^2/2\kappa^2\right] & \text{if } \mu-\nu > 0 \\ \exp\left[-(\mu-\nu)/\kappa'\right] & \text{if } \mu-\nu < 0 \end{cases} \tag{3.6}$$

where all constants are real ($\in \mathbb{R}$). The function $\mathcal{A}(\nu,\mu) := (\mu - \nu + \beta_{\mathrm{r}}) \exp\left[-(\mu-\nu+\beta_{\mathrm{r}})/2\gamma^2\right]$ defines the common synaptic alpha function, μ and ν denote the rows and columns of h, and $(\mu - \nu)$ indicates the relative discretized time difference. The parameters β_d and β_r denote the delay between signal and echo for condition "d" and "r". The constants κ and γ denote how the signals and echos, respectively, get broadened, and κ' is a measure for the decay of the exponential tail in condition "e".

We now want to arrive at a thorough understanding of the receptive fields involved in our optimal model. To this end, we have calculated sensory responses and receptive fields for each of the three echo functions defined by (3.4) – (3.6); cf. Fig. 3.1 and Fig. 3.2. We have also evaluated the robustness of the model by comparing reconstructed and original signal for various conditions, including noise and a "mismatch" condition where echo function and receptive fields do not match; cf. Figs. 3.2 and 3.3. Importantly, we have tested not only the complete optimal receptive fields as derived in the previous section but also an abridged version that does not require an integration window and therefore allows real-time processing. Both abridged and complete receptive fields, with and without noise, work properly; for details see Fig. 3.2.

Moreover, we have tested the collection of receptive fields in case of a mismatch between h and l. In doing so we use the wrong receptive fields for reconstructing the signal and compare reconstruction performance with the "matching" condition. All nine possible cases are depicted in Fig. 3.3 and show that the receptive fields (mainly l^{r}) are able to cope with extreme variations – namely the wrong echo. The signal can be reconstructed by means of reducing the noise and effectively suppressing the echo; cf. Fig. 3.3. This *flexibility* is a prerequisite for any biological system since such a system cannot afford a specific neuronal wiring for every possible echo (suppression-) scenario. It turns out that increasing σ makes the different receptive fields less sensible, more robust, and thus makes them perform better in the "mismatch" condition, cf. Fig. 3.3.

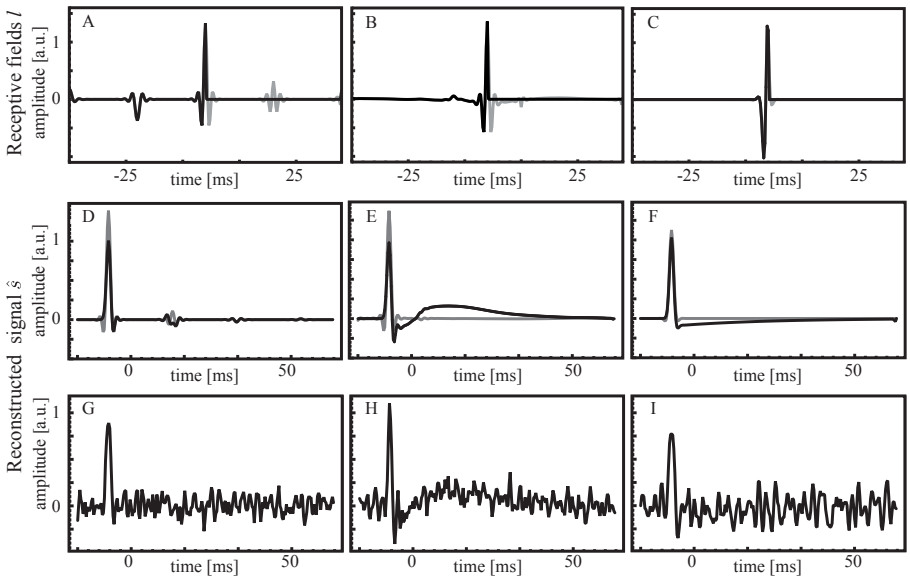

Figure 3.2: *Receptive fields l and reconstructed signals $\hat{\mathbf{s}}$; for condition "d", "r", and "e" from left to right. Graphs $(A-C)$ represent the complete receptive fields $l^{\mathrm{d,r,e}}$ as defined in (3.3) (gray) and the abridged versions (black). The versions d, r, and e always appear as left, middle, and right. The reconstructed signals obtained through both abridged (black) and unabridged (gray) receptive fields are shown in graphs $(D-F)$. The complete receptive fields l lead to higher peak amplitudes and less artifacts in the reconstructed signals. A noisy sensor still allows for a reliable reconstruction as depicted in $(G-I)$ with 10% noise as compared to input strength. All graphs have been expressed in arbitrary units. Original signal and sensory response are as in Fig. 3.1.*

To analyze which part of the receptive fields corresponds to which effect in signal reconstruction, we have set selected entries in the receptive fields to zero and observe the consequences. In result, e.g. for the "realistic" receptive fields, the first, fast inhibitory region sharpens the contour of the signal peak and reduces the exponentially decaying tail of the echo. The second, slow inhibitory region reduces the steepness of the echo onset that is important for auditory object formation.

To summarize the present section, we have derived and analyzed a simple, yet generic archetype of model for auditory stimulus reconstruction. In the next section we briefly review the state-of-the-art approaches in technical auditory processing and relate them to the model at hand.

3.3 Technical prospect

Having derived and discussed a very fundamental model, we now relate it to more recent approaches. From the technical point of view, what we have derived is a single channel de-reverberation algorithm that uses only one frequency channel. De-reverberation is well studied and a not yet completely solved problem in technical audio processing.

Our approach covers a wide range of typical de-reverberation algorithms. For example, ${h_t}^\tau = \delta_t^\tau$ corresponds to a conventional denoising scenario as described by [66, 290]. For a specific echo function and $\sigma = \eta = 0$ the model inverts, or, for a $\sigma \neq 0$ approximately inverts the equivalent RIR in the least-square sense as discussed by Miyoshi et al. [195]. A special case of linear filtering is the parameter set of $\eta = 0$ and $\sigma = 1$ that corresponds to a simple Wiener filter [135, 228, 232, 236]; for a review on linear filtering see [141].

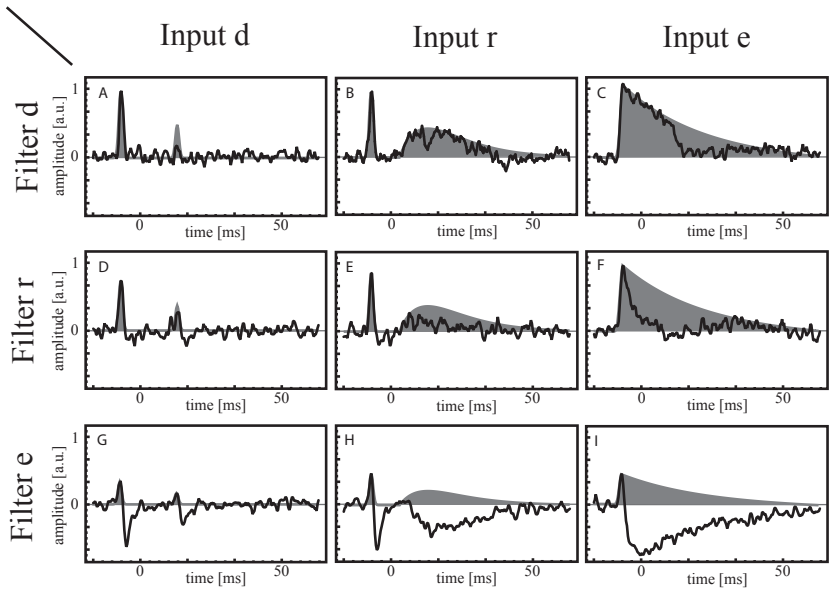

Figure 3.3: *Flexibility of the model. All graphs show the reconstructed signal* $\hat{\mathsf{s}}$ *(black line) and the actual echo function* h *(gray area) in arbitrary units. In each column a different echo function* h *is used to calculate the sensory response whereas the receptive fields* l *are varied in each row. Echo function and receptive fields match on the diagonal* (A, E, I)*. Even using receptive fields mismatching the actual echo function* (B, C, D, F, G, H) *can lead to resonable results if* σ *is chosen appropriately. Here* $\sigma = 5$*. In the case mismatch increased* σ *leads to results much better than the initial* $\sigma = 1$ *which corresponds to the Wiener filter [135, 228, 232, 236].*

For a brief synopsis, we have tried to subdivide the various technical algorithms for de-reverberation in three generations. The first generation consists of straightforward methods, the second combines several methods from the first generation, and the third generation tries to exploit the same signal characteristics that are exploited by biological systems. Single channel (using only one frequency channel) de-reverberation of the first generation can be distinguished into three major groups of algorithms based on de-convolution, blind de-convolution, and suppression. The first group assumes known echo functions (i.e. RIR) and deals with approximate and fast methods for inverting it [105, 195, 206, 215], as real echo functions are usually not exactly invertible [206]. The second group tries to find a filter which accounts for a certain criterion, for instance the mean power spectral density of speech or noise reduction, and indirectly reduces the reverberation [66, 73, 90, 105, 273, 290, 291]. As in the previous case this filter converges to an approximation of the echo function inversion. The third group estimates the amount of echo power spectral density [22, 105, 173] and uses suppression methods to perform, for instance, minimum mean square estimation of the original signal. In the state of art algorithms for echo suppression, the ones we refer to as second generation echo suppression, different methods are combined [79, 105, 110, 288]. An example for the third generation are attempts to rely echo suppression in an unknown environment on the pitch [272, 273].

But still these algorithms do not meet the requirements for artificial cognitive auditory processing such as sound source localization or speech recognition in a dynamic reverberating environment. In contrast, our everyday experience proves neuronal algorithms that allow sophisticated auditory processing exist. Therefore, in the next section, we show how the presented archetype for echo suppression can be transferred to a neuronal system. By contrasting the neuronal realization with actual neurobiology we identify similarities and differences. This allows to pinpoint promising sites of cognitive processing in the auditory pathway, that is, the sites for future research.

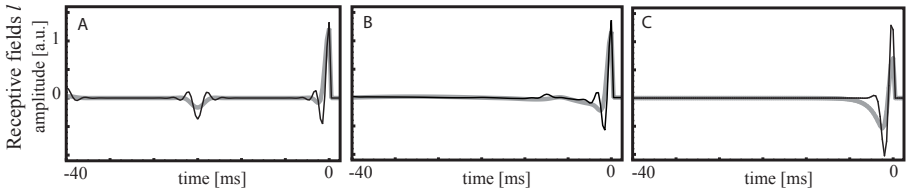

Figure 3.4: *Influence of parameter σ on the receptive fields l for "d", "r", and "e" form left to right. The receptive fields $l^{\mathrm{d,r,e}}$ for $\sigma = 1$ (black) and $\sigma = 5$ (gray) in arbitrary units. As σ increases, the receptive fields get smeared out and lose fine structure. Simultaneously the reconstruction gains generality, which is advantageous for reconstructing a signal deformed by an unknown echo function, see Fig. 3.3. Original signal and sensory response are as in Fig. 3.1.*

3.4 Biological prospect

In the course of this thesis we have stated that the framework of optimal map formation can be implemented to a neuronal setting in a straight forward way. In the following we prove this statement. Our present model of auditory processing allows an easy one-to-one match of the optimal receptive fields we have derived above to the neuronal connections including the explicit values of delay and synaptic strength. In doing so, we map the receptive fields (see Fig. 3.4) to a feedforward network as depicted in Fig. 3.5. Such a neuronal network is then capable of extracting the original signal and suppressing echoes. The performance of the neuronal realization is depicted in Fig. 3.6 for different situations, viz., receptive fields, and varying σ. Just as expected, the signal is extracted reliably. That is, we obtain a neuronal model that suppresses the echo and extracts a signal just as in the mathematical model. We hereby do not state that echo suppression is done in biology by means of a single feedforward network but give an existence proof of our model in neuronal hardware.

A comparative look on the shape of the *optimal* receptive fields reveals that the resulting synaptic strengths of the delay lines follow very plausible patterns. The most obvious feature is the emergence of two distinct time scales of suppression. First, a direct, fast inhibition with a clear maximum (e.g. at about 2 ms lasting ca. 5 ms; cf. Fig. 3.4 B). Second, a delayed shallow inhibitory region (in Fig. 3.4 B from about 10-17 ms) that is therefore called "slow inhibition". These two findings fit very well into today's picture of auditory processing. Gap detection, auditory contrast enhancement, and echo suppression exploit properties of temporal receptive fields, more precisely, delayed inhibition [36, 37, 100, 101, 282]. As for echo suppression, it is understood that it is in part monaural and in part binaural as von Békésy and Koenig already stated more than 50 years ago [20].

The monaural part of the suppression corresponds temporally as well as functionally to the fast inhibition in our model. It has a maximum at similar times (about 2 ms [108]). Physiologically it is located in the cochlear nucleus and hard-wired neuronally without significant variability [283], just as in our model where the fast inhibition is independent of the actual echo scenario. In the model the fast inhibition sharpens the contour of the peak and reduces the exponentially decaying tail of the echo but is not sufficient for complete echo suppression. Similar functioning has been found in and theoretically explained for the cochlear nucleus [36].

As for the binaural part of echo suppression, things are more complicated by nature. First, in binaural processing more nuclei are involved. Second, they have to deal not only with the input from one, but two ears. Third, the involved nuclei receive many efferent fibers, that is, top-down control plays a relevant role. Even if it is hard to disentangle the individual effects, the apparent results are the following. First, binaural echo suppression is slower than the monaural one [283]. Second, it is more flexible. More precise, it has been shown that a binaural inhibitory mechanism suppresses and changes the conscious perceptual processes of echoes depending on the situation

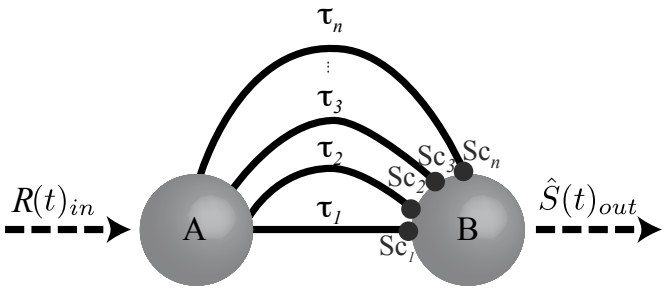

Figure 3.5: *Neuronal setting. A noisy response $R(t)_{in}$ containing an echo is used to stimulate a set of Poisson neurons [115] A, the so-called detector neurons. The resulting spikes propagate through delay lines with different delays $\tau_{1,..,n}$ and stimulate the Leaky-Integrate-and-Fire-neuron (LIF) [88] output population B through the corresponding synaptic connection strengths $Sc_{1,...,n}$. Synaptic connection strengths and delays are given by the* optimal *receptive fields contained in l. The output $\hat{S}(t)_{out}$ of the neurons B represents the original signal with suppressed echo and reduced noise.*

[221]. Compared to the time course of the slow inhibition found in nature, our model shows a similar time course that strongly depends on the situation, cf. Fig. 3.4. Furthermore, it reduces the steepness of the echo onset. This dependency upon the situation fits the biological counterpart where the flexibility of the binaural suppression is controlled by the top-down efferents.

In summary, a neuronal implementation of the archetype model works very well. It connects to the natural echo suppression. The binaural part of echo suppression that is highly adaptable could be a site for future research to identify and mathematically describe cognitive processing steps.

3.5 Discussion

Based on the architectural concept of error minimization and receptive fields, we have applied our framework for optimal map formation to a concrete setting by deducing a unifying archetype model for real-time extraction of an acoustic signal from a corrupted sensory response. We have studied the model performance in extracting the original signal for three exemplary echo scenarios.

On the one hand, we have shown links to common techniques that solve the inverse problem [242], apply linear filtering [141] such as e.g. the Wiener filter [135, 228, 232, 236], pseudo-inverse techniques, and the Maximum Likelihood Estimation [135, 228, 269]. On the other hand, we have implemented the archetype model into a neuronal setting to contrast the *optimal* and the biological approach. Initially, we have posed and answered the question as to which neuronal delays play a role in echo suppression, what are their synaptic coupling strengths, and what are the relevant time scales (see Fig. 3.2 and 3.4). We note that the resulting synaptic strengths of the delay lines follow biological patterns such as the two temporal distinct inhibitory regions and hint to cognitive processing such as the *Clifton Effect* [46]. That is, the auditory system adapts quickly to characteristics (echoes at about $\sim 10\,\text{ms}$) of a particular surrounding. In mathematical terms, the signal processing shows plasticity by means of adapting quickly to a specific transfer function (room impulse response).

Moreover, auditory object formation, both speech recognition and auditory space development, are impacted by the mechanism referred to as binaural echo suppression. This is because the slow inhibition reduces the steepness of the onset of the neuronal response to the echo and auditory object formation strongly relies on the onset of acoustic signals [29, 292]. The slow inhibition is thus responsible for the identification of the echo as a localizable object and a conscious echo perception. In this light, it makes sense that it is realized binaurally with strong top-down control.

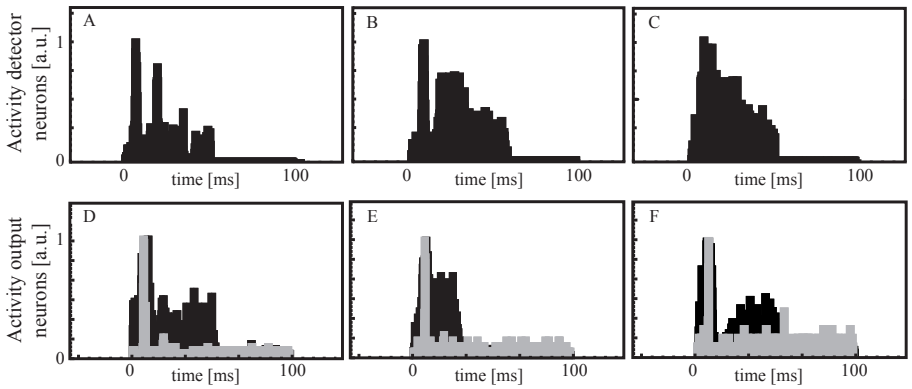

Figure 3.6: *Neuronal input and output activity; for settings "d", "r", and "e" from left to right. According to Fig. 3.5, we feed the sensory response $R(t)_{in}$ with a noise level of 5% (ς_t) to a population of 140 detector neurons A. The resulting spike trains ($A - C$) propagate through the delay lines with delays $\tau_{1,\ldots,n}$ and strengths $Sc_{1,\ldots,n}$ and stimulate a population of 140 LIF output neurons B. The set of delay lines realizes the derived optimal receptive fields contained in l. The upper row depicts the activity of the detector neurons A for the different echo functions $h^{\mathrm{d,r,e}}$, the lower row the corresponding activity $\hat{S}(t)_{out}$ of the output neurons B for two sets of reconstruction parameters σ. More precisely, the lower row depicts the normalized number of spikes of the output population with $\sigma = 10$ (D), 5 (E), and 10 (F) (gray bars). Choosing the right σ enhances the model performance as compared to $\sigma = 1$ (black bars) that corresponds to the Wiener filter. Obviously, neuronal characteristics such as the spontaneous rate, here roughly 15 Hz, do not degrade the performance since the signal is extracted reliably. The threshold of the LIF neurons even enhances echo suppression.*

Coming back to the state-of-the-art algorithms for technical de-reverberation and their potential for improvement, we realize that the natural approach is more powerful. Today's speech recognition and auditory object tracking algorithms cannot cope with noise caused by reverberation in dynamic scenes. Even the identification of one acoustic source in the presence of others is a problem not entirely overcome by technical systems. However, a robust auditory representation as part of a multi-modal representation as found in animals [147, 261] is supposed to be a necessary ingredient for cognitive applications such as robots and cognitive workbenches. Considering the ease with which even not fully developped biological organisms like, e.g., fledglings manage to identify and track an, until then unknown, auditory object such as their mother's voice in a new environment it is obvious that the twofold view on both the biological and the mathematical side of the same problem disembogues into a fruitful merge of cultures.

In summary, our archetype model uniformly ties up with the known physiological, behavioral findings, technical aspects and methods of echo suppression. That is, our framework of optimal map formation is a feasible tool to provide a common converging perspective on sensory processing enabling the multidisciplinary approach – bionics.

Wir können die Welt nicht erkennen, wie sie ist, sondern nur, wie sie für uns ist.

Tor Noerretranders

4. Lateral-line Perception

The lateral-line system is a unique mechano-sensory facility of aquatic animals that enables them not only to localize *prey, predators, obstacles, and conspecifics but also to* recognize *hydrodynamic objects. In this chapter, we present an analytic approach that shows the effect of submerged moving object (SMO) shape to lateral-line perception. Based on these results we derive an explicit model explaining how aquatic animals can distinguish differently shaped SMOs. Our model originates from the hydrodynamic multipole expansion and uses the unambiguous set of multipole components to identify the corresponding object. Furthermore, we show that within the natural range of one fish length the velocity field contains far more information than that due to a simple dipole. Finally, the results we present are easy to interpret both neuronally and technically, and agrees well with available neuronal, physiological, and behavioral data on the lateral-line system.*

4.1 Introduction

All fish and some aquatic amphibians such as the clawed frog Xenopus possess a unique sensory facility, the *lateral-line* system. This system is composed of mechanosensory units called neuromasts located on the trunk of the animal. The neuromasts consist of small cupulae, gelatinous flags protruding into the water, which are sensitive to the local water velocity [21]. The above mentioned animals use their lateral-line system to *localize* predators, prey, obstacles, or conspecifics. The lateral line is so effective that it enables even blind fish to navigate efficiently through their environment by measuring and analyzing the pressure and velocity field of the surrounding water [52, 111]. Moreover, it is supposed that aquatic animals not only localize but also *extract* and *recognize* features from the hydrodynamic structure of the velocity field to determine size, speed, and presumably shape of the object generating it [39, 52]. This can be done even indirectly, for instance, by means of extracting information from the vortexes within the object's wake [52].

The question we now pose, and answer, is whether and how a passive detection system such as the lateral line can both localize a moving object and determine its shape in an incompressible fluid such as water, or air at low velocities. Most studies, both experimental and theoretical [54, 75], used a vibrating or translating sphere as stimulus. Due to the special symmetry of the sphere, the resulting velocity field is exactly that of a dipole – no matter whether vibrating or translating [171]. Of course, not all objects in nature are spheres; cf. Fig. 4.1. Only for vibrating bodies can one identify [107, 142] the influence of their shape on the flow field, showing a reasonable dominance of the dipole. The literature lacks, however, a general explanation in a more global context. Here we show and quantify how much information is available in the velocity field, how one can extract it and to what extent in dependence upon the distance.

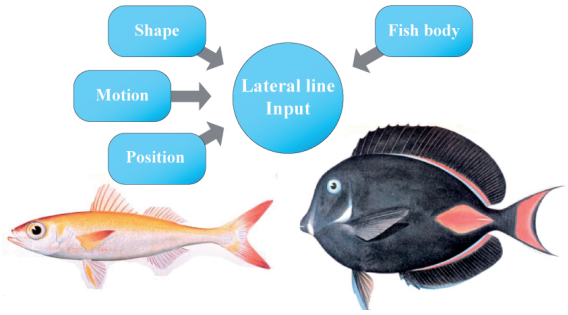

Figure 4.1: *Submerged moving objects (SMOs) appear in widely varying shapes. Furthermore, aquatic stimuli may but need not move at all; for example, vortex structures that are generated in the wake of a swimming fish or at the end of the fins [52]. These stimuli imprint information on the flow field that can be read out by the lateral-line organs. In addition, the input measured at the lateral-line organs is modified due to quantities such as the SMO's position, motion, and shape. The body of the detecting fish also influences the flow field, for instance, by means of damping (From NOAA's Historic Fisheries Collection; IDs: fish3078 and fish3039)*

More in particular, we first derive analytical equations showing that shape information plays an essential role affecting the velocity profile at the lateral line which is the natural stimulus to the mechanosensory system. Second, we demonstrate how hydrodynamic object characterization in terms of a mathematical description in three dimensions is possible by means of a multipole expansion (ME). In addition, we show how an aquatic animal, called detecting animal (DA), such as a predator, can read out the velocity field and reconstruct the shape of a stimulus, viz., a submerged moving object (SMO), such as prey. Finally, we give lines of evidence that there are biologically relevant limits of transmitting and extracting information – for both, the capability to localize and the range up to which shape information is important. The latter has not been explored up to now.

4.2 Analytic description of shape

As explained above, disturbances of the surrounding water form the natural input to this sensory system. The input may, though need not, be generated by submerged moving objects (SMOs) and depends on their position, motion, and shape. The body of the detecting fish also influences the flow field, for instance, by means of damping water disturbances, viz., the signal; cf. Fig. 4.1.

In this chapter we focus on and isolate the influence of the SMO shape. To outline the basic idea of the present section – an analytical description of the influence of stimulus shape – we recall powerful two-dimensional methods based on the complex formulation of potential flow developed for airship design [270]. Potential flow is a valid hydrodynamic description of the lateral-line *perception* although it neglects effects of viscosity [95, 111]. A complex potential flow description of three-dimensional hydrodynamic settings – including even rotational-symmetric bodies – is analytically not possible (see Appendix B). We therefore reduce the objects to circularly and elliptically shaped profiles within the (x, y) plane. In doing so, we gain *analytical* access to the essential effects of object shape on the flow field.

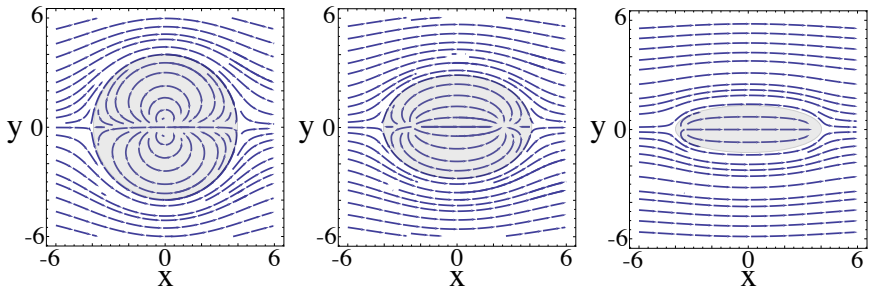

Figure 4.2: *Flow around three differently shaped profiles. From left to right: shape parameter (4.2) $a = 10^{-3}$, 2.5, 3.5; the radius $r = 4$, and $V_0 = 1$. We indicate the profile by gray. One can see how the profile becomes elliptic and the stream lines corresponding to the input to the lateral line adapt accordingly to object shape.*

4.2.1 Mathematical formulation

The complex velocity potential $F(z)^c$ and hence the complex velocity $w(z)^c := \mathrm{d}F(z)^c/\mathrm{d}z$ with $x + \mathrm{i}y =: z \in \mathbb{C}$, around a circular profile [218] is given by

$$F(z)^c := V_0 \left(z + \frac{r^2}{z} \right) \tag{4.1}$$

where V_0 denotes the moving speed and r the radius of the circular profile. In the next step, we use the Joukowski transformation $\zeta(z) := z + a^2/z$ with shape parameter $a \in \mathbb{R}$ [218], i.e., conformal mapping, to obtain the flow field $w^e(z)$ around an elliptically shaped profile. Here the superscripts c and e refer to circular and elliptical, respectively. In contrast to the usual way, we will not study the flow field w^e in the ζ-plane but in the original plane of the sphere, the z-plane. That is, we can identify the coordinates and thus *compare* the flow of the circular and elliptical profile,

$$w^e(z) = \frac{\mathrm{d}F}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}\zeta} = \underbrace{V_0 \left(1 - \frac{r^2}{z^2} \right)}_{w^c(z)} \underbrace{\left(1 - \frac{a^2}{z^2} \right)^{-1}}_{\frac{\mathrm{d}z}{\mathrm{d}\zeta}} . \tag{4.2}$$

In this way we have derived the complex velocity $w^e(z)$ of the flow around an arbitrary elliptic shape. Furthermore, we can compute all interesting quantities such as the velocity potential $\Phi^e(x, y)$ and the velocity field $\boldsymbol{v}^e = (v_x^e, v_y^e)$ by means of (4.2) and $z = x + iy$. All these terms are listed in Table 4.1. The transformation from the flow around circular profiles to the flow around elliptical ones can be described by *only* one parameter $a \in \mathbb{R}$. As $a \to 0$, all elliptical quantities (see Table 4.1) converge to the corresponding quantities of the circular flow.

We summarize these thoughts by stating that the circular flow is only a special case of the more general elliptic flow (4.2). As a consequence, we obviously cannot expect *a priori* statements derived from circular scenarios, i.e., the special case, to hold generally – if at all.

4.2.2 Effects of shape

Having derived a general expression (4.2) for the flow around an elliptically shaped profile, we now exhibit fundamental modifications of the velocity profile as it is measured by the lateral line due to stimulus shape. In the present literature, one distinguishes two classes of theoretical models that explain the detection capability of the lateral-line system [54, 75, 95, 251]. Models of the first category explain the performance of the lateral-line system by using the *whole* velocity profile along the fish's body. The second class of models exploits characteristic points such as extrema and zeros

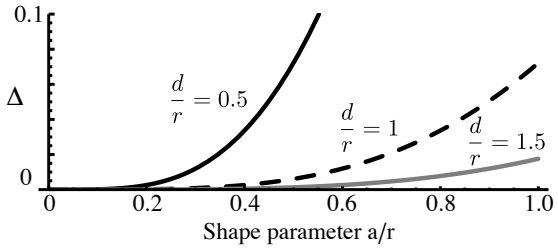

Figure 4.3: *The relative quadratic error Δ of the velocity-profile extrema at the lateral line x_E in dependence upon the distance d/r and shape parameter a/r normalized by the radius r. The graph shows $\Delta := [x_E(0) - x_E(a)]^2 / x_E(0)^2$ for $d/r = 0.5,\ 1,\ 1.5$. For distances d that are comparable to the SMO size (radius r), viz., in mathematical terms ($d \approx r \Leftrightarrow d/r \approx 1$), the positions of the extrema x_E clearly change with the shape parameter a. Present models ignoring shape effects calculating the distance d from a linear relation $d \sim x_E$ therefore overestimate the* real *distance beyond $d \approx r$.*

of the velocity profile along the lateral line since these points are supposed to be easily accessible neuronally.

We first analyze the latter category. Knowing the orientation, i.e, the direction of sensitivity $\boldsymbol{g} = \boldsymbol{g}(x, y)$ and position $\boldsymbol{r} = \boldsymbol{r}(x, y)$ of the lateral-line organs, every lateral-line geometry can be realized by means of $\boldsymbol{v_g}(\boldsymbol{r}) = \mathrm{d}\Phi/\mathrm{dg}$. In other words, we can calculate the water velocity in every direction $\boldsymbol{g}$ and at any point $\boldsymbol{r}$ of the flow field. Here, for ease of comparison with common scenarios in the literature, we study an SMO centered at $(0, 0)$ passing along the x-axis at a distance y to the detecting animal. In this setup, the velocity profile is identical with v_x and y denotes exactly the distance d between the centers of the SMO and the trunk lateral line.

We start by rescaling our expression for v_x by means of normalizing all quantities by the SMO size, i.e., the radius. This results, for instance in $a \to (a/r)r = a'r$. The normalization makes v_x explicitly independent of r. Computing the x-values x_E of the extrema (skipping $x_E = 0$) by means of computing $\partial v_x/\partial x|_{x=x_E} \stackrel{!}{=} 0$ gives $x_E(a) = \pm \left[a'^2 + d' \left(d' \pm 2\sqrt{a'^2 + d'^2}\right)\right]^{1/2}$. Obviously, the extrema depend on the shape; in mathematical terms, on the shape parameter $a' \in [0, 1]$. Only in the case of a circular profile ($a \to 0$) do the extrema encode the pure distance $x_E(0) \sim d$. Common models explain the localization performance of the lateral-line system by means of a linear relation to the zeros $d \sim (x_{E1} - x_{E2})$ [54, 75, 95]. Without taking the shape into account, the present models therefore overestimate the real distance.

At the same time the effect of shape also depends on the distance between SMO and the detecting lateral-line system as well as the SMO body length that directly relates to the parameter radius r in the equations. As we can see in Fig. 4.3, the influence of shape dominates the nearby range defined as the distance up to the SMO body length and vanishes for larger distances $d' \gg 1$.

Next we turn to the other class of models that not only exploit particular points but the *whole* velocity field at the lateral line. To analyze the influence of shape, we expand the complex velocity potential of elliptical profiles $F^e(z)$ in a Taylor series around $a = 0$,

$$F^e(z) = \underbrace{V_0 \left(z + \frac{r^2}{z}\right)}_{F^c(z)} + \left(\frac{r^2 V_0}{z^3} - \frac{V_0}{z}\right) a^2 + O[a]^4 \ . \tag{4.3}$$

Through the substitution $c := r^2/a - a$ we can transform (4.3) into a power series in z,

$$F^e(z) = V_0 z + V_0 c \left[\left(\frac{a}{z}\right) + \left(\frac{a}{z}\right)^3 + \left(\frac{a}{z}\right)^5\right] + O[a]^7 \ . \tag{4.4}$$

Table 4.1: *Hydrodynamic toolbox: Comparison of the flow around circular (denoted by ^{c}) and elliptical (denoted by ^{e}) profiles. Defining $c := r^2/a - a$, we can derive* simple *equations describing the flow around an elliptical profile by means of* only one *shape parameter a.*

Complex velocity potential

$$F^c(z) = V_0\left(z + \frac{r^2}{z}\right)$$
$$F^e(z) = V_0\left\{z + \frac{c}{2}\left[\ln(z+a) - \ln(z-a)\right]\right\}$$

Hydrodynamic potential

$$\Phi^c(x,y) = V_0\left(x + \frac{r^2 x}{x^2+y^2}\right)$$
$$\Phi^e(x,y) = V_0\left\{x + \frac{c}{4}\ln\left[\frac{(x+a)^2+y^2}{(x-a)^2+y^2}\right]\right\}$$

Velocity of the flow in x-direction

$$v_x^c(x,y) = V_0\left(1 + \frac{r^2}{x^2+y^2} - \frac{2r^2x^2}{(x^2+y^2)^2}\right)$$
$$v_x^e(x,y) = V_0\left(1 + \frac{c(x+a)}{2[(x+a)^2+y^2]} - \frac{c(x-a)}{2[(x-a)^2+y^2]}\right)$$

Velocity of the flow in y-direction

$$v_y^c(x,y) = -V_0\frac{2r^2yx}{(x^2+y^2)^2}$$
$$v_y^e(x,y) = -V_0\frac{2(r^2-a^2)yx}{a^4-2a^2(x^2-y^2)+(x^2+y^2)^2}$$

Equation (4.4) shows more clearly than (4.3) that the shape parameter a is the *natural expansion parameter* for the complex multipole series characterizing shape. We will exploit the derived finding that multipoles can be used to define and characterize hydrodynamic object shape later on.

The difference between the flow originating from a circular and an elliptically shaped profile, and thus the impact of shape, can be expressed to first order by

$$\Delta F(z) = [F^e(z) - F^c(z)] = -\frac{a'^2 r^2 V_0}{z} + O[a'r]^3 \tag{4.5}$$

where $a' = a/r$ is again normalized and ranges between 0 and 1. It is easy to see that the complex velocity difference $\Delta w(z) = \mathrm{d}\Delta F(z)/\mathrm{d}z$ and thus the measured velocity difference scales as $(r/z)^2$. The body length of the SMO ($\approx r$) therefore defines a natural length scale up to which the effect of shape contributes substantially to the flow field. Within this range the dipole approximation is only a valid description for circular shapes.

In summary, shape shifts the position of characteristic points of the velocity profile and also modifies it as a whole. Both effects depend on the relation between distance and SMO size. Since under potential flow condition we have $\nabla p \sim v\nabla v$ [95], the SMO shape affects ∇p accordingly and thus also the canal lateral-line processing. Moreover we have given an analytical line of evidence that the natural range up to which information about shape is transmitted is the SMO body length.

Having discussed really simple but nevertheless non-trivial shapes, we now want to give an outlook on possible extensions of our model (4.2). It is straight froward to adapt the equations in Table 4.1 to more fish-like shapes. Substituting $z \to (z + b)$ with $b \in \mathbb{R}$ in the Joukowski transformation leads to more complicated but richer expressions where the additional shape parameter b in return allows for more realistic shapes. We have depicted such shapes and resulting velocity profiles in Fig 4.4.

In the next section, however, we turn to three-dimensional shapes, their influence on lateral-line perception, and, more importantly, how the lateral line can recognize such differently shaped SMOs.

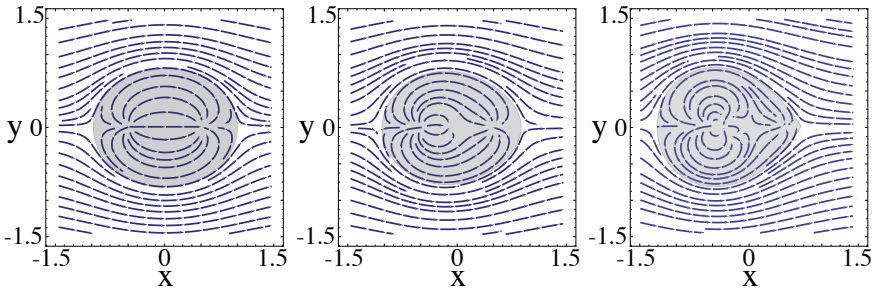

Figure 4.4: *Effect of the second shape parameter b. The graphs depict the flow field around three different profiles with varying the second shape parameter $b = 0, -0.1, -0.3$ (from left to right). The radius of the SMO is $r = 1$ and the shape parameter $a = 0.5$ from (4.2); both r and a are constant. The profile itself is indicated by the gray area. Due to variation of b the profile becomes more drop-shaped and the stream lines adapt accordingly.*

4.3 Hydrodynamic object recognition

In this part, we want to pick up the findings of (4.4), namely that a series of hydrodynamic multipoles is capable to describe the influence of SMO shape, and derive both a mathematical definition and description of hydrodynamic objects. In addition, we derive an *optimal* object recognition algorithm that can localize and reconstruct position and shape of SMO simultaneously in three coordinates. In doing so, we are able to predict the capabilities of the lateral-line system.

4.3.1 Definition of hydrodynamic objects

As we discussed in the previous section, a velocity field represents an adequate and natural stimulus to the lateral-line system and can be described by a multipole expansion (4.4). Now, we leave the two-dimensional space and focus on three dimensions. Because the relevant fluid dynamics are well described by the Euler equation [95] we take the three-dimensional velocity field derived from the real velocity potential $\mathbf{v}(\mathbf{r}) = -\nabla\Phi(\mathbf{r})$ [171] for a starting point. Using the real spherical harmonics $Y_{lm}^{\mathbb{R}}$ and the spherical multipole moments $\mathbf{q} = (..., q_{lm}, ...)$, we can expand the velocity potential Φ,

$$\Phi = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} q_{lm}\Phi_{lm} = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} q_{lm} \frac{1}{2l+1} \frac{Y_{lm}^{\mathbb{R}}(\theta,\varphi)}{r^{l+1}} \tag{4.6}$$

For the sake of simplicity we focus on rotationally symmetrical bodies whose surface S can be described through coordinates $\varsigma \in [0, \alpha]$ and $\eta \in [0, 2\pi]$,

$$S(\varsigma,\eta) = N \cos\left(\frac{\varsigma\pi}{\alpha 3}\right) \times \begin{pmatrix} \frac{\gamma^{-1/3}}{\sqrt{2}\sin(\alpha/2)} \sin(\varsigma)\cos(\eta) \\ \frac{\gamma^{-1/3}}{\sqrt{2}\sin(\alpha/2)} \sin(\varsigma)\sin(\eta) \\ \gamma^{2/3}\cos(\varsigma) \end{pmatrix} \tag{4.7}$$

The surface parameters $\alpha \in (0, \pi/2]$ and $\gamma \in (0, \infty)$ determine the shape of the surface. We note that vortex structures can be described by means of a multipole expansion (ME) as well [171]. Because SMOs such as fish leave characteristic vortices in their wake, the findings we derive in this and the following sections are also applicable for identifying SMOs indirectly by means of characteristic vortices in their wake.

Because of the Euler equation, only the Neumann boundary condition $\mathbf{v} \cdot \mathbf{n}_S = 0$ is realizable where $\mathbf{n}_S$ denotes a normal vector to the surface S. In terms of spherical coordinates we calculate

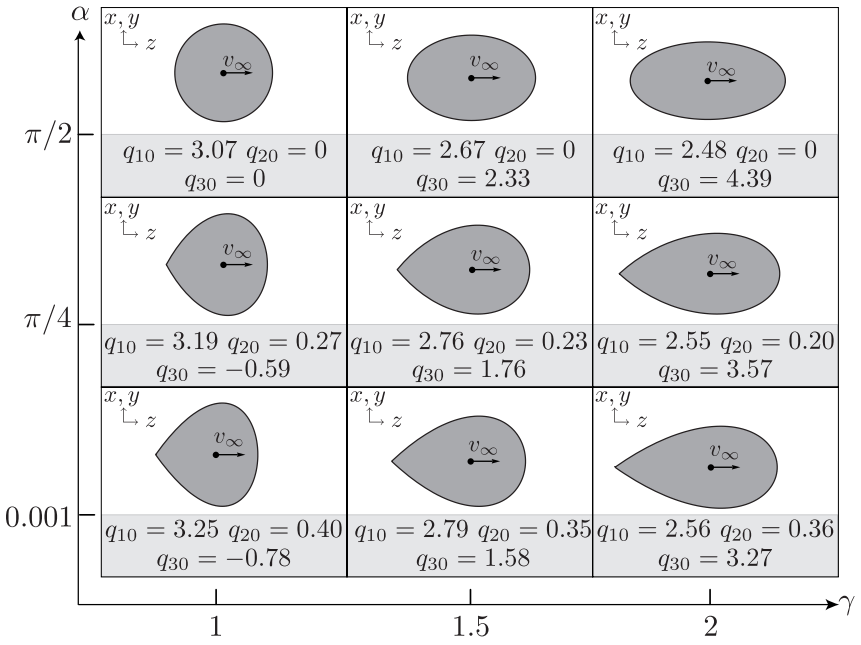

Figure 4.5: *Different shape parameters result in distinguishable sets of multipole components. We depict an approximation for only three multipole moments $l \leq k = 3$ in (4.8). The SMO surface parameter α increases along the vertical axis and the SMO surface parameter γ along the horizontal one. We have calculated the set of multipoles $\hat{\mathbf{q}}$ through a raster of 30,000 randomly distributed positions on each moving object's surface and used objects with lengths of about 5 cm and a matching moving speed $\mathbf{v}_\infty = 0.01$ m/s. Each volume is normalized to the volume of the unit sphere. It is therefore fair to say that the dipole q_{10} stays almost constant while the quadrupole q_{20} varies according to α and the octupole q_{30} according to γ. The relative error $\triangle(k) := |\mathcal{T}\hat{\mathbf{q}} - \mathbf{w}| \,/\, |\mathbf{w}|$ on the SMO surface S as calculated for rotationally invariant elliptic bodies gives $\triangle(k) < 0.4$ for $k = 3$ and $\gamma < 2$. It converges slowly to about 0.15 for $k \to \infty$. For $\gamma \gg 2$ a multipole expansion cannot describe the correct velocity field. This, however, is not too restrictive. For example, the fish in Fig. 4.1 can be described by $\gamma \approx 1$ (left fish) and $\gamma \approx 2$ (right fish).*

the components of $\mathbf{v}(\mathbf{r}) = -\nabla\Phi(\boldsymbol{r}) = (v_r, v_\theta, v_\varphi)$ at $\mathbf{r}$ through coefficients a_{rlm}, etc., so that

$$v_r = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} q_{lm}\, a_{rlm} \;, \quad v_\theta = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} q_{lm}\, a_{\theta lm} \;, \quad \text{and} \quad v_\varphi = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} q_{lm}\, a_{\varphi lm} \;. \tag{4.8}$$

The equations (4.8) can be written in matrix form $\mathbf{v} = \mathcal{A}\mathbf{q}$ with the matrix $\mathcal{A}$ containing the only geometry-depending coefficients a_{rlm}. To explicitly calculate the single multipole moments q_{lm}, we have to specify the boundary condition with respect to $\mathbf{n}_S$ and, for the sake of simplicity, the uniform speed $\mathbf{v}_\infty$ of the SMO in the stationary frame of reference. The boundary condition then reads

$$\mathbf{v}(\mathbf{r}_S) \cdot \mathbf{n}_S = \mathbf{n}_S \cdot \mathbf{v}_\infty = {\mathbf{n}_S}^T \mathcal{A}(\mathbf{r}_S)\mathbf{q} \;\text{ with }\; \mathbf{r}_S \in S \;. \tag{4.9}$$

The influence of the multipoles in (4.8) decreases with increasing l; see (4.6). Hence we expand the velocity field only up to a certain number k and set $q_{lm} = 0$ for $l > k$. We are thus dealing with an approximated set of multipoles, denoted by $\hat{\mathbf{q}}$. To find an appropriate approximation we take several positions $1 \leq i \leq n$ randomly distributed on an SMO surface and calculate the associated surface normals $\mathbf{n}_i$, the matrix $\mathcal{A}_i$, and write it line by line into the transfer matrix

$\mathcal{T} := (\ldots, \mathbf{n}_i{}^T \mathcal{A}_i, \ldots)$. The index i of the velocity components $w_i := \mathbf{v}_\infty \cdot \mathbf{n}_i$ of the vector $\mathbf{w}$ corresponds to $\mathbf{n}_i$ and therefore labels the same position. Here, we note that an increasing moving speed v_∞ of the SMO leads to increasing values of the corresponding multipole moments. In the next section, this will become important in the context of the detector organs signal-to-noise ratio.

To calculate the multipoles we simply solve the linear equation $\mathcal{T}\hat{\mathbf{q}} = \mathbf{w}$. The multipole moments follow from $\hat{\mathbf{q}} = (\mathcal{T}^T\mathcal{T})^{-1}\mathcal{T}^T\mathbf{w}$ [228]. The solution approximates the multipole coefficients up to a given k *optimally* in the sense of minimal quadratic error. Introducing a characteristic length scale λ, say the body-length of the SMO, into equation (4.6) we get

$$\mathbf{v}(\mathbf{r}) = \sum_{l,m} q_{lm} \left(\lambda/r\right)^{l+2} f_{lm}(\theta, \varphi) \tag{4.10}$$

where $f_{lm}(\theta, \varphi)$ are functions depending only on the angular coordinates θ and φ and can be calculated by means of (4.6). In doing so, we obtain dimensionless multipole moments. On the one hand we can compare them independently of the SMO size. On the other hand we see from $\mathbf{v}(\mathbf{r}) = \sum_{l,m} \left(\lambda/r\right)^{l+2} q_{lm} f_{lm}(\theta, \varphi)$ how the influence of each multipole varies in dependence upon the size λ of the SMO and the distance to the DA labeled by the radial coordinate r. This result tells us that the shape of the SMO is only important if the size λ and the distance r are of the same magnitude. It is e.g. obvious that plankton, which is ten to hundred times smaller than the DA, will not transmit any shape information through the flow field – and in this case shape information is not needed either. This result is in perfect correspondency with what we have derived previously starting from different equations in the complex two-dimensional space; cf. Fig. 4.3.

However, as stated before, in schooling, mate finding, or predator-prey behavior object information can make a difference, and as shown in Fig. 4.5, differently shaped objects are represented by different sets of multipole moments. Consequently, multipoles are able to quantitatively define and describe hydrodynamic objects as we have demonstrated for the first time [251] – as far as we know.

4.3.2 Object recognition performance

How, then, can aquatic animals such as fish reconstruct the above multipole moments from water velocity measured through their lateral-line system? Of course the animal has to fulfill this task under the influence of omnipresent noise and the limitations of its neuronal system.

Our reconstruction model for three-dimensional shape recognition is based on a maximum-likelihood estimator [276] and the framework of optimal map formation that we have introduced in Chap. 2. That is, we are looking for the multipoles $\mathbf{q}$ with the highest probability given the measured velocities $\mathbf{w}$ on the DA body. So to speak, this estimator maximizes the conditional probability $p(\mathbf{q}, \mathbf{r}_0|\mathbf{w})$ for the SMO center of mass $\mathbf{r}_0$, which has also to be determined by means of the information available through the DA superficial neuromasts, the water velocities. The velocity size w_i at organ i is given by

$$w_i = \mathcal{T}(\mathbf{r}_0, \mathbf{r}_i)\mathbf{q} + n_i \ . \tag{4.11}$$

We note that the transfer matrix $\mathcal{T}(\mathbf{r}_0, \mathbf{r}_i)$ depends on the SMO position $\mathbf{r_0}$ and the position $\mathbf{r}_i$ of the lateral line organ i (DA). As introduced in the first chapter, the transfer matrix maps the signal onto the detectors. In this case $\mathcal{T}(\mathbf{r}_0, \mathbf{r}_i)$ projects the multipoles q_{lm} onto the velocity components w_i measured at lateral-line organ i. Noise is modeled by adding independent Gaussian random variables n_i with mean 0 and standard deviation σ_n to the w_i (4.11). All positions $\mathbf{r}_0$ are of equal probability. We therefore take a Gaussian probability distribution $p(\mathbf{q}, \mathbf{r}_0) \sim \exp(-\mathbf{q}^2/2\sigma_q^2)$. Instead of maximizing the combined probability based on Bayes law, we maximize its logarithm $L(\mathbf{q}, \mathbf{r}_0) := \ln[p(\mathbf{q}, \mathbf{r}_0|\mathbf{w})]$. Defining the dimensionless parameter $\sigma := \sigma_n/\sigma_q$, we just have to maximize

$$L(\mathbf{q}, \mathbf{r}_0) = -\left\{[\mathbf{w} - \mathcal{T}(\mathbf{r}_0, \mathbf{r}_i)\mathbf{q}]^2 + \sigma^2\mathbf{q}^2\right\} \tag{4.12}$$

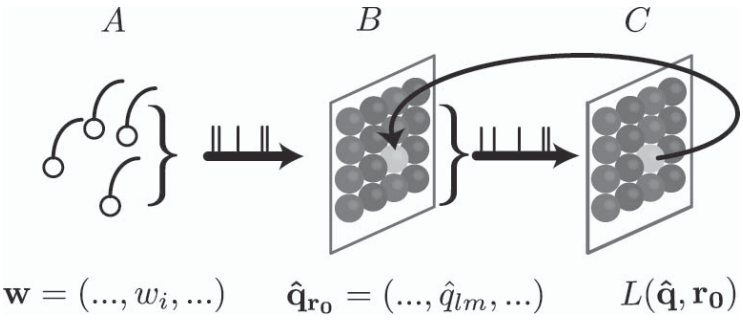

Figure 4.6: *Neuronal implementation of* (4.14) *is straightforward. In a first step* $(A) \rightarrow (B)$ *the multipole components* $\hat{q}_{lm}(\mathbf{r_0})$ (B) *are calculated for different positions* $\mathbf{r}_0$ *from the measured water velocities* w_i *at the neuromasts* (A)*. This corresponds to a network of synaptic connections between the lateral line and the central nervous system. The strength of the individual connections can be computed from the entries of the right-hand side of (4.14) or can be learned neuronally [74]. In a second step* $(B) \rightarrow (C)$*,* $L(\mathbf{q}, \mathbf{r}_0)$ *is calculated from the* $\hat{\mathbf{q}}_{\mathbf{r}_0}$*, which can also be done by feedforward connections; cf. (4.12). The maximum* $L(\hat{\mathbf{q}}, \hat{\mathbf{r}}_0)$ *indicates the correct position of the SMO and selects the optimal* $\hat{\mathbf{q}}_{\mathbf{r}_0}$ *from step* $(C) \rightarrow (B)$*; cf. (4.13).*

with respect to the multipole-moment vector $\mathbf{q}$ and $\mathbf{r}_0$. The necessary condition for a maximum of the likelihood $L(\hat{\mathbf{q}}, \hat{\mathbf{r}}_0)$ at the correctly estimated position $\hat{\mathbf{r}}_0$ is

$$\frac{\partial L}{\partial \mathbf{q}} (\mathbf{q}, \hat{\mathbf{r}}_0) \mid_{\mathbf{q}=\hat{\mathbf{q}}} = \left[\left(\mathcal{T}^T \mathcal{T} + \sigma^2 \right) \mathbf{q} - \mathcal{T}^T \mathbf{w} \right] \mid_{\mathbf{q}=\hat{\mathbf{q}}} = 0 \ , \tag{4.13}$$

which leads to a linear system of equations solvable by means of the pseudo-inverse technique [228],

$$\hat{\mathbf{q}} = \left(\mathcal{T}^T \mathcal{T} + \sigma^2 \right)^{-1} \mathcal{T}^T \mathbf{w} \ . \tag{4.14}$$

We estimate the set of multipoles $\hat{\mathbf{q}}$ given *only* the measured velocities $\mathbf{w}$ at the lateral-line organs. A neuronal implementation of these calculations is straightforward and can be done easily by neuronal hardware; see Fig. 4.6. Thus the position and the appropriate multipole moments can be calculated neuronally.

To test whether it is possible for aquatic animals to determine SMO position and shape using the above method, we have applied noise to the input signal w_i. We have used $\sigma_n = 10^{-4}$m/s as standard deviation of the noise, which corresponds to the velocity threshold of *Xenopus'* lateral line organ [21]. Furthermore, we define a signal-to-noise ratio, $SNR := v_\infty / \sigma_n$. Plausible maxima of the moving speed of SMOs range between one body-length per second and zero. To be fair, in the examples in the figures we have used a velocity as low as $v_\infty = 1$ cm/s corresponding to only 1/5 to 1/10 of the typical maximum moving speed. That is to say, we have dealt with a worse-case scenario since in case of higher moving speed v_∞ the reconstruction performance is even better than presented. Another aspect regarding lateral-line performance is the number of detector organs. Within a biological range the influence of the exact number of lateral-line organs is marginal. We show in Fig. 4.9 that the reconstruction as well as the localization error increases with the distance but stays almost constant for different numbers of neuromasts. Of course, taking, for instance, ten times more neuromasts can compensate a higher SNR value – but only within marginal ranges. In summary, the SNR dominates the performance as compared to the number of neuromasts.

Another issue is, the more multipoles we use the more positions are of high probability because one has more parameters to fit the measured w_i's. Thus the estimated position gets ambiguous but the shape estimate improves. We can choose an appropriate σ for a certain k to compensate this

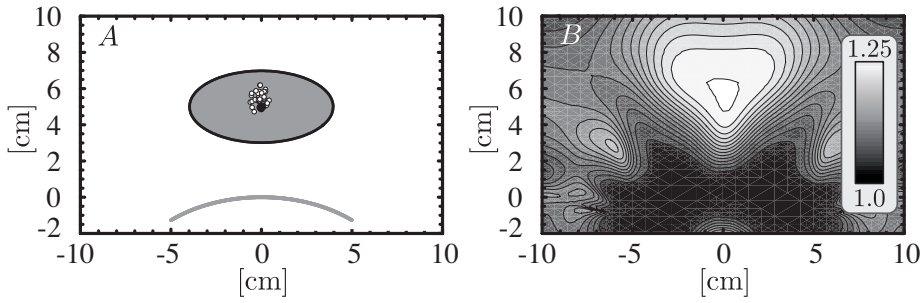

Figure 4.7: *The lateral line (A, gray line) is centered at* $(0,0)$*. The distance between the object (SMO, gray area) center of mass* $\mathbf{r_0}$ *and the lateral line (DA) is just below the length of the lateral line. Using the dipole term* $(k=1)$ *only, we can easily reconstruct the position* $\hat{\mathbf{r}}_\mathbf{0}$ *(A, cloud of black points around* $\mathbf{r_0}$*), a result consistent with that of Franosch et al. [74]. If we take* only *three multipole moments* $(k=3)$ *as an* efficient *minimal model, our method allows a proper estimate of both position (A, small open circles) and form, cf. Fig. 4.8 – even though more positions are now likely to occur. For both cases, 40 reconstructions have been depicted with* $\sigma_n = 10^{-4}$m/s *moderate* $SNR := v_\infty/\sigma_n = 100$ *and 500 neuromasts. B depicts* $\log[L(\hat{\mathbf{q}}, \mathbf{r_0})]$ *corresponding to the neuronal map of position estimation. It shows no ambiguity; cf. Fig. 4.6C.*

effect and enable a faithful localization; for details of how to tune σ consider Chap. 2. In Fig. 4.7 we show the ability of our method to estimate the SMO position $\hat{\mathbf{r}}_0$.

Following the ansatz of the neuronal model (Fig. 4.6) we have calculated the multipole moments at the estimated position, i.e., not only at the correct position but also at $\hat{\mathbf{r}}_0 + \delta\mathbf{r}$ where $\delta\mathbf{r}$ accounts for the continuity of real space, a limited number of map neurons as well as a slight localization uncertainty (Fig. 4.7). Our model can reconstruct $\hat{\mathbf{q}}_{r_0}$ even under noisy conditions and at slightly wrong positions (Fig. 4.8); for example, the fish bodies of Fig. 4.1 (see also Fig. 4.5) can be recognized and distinguished. Thus object *localization* as well as object *recognition* based on the estimation of multipoles is possible simultaneously.

In the subsequent lines we discuss the limitations of the proposed method and likewise the theoretical limitation of aquatic animals' ability of object recognition. To compare the quality of different multipole estimates $\hat{\mathbf{q}}$ at different distances d, we define $\delta_q := \sqrt{\langle(\hat{q}_{lm} - \langle\hat{q}_{lm}\rangle)^2\rangle}/\langle\hat{q}_{lm}\rangle$ as quality measure. Here δ_q labels the normalized standard deviation due to wrongly estimated positions $\hat{\mathbf{r}} + \delta\mathbf{r}$ and noise. For small δ_q the estimation error in q_{lm} is small and, in spite of a slightly wrong position, the shape parameter γ can be recognized. For values $\delta_q > 1$ a shape reconstruction based on the estimated $\hat{q}_{lm}$ is impossible.

The results of our numerical experiments, that is, the lateral-line performance our model predicts, support the previous analytical findings as follows. More precisely, for the octupole q_{30} the critical distance where the error δ_q starts to grow extremely fast is the SMO size itself. For localization the critical distance can be taken as the length of the DA lateral-line system, say a fish length, which is in case of a predator hunting for prey larger than the critical length of shape reconstruction (Fig. 4.9).

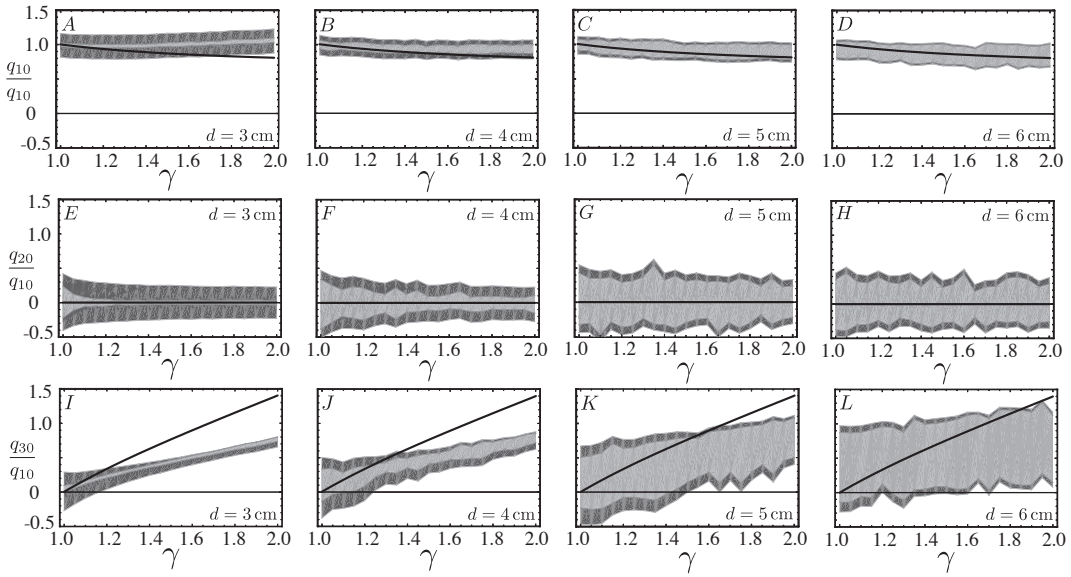

Figure 4.8: *Reconstructed dipole $\hat{q}_{10}$ $(A-D)$, quadrupole $\hat{q}_{20}$ $(E-H)$, and octupole strength $\hat{q}_{30}$ $(I-L)$ normalized by the dipole strength q_{10} for $\alpha = \pi/2$ and variable shapes $1 \leq \gamma \leq 2$ at different distances d. The gray colored region depicts the value of estimated multipole strengths with noise at the most likely position (light grey) and at slightly incorrectly estimated positions (dark grey) as explained in the main text and corresponding to Fig. 4.7. Since we have used an approximation with only three components, the strengths of the multipoles differ from the real q_{lm} (black line). Remarkably, q_{30} is an approximately linear function of γ and thus almost suffices to distinguish differently shaped objects. We see (A, B, E, F, I, J) that for small distances between SMO and DA differently shaped bodies can be distinguished, even in spite of noise and truncated ME. If d approaches SMO size (5 cm), the reconstruction starts getting blurred (B, C, F, G, J, K). At larger distances (D, H, L) no strong correlation between $\hat{q}_{10}, \hat{q}_{20}$ or $\hat{q}_{30}$ and γ can be found, and thus no shape can be recovered. Nevertheless the two fish in Fig. 4.1 with, say, $\gamma = 1$ and $\gamma = 2$ can be distinguished at least at $d = 3$ cm $(A, E,$ and $I)$ and $d = 4$ cm $(B, F,$ and $J)$ by means of estimating the whole set of multipole moments or only the strength of $\hat{q}_{30}$. In both cases $\hat{q}_{30}$ discriminates $\gamma = 1$ and $\gamma = 2$.*

In summary, our multipole expansion (ME) method can quantify stimulus characteristics by means of reconstructing the multipole moments q_{lm}. Because the flow field need not be generated by a submerged moving object alone, the method can also be applied to characterize composite situations such as schooling or vortex structures found in the wake of a fish. In presenting the results depicted in Fig, 4.8 and Fig. 4.9 we have answered the compelling question, whether a detecting animal (DA) or artificial system can *localize* and *recognize* a SMO given *only* the measured water velocities at the DA lateral-line organs and, and if so, how well. That is, the ME is useful only if the distance between the object (SMO) and detector (DA) is approximately of the same size as the SMO, or shorter. Otherwise the ME reduces to its dipole simplification.

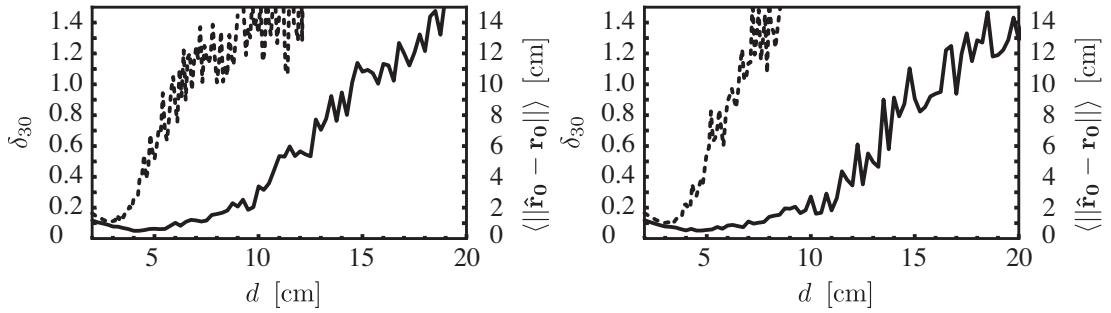

Figure 4.9: *The two graphs depict the relative reconstruction error $\delta_{q_{30}}$ (dashed line) of the octupole moment q_{30} as defined in the main text for different numbers of neuromasts. They show that for distances d larger than a moving object's size (5 cm) multipole reconstruction is not possible. The solid line depicts the localization error $\langle||\hat{\mathbf{r}}_\mathbf{0} - \mathbf{r}_\mathbf{0}||\rangle$, which starts to grow fast at a distance comparable to the length of the detecting animal (10 cm). This agrees well with experimental findings where localization performance starts decreasing at distances beyond the fish body size [54]. For discrimination tasks, fish have to be very close to the object under investigation [39]. The left graph shows the errors employing 250 (left) and 500 (right) neuromasts. As expected, in both plots errors increase with d but the difference due to the number of neuromasts (within biological range) is marginal.*

4.4 Discussion and outlook

In this chapter, we have focused on two major aspects of lateral-line performance. First, no matter whether the animal's capability to localize prey, predators, conspecifics, or obstacles is based on the processing of characteristic points of the velocity profile or the whole velocity field, the SMO shape plays an essential role [253]. Assuming that potential flow is an adequate tool for approximating the corresponding hydrodynamics, we have quantified *analytically* the effect of shape to gain further insights into mechanosensory processing. In addition and for the first time, we have pointed out that differently shaped objects result in different boundary conditions for the hydrodynamic flow field and thus cause distinguishable multipole moments q_{lm}. We have developed a multipole expansion method (ME) that predicts the possible object recognition performance of the mechanosensory sensory system – the lateral line.

Second, in the present literature there is a natural characteristic length-scale up to which the animal can localize objects, viz., its body length, or more precisely, the length of its trunk lateral line [95, 75]. Here, we have given different lines of evidence that there is a second length scale within which shape information is transmitted, namely, the body length of the SMO to be detected. If the distance between the DA and the SMO is below that length scale even the simplest objects transmit measurable shape information. This information can be used not only to localize but even to *recognize* hydrodynamic objects [251].

Promising steps for further enlightening this interesting sensory system are manifold. In concreto, we suggest to experimentally test the ability of the lateral line to recognize shape in different settings to evaluate the predicted recognition performance. One could clarify, for instance, the number of multipoles that is necessary to recognize shape in natural conditions. Another interesting aspect would be to analyze whether the sensory system is "blind" when it comes to distinguish SMOs with special symmetries. Moreover, the possible experimental setups may not be limited to the lateral line, viz., hydrodynamics. Air-dependent mechanosensory systems such as the sensory system of crickets can also be considered.

On the theoretical side, it would be very interesting to extend our model to the time domain. That is, on the one hand models that enhance the present ME model should deal with a temporal varying multipole pattern $\boldsymbol{q}(t)$ to faithfully *identify* hydrodynamic objects. Namely, the moving body or the vortices in the wake of swimming fish produce specific hydrodynamic structures that are time-dependent, which is true for the corresponding multipoles as well. Presumably, the time-dependent multipoles are animal specific. On the other hand, it is possible to extend the present ME model to not only reconstruct the most probable SMO position at one specific point in time but even to estimate the most probable SMO trajectory. Given the trajectory and assumptions on the moving object such as its inert mass it is easy to predict its future positions. The dipole moment includes the actual moving direction and speed of the SMO and would therefore already suffice to start with.

In summary, our object-recognition model agrees well with biological findings and provides a theoretical understanding of hydrodynamic object perception through the lateral line. Furthermore, we now understand the fundamental restrictions for any (sensory) evaluation of lateral-line data. Finally, these findings can also be applied to biomimetics [289]; e.g., to improve passive naval navigation systems.

Figure 4.10: *Sometimes, it makes a difference to know* what *is* where. *From my poster presented at the 2009 Sloan-Swartz Summer School at Harvard University.*

Die Wahrheit triumphiert nie,
ihre Gegner sterben nur aus.

Max Planck

5. Infrared System of Snakes

Some snakes possess an infrared detection system that is used to create a heat-image of their environment. In this chapter, we explore the involved physics of signal mapping, the neuronal characteristics of sensory en- and decoding, and address the question of how the infrared system can create a functional representation of the thermal properties of the environment.

5.1 Introduction

Two groups of snakes, pit vipers and boids, possess a sensory system to detect infrared (IR) radiation [198] which allows them to perceive a two-dimensional image of the heat distribution in their surroundings. The detection system consists of a set of cavities called pit organs. In pit vipers, a pit organ is located on each side of the snake's head near the eyes; cf. Fig. 5.1 A. Suspended in each cavity is a heat detecting membrane, which is sensitive to mK temperature differences [33, 63]. The optical principle underlying the detection of IR radiation is that of a pinhole camera with large aperture; cf. Fig. 5.1 B.

Radiation entering through the pit hole hits the membrane at a certain spot depending on the source direction; see Fig. 5.2 A. The resulting temperature change at this spot is detected by heat

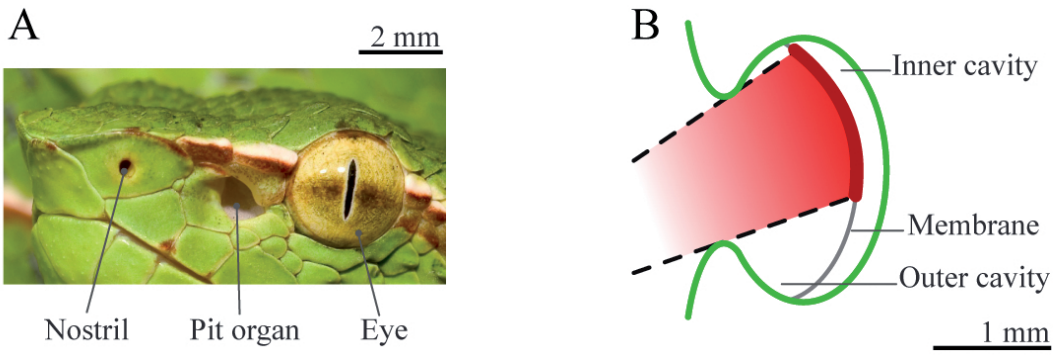

Figure 5.1: *(A) Head of a pit viper with nostril, (large) pit hole, and eye, left to right. Photograph courtesy of Guido Westhoff; adapted from [252]. (B) A pit viper's infrared-sensitive pit organ works like a pinhole camera. Radiation entering through the opening (left) hits a heat-sensitive membrane, suspended freely so as to minimize heat loss to the surrounding tissue. To ensure a large enough energy influx, the aperture has to be quite large (∼ 1mm), comparable to the distance aperture-membrane. The image of a point-like source thus forms a disc-shaped image on the membrane.*

sensitive cells, free nerve endings (FNEs) that are organized within dense architectures called the terminal nerve masses (TNMs) [6, 121]. The TNMs represent the heat sensitive sensory areas that are distributed throughout the membrane; cf. Fig. 5.2 B. There are about 40×40 sensory areas on the membrane [198], and the field of view is about 100° wide, which implies that input to the organ could be represented in the brain with a resolution of approximately 2.5°. Since the radiation flux entering the organ must be large enough to quickly detect moving prey, the aperture of the organ is wide, approximately 1 mm, and comparable to the organ depth. Thus incoming radiation from a point source does not strike a point-like region on the membrane, as in an ideal pinhole camera, but rather much larger disc-shaped region; cf. Fig. 5.1. It is nonetheless possible to determine the direction of the incoming radiation if the boundary of the disc-shaped region remains narrow enough. This approach breaks down if multiple or non-pointlike heat sources are present. Then the resulting heat distribution on the membrane will be heavily blurred (see Fig. 5.10), precluding a direct evaluation of the input.

The IR-detection system of snakes presents a paradox. The optics of the pit organ as pinhole camera ensures that images on the membrane will be distended and blurry. Yet the information from the IR system, combined with input from the "normal" visual system, allows formation of a rather precise *neuronal map* [116] in the brain's *optic tectum* [109, 210, 211]; cf. see Fig. 5.2 C. This map is sharp enough to serve as a *topological* representation of the outside world in which neighboring neurons represent neighboring regions of the outside world. Even a multimodal interaction between the infrared map and the visual map has been found [109, 210, 211]. Experimental studies have shown that a snake's orientation to a point source of heat varies but that an accuracy of 5° is nevertheless attainable [211], which approximates our rough estimate above. We are thus faced with a paradox, in that the optical quality of a pit organ is low but the neuronal performance is high. How does that happen?

The paradox involves the optical quality of IR "vision". We therefore explore the optical quality of the pit organ and address the thermodynamic phenomena involved. This enables us to calculate the heat distribution on the membrane for a given heat distribution in space. In a second step, we address the question of how the mapped heat distribution is encoded neuronally. Based on the membrane image and the knowledge of sensory encoding, we continue with the third step of sensory processing, that is to neuronally decode the original spatial heat distribution from the measurements on the membrane. We develop two reconstruction models which are able to explain how the neuronal system can compensate the oversized pinhole. Since a snake has limited computational resources (all "calculations" must be realizable in neuronal "hardware") any reconstruction model must be simple. Accordingly, our models use only one computational step (they are *non*-iterative) to estimate the input image only from the measured response on the pit membrane. Both resemble a Wiener filter and are akin to but different from some of the algorithms used in image reconstruction [229]. It has to be constantly born in mind that the optical device discussed here has *very* limited imaging capabilities in comparison to human-devised detection systems. Nonetheless, as we will show, it is still possible to reconstruct the spatial heat distribution in a biologically plausible way. In addition, we quantify the performance of our models and give reasonable limitations for localizing warm objects by means of this sensory system. Beyond various specific issues of the pits that are considered, we extend the current understanding of this sense by means of enlightening the question of what the appropriate stimulus to the IR-detetction system of snakes is. Finally, we present an outlook and discuss still open questions that should be answered.

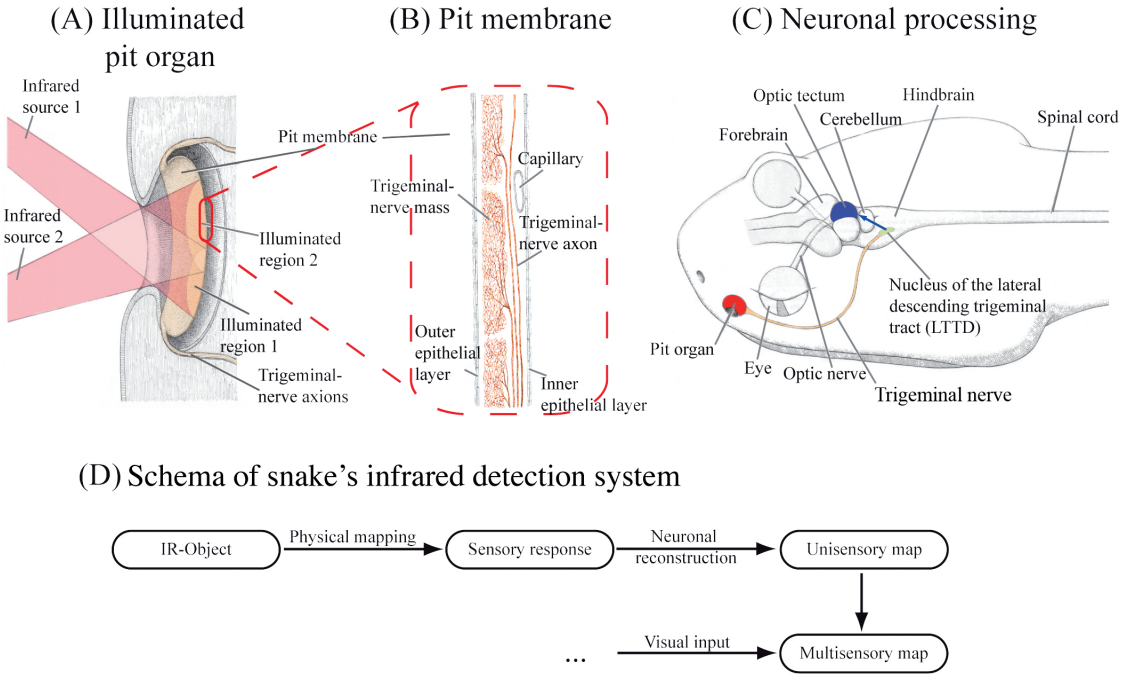

Figure 5.2: *Schematic drawing of the essential parts of the snake's IR-detection system; adapted from Newman and Hartline [211]. (A) depicts a pit organ that is illuminated by two infrared sources. The illuminated areas 1 and 2 on the pit membrane heat up. (B) Sensory responses are generated within the free nerve endings of the branching trigeminal nerve axons that form the terminal nerve mass (TNM). The spikes are transmitted along the trigeminal nerve to the nucleus of the lateral descending trigeminal tract (LTTD). (C) Within the brain, the infrared information is, i.a., represented in the optic tectum by means of a neuronal map. (D) Within the optic tectum several computational steps between the visual map and the infrared map can be found. The information of these two sensory systems fuses into a comprehensive percept of the environment. For a detailed review of the sensory pathway we refer to elsewhere [162, 163, 182, 191, 197, 198, 199, 209, 211, 210, 246, 255, 256, 266].*

5.2 Warming up: Signal mapping

In this section, we derive equations of how to map the stimulus, the spatial heat distribution of the external world onto the sensory cells within the pit membrane of the snake as depicted in Fig. 5.2. Initially, we clarify which the appropriate stimulus for the infrared detection system is and what its characteristics are. In so doing, we derive an appropriate mathematical description of the signal mapping process.

5.2.1 A thermodynamic vista

To start with the latter issue, we follow the widely accepted opinion that the pit membrane receives thermal radiation from the external world and hence heats up or cools down, viz., it is a warm receptor organ [48, 49, 168]. On the other hand, the membrane does not respond to specific wave lengths like the retina [93, 196]. We hence have to analyze the mechanisms of how the membrane heats up and cools down. We derive a simplified model of the complex heating and warming process to estimate the importance of the different mechanisms involved.

At the outset, we conceptually divide the heat transport phenomena into two categories, the heat transport within the pit membrane and the heat transfer through the membrane-air surfaces. That is, we consider heat conduction through the air, energy gains and losses through radiation,

blood perfusion, heat conduction within the tissue, and convection. All these processes affecting the temperature of a small pit membrane slice are schematically depicted in Fig. 5.3.

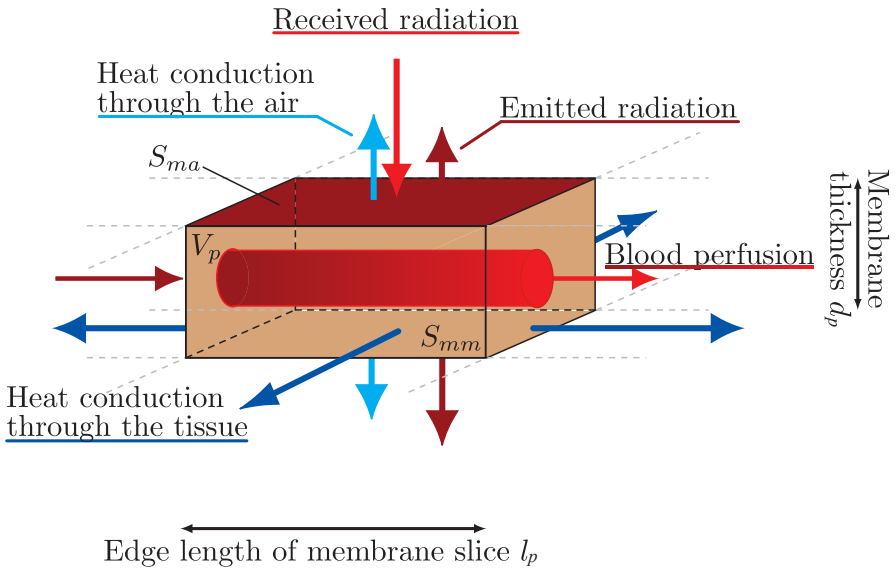

Figure 5.3: *Slice of the center of the pit membrane. We can see the different heat flows that go through the different surfaces, i.e, the membrane-membrane surfaces $S_{mm} = l_p d_p$ and the membrane-air surfaces $S_{ma} = l_p^2$. The edge length l_p of the quadratic slice is 25×10^{-6}m and the thickness of the membrane d_p is assumed to 15×10^{-6}m.*

The starting point of our study is the basic heat equation that describes the heat development within the tissue of the pit membrane. It reads

$$\rho_p C_p \frac{\partial T(\boldsymbol{x}, t)}{\partial t} = k_p \nabla^2 T(\boldsymbol{x}, t) + \sum_i Q_i \ . \tag{5.1}$$

In this section, we denote quantities associated with the pits with a subscript p, for instance, the pit membrane density $\rho_p (\approx 1080\,\mathrm{kg\,m^{-3}})$. The other quantities are $C_p (\approx 3850\,\mathrm{J\,kg^{-1}K^{-1}})$, the heat capacitance and $k_p (\approx 0.49\,\mathrm{Wm^{-1}K^{-1}})$, the thermal conductivity. Both refer to the characteristics of the tissue. As reasonable values for the parameters of the pits are not available, we have taken, for the sake of consistency, values for neuronal tissue, i.e., the brain, from Fiala et al [72]. The first term in the right-hand side of equation (5.1), $k_p \nabla^2 T_p(\boldsymbol{x}, t)$ describes heat conduction through the tissue. The fundamental solution, viz., in case $Q_i = 0$ is an exponential diffusion $T(\boldsymbol{x}, t) \sim \exp[-x^2/ct]$ of the heat profile with $c = \text{const.}$. The effect of this term is that differences in the heat distribution get smeared out [70]. We will come back to this term later on and, in so doing, we specify and discuss the different sources and sinks of energy Q_i.

The second transport phenomenon within the tissue is the heat transport through blood. Many experimental findings [6, 92, 121, 205] have reported a dense capillary system within the pit membrane. Amemiya et al. [5] even suggested the capillaries function as a "heat exchanger"; see Fig. 5.7. Pennes [222] first formulated and subsequent investigators validated [50, 72, 207, 286] the so-called bio-heat equation describing the heat transport through the blood Q_b for parts of the

body such as an arm,

$$Q_b = \rho_b V_b C_b (1-\kappa)(T_a - T) = \omega C_b (1-\kappa)(T_a - T) \ . \tag{5.2}$$

Here Q_b is the rate of heat transfer per unit volume of the tissue. The perfusion rate per unit volume of the tissue is denoted as V_b, ρ_b is the density of blood, $C_b (\approx 4000\,\mathrm{J\,kg^{-1}K^{-1}})$ is the specific heat of the blood, κ is a factor that accounts for incomplete thermal equilibrium between blood and tissue. Typically, it is set $\kappa = 0$. The temperature of the arterial blood is denoted as T_a, and T describes the local tissue temperature. Despite the large number of mitochondria around the TNMs, we do not assume in the above equation (5.2) that the pit membrane itself produces essential thermal energy according to biochemical processes.

Next, we turn to the phenomena that appear at the membrane–air interfaces; cf. S_{ma} in Fig. 5.3. Suspended within the cavity of the pit organ the pit membrane divides the pit organ into an inner and an outer chamber [9, 198, 211]; see Fig. 5.2. In other words, the membrane is surrounded by air. Since we deal with *very* small temperatures and temperature differences, air cannot be treated as an ideal insulator. The heat transport through the air can be divided into convection and conduction. The pit membrane is almost perfectly protected from forced convection because of its location within a cavity. Free convection within an enclosure only appears if the Grashof number $Gr := g\beta L^3 (T - T_\infty)/\nu^2 \gg 1000$ [64]. Here $g = 9.81\mathrm{ms}^{-2}$, β is the coefficient of thermal expansion, L is a characteristic length, T describes the temperature, T_∞ the steady-state temperature, and ν denotes the kinematic viscosity. The Grashof number describes the relation between temperature induced lifting force and viscosity. Because of the small dimensions of the pit organ $Gr \ll 1000$ we can neglect free convection as Bakken and Krochmal [9] have already pointed out. With the help of Fourier's law we approximate the amount of heat energy lost by conduction through air Q_a by

$$Q_a = k_a \frac{T - T_p}{d_i} + k_a \frac{T - T_p}{d_o} = k_a \frac{d_i + d_o}{d_i d_o} (T - T_p) \ . \tag{5.3}$$

Here d_i and d_o denote the effective distances from the pit membrane to the wall of the inner (d_i) and outer cavity (d_o), respectively. The thermal conductivity of air is given by $k_a = 0.026\,\mathrm{Wm^{-1}K^{-1}}$, and T_p denotes the temperature of the pit cavity walls.

The remaining transport process, namely the radiation process, is the most important one since it maps all information of the external world onto the membrane. In the first place, radiation heats up the membrane, and this is the focus of this section. The second problem – how information is transmitted and extracted by the pit organs – will be discussed later. We describe the energy transport due to radiation processes Q_r by

$$Q_r = \underbrace{\underbrace{-2\sigma T^4}_{\text{energy losses}}}_{\text{both sides}} + \underbrace{\underbrace{\mu_p \sigma T_e^4}_{\text{external world}} + \underbrace{\mu_o \sigma T_o^4}_{\text{warm object}}}_{\text{front side}} + \underbrace{\underbrace{\sigma T_p^4}_{\text{cavity wall}}}_{\text{back side}} \tag{5.4}$$

where T describes the local temperature of the pit membrane, as always in this section. The mean temperature of the snake, viz. its cavity walls, is described by T_p, and T_e refers to the mean temperature of the external world projected onto the front side of the pit membrane. In the following, we assume $T_e = T_p$ because the snake is a cold-blooded animal. The dimensionless parameters μ_p and μ_o denote the ratio under which a small surface of the pit membrane "sees" the areas with mean temperature T_p and the warm target object with temperature T_o; obviously $\mu_p + \mu_o = 1$. We calculated $\mu_o := A/(2\pi d^2)$ with A, the visible surface of the warm object, and the distance d between membrane and object. We note that the information of the external world is contained in the exact values of μ_p and μ_o for each sensory area on the membrane but for the

heating process; it suffices to assume, for instance, $\mu_o \approx 0.025$ ($d = 0.5\,\mathrm{m}$) or 0.1 ($d = 0.25\,\mathrm{m}$). Furthermore, we assume all surfaces to be black body radiator with the constant σ denoting the Stefan-Boltzmann factor ($\sigma = 5.67 \times 10^{-8}\mathrm{Wm}^{-2}\mathrm{K}^{-4}$) emitting their energy into a half space. We can summarize the terms and rewrite (5.4),

$$\begin{aligned} Q_r &= -2\sigma(T^4 - T_p^4) + \sigma\mu_o(T_o^4 - T_p^4) \\ Q_r &\approx -8\sigma T_p^3(T - T_p) + 4\sigma\mu_o T_p^3(T_o - T_p)\ . \end{aligned} \tag{5.5}$$

It can clearly be seen in (5.5) that the energy input and the energy output depend on *heat contrasts*, namely, the external and the internal heat contrast.

The pit-heat equation and its simplification. We can now combine the different sources and sinks of heat energy we have derived above into a pit heat equation that reads

$$\rho_p C_p \frac{\partial T}{\partial t} = k_p \nabla^2 T + Q_b + Q_a + Q_r\ . \tag{5.6}$$

Given all parameters, the above equation can be solved numerically. Next, however, we follow an analytic approach to derive the characteristics of and dependencies upon the heat stimulus.

As we have seen previously, all of the Q_i terms depend on temperature *differences* involving T. We can reasonably simplify the situation by assuming all parts of the snake and the background of the external world, except the pit membrane and the target object, to have the same mean temperature T_p. This reduces (5.6) to an equation with only two variables, namely, the heat contrast in the *ext*ernal world $\xi_{ext} := T_o - T_p$ and the *int*ernal heat contrast $\xi_{int} := T - T_p$ on the pit membrane. Furthermore, we can identify $\partial\xi_{int}/\partial t = \partial T/\partial t$. In addition, we leave the differential description (5.6) and consider a small piece of the pit membrane located in the center of the membrane with a squared base of $l_p \times l_p$ and a thickness d_p; cf. Fig. 5.3. We take $l_p = 25\,\mu\mathrm{m}$ and $d_p = 15\,\mu\mathrm{m}$. The length l_p corresponds approximately to the size of a single independent receptor area (terminal nerve mass). Figure 5.3 depicts the heat flows through different surfaces of the membrane cuboid. The membrane–air surfaces are $S_{ma} = l_p^2 = 625 \times 10^{-12}\,\mathrm{m}^2$, the membrane–membrane surfaces are $S_{mm} = l_p d_p = 375 \times 10^{-12}\,\mathrm{m}^2$, and the volume of the cuboid is $V_p = l_p^2 d_p = 0.01 \times 10^{-12}\,\mathrm{m}^3$. The simplified pit-heat equation (5.6) for our small cuboid now reads

$$\tau_p \frac{\partial \xi_{int}}{\partial t} = -Q_{out}\xi_{int} + Q_{in}\xi_{ext}\ , \tag{5.7}$$

where $\tau_p = \rho_p C_p V_p (\approx 40 \times 10^{-9}\mathrm{JK}^{-1})$. We can estimate the values of the terms Q_{out} and Q_{in} by considering the corresponding membrane cuboid; cf. Fig. 5.3.

$$\begin{aligned} Q_{out} &= \underbrace{S_{ma}\sigma 8 T_p^3}_{\approx 8\times10^{-9}\mathrm{WK}^{-1}} + \underbrace{S_{ma}k_a\frac{d_i + d_o}{d_i d_o}}_{\approx 15\times10^{-9}\mathrm{WK}^{-1}} + \underbrace{\left(S_{mm}\frac{4k}{\hat{r}_p}\right)}_{0-10\times10^{-6}\mathrm{WK}^{-1}} \\ &\quad + \underbrace{\omega C_b V_p}_{\approx 0.4\times10^{-9}\mathrm{WK}^{-1}} \gtrapprox 25 \times 10^{-9}\mathrm{WK}^{-1} \\ Q_{in} &= \underbrace{S_{ma} r_o \sigma 4 T_p^3}_{\approx 0.1-0.4\times10^{-9}\mathrm{WK}^{-1}} \lessapprox \frac{1}{100} Q_{out} \end{aligned} \tag{5.8}$$

The static pit-heat equation and "optical" contrast. We now set $\partial \xi_{int}/\partial t = 0$ in (5.7) and compute similarly to [9] the maximal value of the internal heat contrast ξ_{max} and the steady-state relation between internal and external heat contrast

$$\xi_{max} = \frac{Q_{in}}{Q_{out}} \xi_{out} \lessapprox \frac{1}{100} \xi_{ext} \quad \text{and} \quad \frac{\xi_{max}}{\xi_{ext}} \lessapprox \frac{1}{100} \ . \tag{5.9}$$

We can see clearly from (5.9) that the quality of the "optical system" pit organ is poor, i.e., it represents external contrasts roughly 100 times less strongly than they are in the external world. In other words, to detect objects with about 1 K temperature contrast, snakes have to measure temperature contrast on the pit membrane in the order of 10 mK. This purely theoretically derived result corresponds to, and explains, many experimental and behavioral findings [9, 33, 63, 211].

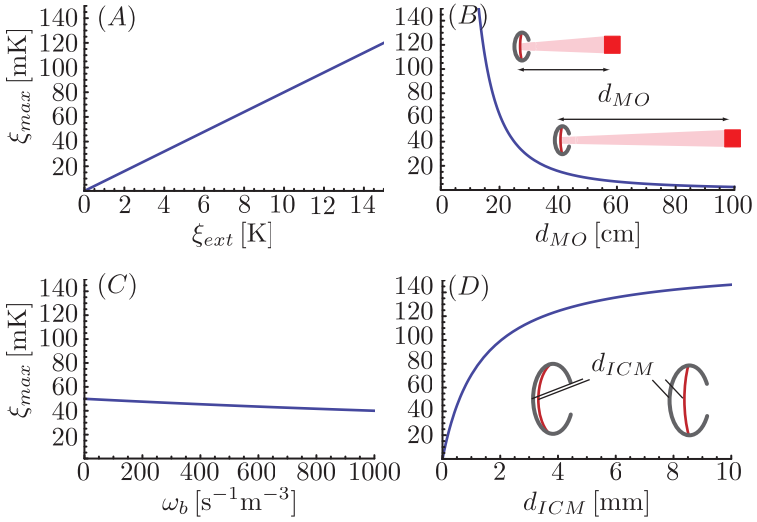

Figure 5.4: *The steady-state internal heat contrast ξ_{max}. Above we have studied the steady-state of the internal heat contrast ξ_{max} of a small slice in the center of the pit membrane (see Fig. 5.3) by means of varying (A) the external heat contrast ξ_{ext}, (B) the distance between membrane and a warm object d_{MO}, (C) the blood perfusion rate of the membrane ω_b, and (D) the distance between the inner cavity wall and the back side of the pit membrane d_{ICM}. Except if stated otherwise, parameters are $Q_{in} = 0.4 \times 10^{-9}\,\mathrm{WK}^{-1}$, $Q_{out} = 50 \times 10^{-9}\,\mathrm{WK}^{-1}$, and $\xi_{ext} = 5\,\mathrm{K}$.*

To get a deeper understanding of the importance of the different effects, we next analyze and discuss the terms in (5.8). Initially, the heat conduction term $\sim \operatorname{grad} T$ is strongly simplified to $\sim 1/\hat{r}_p$. If we use this simplification and take $\hat{r}_p$ to be the radius of the pit membrane heat conduction within the tissue would then clearly dominate all other terms. This approach, however, is too simple. For a numerical calculation $\hat{r}_p$ refers to the distance l_p between the center of the cuboid and the neighboring membrane cuboids. In real situations, the neighboring cuboids will show a similar temperature as the observed one. Hence the temperature differences will not be ξ_{int} but almost zero. Nevertheless, in the case there is an external heat source, the radiated heat energy will heat up the membrane not homogeneously; cf. Fig. 5.10. The warming process generates heat gradients at the edges of the illuminated membrane area that will be erased quickly by the heat conduction term. In absence of steep temperature gradients on the pit membrane, i.e., between two receptor areas in the center of the illuminated membrane area, we can neglect the membrane heat conduction term.

The second major term of Q_{out} in (5.8) describes heat conduction through the air. This term depends on the anatomy of the pit organ. The greater the distance between the inner cavity wall

and the back side of the pit membrane the smaller this term and the better the heat contrast on the membrane [9]; see Fig. 5.4 D.

The next important term of Q_{out} in (5.8) is due to radiation losses. Amemiya and others [4, 38], for instance, have reported a special surface structure of the pits. It seems to be plausible that transmittance and reflection could be minimized through such a surface structure. However, conservation of energy, described by Kirchhoff's law, equals thermal radiation emittance and absorbance. Since they are assumed to be nearly 1 [9] we set both the thermal radiation emittance and the absorbance to 1. Accordingly, the possible variation of these two parameters is very small. Moreover, experimental quantification is missing.

The last term of Q_{out} describes the heat losses through blood flow. If we take the values for brain tissue [72], this term is quite small and unimportant as we have depicted in Fig. 5.4 C. But as reported in many experiments, the membrane shows a great amount of capillaries and it is of very small scale compared to the tissue used for deriving the bio-heat equation [222]. Moreover and unfortunately without giving numbers on the interesting time scale, Goris and colleagues [92, 121, 205] have reported that the blood flow within the capillaries follows the temperature development. They also speculated about a re-cooling mechanism of the TNM by means of the blood flow, which even seems to be controlled locally by the TNM. Accordingly, we either have to modify the equation by means of considering $\omega = \omega(T)$ (see Fig. 5.4 C) or we have to model the effect of blood flow other than Penne's bio-heat equation because the dimensions of our problem are more two-dimensional instead of three-dimensional.

On the other side, Q_{in} and thus ξ_{max} is influenced by the external temperature contrast ξ_{ext} as we have depicted in Fig. 5.4 A. The geometry of the setting, in mathematical terms μ_o, is influenced by the size of the visible surface A and the distance d_{MO} between the membrane and the warm object of interest. The distance d_{MO} limits the maximal heat contrast ξ_{max} and thus the detection range as we have illustrated in Fig. 5.4 B.

The time-development of the pit-heat equation. Another important aspect of (5.7) is the time development of the heat contrast and thus a theoretical approach to the response time of the infrared system. We can rewrite (5.7) to the differential equation

$$\frac{\partial \xi_{int}}{\partial t} = -\frac{1}{\tau}\xi_{int} + \hat{Q}_{in} \ , \tag{5.10}$$

where we have substituted $\tau := \tau_p/Q_{out}$ and $\hat{Q}_{in} := Q_{in}\xi_{ext}/\tau_p$. The general solution of the above equation is given by a superposition of the homogeneous solution ($\hat{Q}_{in} = 0$) and a particular solution. Combined, the solution reads

$$\xi_{int}(t) \quad = \quad \xi_{max}\left[1 - \exp\left(-\frac{t}{\tau}\right)\right] \ , \tag{5.11}$$

with $\xi_{max} := \tau\hat{Q}_{in} = (Q_{in}/Q_{out})\xi_{ext}$ – the result for the steady-state condition (5.9). Under the assumptions and simplifications we have made above, we receive a time-dependend equation that allows us to study the question of what influences the warming process of the pit membrane (5.11). In Fig. 5.5 we have illustrated the dependencies of the warming process upon the influencing variables. As a spin-off we find physical evidence for experimental results [33] that suppose a neuronal response below 100 ms using mK temperature resolution. The signal, however, i.e., the temperature contrast on the membrane, is not constant during that period of time but increasing.

In this section, we have for the first time derived an equation including all aspects of energy in- and output of the pit membrane warming process (5.6). By simplifying this equation to (5.7) we have quantified both the dependencies of the heat contrast (5.9) and the time-dependent warming process (5.11). Thus we are able to identify the important effects and show which data are currently missing. From the experimental point of view, we have arrived not only at a theoretical validation

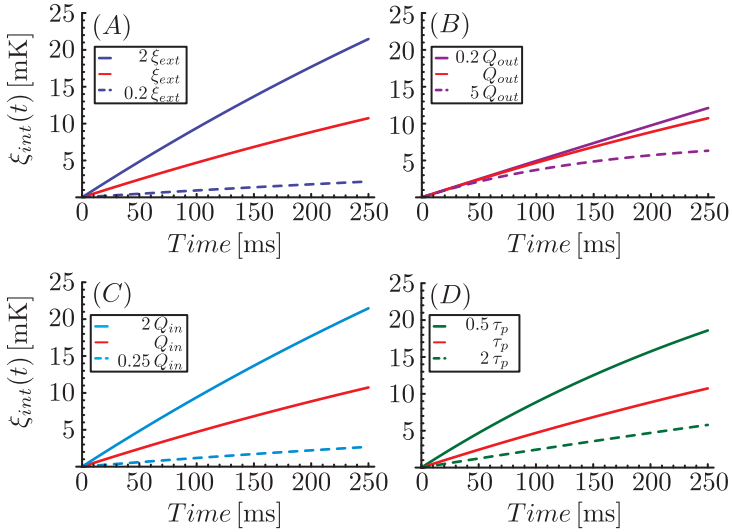

Figure 5.5: *Transient warming process of the pit membrane. Above we have studied the warming of a small slice in the center of the pit membrane (see Fig. 5.3) by means of varying* (A) *the external heat contrast* ξ_{ext}, (B) *the amount of energy lost by the membrane* Q_{out}, (C) *the amount of energy received by the membrane* Q_{in}, *and* (D) *parameters of the pit membrane* τ_p *including for instance the membrane thickness. The basic graph (red) corresponds in all panels to the parameters* $Q_{in} = 0.4 \times 10^{-9}\mathrm{WK}^{-1}$, $Q_{out} = 50 \times 10^{-9}\mathrm{WK}^{-1}$, $\tau_p = 40 \times 10^{-9}\mathrm{JK}^{-1}$, *and* $\xi_{ext} = 5\ K$.

and, more importantly, at a clarification of the possible sensory performances but have also provided mathematical tools for further investigations.

In the next section, we turn to the information content that is transmitted to the pit organ. That is, we will clarify the radiation process in more detail by means of disentangling μ_o for every sensory area on the whole membrane.

5.2.2 Mapping the heat

Having clarified that only the radiation of the external warm objects can cause a reasonable stimulus and that the accuracy of the heat contrast plays a critical role, we have to analyze carefully which physical description fits best to map the external heat contrast to the the pit membrane. The integrated spectral radiance $u(\lambda, T) = \pi I(\lambda, T)$ derived by Max Planck in 1901 [226]

$$u(\lambda, T) = \frac{2hc^2\pi}{\lambda^5} \frac{1}{\exp\frac{hc}{\lambda kT} - 1} \quad \left[\mathrm{Js}^{-1}\mathrm{m}^{-3}\right] \ , \tag{5.12}$$

describes the amount of energy a black body with temperature T emits into the half space per time, per wavelength λ, and per unit surface. As we can see in Fig. 5.6, almost all (99,5%) of the radiated energy from a black body with $T = 303\,\mathrm{K}$ is emitted below 100 μm which is about one tenth of the typical pit organs aperture [47, 137]. We can therefore ignore effects that are due to the wave character of light and instead use geometric optics to describe the physics of our problem, that is, the mapping of the spatial heat distribution. As a side-remark, atmospheric transmittance can be ignored over distances of meters [8]; so we can neglect this effect, too.

Next, we condense the physics of geometrical optics into a formalism that maps the emitted signals from the external environment, for instance, a prey onto the pit membrane of a snake. In the previous subsection, we saw in (5.7) that the external heat contrast ξ_{ext} is an appropriate

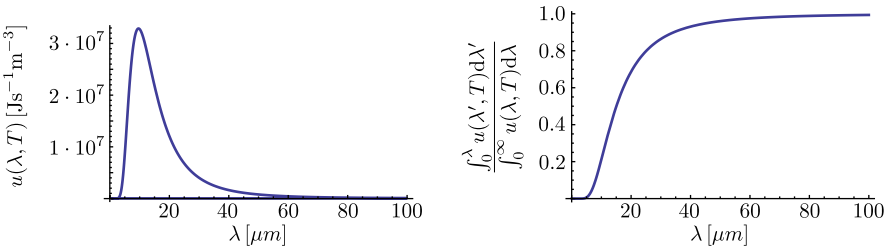

Figure 5.6: *The spectral radiance $u(\lambda, T)$ (left) and the normalized integral of $u(\lambda, T)$ (right) as a function of the wavelength λ at temperatures of $T = 303$ K. We see that the wave character of light can be ignored and we can use geometric optics.*

description for the stimulus because Q_{in} – a constant, mainly geometric factor – times ξ_{ext} is the received energy on the membrane. We now leave the simplified Q_{in}, which is a description for one area of the pit membrane by means of μ_o and μ_p in (5.5), and derive a description for all sensory areas i on the pit membrane accounting for an arbitrary spatially distributed external heat contrast. We therefore discretize the external space in front of the pit into plain surfaces j of equal space S_{ext} with the heat contrast $\xi_{ext;j}$ from now on denoted by ξ_j. Using the general formalism derived in Chap. 2, the received energy r_i due to radiation of a sensory area i on the pit membrane now reads

$$r_i = h_i{}^j \xi_j \ . \tag{5.13}$$

The mapping is mathematically modeled by means of the transfer function $h_i{}^j$ component-by-component where the components of the transfer function are

$$h_i{}^j = \begin{cases} S_{ma} S_{ext} \sigma T_p^3 1/d_{ij}^2 \cos(\varphi), & \text{if } i \text{ is visible from } j \\ 0, & \text{otherwise .} \end{cases} \tag{5.14}$$

To account for the conservation of energy, the emitted energy scales as $1/d_{ij}^2 \cos(\varphi)$ where d_{ij} denotes the distance between the surfaces j in the external world and i on the membrane. We refer to φ as the angle between the normal vectors of the surfaces i and j, and therefore the cosine term compensates for different orientations of the two surfaces i and j. The tensor $h_i{}^j$ is only determined by the geometry of the pit organs and the mean temperature. We can assume both to be constant. So, the tensor projects every heat distribution of the external world in the *same* way onto the pit membrane.

To explicitly calculate the resulting internal heat contrast on the membrane, one has to solve (5.7) for each sensory area i given r_i, or one could linearize (5.7) and just multiply r_i by the appropriate factor derived from (5.7). As we want to study information contents, we will use the pure heat intensity distribution r_i as sensory input in the following; cf. Chap. 1.

For realistically modeling the mapping process of the heat intensity distribution onto the surface of the sensory cells i of the membrane, we also have to consider uncertainties and noise. Two sources of noise can be identified. Noise of the signal and measurement errors of the detector, respectively. We model both by means of adding stochastic variables α_j and χ_i to the input and the detector response

$$r_i = h_i{}^j (\xi_j + \alpha_j) + \chi_i \ . \tag{5.15}$$

The noise of the signal can be caused by means of disturbances of other objects such as grass and fast movements. Since every detection process is blurred by noise, it is reasonable to assume the same here, too; cf. Chap. 2 and Chap. 3.

In summary, we have identified the relevant physics, i.e., geometrical optics, and derived a model that describes the heat intensity distribution across the pit membrane. In the next section we turn to the detection process itself.

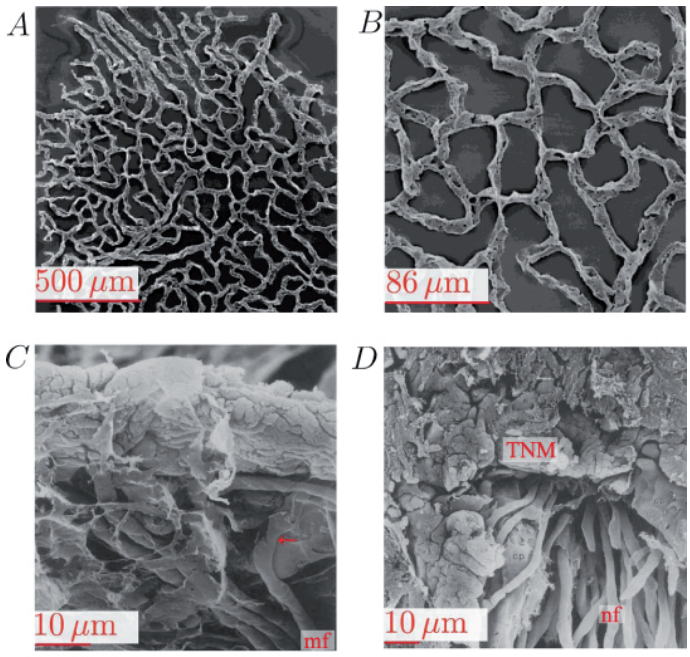

Figure 5.7: *Capillary network and terminal nerve mass (TNM) of the pit membrane; adapted from [5, 6]. The upper panels (A – B) depict scanning electron micrographs showing the capillary vasculature of the pit membrane. The lower panels illustrate (C) the myelinated nerve fibers (mf) losing their sheath (small arrow) and (D) the free nerve endings (nf) of a terminal nerve mass array. A collection of free nerve endings forms a receptor unit and responds to temperature stimuli as modeled in Sec. 5.3. Experimental data on the nervous-controlled capillary network and its effect on the warming process are missing.*

5.3 Sensory encoding

In this section we want to slightly touch upon the question as to how the stimulus is encoded neuronally to extract important ideas for the overall functioning and performance of the infrared detecting system of snakes.

Free nerve endings – a multi-purpuse detector. We know that the location where the stimulus is detected is the pit membrane. More in particular, the sensory responses are generated within the free nerve endings (FNE) of the terminal nerve mass (TNM) of pit membrane. In other words, the free nerve endings encode the temperature of their surrounding (static response) and they respond to temperature changes (dynamic response) [33, 117]. Free nerve endings are also found in other animals such as mice, vampire bats, or cats [24, 118, 243, 244]. In these settings, FNEs [41] are known as (mammalian) cold or warm receptors. Two findings seem to be of universal validity. First, the static response curves (neuronal activity vs. temperature; cf. inlay Fig. 5.8) of the thermo receptors to different levels of constant temperature are similar; compare [117] and [24, 235]. Second, in the animals above – although not as pronounced as in the snake's pit

membrane – there is also a dynamic response to time-dependent thermal stimuli; cf. Fig. 5.8 and e.g. [243, 244]. In passing we note that FNEs seem to be used also in specific water snakes for detecting water movements [43].

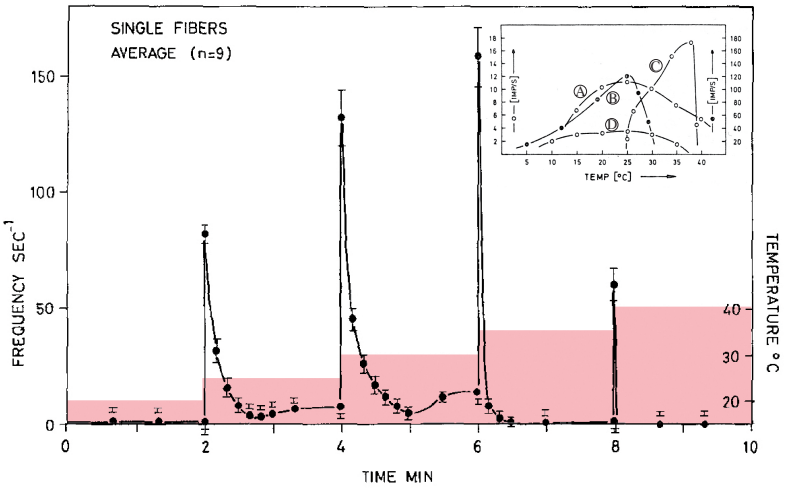

Figure 5.8: *Average impulse frequency of a population ($N = 9$) of warm fibers from the trigeminal branch when applying thermal stimuli to the skin of a Boa Constrictor; adapted from [117] and [25]. During this experiment, Hensel et al. have increased the temperature (red colored area and right y-axis) step-wise ($\Delta T = 5°C$), starting at $T = 20°C$ and ending at $T = 40°C$. As we can see, the maxima of the dynamic response as well as the steady-state values of the static response (see inlay graph C) follow a bell-shaped curve. The inlay graph shows the steady-state response curves to static temperature stimulus of (A) a cat (N.lingualis, bursting), (B) a dogfish, (C) a Boa Constrictor, and (D) a cat (N.infraorbitalis, non-bursting). Not only the resulting detector activity is qualitatively comparable also the underlying detection principle seems to be the same: free nerve endings.*

A model for thermoreception. The static response curves of mammalian thermoreceptors, the so-called "cold fibers", and the underlying biological detection mechanism – free nerve endings – seem to be the same or at least comparable to what we find in the pit membrane; cf. Fig. 5.8. We have therefore adapted a common ionic model for mammalian cold fibers, the Huber-Braun model [26, 27, 28, 124], to study FNE responses. Not surprisingly, we can reproduce the experimentally observed static responses, a bell-shaped frequency vs. temperature curvature. Even more interestingly and for the first time we have been able to reproduce the dynamic responses; see Fig. 5.9. To our knowledge, no other thermo-receptor model has been able to do so.

The Huber-Braun model for thermoreceptors is based on a simplified version of the Hodgkin-Huxley model. The model consists of two sets of ionic conductances that operate at two different voltage levels and time scales. The temperature dependence is modeled by two scaling factors for the rate constants $\rho(T)$ and the maximum of the conductances $\Phi(T)$. The main observable is the membrane potential V that is controlled by five currents according to the membrane equation

$$C_M \frac{\mathrm{d}V}{\mathrm{d}t} = -I_{Na} - I_K - I_{sd} - I_{sr} - I_l \ , \tag{5.16}$$

where C_M is the membrane capacitance. The ionic currents of the right hand side can be grouped into three categories. The first two fast currents $I_{Na/K}$ generate the action potentials. The next

two currents $I_{sd/sr}$ are slow, subthreshold *d*e- or *r*epolarizing, and their interplay causes membrane potential oscillations. Finally, the last current I_l is the passive leak current. All currents are modeled in the following way

$$I_i = \rho g_i a_i (V - V_i) \ , \tag{5.17}$$

with ρ a temperature scaling factor, g_i the maximum conductance, a_i the activation variable, and V_i the reversal potential. The activation variable a_i itself can be described by a differential equation. For a detailed description of the Huber-Braun model and its numerical parameters we refer to elsewhere [26, 27, 28, 124]. In the light of our focus, the Huber-Braun model can reproduce the neuronal responses to static temperature levels, e.g., the bell-shaped activity vs. temperature characteristic; see Fig. 5.8. Yet the model cannot reproduce the observed neuronal responses to changing temperatures.

The Huber-Braun model and its extension. We modify the repolarizing subthreshold current $I_{sr} = \rho g_{sr} a_{sr} f(V, T)$ in (5.17)(corresponds to equation (5) in [27]) by introducing a temperature-gradient-dependent activation variable a_{sr} through

$$a_{sr}(t > t_0) = a_{sr}(t) - \Delta a_{sr} \ \text{ if } \ \operatorname{grad} T \mid_{t=t_0} \neq 0 \ . \tag{5.18}$$

This means that if a temperature step occurs, the activation variable a_{sr} is reduced by $\Delta a_{sr} = 0.5$. It is mathematically unimportant whether we modify a_{sr} or the conductance variable g_{sr}. The result is that I_{sr} is reduced shortly after a change in temperature, and the neuron fires a burst of spikes. This modification corresponds in principle to a transient ionic gating mechanism due to, e.g., thermo-sensitive TRP (transient response protein) channels [23, 176, 189, 212, 275] which results in an additional current occurring at temperature changes. Both a temperature-depending ionic gating mechanism by means of TRP channels [97] and an additional transient current sensible for temperature changes [219] have been found in the snake recently.

The result of our numerical experiment is depicted in Fig. 5.9 and agrees well with biological findings of the snake's neuronal responses; cf. Fig. 5.8. That is, the extended Huber-Braun model can now reproduce the neuronal responses to temperature changes. More precisely, our findings hint towards three important issues.

- First, even if the mechanism of the dynamic answer is much more complex – as we believe – it can be understood at first sight by means of ionic models including an additional thermo-gradient-sensitive current or channel.
- Second, we want to highlight that a temperature change modeled in the way described above synchronizes the different FNEs. We predict that synchronizing effects of FNE responses will enhance reliability, just as in many other neuronal sensory systems [80, 81]. A precise quantification, however, will need a more detailed analysis.
- Third, we can see a prominent difference in the quality of the static and the dynamic answer. The temporal precision and the sudden change in frequency ($\approx 80\,\mathrm{Hz}$) at a temperature step, i.e., the dynamic answer, as compared to the increase ($\approx 5\,\mathrm{Hz}$ for $\Delta T = 5°\mathrm{C}$) in the corresponding static response is very high. In other words, we can define the relative quality Ω of how precisely the FNEs system can encode a temperature change by means of a change of the corresponding neuronal activity from A_T at T to $A_{T+\Delta T}$ at $T + \Delta T$,

$$\Omega = \frac{A_{T+\Delta T} - A_T}{A_T} \ . \tag{5.19}$$

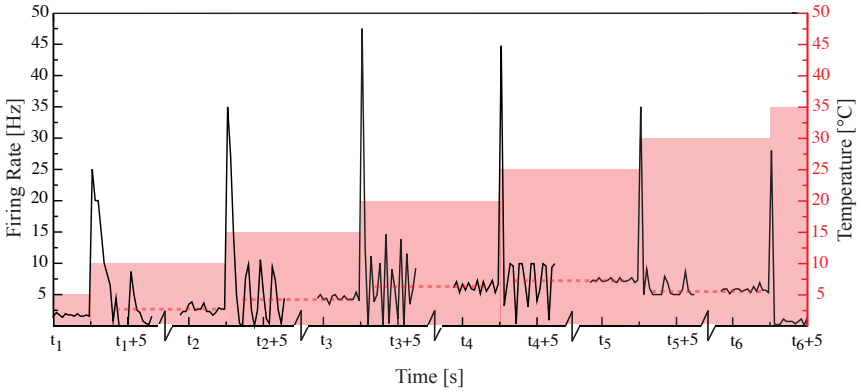

Figure 5.9: *Simulation of FNE responses, the sensory encoding step. The neuronal activity of uncoupled neurons ($N = 26$), the averaged (time intervals are 200 ms) firing rate, is plotted on the (left) y-axis. The neurons are modeled according to the modified Huber–Braun model of thermo-sensitive cold receptors [27, 28, 124] as described in the main text. The neurons sequentially start 50 ms after each other. During our numerical experiment, we have increased the temperature (red colored area and right y-axis) step-wise ($\Delta T = 5^\circ C$), starting at $T = 5^\circ C$ and ending at $T = 30^\circ C$. As we can see, the maxima of the dynamic response as well as the steady-state values of the static response follow a bell-shaped curve comparable to the biological data; cf. Fig. 5.8. The x-axis is presented piece-wise (e.g. from $t = t_1$ till $t = t_1 + 5\,s$) around the temperature steps because the temperature steps synchronize the output of the FNEs and prevent a reasonable averaging. We conclude that the modified (see main text) Huber–Braun model can reproduce the sensory encoding (FNEs responses) of temperature dynamics. That is TRP channels resulting in a sub-threshold currents that are sensitive to thermal gradients can cause the dynamic response.*

We can now see that the dynamic answer encodes temperature changes: temporally precisely and very reliably; $\Omega_{\text{dyn}} \approx 80\,\text{Hz}/20\,\text{Hz} = 4$. In contrast, the static response to absolute temperature is less precise, relatively weak, and even ambiguous on the level of frequency; $\Omega_{\text{sta}} \approx 5\,\text{Hz}/20\,\text{Hz} = 1/4 = \Omega_{\text{dyn}}/16$. This becomes even more evident if we additionally consider the amount of time in which the change of activity ΔA occurs.

Accordingly, we can speculate that the snake does not use the absolute temperature contrast encoded by the static response but instead uses *changes* of the temperature contrast encoded by the dynamic neuronal response. The present paragraphs have shown that the dynamic answer and temporal changes of the heat contrast on the membrane possibly play a more important role than yet believed. Next, we will exploit the results of the present two sections. That is, we combine our knowledge on the physical transmission process and the findings of the sensory detection process to derive a plausible model for the neuronal reconstruction mechanism.

5.4 Neuronal decoding

In this section, we start with the simplest model we can imagine that is able to reconstruct the heat contrast distribution ξ in the external world. This *minimal model* accounts for the limitation of a neuronal system by means of restricted computational power. We extend the minimal model by means of incorporating properties of the sensory encoding process so as to arrive at a plausible and meaningful description of the infrared-vision system of snakes. In addition, we test our models in various conditions with regard to external heat contrast and omnipresent noise. Finally, we show that and predict how feature extraction is possible using both the static (minimal model) and dynamic (extended model) responses of the infrared detectors of the pit membrane.

5.4.1 Minimal model

We have to think of simple models because the snake has limited computational resources in view of the fact that all calculations must be realizable in neuronal hardware and fast. We therefore propose a model that is non-iterative and uses only one computational step that can be done by a feedforward network. The reconstructed external heat contrast distribution $\hat{\xi}_j$ at location j can therefore be written as a linear combination of the measured intensities r_i,

$$\hat{\xi}_j = l_j{}^i r_i \ . \tag{5.20}$$

The optimal reconstruction. To find the optimal coefficients of the reconstruction tensor $l_j{}^i$ we follow the framework of optimal map formation presented in Chap. 2 and define the expectation value E of the quadratic error of our estimate $\hat{\xi}_i$ as

$$E \ = \ \left\langle (\hat{\xi}^i - \xi^i)(\hat{\xi}_i - \xi_i) \right\rangle \ . \tag{5.21}$$

In the above equation we insert the expression $\hat{\xi}_i = l_i{}^j r_j = l_i{}^j \left[h_j{}^k \left(\xi_k + \alpha_k \right) + \chi_j \right]$ from (5.15) and (5.20). Furthermore, we compute the expectation value over all possible input α_i and detector χ_i noises. The noise-free part of the input ξ_i is deterministic. However, we want to *optimally* reconstruct not only one specific situation but the *typical* IR environment. In other words, biologically relevant signals belong to a class of signals that we denote by "typical". Consequently a specific sensory signal is a concrete realization of a typical biologically relevant signal; hence it is a stochastic quantity. We therefore have to deal with the *expectation value* of the signal; for a detailed discussion of why we have to take the expectation value we refer to Sec. 2.2.2.

In the following calculations of $\delta E / \delta h_j{}^i = 0$, we assume all appearing cross correlations to be zero because the noise terms are independent of each other and of the input strengths. The correlation terms can be approximated to first order

$$\langle \chi_i \chi_j \rangle = \sigma_\chi^2 \delta_{ij}, \ \langle \alpha_i \alpha_j \rangle = \sigma_\alpha^2 \delta_{ij}, \ \text{and} \ \langle \xi_i \xi_j \rangle = \sigma_\xi^2 \delta_{ij} \tag{5.22}$$

where δ_{ij} denotes the Kronecker delta. All three equalities define autocorrelations.

The first two terms denote the noise terms on the input and the receptor manifold. Here we refer to noise as a spatially independent additive random variable with zero mean and standard deviation $\sigma_{\chi/\alpha}$ that models the following points. Even if the detection mechanism on the pit membrane is fast, a signal in the external world can be faster. So we can define an integration window within the signal will be averaged to the relevant signal that can be reconstructed. This low-pass filter effect is inherent to any sensory system, no matter if technical or biological. The input noise α_i accounts for this effect and therefore blurs the true signal ξ_i. Strictly speaking, the noise would then be spatially correlated due to a common movement of the real IR object. However, we will

ignore this knowledge and thus deal in the following with a worst-case scenario of knowing nothing. The second noise term describes the detector noise on the pit membrane that may originate from physiological processes in the heat-sensitive cells.

The third term in (5.22) describes the autocorrelation of the input that we have motivated above. In general, it can be written as $\langle \xi_i \xi_j \rangle = (\mu_\xi^2 + \sigma_\xi^2)\delta_{ij}$ with mean μ_ξ and the standard deviation σ_ξ of the input $\boldsymbol{\xi} = (\dots, \xi_i, \dots)$; see Chap. 2. We will only use one term containing the typical information about the IR surrounding, viz., either the mean or the standard deviation. It now depends on the perspective of whether the *relevant* stimulus information is defined as the typical heat contrast or as the variation of a typical heat contrast relating to an object of interest, e.g., prey. The former case refers to the mean μ_ξ and the latter case to the standard deviation σ_ξ. For the sake of uniformity in our mathematical description, we can denote the important quantity without loss of generality by σ_ξ and skip for the time being the above question of what is the relevant stimulus to the pit organ. In doing so, we assume that the correlation between inputs at different location in space $i \neq j$ vanishes. That is, a worst-case ansatz again because knowing the value of the input at one particular point now does not tell us something about its value at other positions. This assumption of course will not hold in reality because neighboring points tend to be from the same object or the same surrounding. Heat transport phenomena eliminate steep heat gradients, and the temperature level of neighboring points will therefore be "similar". Assuming, however, less than possible makes reconstructing the input only more difficult and suffices for a minimal approach.

Having clarified the mapping of the input onto the pit membrane (5.15) and how we can describe the noise and input characteristics (5.22), we next proceed finding the optimal reconstruction tensor ${l_j}^i$ as defined in (5.20). That is, we have to minimize the error (5.21) with respect to the coefficients of the reconstruction tensor, i.e., $\delta E / \delta {h_j}^i = 0$. Defining the dimensionless parameter $\sigma := \sigma_\chi / \sigma_\xi$ and $\tau := \sigma_\alpha / \sigma_\xi$ we can compute the components of the reconstruction tensor ${l_j}^i$ so as to find

$$ {l^i}_k \left[{h_j}^v {h^k}_v (1 + \tau^2) + \sigma^2 \delta_{jk} \right] = {h_j}^i \ . \tag{5.23} $$

Next we take a brief look at the two model parameters τ and σ, their meaning, and their relation to real quantities. The parameter τ denotes the inverse signal-to-noise ratio of the input which in general will be small. The second parameter σ relates noise on the pit membrane to the mean input intensity in the external world. Since σ directly modifies the diagonal of the matrix $({h_j}^v {h^k}_v)$, i.e., it increases its eigenvalues, the value of σ influences the reconstruction tensor significantly.

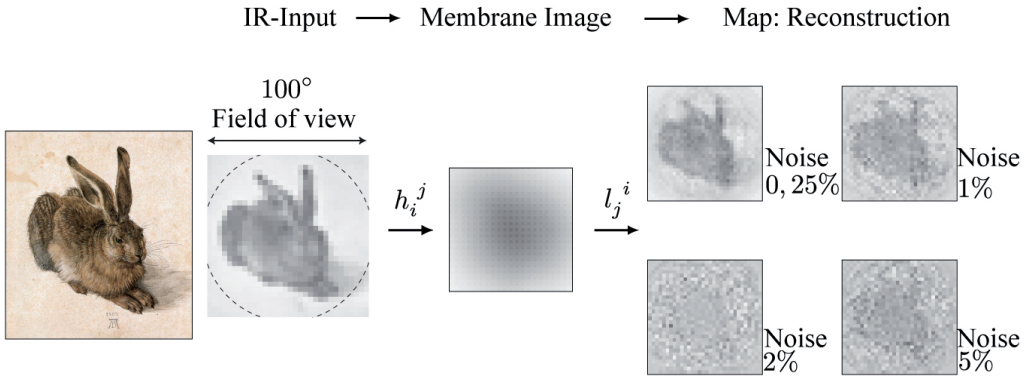

Figure 5.10: *Minimal-model reconstruction performance; figures are adapted from [252]. We discretize Albrecht Dürer's 1502 painting of a hare (left) into a pixelized version with an 8-bit gray scale intended to represent the hare's surface temperature. In a computer model, this IR-source image was presented to an idealized snake pit organ. The dashed circle indicates the organs field of view (100°). Mapping the IR source by means of the transfer function h given by (5.14) onto the pit membrane results in a heavily blurred membrane heat image (middle). In the next step, the model reconstructs the IR input, i.e., the hare from the IR image on the membrane of the model pit organ. The right panels show the reconstructed image by means of the reconstruction matrix l of (5.23) calculated for different noise levels of the membrane sensor cells (0.25%, 1%, 2%, and 5% of the mean value of the membrane intensities). For further model parameters see [252]. Despite the poor optical quality of the pit organ (center image), the original information can still be obtained neuronally from the membrane image.*

The reconstruction performance of the minimal model. To illustrate the abilities of our model, we present an explicit example of how precisely the snake can reconstruct its IR surroundings. We use the famous 1502 drawing of Albrecht Dürer of a hare converted into 8-bit gray levels at a resolution of 32×32; cf. left part of Fig. 5.10. In the next step, we calculate the heat intensity distribution r_i on the membrane using the transfer function $h^k{}_v$ given by (5.14). We have chosen conservative but realistic values that combine a relatively wide organ aperture (radius 0.4 mm) and a moderate number of heat receptors [252]. The calculated membrane heat distribution is shown at the center of Fig. 5.10.

The results of applying the reconstruction algorithm to the membrane image are shown in Fig. 5.10. The quality improvement is spectacular, provided there is not too much detector noise. The most important result is that the information needed to reconstruct the heat panorama is still present although invisible to our eye in the membrane image. The present model has a fairly high *input* noise tolerance. For input noise levels up to 50%, the hare is recognizable. Sensitivity to measurement errors is larger. In our calculations, one pixel of the reconstructed image corresponds to about 3°. For detector noise levels up to about 1% of the membrane heat intensity, a good reconstruction is possible, meaning that the edge of the hare may be determined with about one pixel accuracy. At detector noise levels beyond about 1%, the image is not so easily recognizable. Choosing a higher σ can compensate the membrane noise until a certain noise-level revealing at least the presence of an object; cf. Fig. 5.10. The drawback of choosing high σ values is that details, viz., variations of the heat image are "treated" as noise and will be suppressed. In addition, in the case of a simple and clear input – say, a small warm animal in front of a cold background – reconstruction may still be possible at higher detector noise levels as we will see later.

Predicted neuronal characteristics. Our model is not only able to predict the sensory performance of the snake's IR system but can also shed light onto the neuronal decoding mechanism used to sharpen the membrane image. The neuronal implementation of the model is fairly straightfor-

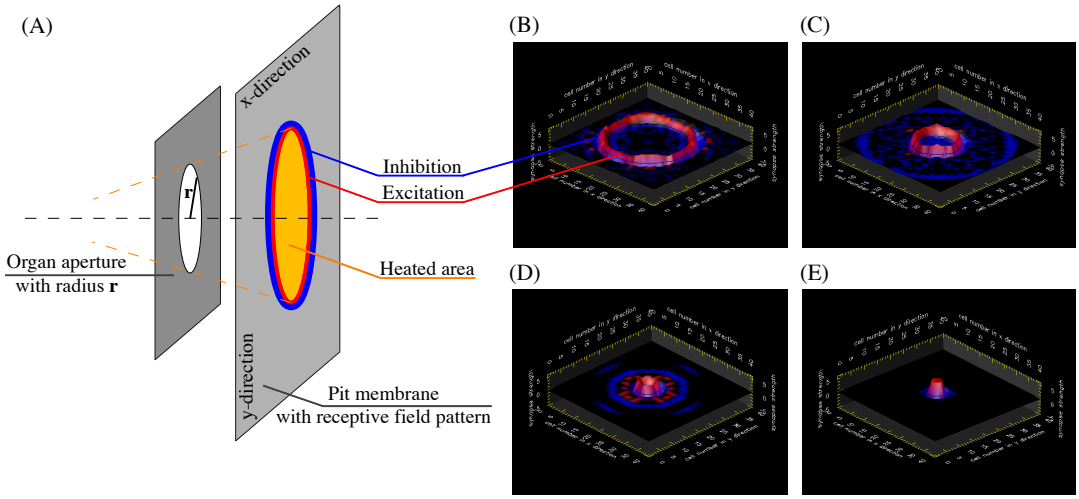

Figure 5.11: *Predicted receptive fields. As explained in the main text, the strength of the individual connections corresponds to the entry* ${l_j}^i$*. The neurons in the topological map receive excitatory (positive, red) as well as inhibitory (negative, blue) input from the membrane receptors. The spatial distribution of these excitatory and inhibitory regions on the membrane is called the receptive field of a corresponding map neuron. As we can see in panel* $(A-E)$ *the receptive field for a single map neuron consists of a excitatory ring located at the edges of the heat profile surrounded by an inhibitory ring. This construction constitutes an edge detector. We can shrink the radius* r *of the organ aperture (*B*:*$0.4\,mm$*,* C*:*$0.2\,mm$*,* D*:*$0.1\,mm$*, and* E*:*$0.05\,mm$ *) until we receive a pin-hole-like geometry that is comparable to the setup in optical systems employing a lens. In both sensory systems, the IR sense and the visual system, we find the same mechanisms that follow the same mathematical principles of optimality. Reconstruction parameter* $\sigma = 3$*, distance between aperture and object is* $1\,m$*, and distance between aperture and membrane is* $0.8\,mm$*.*

ward. The reconstruction tensor ${l_j}^i$ corresponds to a network of synaptic connections between the membrane IR detectors and neurons building a map in, for instance, the optic tectum [109, 210, 211]. The strength of the individual connections would just be the value of the corresponding entry ${l_j}^i$, based on a rate coding. The neurons in the topological map receive excitatory (positive) as well as inhibitory (negative) input from the membrane receptors. The spatial distribution of these excitatory and inhibitory regions on the membrane is called the receptive field of a corresponding map neuron; cf. Chap. 1. It can be calculated by our model as illustrated in Fig. 5.11. Ring-like structures arise that "detect" the edges of projected images corresponding to the position that the map neuron encodes. A similar system (called lateral inhibition) is found in mammalian neurons receiving input from the retina; cf. Chap. 2. Although the excitation-inhibition pattern there is disc-shaped ("small" aperture) rather than ring-like (large aperture) the excitation-inhibition pattern naturally follows the same principle of mathematical optimality as the receptive field we have derived here.

In summary, the snake does need accurate detectors to get a reasonable reconstruction quality. Indeed, the experimentally determined membrane precision is high. The pit membrane responds to temperature changes down to about 1 mK [33, 63]. As we discussed in part 5.3, the response of the detectors (FNEs) to temperature changes is much more prominent and thus reliable than to static temperature values. We now include this finding and adapt the minimal model derived above accordingly by means of incorporating the dynamic detector responses.

5.4.2 The extended model

Here we extend the minimal model with respect to the major finding of Sec. 5.3 concerning the neuronal characteristics of sensory encoding. The major result is that temperature changes could be encoded neuronally more reliably, i.e., with a much higher temporal precision, and more clearly than the static membrane temperature (5.19). We next discuss this finding in the light of neuronal decoding and adapt our minimal model derived above accordingly.

Detector characteristics in the light of neuronal decoding. The minimal model clearly demonstrates that the information transmitted to the pit membrane is still there and that the information can be decoded neuronally. The minimal model, however, faces a big drawback. It needs highly accurate steady-state temperature measurements on the membrane and the literature lacks an explanation of where the necessary precision could originate from. To give some numbers, we refer to the findings of Sec. 5.2.1 where we could see that the temperature contrast is in the range of mK within the first 100 ms of a stimulus onset. In addition, the measurement errors need to be less than 10% to reconstruct a heat panorama as we have seen previously; cf. Fig. 5.10. That is, the FNEs have to detect absolute temperature contrasts of about 1 mK up to 20 mK with at least 0.1 mK accuracy. Given a noiseless detection process, the detectors would then need 200 sampling intervals to encode the relevant temperature values. Looking at the neuronal responses found in the ascending nerves of the pit membrane (Fig. 5.8) we find a frequency range of say 20 Hz (see Fig. 5.8) to encode the range of $\Delta\xi = 20\,\mathrm{K}$; even if we consider the temporal pattern of the corresponding spike trains a high accuracy in representing the stimulus seems very challenging for the sensory system and almost unbelievable in case of a noisy detection process. We therefore now focus on the other neuronal response, the *dynamic* answer of the TNMs to temperature changes.

The extension of the minimal model. We can think of two different origins of temperature changes of the membrane. The first possibility is the sudden (dis)appearance of a heat-emitting object causing the rise or fall of the membrane temperature. The second and most natural is due to movements. Warm objects in the external world translating relatively to the pit membrane lead to temporally changing temperatures at the sensory cells on the pit membrane. If such a warm object moves, temperature changes appear at the edges of the resulting temperature profile on the membrane. In encoding temperature changes, the FNEs do nothing but effectively *filtering* the *edges* of the pit membrane intensity distribution r_i.

We can calculate an edge-filtered version f_i (see Fig. 5.12) of the membrane intensity distribution r_j by means of linear filter algorithms such as the Laplace filter matrix ${t_i}^j$ [91],

$$f_i = {t_i}^j r_j = {t_i}^j ({h_j}^v \xi_v) = {\hat{h}_i}^v \xi_v \ . \tag{5.24}$$

The above equation defines similarly to (5.14) a transfer function ${\hat{h}_i}^v$ that now maps the external heat contrast ξ_v onto an edge-filtered intensity profile f_i on the membrane. Since the edge-filtered image approximates the temporal derivative of the intensity profile on the membrane, we have extended our model of the snake infrared sense in a simple but reasonable way that now includes essential properties of the detection mechanism. Using the same methodology as described in (5.20)

– (5.23), we can calculate correspondingly to (5.23) an optimal reconstruction tensor $\hat{l}_i{}^j$ using the new transfer function $\hat{h}_i{}^v$ of(5.24)

$$\hat{l}^i{}_k \left[\hat{h}_j{}^v \hat{h}^k{}_v (1+\tau^2) + \sigma^2 \delta_{jk} \right] = \hat{h}_j{}^i \ . \tag{5.25}$$

Arriving at the above equation we have implemented a very plausible second step of sensory processing (see Chap. 1), i.e., we have included the pit membrane detector characteristics. Even this seems to be a small step – as we have left out the derivation steps analogue to (5.20) – (5.23) for the minimal model – it results in a big advantage as we will see in a minute.

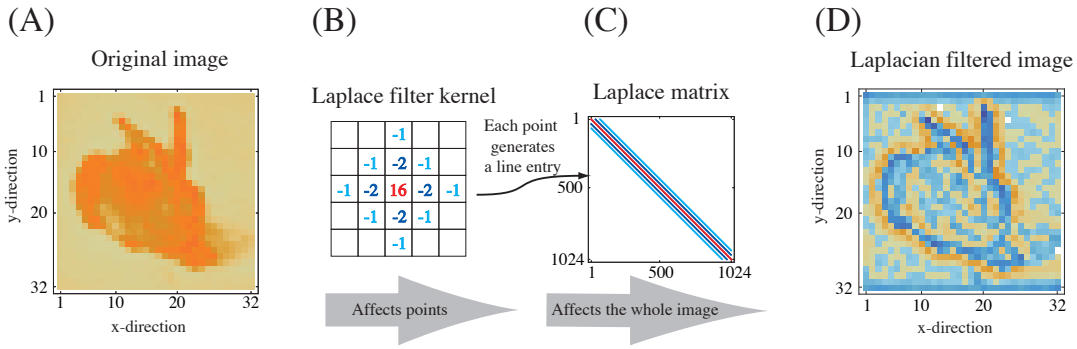

Figure 5.12: *The Laplace filter. As explained in the main text, we use the Laplacian filter that can "detect" the edges of the membrane intensity distribution to extend the minimal model. The Laplacian filter mechanism works in the following way: (B) The filter kernel sequentially takes every point of the original image, scales it, and subtracts the weighted value of its neighboring points. Depending on the Laplace filter kernel at hand, it can affect the area of* 3×3*, as in the figure* 5×5*, or more pixels centered around the evaluated point. (C) To use this method in the context of the minimal model we have to write the filter kernel accordingly to the indexing of the original image into the lines of a matrix* $t^i{}_j$ *for each image point. The Laplace matrix* $t^i{}_j$ *now maps the original image (A) onto the Laplacian filtered image (B) by means of one matrix multiplication.*

Advantages of the extended model. The extended model features the following conceptual advantages that arise from using the dynamic instead of the static response. First, the change in neuronal activity $\Delta A_{\mathrm{dyn}} \approx 80\,\mathrm{Hz}$ of the dynamic response caused by an increasing temperature contrast ξ is about 16 times larger than $\Delta A_{\mathrm{sta}} \approx 5\,\mathrm{Hz}$ of the static response caused by an increasing temperature contrast; cf. Sec. 5.3. Due to this fact, using the dynamic response allows the snake to employ 16 times more sampling points to decode the signal. That means, the *detection accuracy* improves.

Second, not only the absolute change of neuronal activity ΔA of the free nerve endings but also Ω, the relative change of frequency compared to the actual frequency of the detector, is increased (5.19). That is, the quality and reliability of the *neuronal response* even more than improves.

Third, a changing IR environment – moving objects – does not hinder but actually enable the detection process. In case of using the absolute temperature contrast ξ of the IR environment (minimal model), ξ has to be static. That is, in case of the minimal model, movements are treated as noise. In contrast, the extended model is based on the dynamic response, i.e., detector responses to changes of the IR environment, $\mathrm{d}\xi/\mathrm{d}t$. Thus, movements are not related to noise but constitute the signal to the model. This decreases the noise level in the signal *mapping process*. Of course, there is also a temporal "integration window" during which a signal has to be constant. Since the

dynamic response, i.e., the sudden (about 100 ms) increase in frequency of the neuronal-detector activity, is temporally quite precise and fast, the integration window is narrow, and thus the detectors are precise.

Fourth, we remember two findings. First, in Sec. 5.2.1 we have derived that the temperature contrast ξ on the membrane has not reached its steady-state value within the relevant reaction time of the snake. The signal ξ can thus never be constant and self-averaging, respectively. Second, the reconstruction mechanism needs very accurate sensory data. A signal that is constant during a certain period of time can be determined more precisely by means of making more measurements and averaging the results. In the case of the minimal model, ξ is not constant and thus cannot be improved by statistical methods. On the contrary, the time derivative of the temperature contrast $\mathrm{d}\xi/\mathrm{d}t$ is almost static within the relevant time frame. A static physical *signal* provides averaging and thus reducing additive detector noise. We can at least improve the measurement by a factor $\sim \sqrt{t}$ with more measurements. That is, the accuracy of how we can detect and encode the signal is higher in case of the extended model.

In summary, we have derived an extension of the minimal model that connects smoothly to a plausible understanding of the dynamic response pattern found in the FNEs; see Fig. 5.8. To clarify the overall sensory performance of the extended model, we next have to quantify its dependency upon noise. Since the underlying mathematics of the extended is akin to a Laplacian-filtered version of the membrane image one cannot expect a good noise robustness beforehand. This fact could cancel out the above advantages. In the following section, we therefore compare the ability of the two models at hand to reconstruct the IR environment and to extract relevant features such as the object position in various settings.

5.4.3 Feature extraction

After having seen above that the IR-sensory system of snakes can reconstruct the heat environment, we now address the question of how precisely snakes, that is our two models, can localize warm objects and which quantities influence the sensory performance.

The numerical "experiment" setup. In the following numerical experiments we present a warm object within a non-uniform and non-zero temperature background to an artificial pit organ. We assume an IR environment with 25°C mean temperature. As we derived in Sec. 5.2.1 and 5.3, the external heat contrast ξ_{ext} and the corresponding time derivative $\mathrm{d}\xi_{ext}/\mathrm{d}t$, respectively, are the appropriate stimuli for the pit organ. We add a Gaussian-distributed temperature background contrast with Γ°C mean and Γ°C standard deviation to study the effects of external noise and signal contrast. The warm object shows a Gaussian-shaped temperature profile with maximal heat contrast ξ_{max} (in respect to 25° C) and standard deviation 6.25 cm that is located at $\boldsymbol{x} = (x, y)$ with $x, y \in [-1\,\mathrm{m}, 1\,\mathrm{m}]$, 1 m in front of the pit organ; cf. Fig. 5.13. For every run we randomly position the object center at $\boldsymbol{x}$, add a noisy background, map the whole infrared scene onto the pit membrane and reconstruct the environment. We use simple shapes because we now can extract the estimated position $\hat{\boldsymbol{x}} = (\hat{x}, \hat{y})$ easily by means of taking the average of the 50% warmest locations of the reconstructed environment and compare the estimated position $\hat{\boldsymbol{x}}$ to the real position $\boldsymbol{x}$; cf. Fig. 5.14. During the numerical experiments we monitor the dependence of the localization performance upon the background characteristics, namely, Γ and the detector noise χ; see Fig. 5.15. We have to note that the focus of this work is on the IR sensory system and not on the most cognitive algorithm for object recognition given sensory data, i.e., a post-processing step. We have already shown [251] how to estimate both, position and shape of objects, given sensory data of a natural sensory system. So, our findings present a bottom-line performance of the snake's sensory abilities that can be better than derived here – in the aspect of object recognition.

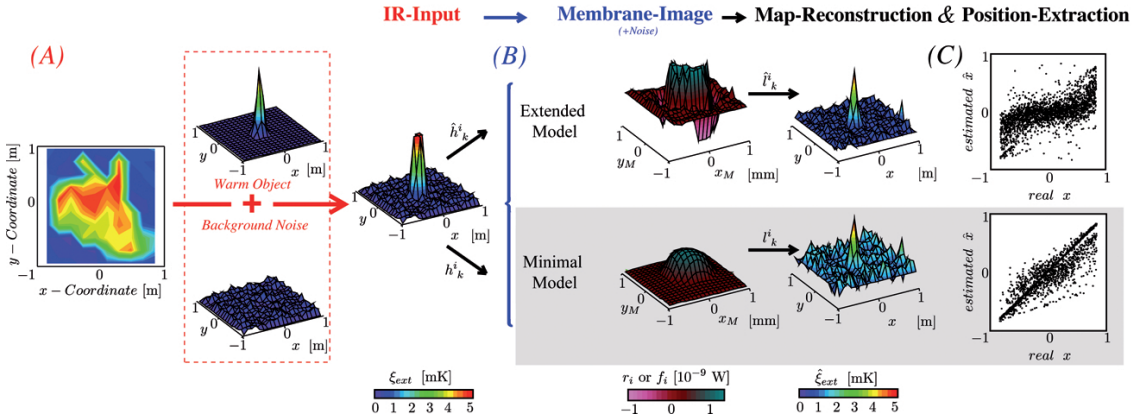

Figure 5.13: *Setup for localization performance experiments.(A) We simplify the warm object to be a two-dimensional Gaussian-shaped profile with standard deviation 6.25 cm, maximal amplitude of $\xi_{max} = 5^\circ C$, and position $\boldsymbol{x}$, and add a Gaussian random background temperature contrast distribution with mean and standard deviation of $\Gamma^\circ C$. (B) This parameterized IR environment is mapped (minimal model: $h^i{}_k$ or extended model: $\hat{h}^i{}_k$) onto the pit membrane, and detector noise χ_i is added; cf. (5.15). Now it depends on which model we consider. Either the intensity distribution r_i and the corresponding reconstruction tensor $l^i{}_k$ (minimal model) or the edges of the intensity distribution f_i and $\hat{l}^i{}_k$ (extended model) are used to reconstruct the IR environment. In the latter case the edge-filtered membrane image f_i approximates the time-derivative $f_i \approx dr_i/dt$ due to object movements; see Sec. 5.4.2. (C) In the final step, we extract the estimated position $\hat{\boldsymbol{x}}$ of the warm object from the reconstructed IR image and compare it to the real position $\boldsymbol{x}$ during n independent runs: see Fig. 5.14. To quantify how the parameters affect the sensory performance, we present the root-mean-square error between $\boldsymbol{x} = (x, y)$ and $\hat{\boldsymbol{x}} = (\hat{x}, \hat{y})$ in Fig. 5.15.*

Sensory performance "experiments". Within the setup outlined in the last paragraph, we vary the signal-to-nosie (SNR) ratio of the external heat contrast ξ by means of increasing Γ given a constant $\xi_{max} = 5\,\mathrm{K}$. In addition, we study how the detector noise χ_i on the pit membrane as given by (5.15) affects sensory performance. The results (Fig. 5.14) illustrate how precisely the position can be estimated from the IR environment. In case of moderate noise both the minimal model (left panels in Fig. 5.14) and the extended model (right panels in Fig. 5.14) can reconstruct the position of the warm object quite precisely.

In this section, the standard deviation of the membrane noise σ_χ is related to the maximal value of either the intensity distribution (minimal model: $\sigma_\chi \sim r_{\max}$) or the edges of the intensity distribution (extended model: $\sigma_\chi \sim f_{\max}$) on the membrane. The previous definition of membrane noise presents the standard deviation σ_χ of the additive Gaussian noise variable χ_i in percentages of the mean value of the membrane intensities: $\sigma_\chi \sim 1/n \sum_i^n r_i$; see Fig. 5.10. Since a higher temperature background – viz., increasing Γ – also increases the mean value of intensities r_i on membrane, the two different kinds of noise would not be independent in case of the previous definition. So we redefine the noise level to be now related to the maximal value of the membrane intensity r_i or the edges of the membrane intensity f_i. This re-definition almost completely decouples χ and Γ .

Furthermore, we see in Fig. 5.14 that in both models the capability to localize a warm object is still available beyond the membrane noise level of 1% but decreases fast. Independently of each other both the external temperature background Γ and the membrane detector noise levels χ, play an important role in predicting the sensory performance and in evaluating experimental results. Surprisingly, the estimated positions $\hat{x}$ do not only become more vague but also clustered around the middle of the pit aperture due to increasing noise levels. The latter effect is more prominent in

case of the minimal model; cf. left panels in Fig. 5.14. The extended model employing the dynamic responses to temperature changes in principle access sensory data with higher accuracy as compared to the minimal model. On the contrary, since Laplacian filters typically enhance image noise [91] we could expect that the extended model that is based on a Laplacian filter is less tolerant to noise as compared to the minimal model. If this is the case, the less noise tolerant decoding algorithm would cancel out the advantages gained through a more accurate sensory encoding mechanism. At first glance, the extended model, however, is as robust to noise as the minimal model; see Fig. 5.14.

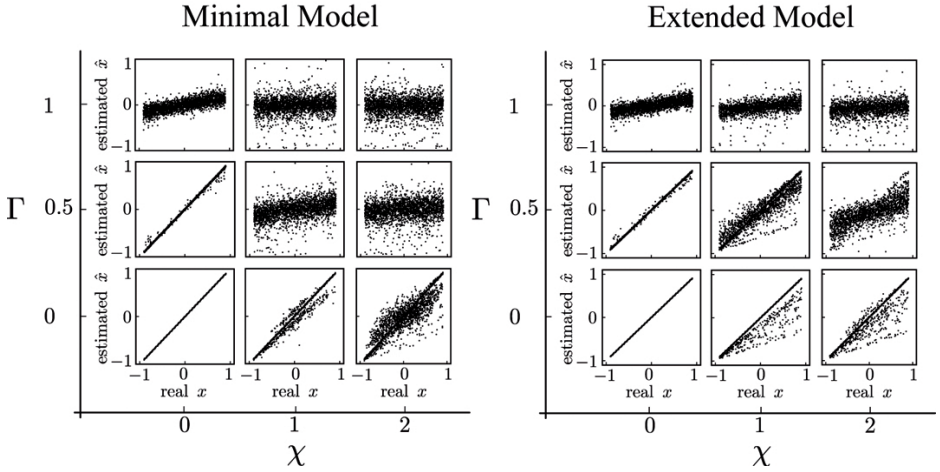

Figure 5.14: *Localization performance experiments. The two collections of nine panels illustrate the effects of the two parameters χ_i, the membrane noise (x-axis) and Γ, parameterizing the external IR environment (y-axis) for both the minimal model (left) and the extended model (right). In each of the 18 panels we have depicted real positions x (panel x-axis) versus estimated positions $\hat{x}$ (panel y-axis) of 2500 trails corresponding to the setting described in Fig. 5.13. We see clearly that both models can reconstruct the position of the infrared object very precisely in case of moderate noise levels. In both models, localizing a warm object is possible beyond the membrane noise level χ of 1% but decreases fast. Furthermore, the level of contrast Γ within the external temperature background plays an important role in predicting the sensory performance as the localization gets worse with increasing Γ (see y-axis). Interestingly, not only does the estimation accuracy decrease but for high Γ the estimated positions tend to be more clustered around the center as well. Above all, the extended model robustness to noise seems at least to be comparable with the minimal model – it is even better in the parameter range depicted above; but see also Fig. 5.15.*

To mathematically quantify the robustness with respect to both temperature background and detector noise – the performance of the two models – we define the root-mean-square error E_{rms} between real x and estimated coordinate $\hat{x}$

$$E_{rms} = \sqrt{\sum_n (x - \hat{x})^2 / n} \tag{5.26}$$

through n independent runs. In Fig. 5.15 we see that the minimal model shows a higher noise robustness as compared to the extended model. The noise tolerance of the extended model, however, is good enough to allow, in combination with the advantages of the sensory encoding step, i.e., the higher quality of the neuronal response (5.19), the conclusion that the extended model is the model of choice.

Consequently, we again pose the question of what is the stimulus to the infrared system since we have derived the superiority of the extended model employing the dynamic response. In Sec. 5.2.1

we have derived that the stimulus has to be the external heat contrast ξ_{ext} according to the thermodynamics of the system instead of the external absolute temperature. The findings of sensory encoding on the level neuronal compartments in Sec. 5.3 have then hinted towards the time derivative of the external heat contrast $\mathrm{d}\xi_{ext}/\mathrm{d}t$ to be the advantageous and appropriate input for the pit organ instead of ξ_{ext} or the absolute temperature . The numerical performance experiments have now underlined this assumption. Although in the literature the absolute external temperature is referred to being the stimulus to the sensory system we can also find experimental statements that snakes respond best to *moving* instead of stationary warm objects [94]. As we have discussed above, the latter statements correspond to $\mathrm{d}\xi_{ext}/\mathrm{d}t$ and hence provide an experimental evidence for our theoretically derived finding.

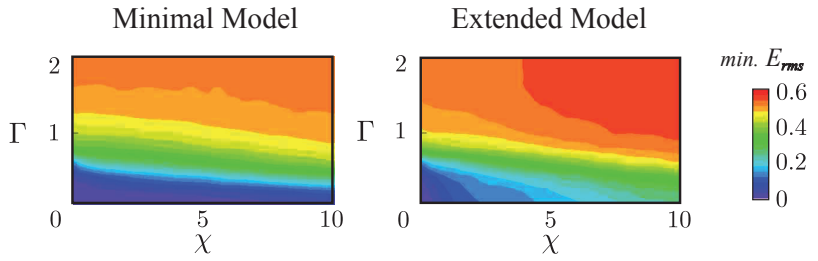

Figure 5.15: *Quantification of the models' localization performance. The plots show E_{rms} for the minimal (left) and the extended model (right) in dependence upon the noise parameters Γ (x-axis) and χ (y-axis) using an optimal σ. For each set of noise parameters Γ and χ, we have varied σ (2500 runs for each combination) to find the optimal σ that minimizes the localization error E_{rms}. The minimal model employs a larger parameter space in which localization of a warm object is quite well possible, viz., in which E_{rms} is small. The noise tolerance of the extended model is at least good enough to not cancel out the advantages of using the dynamic response such as the higher quality of encoding a temperature stimulus.*

In summary, we have shown mathematically that the IR signals originating from an external heat distribution and getting projected onto the pit membrane still suffice to reconstruct both, an image of the external heat distribution and the position of warm objects. Furthermore, we have clearly demonstrated that including characteristics of the detection process into the reconstruction model, viz., focussing on the *dynamic* response, enables a reasonable understanding of the remarkable performance of the infrared system of snakes.

5.5 Discussion and Outlook

In this chapter we have analyzed the infrared system of snakes by studying the three steps of sensory processing that we had introduced in Sec. 1.1. First, using the eye glasses of physics we have looked at the membrane heating process and corresponding experimental findings of the pit organs so as to identify the relevant equations describing the first step of sensory processing, viz., the signal mapping. In doing so, we have quantified how the heating process depends on external quantities – such as distance or temperature contrast of a warm object, anatomical data, and time. Second, we have presented an ionic model that combines recent anatomical findings and present models of heat receptors. That is, by taking the effects of transient response protein (TRP) channels and time-dependent sub-threshold currents into account we were able to reproduce the neuronal responses of the free nerve endings (FNEs) not only to static temperatures but also to temperature changes. Our findings of sensory encoding, furthermore hint towards synchronization effects of the FNEs that could be responsible for the outstanding temperature sensitivity of this sensory architecture. Third, for the last step of sensory processing, i.e., neuronal decoding, we have provided both a minimal and an extended model that are able to reconstruct the external heat distribution, i.e., the infrared image of the environment, given only the heavily blurred membrane image. A comparison of both models with their response to detector and external noise has revealed the abilities and limitations of this IR sense. Our findings both fit well into today's picture and extend the understanding of the involved processes.

Moreover, we have posed and answered the question of what is the true stimulus to this sensory system. In the first section of this chapter we physically derived that it is not the absolute temperature but the temperature *contrast* of the external world that effectively affects the membrane temperature. In addition, the equations have shown that the temperature contrast on the membrane does not reach its steady-state value during the relevant response time. Since the time derivative of the temperature contrast on the membrane stays almost constant, we have started to consider the time derivative of the external heat contrast to be *the* stimulus to the IR system. In this chapter we have found more evidence to support this thesis. The superior quality of the sensory response to temperature changes as compared to static temperatures is one of them. To test the assertion, we have extended the minimal ansatz modeling the reconstruction performance. In summary, we have not been able to find or derive reasonable arguments against but only in favor for the claim that the temporal derivative of the external heat contrast is the appropriate stimulus for the snakes' IR system.

Within this work, not only are new insights presented but also new questions for further research have arisen. In the context of signal mapping, experimentally gained values of the involved quantities such as the blood flow kinematics or the heat conductance of the membrane tissue are needed. Systematic and quantitative measurements of the warming process dynamics would allow for understanding the functions of the individual phenomena such as blood flow and surface structure one-by-one and in interplay with one another.

On the level of sensory encoding models incorporating experimental data of TRP channel dynamics are needed to further enlighten the question of what makes the pit membrane so sensitive. A promising issue could be synchronization effects of the FNEs.

In the context of reconstruction models we see two future approaches. On the one hand, theoretical models should become more biological. That is, they should be based on neuronal data – if present – to describe sensory performance on the level of spiking neurons. The extended model based on characteristics of the sensory encoding process, i.e., the dynamic response, provides a good starting point. On the other hand, nobody has ever addressed the question of how the accuracy of the IR sense and thus the topography of the corresponding neuronal synaptic connections, arises. There is a clear lack of experimental data on this issue. From the theoretical point of view, one can imagine instructive learning as found in the auditory pathway, or some kind of auto-focus algorithm

that uses image stochastics to improve itself, to cause the development of this sensory system. In the next chapter we will study a model, called integrated Multimodal Teaching, that could be a reasonable explanation of how the infrared map and the visual map in the optic tectum get aligned.

The findings derived in this chapter are not purely of biological interest. As our mathematical models show a good performance without employing complex calculations, we have already started to set up a project to implement our findings into hardware; cf. Fig. 5.16. The overall goals are to evaluate our models in real settings, to isolate individual aspects of the models and the involved physics that are not possible to study in a biological context, and to develop and test learning algorithms. On the hardware side the goals are to clarify the possibility of implementing bio-inspired cognitive circuits in silicon to improve optical IR instruments. First steps are promising, and we are looking forward to this interdisciplinary, bionic approach.

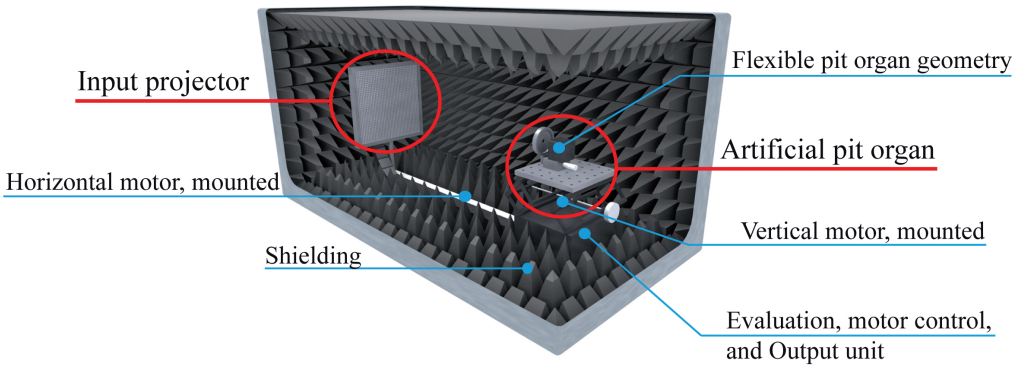

Figure 5.16: *Bionic experiment setup. As described in the main text, we are going to set up an artificial pit organ experiment in cooperation with Erik Jung and Alexander Hilgarth from the Fraunhofer IZM, Berlin. The input projector maps the signal onto the detectors of the artificial pit organ. We realize a fully flexible motor-controlled geometry (flexible position of the pit organ with respect to the input, variable organ aperture, adjustable distance between aperture and detector) to evaluate the predictions we have derived in this work. Moreover, we will develop and implement "reality-feedbacked" learning algorithms and miniaturize this setup in the progress of the project.*

A final remark regarding the sensory performance of the infrared system of snakes: It seems that not one big development but rather the complex interplay of several specific steps such as a very thin and isolated membrane, a dense capillary network, FNEs employing temperature sensible ionic channels that are organized to a compact network called terminal nerve mass, neuronal decoding mechanisms, etc. have led to one of the most specialized and remarkable sensory systems found in nature. We were able to answer some but not all of the questions connected with these steps.

In the next chapter, we focus on the general question of how a precise alignment of different sensory maps can develop *intrinsically*. The alignment of different maps is fundamental for multimodal integration that is also found in the snake's optic tectum. There, the infrared map and the visual map interact faithfully, for instance, by means of performing "AND" (e.g. visible and warm) and "OR" (e.g. visible but not warm) operations [109, 210, 211].

Millions saw the apple fall, but Newton was the one who asked why.

Bernard Mannes Baruch

6. Integrated Multimodal Teaching

In this chapter we come back to the topic of learning and address global structures underlying the evolvement of sensory systems and their precise interaction. We introduce the concept of integrated Multimodal Teaching *(iMT) that is based on supervised spike-timing-dependent plasticity (sSTDP) and multimodal integration. We show that a multimodal teacher can ensure proper* intrinsic *map formation and alignment. The dominance of vision in guiding map alignment is demonstrated to be a natural consequence of the exceptional precision of the visual system.*

6.1 Introduction

Our ability to act in a dynamic and complex environment relies first and foremost on the *precise* interaction of the different sensory modalities that form a congruent perception of the external world. This precise interaction of the different modalities is not present at birth but has to be *learned* during the first period of sensory experience in life [271, 280].

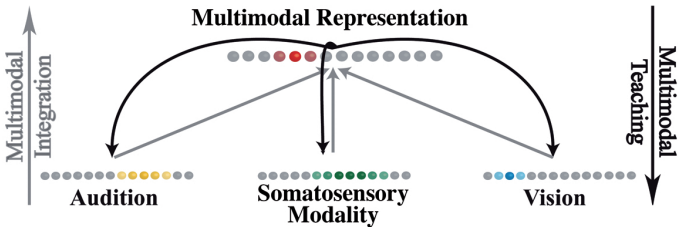

Figure 6.1: *The basic idea of integrated multimodal teaching (iMT). All modalities are integrated into an integrated multimodal map (gray arrows). The activity of the multimodal map then induces a teaching signal (black arrows) that leads to an adaptation of the unimodal maps. We explain the detailed model setup employed in the present chapter in Sec. 6.2.2.*

The process of sensory calibration is commonly seen as vision-guided; cf. King's review [148]. Several remarkable experiments show that destruction or disturbance of the visual pathway leads to disorganized and abnormal sensory maps in non-visual modalities. These findings, obtained in hamster [200], cat [278, 280], clawed frog [44], ferret [150], barn owl [159], and in snakes [96], support the idea of vision as teaching modality. However, vision-guided map calibration is not as evident as it seems at first glance. First, sensory maps, although imprecise, can evolve without any visual input [146, 158, 271]. Experiments on visually deprived animals and humans have shown that despite a missing visual system, the auditory and somatosensory systems can develop normally; see [152, 231, 234] for details. Second, behavioral [30, 68, 77, 249, 258] and physiological [202] studies

have shown that vision can also be influenced by other modalities. Third, and most importantly, Wallace and Stein [277, 279] have pointed out that receptive fields in the superior colliculus (SC) already shrink at times where *no* visual neurons are present. Furthermore, they have observed that all modalities show a *temporally coinciding* development in which even vision improves. But if vision should act as teaching modality, how can it guide calibration without being present and how can it improve itself? Together, the above paradoxical experimental and theoretical findings question the current picture of vision as the predominant guiding modality. That is, what is a coherent alternative to vision as *the* teaching modality?

In this chapter, we take two findings as starting points. First, the formation of different modalities begins at the time of appearance of multimodal integration [279]. Second, the motor, i.e., a multimodal map in the SC as well as in its non-mammalian homolog, the optic tectum (OT), has been suggested to provide a teaching signal for auditory space alignment [180]; cf. Fig. 1.4 for SC connectivity. Therefore, we suggest an *integrated Multimodal Teaching* concept (iMT); see Fig. 6.1. We explain the well-proven dominance of the visual system and unify contradictory findings as to how different modalities can be aligned with respect to each other. In doing so, we introduce an integrated multimodal teacher and discuss its intrinsic characteristics and implications on a conceptual level. To this end, we focus on map formation within the SC/OT by means of long-term potentiation (LTP) and long-term depression (LTD) of synapses that can be described by sSTDP. We leave out the question of dynamic adaptation as appearing in many animals that can move their sensory organs (eyes, ears, whiskers, etc.) independently of their body. The mechanism underlying this adaptation is not known, but retinotopic coordinate systems seem to play a key role [83]. Attempts to model this issue can be reviewed in [82, 99, 240, 241]. Within the SC/OT, however, multisensory and predominantly monosensory layers are mutually aligned to gain a unified multisensory representation of sensory space [147, 261]. The combined sensory information can then be used to generate directional responses in the SC/OT motor map [166, 167, 179, 216, 260]. In other words, at this stage all sensory maps share a common representation system. Similar to other theoretical models [78, 287], the suggested iMT concept concentrates on the question of map calibration appearing in the context of the very common sensory space. In addition to deriving the general concept for iMT, we apply our model to concrete examples. Using this model, we reanalyze in detail the concrete experiments of Knudsen et al. [159] and compare them with the findings of Wallace and Stein [277, 279] in the light of our iMT concept. Finally, we conclude by giving experimentally testable predictions of our approach.

6.2 Integrated Multimodal Teaching (iMT)

Our concept of intrinsic sensory map calibration structurally consists of a stack of unimodal maps, each connected with the multimodal map (cf. Fig. 6.1). In our iMT concept, different unisensory modalities are first merged into the multimodal map. Such a multimodal integration of unimodal maps can currently be found within the SC/OT, where this situation is experimentally well described [261, 279]. In the next step, the integrated multimodal map projects back to the unimodal maps. Thus, the multimodal map induces a teaching signal that guides the calibration (formation and alignment) process of the different unimodal maps. Therefore, the teacher input does not originate from a single modality, for example, vision, but from the integrated multimodal map. For a detailed understanding of our model design, please see Sec. 6.2.2.

In general, map calibration can be divided into two different classes. The first describes the situation in which sensory maps have to be learned from scratch. Connections between sensory systems (input maps) and representing maps contain no information and are randomly distributed. The connections have to evolve during the calibration process by means of LTP and LTD of the synapses; in mathematical terms described by STDP. We denote this case as *formation* of the maps.

The second class requires a collection of already pre-wired maps [271] and corresponds to

an *alignment* of maps, for instance, during growth or in shifting experiments [158, 159]. The stimulus position is here displaced in comparison to the position of the teacher input. By means of activity-based synaptic plasticity, the synaptic pattern of the map is modified so as to compensate for the misalignment between the map and teacher input.

To carefully analyze the consequences of applying an iMT signal we have to understand map formation and alignment from a procedural point of view. To this end, we note that the iMT process (as shown in Fig. 6.2) constitutes a closed feedback loop that consists of three repeating steps. First, during *multimodal integration*, different sensory modalities are optimally merged into an integrated multimodal map. Second, during *multimodal teaching*, the integrated multimodal map induces an inhibitory teaching signal that guides the calibration process, i.e. formation and alignment, of the different sensory maps. Third, during *unimodal map adaption*, the adapting and adapted unimodal maps again modify the integrated multimodal teacher.

The three phases we have described above are inherently connected by means of three major questions: 1) How do unisensory maps determine iMT? 2) How do the characteristics of iMT influence map adaptation? and 3) How does iMT calibrate different unimodal maps? Taken together, these three questions provide a structure for our iMT concept and will additionally guide us through the analysis of the iMT process in the following.

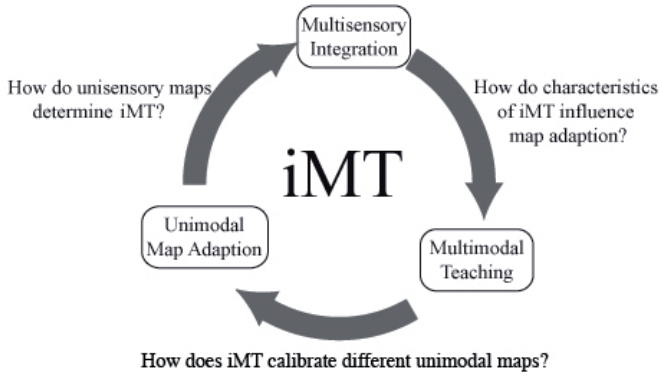

Figure 6.2: *Main procedural aspects of iMT. Combining different unimodal maps into a multimodal map leads to certain characteristics. What are these characteristics and how do they form an instructive teaching signal for the map calibration process that can be described by means of supervised spike timing-dependent plasticity (sSTDP)? To close the feedback loop, we discuss how different unimodal maps adapt to a common teaching signal.*

6.2.1 How do unisensory maps determine iMT?

To answer the first major question, we initially have to clarify the essential terms. We refer to a sensory modality as its neuronal representation on a spatial map; see Sec. 1.2 . On such a map one can interpret the firing rate as the *likelihood* to find a stimulus at the position that the neurons encode. In this manner, sensory maps represent position estimators or population codes [60, 116, 131, 227]. Depending on the statistics of the firing profile, we can attribute a mean μ corresponding to the estimated position and a standard deviation σ describing the estimator's accuracy to each sensory estimator. The integration of different sensory maps can therefore be seen to be equivalent to a combination of position estimators [2, 67, 102, 113, 126, 165, 202].

Although the exact neuronal mechanism as to how different modalities are integrated into a multimodal map [60, 61] is still largely terra incognito, the family of Bayesian combination schemes mathematically describes the result of optimal multimodal integration. For the sake of simplicity, we will focus on the maximum likelihood combination of only two modalities in the following text,

namely, *vision* (V) and *audition* (A). The case of three modalities is described in Appendix C.2. The visual and auditory maps are characterized by their means,i.e. the position estimates, μ_V for vision and μ_A audition. Furthermore, the accuracy of the position estimates is given by σ_V (vision) and σ_A (audition), and illustrated by Fig. 6.3.

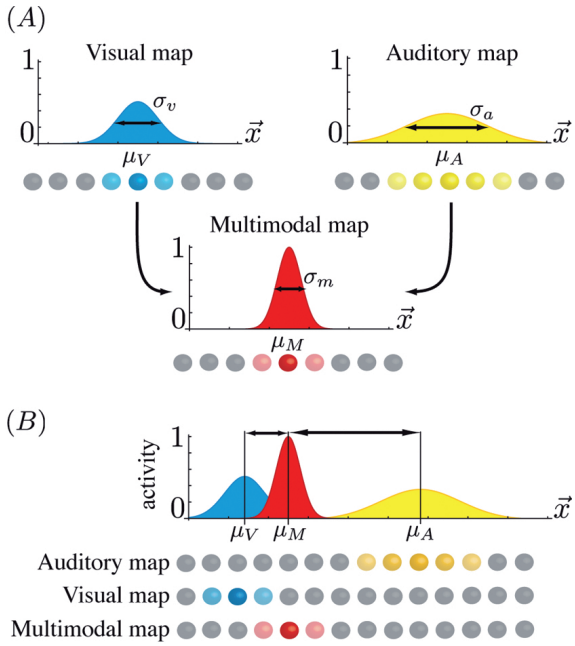

Figure 6.3: *Multimodal integration of two unisensory maps. The firing profiles of neuronal maps for audition, vision, and the multimodal map assumed as Gaussian functions with mean $\mu_{A/V/M}$ and standard deviation $\sigma_{A/V/M}$ can be interpreted as position estimators of the stimulus. The integrated map is assumed to be integrated by means of a maximum likelihood method represented by (6.1) and (6.2). (A) The multimodal map shows the smallest standard deviation $\sigma_M < \sigma_V < \sigma_A$ and is thus the most precise of all maps. (B) If the unisensory maps are shifted $\mu_V \neq \mu_A$, the multimodal map μ_M (red profile) is located more closely to the more precise modality (blue profile) (see (6.3).*

The multimodal position estimate μ_M as a maximum likelihood combination from coinciding visual and auditory inputs is given by [135, 144]

$$\mu_M = \frac{\sigma_V^2}{\sigma_A^2 + \sigma_V^2}\,\mu_A + \frac{\sigma_A^2}{\sigma_A^2 + \sigma_V^2}\,\mu_V \tag{6.1}$$

with the standard deviation σ_M of the multimodal position estimate. As a consequence of optimal integration, the multimodal estimator features the smallest possible standard deviation σ_M given by

$$\sigma_M = \frac{\sigma_A \sigma_V}{\sqrt{\sigma_A^2 + \sigma_V^2}}\,. \tag{6.2}$$

For a detailed derivation of the two above equations, please see Appendix C.1.

We can also compute the distance between the multimodal position μ_M and the unimodal positions $\mu_{A/V}$ by means of (6.1) to

$$|\mu_M - \mu_{A/V}| = \sigma_{A/V}^2\,|\mu_A - \mu_V| \,/\, \left(\sigma_A^2 + \sigma_V^2\right)\,. \tag{6.3}$$

That is, the multimodal and therefore the teaching signal is located *more closely* to the position of the more precise modality, e.g., vision in comparison to audition, as shown here.

The above equations (6.1) to (6.3) illustrate how we can describe an integrated multimodal map at a specific position in space and time. Next we have to take into account that the integrated multimodal map will provide the guiding signal for unimodal map formation and alignment. The equations of optimal combination, however, do not describe how unimodal maps change during the learning process. Map formation and alignment are an effect of synaptic plasticity (mathematically described by STDP), as we will discuss in Sec. 6.2.2. In this section we focus on fundamental characteristics of the multimodal map that follow directly from *optimal* multimodal integration. For now, we presume that the unimodal maps that form the multimodal map are variable in time. More precisely, on the one hand, the position of the unimodal maps $\mu_{A/V}(t)$ can and will vary during the calibration process. On the other hand, the maps can improve in precision such that $\sigma_{A/V}(t)$ decreases as well. These two effects lead to a variable iMT with time-dependant $\mu_M(t)$ and $\sigma_M(t)$ during the calibration process of the unimodal maps.

In short, the multimodal map is *positionally flexible* over the course of the alignment process because it is the combination of two or more sensory maps (6.1) to (6.3) that themselves change during the learning process. We will now discuss in greater detail what positional flexibility means. We have characterized each map in this paper by its standard deviation σ and estimator position, i.e., the mean μ of its firing profile. Thus, we first change σ_V (Fig. 6.4 A) and see that a sensory map that becomes more precise (blue profile) "attracts" the position of the multimodal map, i.e., the learning signal. Sensory maps that increase in accuracy get more stable positionally during sSTDP learning processes. Second, a shifting sensory map, for instance, vision (blue profile) in Fig. 6.4 B, induces a shift of the multisensory map towards the stationary sensory map (yellow profile). The upshot of iMT even without including effects of sSTDP in the context of map alignment is already that a more static sensory map attracts the teaching signal and will therefore hardly adapt at all.

In summary, we have analyzed characteristics of the optimally integrated multimodal map and have thereby shown three essential aspects. First, the multimodal map is always more precise than any unimodal map; cf. (6.2). Second, the optimal multisensory instructive signal is located more closely to the position of the more precise modality; see (6.3). Third, on the basis of an adaption of the monosensory maps through STDP learning rules, (6.1) and (6.2) describe how position and precision of the multimodal teacher varies for given $\mu_{A/V}(t)$ and $\sigma_{A/V}(t)$ of the input maps. Equipped with these characteristics of the teacher signal, we now analyze how the characteristics of iMT influence unimodal map adaptation on the level of synaptic plasticity, that is, in physiological and mathematical terms for sSTDP map alignment.

6.2.2 How do iMT characteristics influence map adaptation?

In the previous section we could see that an integrated multimodal teacher has a specific variance and position that both depend on the characteristics of the input maps; cf. (6.1) and (6.2). In addition, these characteristics will vary during map calibration. In this section, we therefore discuss how teachers with different accuracies and teachers that are shifting during the learning phase affect map alignment.

The results as presented in the following are based on numerical simulations of an *inhibition*-mediated STDP map alignment process. The basic structure of our model consists of the three modalities audition A, somatosensory S, and vision V that adapt to the multimodal inhibitory teacher modality T via STDP learning rules; see Fig. 6.5. The fundamental principles of inhibition-mediated STDP map alignment have been described elaborately by Friedel and van Hemmen [78]. All sensory maps are one-dimensional maps of N Poisson neurons [145]. Each modality consists of an input map with a Gaussian firing profile

$$f_i^{A/S/V}(y) := S_{A/S/V} \exp\left[-(y - x_i)^2 / 2\sigma_{A/S/V}^2\right] \tag{6.4}$$

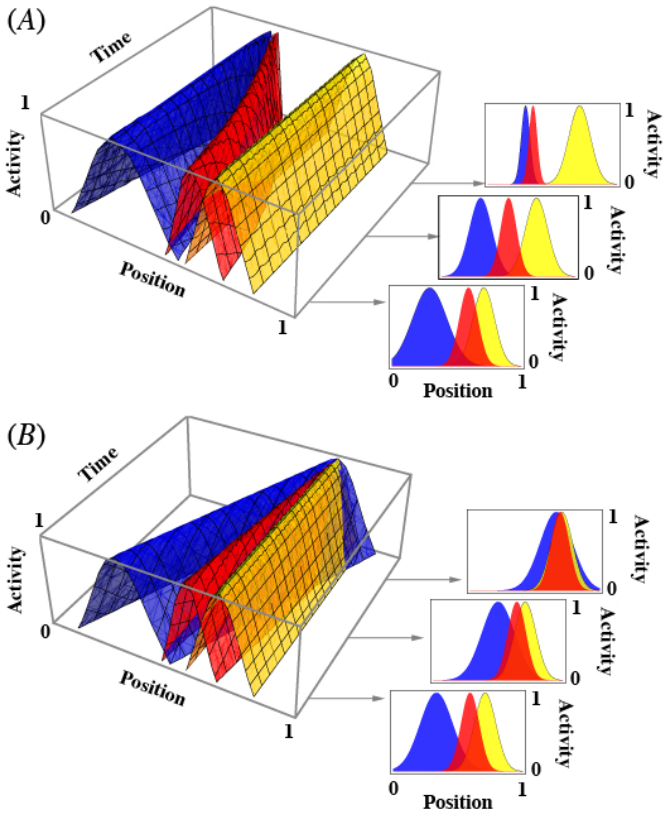

Figure 6.4: *Characteristics of an integrated multimodal map. During the iMT learning process the different modalities ($\mu_{A/V}$ changes) align with the instructive signal provided by the multimodal map, and their receptive fields become more precise ($\sigma_{A/V}$ decreases). Given that the multimodal map is a combination of the unimodal sensory maps, its precision and position are determined by the sensory maps and consequently change as well. Therefore, we vary one modality (blue profile) in precision (A) and position (B) and leave the other modality (yellow profile) unchanged. The resulting development of the multimodal teacher (red profile) is depicted in the upper graphs. (A) The modality (blue profile) that gets more precise attracts* (6.1) *and sharpens* (6.2) *the multimodal teacher (red profile). (B) In addition, the modality (blue profile) that aligns with the teacher (red profile), induces the multimodal map to shift as well. The distance between the multimodal teacher and the* stationary *map (yellow profile) therefore changes as well (cf.* (6.3)*).*

for $0 \leq i \leq N$, with stimulus position y, the i-th map neuron's preferred position x_i, standard deviation $\sigma_{A/S/V}$, and $S_{A/S/V}$ being the maximal amplitude of the activity. For two different maps, e.g. audition (A) and vision (V), the accuracies differ and $\sigma_A \neq \sigma_V$. The input layer projects via all-to-all synapses $\boldsymbol{J}^j_{A/S/V}(t)$ onto neuron j of the map layer. In addition, each neuron of the map layer receives inhibitory input from one corresponding neuron of the teacher modality T. The synapses connecting teacher and map are modeled as static one-to-one synaptic connections.

We interrupt the multimodal feedback loop compared to Fig. 6.1 and 6.2 by simulating an artificial teacher that is not generated from the multimodal representation but shows the essential characteristics that we derived in Sec. 6.2.1. A neuron p of the teacher map therefore responds to a stimulus at position y in form of an "inverted" Gaussian function

$$f_p^T(y) := S_T \left\{1 - \exp\left[-(y - x_p^T)^2 / 2\sigma_T^2\right]\right\} \tag{6.5}$$

with x_p being the preferred position of neuron p, S_T being the maximal firing activity of the teacher

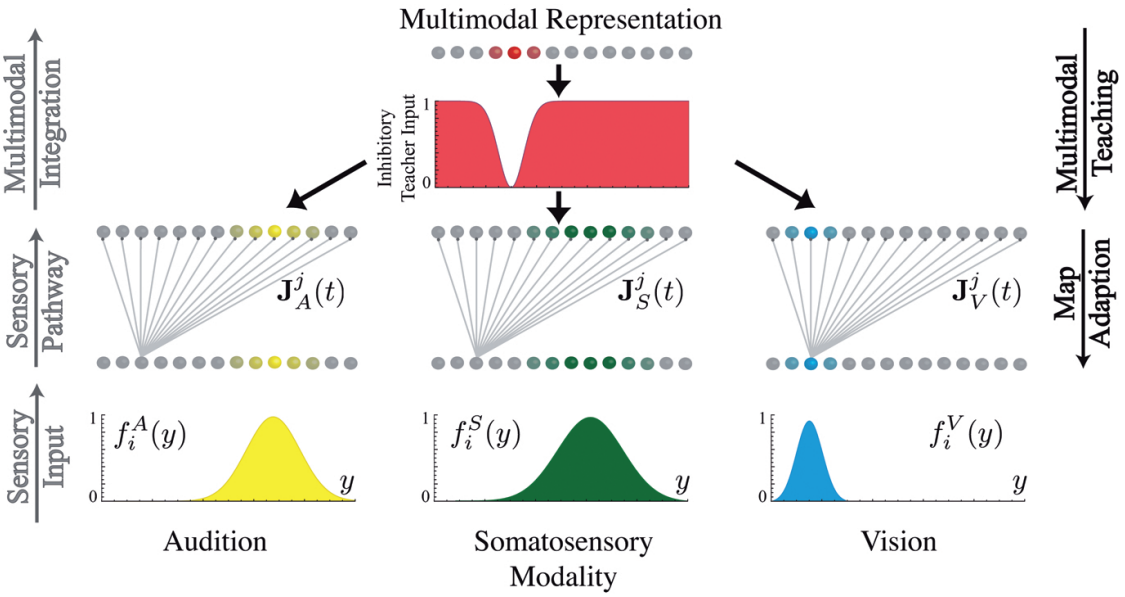

Figure 6.5: *Model architecture. A multimodal object such as a water drop hitting the skin produces sensory input for audition, somatosensory, and vision. We model the input by means of a Gaussian* $f_i^{A/S/V}(y)$ *so as to stimulate neuron* i *of the input layer. Neurons of the input layer project through all-to-all synaptic connections* $\mathbf{J}^j_{A/S/V}(t)$ *to neurons* j *of the map layer. These synaptic connections will adapt to align the output of the map layer to the inhibitory teacher input by means of sSTDP. Here we use this model setup to study the consequences of an iMT, which is the multimodal representation of all sensory modalities.*

neurons, and σ_T being the width of the teacher influence. The process we have just described mathematically, where a teacher suppresses the activity in the map layers at locations different from the real object's position y, is called *selective disinhibition* [78, 127].

To analyze the alignment process we perform numerical experiments. We introduce an artificial shift Δ in the sensory position of the modality, e.g., realized through prisms in the Knudsen [158, 159] experiment on the barn owl. In mathematical terms, we replace y by $y + \Delta$ in (6.5). The sensory modalities, say, audition A, somatosensory S, and vision V, subsequently change their synaptic connections between the input and the map layer to compensate (follow) the shift; see Fig. 6.5.

The situation in our numerical experiments corresponds to the situation of shifting vision by lenses [158, 159]. As explained in Sec. 6.3.1, a shift of vision induces a displacement of the multimodal teacher. Consequently, a shift of all sensory maps with respect to the position of the combined multimodal teacher occurs, that is, against intuition, lenses shifting vision hardly displace vision with respect to the teacher but displace all other sensory maps with respect to our multimodal teacher. Therefore, the biological experiments and our numerical experiments agree.

The changes within the synaptic connectivity pattern during the alignment process can be described and explained by STDP [78, 145]. To quantify changes within the synaptic pattern we introduce the *map estimation error* E as a least-square error. It measures how well a map can reproduce positional information of m systematically varied incoming signals,

$$E = \sqrt{\sum_{i=0}^{m} \left(\overline{x}_i - x_i^T\right)^2 / m} \; . \tag{6.6}$$

Here, $\overline{x}_i$ denotes the mean position of the firing profile of the map, and x_i^T represents the theoretical position. The smaller the value for the error E, the better the map reproduces the real (theoretical) position of an object. The parameters used in our numerical experiments are given in the Table C.1.

Multimodal teacher with varying variance

One of the essential characteristics of the multimodal teacher is its ability to improve in precision during the alignment process as a consequence of dynamic unimodal maps; see (6.2). After birth, most animals first have to learn how to interpret the sensory information they receive. All available maps as given by genetics are very inaccurate [181, 271]. Furthermore, Wallace and Stein [279] observed that multisensory neurons in cats younger than 28 days reliably respond to inputs from more than a single modality but do not show multimodal enhancement or depression. These observations, however, are measurements on the level of single cells. Subsequent conclusions in the context of neuronal maps and estimators are elusive. Therefore, it is fundamentally necessary to show that learning still functions or at least starts with an inaccurate teacher. Before we study the effect of a dynamic teacher on sSTDP map alignment we will clarify how teacher accuracy is defined. We note that the less accurate the unimodal maps, the larger the variance of the multimodal teacher (6.2), that is, the estimated position and thus the teaching signal is *spatially* less accurate. Consequently, we model the teacher accuracy by adding a Gaussian distributed random shift Δ_T to the correct teacher position. As a result of the numerical experiments with a varying teacher shift Δ_T, we see that the unimodal maps adapt more slowly and with a larger map estimation error to a less accurate teacher signal; cf. Fig. 6.6. Within a given time window, an inaccurate teacher therefore slows down but does not prevent the map alignment process. We can therefore conclude that the iMT concept does not depend crucially on a concrete realization of the multimodal integration.

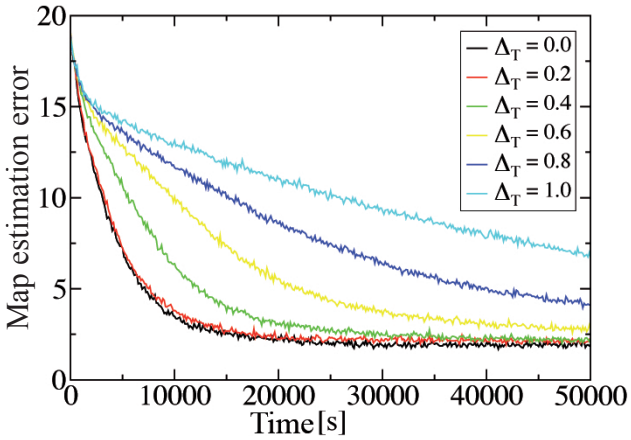

Figure 6.6: *Influence of teacher precision Δ_T on map alignment. We depict the map estimation error E (6.6) for map alignment with a multimodal teacher varying in precision. The map estimation error between teacher position and represented position on the map drops faster and to a lower level with a more precise (lower Δ_T) teacher. The precision of the teacher therefore influences the temporal development of map alignment.*

Multimodal teacher with dynamically changing position

As illustrated in Fig. 6.4, the integrated multimodal teacher position shifts during the alignment process. Therefore, we model teacher activities that continuously shift in space and analyze how their shifting velocities influence map alignment. For the time step $t = 2000$ s, Fig. 6.7 shows the temporal development of sensory maps, i.e. their synaptic pattern, modified by means of teaching

signals shifting with different velocities. We note that the dynamics of the teacher essentially determines to which degree a map can adapt to the teacher. The slower the teacher position shifts, the better a map can adapt to the teacher. Moreover, the critical velocity up to which map adaption works properly depends on the precision of the sensory modality that is reflected in the precision of the input map. A less precise modality can still adapt to a teacher with a given velocity, whereas a precise map cannot follow anymore; cf. Fig. 6.7, change from top to bottom.

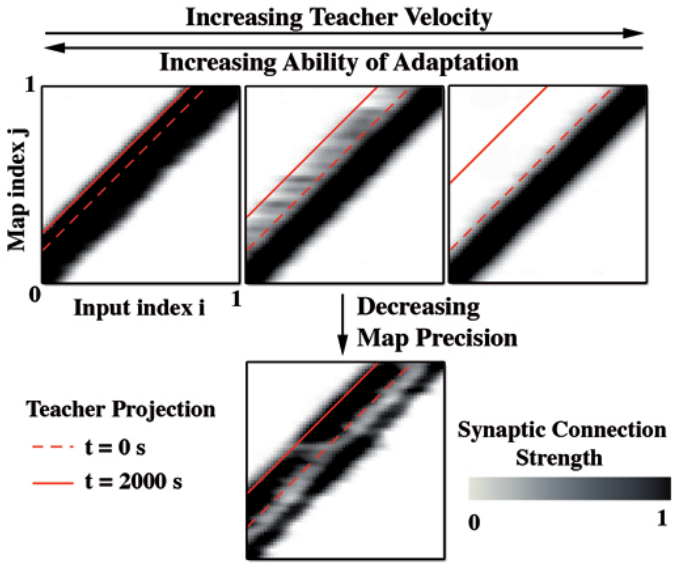

Figure 6.7: *Influence of dynamically changing teacher positions on map alignment. Simulations performed with input standard deviation* $\sigma = 0.03$ *(top line),* $\sigma = 0.05$ *(bottom) and teacher standard deviation* $\sigma_T = 0.02$*. For details regarding the parameters, see Appendix C.3. We shift the teacher signal, that is, the inverse Gaussian profile, in space. The red lines indicate the mean position of the teacher. In our numerical experiments the teacher position starts shifting from the dashed red line (*$\Delta = 0.1$*). The straight red line indicates how far the teacher has shifted within the simulated time of* 2000 *s. The distance between the two red lines depends on the different shifting velocity of the teacher signal. The synaptic pattern of the sensory maps are initially located around the diagonal. The box at the top left shows the synaptic pattern caused by a slowly shifting teacher. With a certain delay, the synaptic pattern of the map has adapted to the teacher map. By stepping to the right, the teacher velocity doubles. As a consequence, the map adapts much less until the velocity reaches a critical level (top right box) where the map does not adapt at all. The velocity of the teacher therefore essentially determines to which degree a map adapts to the teacher. This effect can be compensated by decreasing map precision as realized from the top to the bottom line. We then see that the synaptic pattern of a less accurate map still adapts very well to the teacher map. The critical velocity of non-adaptation is therefore higher for a low-precision map than for a high-precision map.*

We conclude that different teacher accuracies will not prevent alignment of different maps by means of sSTDP. However, the teacher accuracy influences the temporal development and, to some degree, the accuracy of the final state. We have seen that the shifting velocity of the teacher position determines to which degree a map can adapt to the teacher. A less accurate map is more flexible than an accurate map because it can adapt to a faster dynamically changing teacher. The precision of the different contributing unimodal maps therefore seems to determine how they adapt to the teacher signal. In the next section we will study these effects in detail.

6.2.3 How does iMT calibrate different unimodal maps?

Having analyzed first the combination of an integrated multimodal map as it arises from unimodal maps and the effect of a dynamically changing teacher on map alignment, we simulate in the third and last step (see Fig. 6.2) how the essential characteristics of the different modalities, in particular, the different accuracies of the sensory systems, influence map formation and alignment.

Results of our simulations of map formation and alignment are shown in Fig. 6.8. Based on these numerical experiments we observe that map formation and alignment by means of the iMT is, in general, possible for inputs of different accuracies. However, the quality and temporal progress of map formation and alignment depend on the accuracy of the sensory system. In passing, we note that within the alignment process, a high-precision map reaches a modified synaptic pattern that is again more precise than that of a low-precision map. More importantly, a high-precision map adapts more slowly to the teacher input than a map with low-precision. In conclusion, the higher the precision of a sensory map, the more static the map. We will see in the next section how these findings can explain experiments of sensory map formation and adaptation.

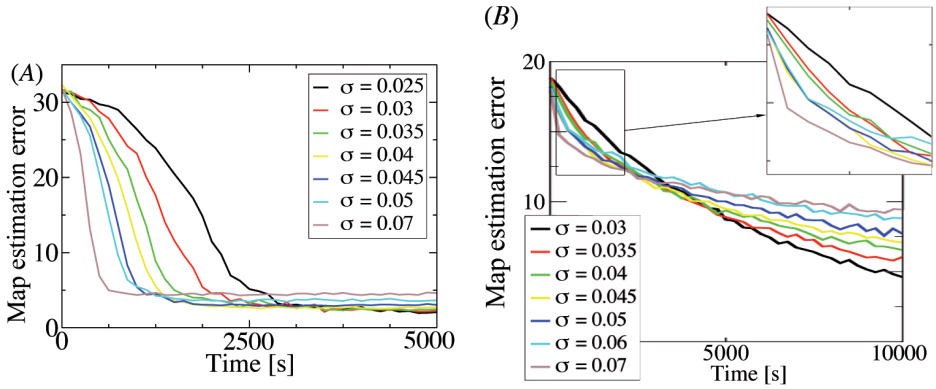

Figure 6.8: *Influence of varying input map precision (σ) on map formation and alignment. The plots show the map estimation error E as given by* (6.6) *from unimodal maps in comparison to the teacher. The maps vary in precision (input standard deviation σ) as indicated. The graphs show the map estimation error E for map formation (A) and for map alignment (B). The boxed region is enlarged to clearly show which map adapts faster. Both plots show that maps with a higher precision adapt more slowly to the teacher input than lower-precision maps, but regain higher precision after the map calibration.*

6.3 Model results and experimental findings

After detailed analysis and being equipped with a deeper understanding of the iMT concept, we return to the experiments presented in the introduction and review them in the light of our iMT concept. In doing so, we distinguish between the previously mentioned experiments that can be explained by a vision-guided concept of map calibration and those that cannot. The iMT concept we present here offers explanations for both.

In the following we will analyze a group of experiments that motivated the idea of vision as the dominant guiding sensory modality: the prism glass experiments of Knudsen and coworkers [127, 158]. We review the remaining experiments while denying a vision-guided map alignment. We then illustrate using experimental evidence as examples from Stein et al. [277, 279] how an iMT concept could work. Finally, we make suggestions and testable predictions for new experiments based on the iMT concept.

Animal	σ_A	σ_V	$\left\lvert\frac{\mu_A-\mu_M}{\mu_V-\mu_A}\right\rvert$	$\left\lvert\frac{\mu_V-\mu_M}{\mu_V-\mu_A}\right\rvert$
Human	1°	0.02°	99.96 %	0.04 %
Barn owl	2°	0.3°	97.8 %	2.2 %
Cat	8°	0.2°	99.93 %	0.06 %

Table 6.1: *Visual and auditory localization capabilities of human [57, 194], barn owl [106, 157], and cat [19, 42]. The visual resolutions are originally given in cycles per degree, i.e., how many lines can be distinguished in one degree of the visual field. We have assumed the resolutions to be the inverse of these values. On the basis of these variances we combine the two unimodal maps to a multimodal teacher and calculate the distance of each map to the multimodal map due to* (6.3) *in comparison to the distance between the two modalities.*

6.3.1 Experiments pro vision-guided map formation

The most prominent experiments supporting vision-guided map formation are the shifting experiments by Knudsen et al. [127, 158]. Here, prisms shift the visual system of owls by a certain angle. After a learning period the neuronal projections from the auditory map are rearranged to compensate for the misalignment between visual and auditory maps; see Sec. 1.3 in the introduction. The visual map, in contrast, remains constant during the whole alignment process.

Can our iMT concept reproduce the *dominance of vision* within the learning process? To answer this question we have simulated alignment of maps with different variances and distances to the teacher map as given by (6.3). We note that the visual system usually has a much higher spatial resolution (about a few arc minutes) than all other sensory systems, e.g. the auditory system (about a few degrees); see Table 6.1 for details. We now revisit synaptic plasticity during a prism shift experiment from the point of view of iMT.

- **Initial state** (Fig. 6.9, top)
  Auditory and visual maps are shifted with respect to each other because of an artificial shift of the visual system that is introduced by prisms. Given that the multimodal teacher map is a combination of the unimodal maps, its activity mean has shifted as well. An adaptation process of *all* unimodal maps is induced. Against normal intuition, shifting vision does not only result in a shift between vision and the multimodal teaching signal; it instead results in a shift of all sensory maps compared to the multimodal teacher map. As a consequence of the greater precision of the visual system, the multimodal teacher map is mainly determined by the characteristics of the visual map, that is $\sigma_{ML} \approx \sigma_V$ and $\mu_{ML} \approx \mu_V$; cf. (6.1) and (6.2)). Equation (6.3) neatly illustrates that the shift between vision as the most precise sensory system and thus the multimodal teacher map is smaller than for any other system.

- **Intermediate state** (Fig. 6.9, middle)
  Synaptic connections to the auditory map start changing to compensate for the shift between audition and the multimodal teacher. Based on our previous studies as presented in Sec. 6.2.2 and 6.2.3, we know that because of its higher precision, the visual map adapts much more slowly than the auditory map. The auditory map therefore starts adapting to the multimodal teacher. The teacher as a combination of the two modalities thus shifts toward the visual map. The auditory map adapts again to the teacher input and hence continues to shift toward the visual map, which stays almost static; see Fig. 6.4, B.

- **Final state** (Fig. 6.9, bottom)
  The two maps have been realigned. The position of the visual map has hardly changed, whereas

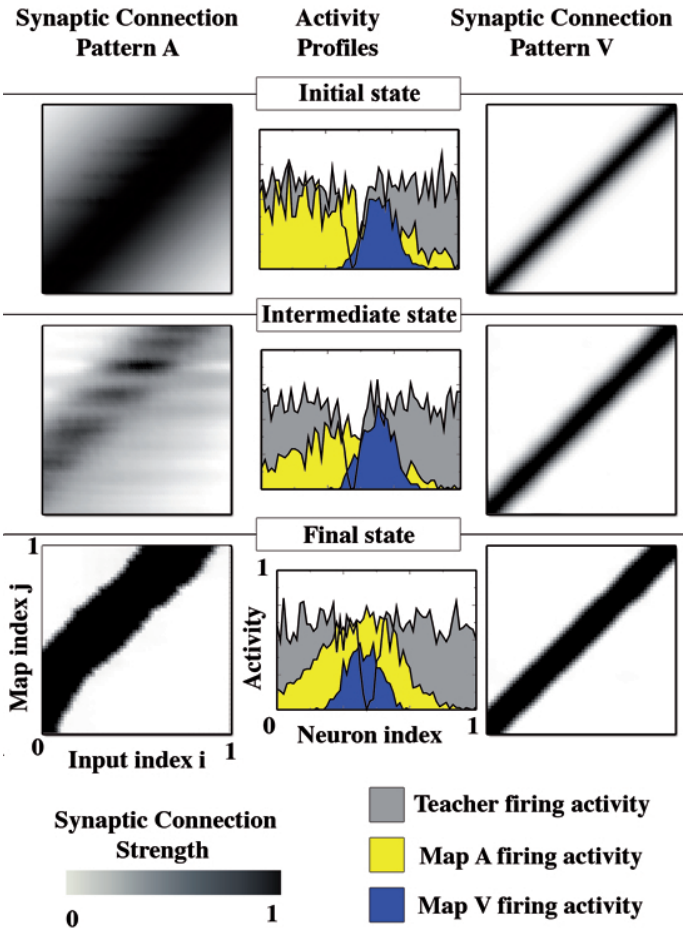

Figure 6.9: *Numerical realignment experiments. We illustrate the shifting process for the auditory map (A) with variance $\sigma_A = 0.07$ and the visual map (V) with $\sigma_V = 0.03$. The distances between teacher and unimodal map positions are chosen accordingly to* (6.3) *for the corresponding variances. The array plots show the synaptic connectivity pattern for vision (left column) and audition (right column). The synaptic strengths hereby range between* 0 *and* 1*. The activity profiles of the auditory map A, the visual map V, and the teacher map are shown in the middle column. The color code is specified in the bottom right corner. The plot illustrates how synaptic plasticity depends on the accuracy of a modality. In concrete terms, the more precise visual map V shifts more slowly and far less than the less accurate auditory map A.*

the auditory map has shifted almost the whole distance induced by the prisms at the beginning of the experiment.

In summary, our iMT concept can reproduce the different steps of the shifting experiments of Knudsen and others [127, 158]. However, similar numerical results can be obtained for any two arbitrary sensory maps with differing variances. Therefore, the observed dominance of the visual system only reflects the more general concept of a *dominance of accuracy* following naturally from the iMT concept.

6.3.2 Experiments contra vision-guided map formation

In the introduction of this chapter we reviewed some experiments that question a vision-guided learning process. The analysis of our concept has shown that vision plays a dominant role because of its high precision but that vision is not static. It can be modified by a dynamically changing multimodal teacher map and therefore by *any* unimodal map [68, 77, 202, 249, 258].

Furthermore, the iMT concept should be able to reproduce formation of unimodal maps in visually deprived animals. Studies [152, 231, 234] show that blind humans, cats, and ferrets can localize sound as precisely as individuals with normal vision. Measurements of Wallace and Stein [277, 279] support the critical view of vision-dominated learning by showing that receptive fields already shrink at times when *no* visual neurons are present. Learning is possible without vision but is dominated by vision in cases in which it is available.

In the context of our iMT concept a multimodal teacher map can exist without a visual map, but it has a much lower precision. Formation of the other maps is therefore still possible within certain limits; cf. Fig. 6.6 and [146, 158, 271].

In the following text, we will reproduce formation of unimodal maps in a visually deprived animal, thereby studying whether somatosensory and auditory maps can develop *without vision*, i.e., with a very imprecise teacher.

- **Initial state** (Fig. 6.10, top)
  Somatosensory and auditory maps are only coarsely prewired, and the visual map does not exist at all. The multimodal map as a combination of somatosensory and auditory maps is very imprecise as well; see Sec. 6.2.2 for details.

- **Intermediate state** (Fig. 6.10, middle)
  Even with a very imprecise teacher map the auditory and somatosensory maps start to improve while their variances decrease. The earlier simulations illustrate that the learning process with an imprecise teacher starts very slowly, corresponding to the topmost curve in Fig. 6.6. Nevertheless, the unimodal maps improve slightly, as does the multimodal map that is a combination of the two sensory maps as a consequence. With the improving teacher (see Fig. 6.6), the learning process accelerates.

- **Final state** (Fig. 6.10, bottom)
  The two maps are aligned and have reached their best-possible resolutions.

Together, multimodal teaching illustrates that the iMT concept is a successful concept for calibrating different unimodal maps. In our intrinsic learning process all maps develop and improve in parallel, as reported in experiments such as [277, 279].

6.4 Discussion and Outlook

In the present chapter we have introduced the concept of integrated Multimodal Teaching. The application of iMT to map formation and alignment leads to three fundamentally new ideas. First,

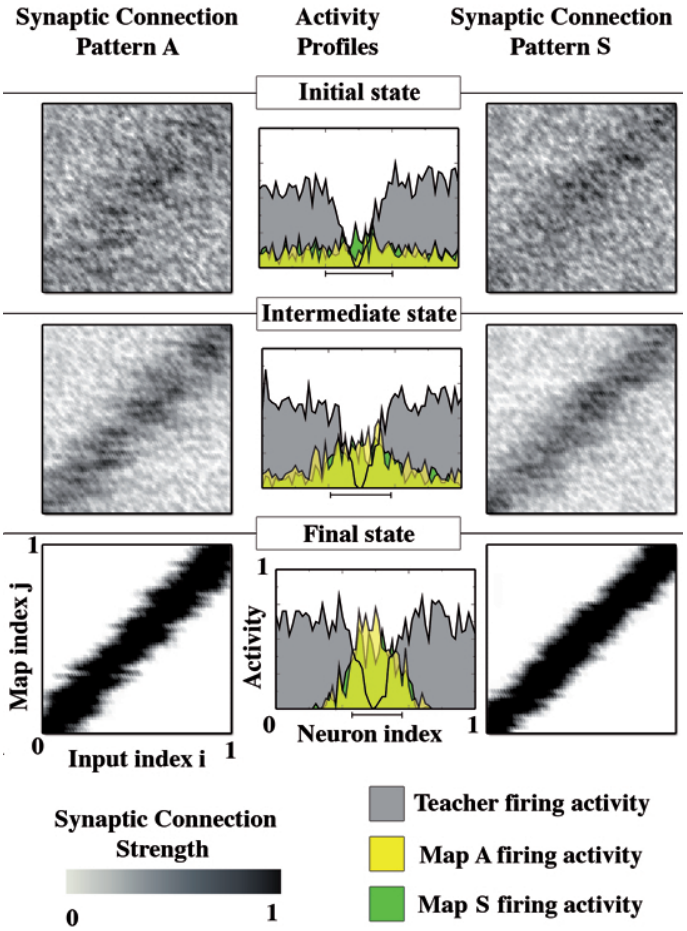

Figure 6.10: *Numerical map formation experiments with an improving teacher. Activity profiles of the auditory map A ($\sigma_A = 0.03$), the somatosensory map S ($\sigma_S = 0.045$), and the teacher map are shown in the middle column. The blurred teacher projects onto the maps with a precision that starts at $\Delta_T = 0.1$ and then reduces linearly in time to $\Delta_T = 0.02$ for $t = 7500$ s. The black bars below the firing-activity plots (middle column) denote the regions within the teacher is projecting for the corresponding time step. The array plots show the synaptic connection pattern J_A for audition (left column) and J_S for the somatosensory modality (right column). The synaptic strengths range between 0 and 1. The initial synaptic pattern is only coarsely prewired and it is additionally blurred by means of randomly distributed synapse strengths. The color code is specified in the bottom right corner. Despite an imprecise teacher, both maps can develop properly.*

the *integrated* teacher contains all available information of the maps involved. Second, the teacher is intrinsic to the system because it is generated only from information available within the system of maps. Third, as a consequence, the teacher itself is *dynamic* with the potency to shift in position and to improve in precision during the learning process.

Moreover, the presented results provide the basic insight that the more precise a sensory system is, the more slowly it adapts to instructive signals. This natural consequence, which we denote as *dominance of accuracy*, explains observations where a dominance of vision within the alignment process may be wrongly interpreted as purely visually driven.

Comparison to the state-of-the-art

Here, we have simulated multimodal teacher characteristics for map calibration on the level of spiking neurons and synaptic plasticity, namely, sSTDP. The teacher signal influences the unimodal maps by selective disinhibition as proposed by an increasing number of experimental findings, for instance, in the ICC-ICX-OT pathway [14]; see also Sec. 1.3 of the introduction. Taken together, our iMT concept therefore connects two fundamental neurophysiological phenomena: inhibition-mediated map calibration and multimodal integration.

Previous theoretical studies have focused on only parts of these aspects. Gelfand and Mysore [86, 204] proposed models for Hebbian map calibration based on the assumption that vision serves as teacher signal to modify, for instance, synapses of the auditory projection pathway. As already pointed out in the introduction of this chapter, vision as teacher signal nevertheless has to be questioned. Stein and Wallace [277, 279] therefore suggested a connection between multimodal map and map formation in cats. The latest experimental studies on barn owls from Bergan and Knudsen [14] also show that cross-modal effects exist in the ICX and may influence experience-dependent calibration of converging representations.

Including these experimental findings, Witten et al. [287] developed a theoretical model for bimodal map alignment. In their model, the synapses of both the auditory *and* the visual map are adapted due to Hebbian learning. By these means, the model can reproduce the dominance of vision such that the channel with the weaker or broader receptive fields always exhibits most or all of the plasticity. However, the theoretical model of Witten et al. is exclusively based on excitatory inputs and rate codes. The latter assumptions are both quite implausible in the context of map alignment in ICX and OT. Concretely, experimental data from Bergan and Knudsen [14] suggest that instructive signals to the ICX are gated by inhibition. In addition, Kempter et al. [145] demonstrated that the assumptions underlying rate-based Hebbian learning are not necessarily valid, especially in the auditory system of the barn owl.

Including the last mentioned experimental results, Friedel et al. [78] developed a theoretical model of map alignment by means of selective disinhibition on the level of spiking neurons. The model shows that an inhibitory teacher is essential to realign already calibrated unimodal maps. However, the model only contains plasticity within the auditory system, that is, the visual system is static and provides the guiding signal for auditory space alignment.

The present iMT model combines the fundamental aspects of the models from Witten et al. [287] and Friedel et al. [78]: spike-timing-dependent learning based on both selective disinhibition and full plasticity within all contributing modalities, that is, all available information is integrated into a multimodal map as a first step. In a second step, the multimodal map provides an instructive signal for map calibration to all modalities. Therefore, the iMT model can explain the dominant role of vision in map alignment, plasticity in modalities without vision, plasticity within the visual modality guided by other modalities, plasticity of non-visual modalities guided by another non-visual modality, and any combination of these settings.

Testable predictions of the iMT concept

Given that we have shown that our iMT concept can reproduce both experiments supporting and questioning vision-dominated learning, it is now time to offer some suggestions regarding how the

iMT concept can be experimentally verified. To distinguish vision-dominated teaching from iMT we have to bypass the dominance of vision. We can either study artificial situations in which the precision of sensory maps is equal, investigate animals with a poorly developed visual system, or observe map formation while multimodal maps are experimentally deactivated.

To first equalize the precision of sensory maps, one could, for instance, perform shifting experiments with prisms that displace *and blur* the visual input at the same time. Second, studying map formation and alignment for animals with a more precise sensory system than vision could also identify if there is a general principle such as dominance of accuracy. A third possibility would be to deactivate the multimodal integration during the learning process. Given iMT, deactivation of the multimodal map(s) correspond(s) to a deactivation of the teaching modality. Map formation and alignment should fail in this case. In contrast, for a vision-dominated teaching scenario, e.g. vision-guided map alignment, no negative effect would be observed on the learning process.

Outlook

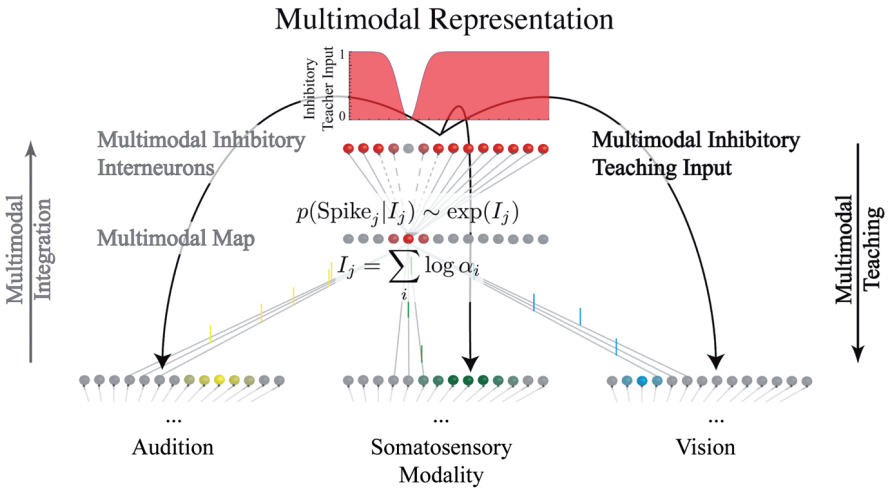

Figure 6.11: *Outlook: iMT network setup with feedback connections. One can extend the current iMT setup shown in Fig. 6.5 by means of including the multimodal integration process on the level of spiking neurons. That is, we automatically incorporate the feedback mechanism discussed previously in Sec. 6.2 and depicted in Fig. 6.2. First, the post-synaptic currents α_i generated by the incoming spikes from the neurons i of the different sensory maps $A/S/V$ are logarithmically summed up; $I_j = \sum_i \log \alpha_i$. Second, the probability to evoke a spike in neuron j of the multimodal map can be modeled to relate exponentially to the complete input current I_j injected into neuron j of the multimodal map. Third, we use an interneuron layer to "invert" the excitatory activity profile to an inhibitory disinhibition signal that, fourth, guides map calibration of the sensory maps; sensory pathway and input is faded as compared to Fig. 6.5. In other words, with the network setup described above, we can neuronally calculate an optimal multimodal map employing all the characteristics we have used for the iMT in this chapter by multiplying the activity profiles of the sensory maps on the level of spiking neurons; for details see Appendix C. We have already implemented and tested this setup and the results look promising. Map formation and alignment is in principle possible but the feedback mechanism needs further clarification.*

In summary, we have introduced the new idea of an *intrinsic* multimodal learning process and in so doing have revealed interesting new aspects of multimodal map formation. The question of (multimodal) map formation, however, is not yet fully understood. On the theoretical side, future models should include a two-way feedback between unimodal maps and the multimodal teacher. We have already implemented and started to test a more sophisticated model on the level

of spiking-neurons; see network setup in Fig. 6.11. In short, spikes from the sensory maps A/V are summed logarithmically providing the input current to the neurons of the multimodal map M that are modeled to respond exponentially to the input current. This corresponds to a multiplication of spikes and models optimal multimodal integration on a level of spiking neurons; see (C.6) in Appendix C. The activity of the multimodal teacher is then "inverted" via inter-neurons and projected back to the map neurons of the sensory systems A/V to provide an inhibitory teaching signal. We are looking forward to derive an explicit model of iMT including full neuronal feedback dynamics without artificial normalization terms.

On the experimental side, experiments may well deliver convincing evidence for or against a connection between multimodal processing and map calibration. It will be interesting to see whether a system of maps such as what is found in the SC/OT (see Fig. 1.4) is able to calibrate intrinsically, and whether multimodal maps play a key role in calibration.

In presenting the concept of iMT, we have not only addressed the question of how global structures can enable local mechanisms such as synaptic plasticity to form complex brain structures, e.g., the SC/OT, we have also given another line of evidence that sensory experience is *the* driving force for brain development. To give a final statement beyond this chapter, this is really remarkable since the brain exists to process the quantity which determines its functioning, viz., sensory experience. That is why we believe it is so important to understand sensory processing.

Appendices

A. Appendix: Optimal Map formation

A.1 A recipe for making maps

To bring to life the mathematical framework of Sec. 2.2, we present an easy step-by-step "recipe" to find the optimal connectivity in a realistic biological setup:

- First, we derive the transfer function $h_i^{\boldsymbol{x}}(t)$ that determines the response of the detector i to a stimulus pulse that occurred t time units ago at position $\boldsymbol{x}$.
- Next, we calculate the Fourier transform $H_i^{\boldsymbol{x}}$ of the transfer function $h_i^{\boldsymbol{x}}(t)$.
- We choose suitable values of τ and σ. In general the noise-to-signal ratio τ can be assumed to be much smaller than 1 for any measurable signal. In contrast, σ needs to be estimated in dependence upon the situation at hand [76, 75, 252].
- We then calculate the matrix entries M_{ij} as given by Eq. (2.19) and invert the model matrix $\mathcal{M}$.
- We multiply the inverted matrix $\mathcal{M}^{-1}$ by the vector $\overline{H_i^{\boldsymbol{x}}}$ so as to find the input connection strengths $L_i^{\boldsymbol{x}}$.
- Finally, we calculate the inverse Fourier transform of $L_i^{\boldsymbol{x}}$ so as to find the connection strengths $l_i^{\boldsymbol{x}}(t)$.

A.2 Nonlinearities in information processing

The presented model assumes a linear relation between stimulus and detector response. For a number of sensory systems, however, we find non-linearities in the mapping process. First, the transfer function h can be a non-linear function $\tilde{h}$. Second, the neuronal detector response can be non-linear, typically logarithmic [169, 213, 172, 136, 53]. In case of a logarithmic response and a non-linear transfer function the biological detector response $\tilde{r}$ has to be rewritten from (2.1) as

$$\tilde{r}_i(t) = \log\left[\int_{\text{all space}} \mathrm{d}\boldsymbol{x} \int_{-\infty}^{t} \mathrm{d}\tau \; s^{\boldsymbol{x}}(\tau)\tilde{h}_i^{\boldsymbol{x}}(t-\tau)\right] . \tag{A.1}$$

To apply our model we first have to incorporate an additional computational step canceling the logarithm. In a biological system this can be realized, e.g., by neurons with exponential firing behavior. Assuming such a neuronal step $\tilde{r}_i(t)$ as in (A.1) reduces to (2.1) with a non-linear $\tilde{h}$. We can linearize a non-linear transfer function by a redefinition of the signal $s \to \tilde{s}$. That is, we identify appropriate characteristics of the stimulus that are linearly related to r. For example, instead of looking at the heat distribution $T(\boldsymbol{x}, t)$ we can consider the intensity distribution of the corresponding radiation $\sim T^4(\boldsymbol{x}, t)$ due to the Stefan-Boltzmann Law. In this way, a *reasonable* redefinition of detector response and signal can allow for an optimal linear stimulus reconstruction.

A.3 Remaining derivation steps leading to (2.16)

In the following we elaborate some steps skipped in the derivation of (2.16) in the main text. In doing so we take advantage of ideas due to the calclulus of variations [138]. We therefore start from (2.6) as a condition to minimize the expectation value of the quadratic error with respect to the optimal reverse transfer functions, the connection strengths $l_j^{\boldsymbol{x}}(t)$. This leads to

$$\begin{aligned} & \frac{\partial \left\langle [s^{\boldsymbol{x}}(t') - \hat{s}^{\boldsymbol{x}}(t')]^2 \right\rangle}{\partial\, l_j^{\boldsymbol{x}}(t')} &= 0 \qquad \text{for every } j. \\ \Leftrightarrow \qquad & \left\langle [s^{\boldsymbol{x}}(t') - \hat{s}^{\boldsymbol{x}}(t')] \frac{\partial\, \hat{s}^{\boldsymbol{x}}(t')}{\partial\, l_j^{\boldsymbol{x}}(t')} \right\rangle &= 0\,. \end{aligned} \tag{A.2}$$

In order to solve (A.2), we expand the estimate $\hat{s}^{\boldsymbol{x}}(t)$ using Eqs. (2.3) and (2.4) from the main text giving

$$\begin{aligned} \hat{s}^{\boldsymbol{x}} &= \sum_i \Big\{ \chi_i \star l_i^{\boldsymbol{x}} + \chi_i \star \lambda_i^{\boldsymbol{x}} + \int \mathrm{d}\boldsymbol{y}\, \big[s^{\boldsymbol{y}} \star h_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}} + s^{\boldsymbol{y}} \star h_i^{\boldsymbol{y}} \star \lambda_i^{\boldsymbol{x}} + s^{\boldsymbol{y}} \star \eta_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}} \\ &+ s^{\boldsymbol{y}} \star \eta_i^{\boldsymbol{y}} \star \lambda_i^{\boldsymbol{x}} + \xi^{\boldsymbol{y}} \star h_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}} + \xi^{\boldsymbol{y}} \star h_i^{\boldsymbol{y}} \star \lambda_i^{\boldsymbol{x}} + \xi^{\boldsymbol{y}} \star \eta_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}} + \xi^{\boldsymbol{y}} \star \eta_i^{\boldsymbol{y}} \star \lambda_i^{\boldsymbol{x}} \big] \Big\}\,. \end{aligned} \tag{A.3}$$

The variation of $\hat{s}$ in (A.3) leads to

$$\frac{\partial\, \hat{s}^{\boldsymbol{x}}(t')}{\partial\, l_j^{\boldsymbol{x}}(t')} = \Big[\chi_j + \int \mathrm{d}\boldsymbol{y}\, \left(s^{\boldsymbol{y}} \star h_j^{\boldsymbol{y}} + s^{\boldsymbol{y}} \star \eta_j^{\boldsymbol{y}} + \xi^{\boldsymbol{y}} \star h_j^{\boldsymbol{y}} + \xi^{\boldsymbol{y}} \star \eta_j^{\boldsymbol{y}} \right) \Big](0)\,. \tag{A.4}$$

As before, we assume on the one hand that all noise terms as well as the expectation of the input are stochastically independent of each other. On the other hand we use the fact that all noise terms have zero mean and the signal is self-averaging. With these two assumptions, the expectation values $\langle spartial\hat{s}/\partial l \rangle$ and $\langle \hat{s}\ \partial\hat{s}/\partial l \rangle$ from (A.2) can be written

$$\left\langle s^{\boldsymbol{x}}(t') \frac{\partial\, \hat{s}^{\boldsymbol{x}}(t')}{\partial\, l_j^{\boldsymbol{x}}(t')} \right\rangle = \int \mathrm{d}\boldsymbol{y}\ \left\langle s^{\boldsymbol{x}}(t')\ (s^{\boldsymbol{y}} \star h_j^{\boldsymbol{y}})(0) \right\rangle \tag{A.5}$$

and

$$\begin{aligned} \left\langle \hat{s}^{\boldsymbol{x}}(t') \frac{\partial\, \hat{s}^{\boldsymbol{x}}(t')}{\partial\, l_j^{\boldsymbol{x}}(t')} \right\rangle &= \sum_i \Big\{ \langle (\chi_i \star l_i^{\boldsymbol{x}})(t') \chi_j(0) \rangle + \int \mathrm{d}\boldsymbol{y} \mathrm{d}\boldsymbol{y'} \Big[\left\langle (s^{\boldsymbol{y}} \star h_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}})(t') (s^{\boldsymbol{y'}} \star h_j^{\boldsymbol{y'}})(0) \right\rangle \\ &+ \left\langle (s^{\boldsymbol{y}} \star \eta_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}})(t') (s^{\boldsymbol{y'}} \star \eta_j^{\boldsymbol{y'}})(0) \right\rangle + \left\langle (\xi^{\boldsymbol{y}} \star h_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}})(t') (\xi^{\boldsymbol{y'}} \star h_j^{\boldsymbol{y'}})(0) \right\rangle \\ &+ \left\langle (\xi^{\boldsymbol{y}} \star \eta_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}})(t') (\xi^{\boldsymbol{y'}} \star \eta_j^{\boldsymbol{y'}})(0) \right\rangle \Big] \Big\}\,. \end{aligned} \tag{A.6}$$

In the next step we substitute the correlation terms as given in (2.10) of the main text. To illustrate the calculations, which simplify (A.5) and (A.6), we analyze two isolated terms from (A.6) as an example. The other terms are treated in a similar way. We first simplify

$$\begin{aligned} & \sum_i \int \mathrm{d}\boldsymbol{y} \mathrm{d}\boldsymbol{y'} \left\langle (s^{\boldsymbol{y}} \star h_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}})(t') (s^{\boldsymbol{y'}} \star h_j^{\boldsymbol{y'}})(0) \right\rangle \\ &= \sum_i \int \mathrm{d}\boldsymbol{y} \mathrm{d}\boldsymbol{y'} \mathrm{d}\tau \mathrm{d}\tau' \mathrm{d}\tau'' \left\langle s^{\boldsymbol{y}}(t' - \tau - \tau') h_i^{\boldsymbol{y}}(\tau') l_i^{\boldsymbol{x}}(\tau) s^{\boldsymbol{y'}}(-\tau'') h_j^{\boldsymbol{y'}}(\tau'') \right\rangle\,. \end{aligned} \tag{A.7}$$

Exploiting the correlation assumptions this expression becomes

$$\begin{aligned}
&\mu_s^2 \sum_i \int \mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{y'}\mathrm{d}\tau\mathrm{d}\tau'\mathrm{d}\tau''\delta(\boldsymbol{y}-\boldsymbol{y'})\delta(t'-\tau-\tau'+\tau'')h_i^{\boldsymbol{y}}(\tau')l_i^{\boldsymbol{x}}(\tau)h_j^{\boldsymbol{y'}}(\tau'') \\
&= \mu_s^2 \sum_i \int \mathrm{d}\boldsymbol{y}\mathrm{d}\tau\mathrm{d}\tau''h_i^{\boldsymbol{y}}(t'-\tau+\tau'')l_i^{\boldsymbol{x}}(\tau)h_j^{\boldsymbol{y}}(\tau'') \\
&= \mu_s^2 \sum_i \int \mathrm{d}\boldsymbol{y}((h_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}}) \circ h_j^{\boldsymbol{y}})(t') \qquad \text{(A.8)}
\end{aligned}$$

with the open circle $\circ$ denoting the autocorrelation integral defined in (2.13). We focus on the second term in the right-hand side of (A.6). That is,

$$\begin{aligned}
&\sum_i \int \mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{y'} \left\langle (s^{\boldsymbol{y}} \star \eta_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}})(t')(s^{\boldsymbol{y'}} \star \eta_j^{\boldsymbol{y'}})(0) \right\rangle \\
&= \sum_i \int \mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{y'}\mathrm{d}\tau\mathrm{d}\tau'\mathrm{d}\tau'' \left\langle s^{\boldsymbol{y}}(t'-\tau-\tau')\eta_i^{\boldsymbol{y}}(\tau')l_i^{\boldsymbol{x}}(\tau)s^{\boldsymbol{y'}}(-\tau'')\eta_j^{\boldsymbol{y'}}(\tau'') \right\rangle \ , \qquad \text{(A.9)}
\end{aligned}$$

which simplifies to

$$\begin{aligned}
&\mu_s^2\sigma_\eta^2 \sum_i \int_{\substack{|\boldsymbol{y}|<y^{\max} \\ 0<\tau'<t^{\max}}} \mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{y'}\mathrm{d}\tau\mathrm{d}\tau'\mathrm{d}\tau''\delta(\boldsymbol{y}-\boldsymbol{y'})\delta(t'-\tau-\tau'+\tau'')\delta_{ij}\delta(\boldsymbol{y}-\boldsymbol{y'})\delta(\tau'-\tau'')l_i^{\boldsymbol{x}}(\tau) \\
&= \mu_s^2\sigma_\eta^2 \int_{\substack{|\boldsymbol{y}|<y^{\max} \\ 0<\tau'<t^{\max}}} \mathrm{d}\boldsymbol{y}\mathrm{d}\tau\mathrm{d}\tau'\delta(-\tau+t')l_j^{\boldsymbol{x}}(\tau) \\
&= \mu_s^2\sigma_\eta^2 \int_{\substack{|\boldsymbol{y}|<y^{\max} \\ 0<\tau'<t^{\max}}} \mathrm{d}\boldsymbol{y}\mathrm{d}\tau' l_j^{\boldsymbol{x}}(t') \ . \qquad \text{(A.10)}
\end{aligned}$$

Altogether the final expressions for the expectation values become

$$\left\langle s^{\boldsymbol{x}}(t')\frac{\partial\,\hat{s}^{\boldsymbol{x}}(t')}{\partial\,l_j^{\boldsymbol{x}}(t')} \right\rangle = \mu_s^2 h_j^{\boldsymbol{x}}(-t') \qquad \text{(A.11)}$$

and

$$\left\langle \hat{s}^{\boldsymbol{x}}(t')\frac{\partial\,\hat{s}^{\boldsymbol{x}}(t')}{\partial\,l_j^{\boldsymbol{x}}(t')} \right\rangle = \sigma_\chi^2\ l_j^{\boldsymbol{x}}(t') + \sigma_\eta^2(\mu_s^2+\sigma_\xi^2) \int_{\substack{|\boldsymbol{y}|<y^{\max} \\ 0<\tau<t^{\max}}} \mathrm{d}\boldsymbol{y}\mathrm{d}\tau\ l_j^{\boldsymbol{x}}(t') + (\mu_s^2+\sigma_\xi^2) \sum_i \int \mathrm{d}\boldsymbol{y}\ \left[(h_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}}) \circ h_j^{\boldsymbol{y}}\right](t') \ . \qquad \text{(A.12)}$$

Equation (A.2) therefore transforms into

$$\begin{aligned}
&l_j^{\boldsymbol{x}}(t)\left[\sigma_\chi^2 + (\mu_s^2+\sigma_\xi^2) \int_{\substack{|\boldsymbol{y}|<y^{\max} \\ 0<\tau<t^{\max}}} \mathrm{d}\boldsymbol{y}\mathrm{d}\tau\ \sigma_\eta^2\right] && \text{(A.13)} \\
&+(\mu_s^2+\sigma_\xi^2) \sum_i \int \mathrm{d}\boldsymbol{y}\ \left[(h_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}}) \circ h_j^{\boldsymbol{y}}\right](-t) && \text{(A.14)} \\
&= \mu_s^2 h_j^{\boldsymbol{x}}(-t) \ . && \text{(A.15)}
\end{aligned}$$

Inserting the parameter σ and τ and applying a Fourier transformation finally leads to (2.16).

In order to test whether the extremum is indeed a minimum, we have to calculate the second variation, which reads

$$\begin{aligned}
&\frac{\partial^2 \left\langle \left[s^{\boldsymbol{x}}(t') - \hat{s}^{\boldsymbol{x}}(t')\right]^2 \right\rangle}{\left(\partial\, l_j^{\boldsymbol{x}}(t')\right)^2} \\
&= 2\, \frac{\partial}{\partial\, l_j^{\boldsymbol{x}}(t')} \left[\left\langle s^{\boldsymbol{x}}(t') \frac{\partial\, \hat{s}^{\boldsymbol{x}}(t')}{\partial\, l_j^{\boldsymbol{x}}(t')} \right\rangle - \left\langle \hat{s}^{\boldsymbol{x}}(t') \frac{\partial\, \hat{s}^{\boldsymbol{x}}(t')}{\partial\, l_j^{\boldsymbol{x}}(t')} \right\rangle \right] \\
&= 0 + 2 \left[\sigma_\chi^2 + \sigma_\eta^2(\mu_s^2 + \sigma_\xi^2) \int_{\substack{|\boldsymbol{y}|<y^{\max} \\ 0<\tau<t^{\max}}} \mathrm{d}\boldsymbol{y}\mathrm{d}\tau + \right. \\
&\left. (\mu_s^2 + \sigma_\xi^2) \int \mathrm{d}\boldsymbol{y} \int \mathrm{d}\tau \, \left[h_j^{\boldsymbol{y}}(\tau)\right]^2 \right] .
\end{aligned} \tag{A.16}$$

Since the squares are positive, so is the second derivative, and thus the extremum is a minimum.

A.4 Gaussian blurred signal

In this subsection we present an equation equivalent to (2.16) of the main text so as to derive an expression for a Gaussian blurred signal. As in (A.17), a realistic signal would fulfill some kind of Gaussian relation for the expectation value

$$\begin{aligned}
&\left\langle s^{\boldsymbol{x}}(t) s^{\boldsymbol{x}'}(t') \right\rangle \\
&= A \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{x}'|^2}{2\sigma_x^2} \right) \exp\left(-\frac{|t - t'|^2}{2\sigma_t^2} \right) \, .
\end{aligned} \tag{A.17}$$

For this case we can, analogously to appendix A.3, derive an equation like (2.16). Since for the signal the Gaussian correlations, however, replace the delta functions, e.g., in (A.10), integrals over space and time cannot be evaluated directly. Instead they can only be restricted to the region where the Gaussian is non-negligible. Denoting these temporal and spatial limits by ϵ_t and ϵ we can derive the analogue to (2.16), viz.,

$$\begin{aligned}
&\int \mathrm{d}\boldsymbol{\epsilon} \, \mathrm{d}\epsilon_t \; A \exp\left(-\frac{|\epsilon|^2}{2\sigma_x^2} \right) \exp\left(-\frac{|\epsilon_t|^2}{2\sigma_t^2} \right) h_j^{\boldsymbol{x}+\boldsymbol{\epsilon}}(t'' + \epsilon_t) \\
&= \sigma_\chi^2 \, l_j^{\boldsymbol{x}}(t'') + \sigma_\eta^2 \sigma_\xi^2 \int_{\substack{|\boldsymbol{y}|<y^{\max} \\ 0<\tau<t^{\max}}} \mathrm{d}\boldsymbol{y}\mathrm{d}\tau \; l_j^{\boldsymbol{x}}(t'') + \sigma_\xi^2 \sum_i \int \mathrm{d}\boldsymbol{y} \left[(h_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}}) \circ h_j^{\boldsymbol{y}} \right](t'') \\
&+ \sigma_\eta^2 \int_{\substack{|\boldsymbol{y}|<y^{\max} \\ 0<\tau<t^{\max}}} \mathrm{d}\boldsymbol{y}\mathrm{d}\epsilon\mathrm{d}\tau' A \exp\left(-\frac{{\epsilon_t}^2}{2\sigma_t^2} \right) l_j^{\boldsymbol{x}}(t'' + \epsilon_t) \\
&+ \sum_i \int \mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{\epsilon}\mathrm{d}\epsilon_t A \exp\left(-\frac{|\boldsymbol{\epsilon}|^2}{2\sigma_x^2} \right) \exp\left(-\frac{\epsilon_t^2}{2\sigma_t^2} \right) ((h_i^{\boldsymbol{y}} \star l_i^{\boldsymbol{x}}) \circ h_j^{\boldsymbol{y}+\boldsymbol{\epsilon}})(t'' + \epsilon_t) \, .
\end{aligned} \tag{A.18}$$

The effect of the additional remaining spatio-temporal integrals as compared to (2.16) is a smoothening of the final reconstruction. Not only is the value at a specific point in space and time $(\boldsymbol{y}, t'')$ taken into account but neighboring points in a nearly area surrounding it are included as well.

A.5 Pseudoinverse

A pseudoinverse matrix $\mathcal{B}$ of an arbitrary $m \times n$ matrix $\mathcal{A}$ with its elements $\in \mathbb{R}, \mathbb{C}$ is a generalization of the inverse matrix. Following Moore and Penrose [201, 223] the pseudoinverse $\mathcal{B}$ is defined by four equations,

$$\mathcal{A}\mathcal{B}\mathcal{A} = \mathcal{A} \ , \tag{A.19a}$$

$$\mathcal{B}\mathcal{A}\mathcal{B} = \mathcal{B} \ , \tag{A.19b}$$

$$\overline{(\mathcal{A}\mathcal{B})}^T = \mathcal{A}\mathcal{B} \ , \tag{A.19c}$$

$$\overline{(\mathcal{B}\mathcal{A})}^T = \mathcal{B}\mathcal{A} \ , \tag{A.19d}$$

where the overbar denotes the complex conjugation. The pseudoinverse is used for a matrix $\mathcal{A}$ that is of incomplete rang and therefore cannot be inverted directly but only due to auxiliary constructions (A.19 a, b). The last two equations (A.19 c, d) tell us that the product of matrix $\mathcal{A}$ with its pseudoinverse $\mathcal{B}$ is Hermitian.

One of the most famous applications of the pseudoinverse is to calculate the least-square solution of a system of linear equations.

$$\mathcal{A}\boldsymbol{x} = \boldsymbol{b} \rightarrow \hat{\boldsymbol{x}} = \mathcal{B}\boldsymbol{b} \tag{A.20}$$

Given a matrix $\mathcal{A}$ and a vector $\boldsymbol{b}$, the above solution $\hat{\boldsymbol{x}} = \mathcal{B}\boldsymbol{b}$ minimizes the Euclidean norm $||\mathcal{A}\hat{\boldsymbol{x}} - \boldsymbol{b}||$. If the inverse matrix exists, the pseudoinverse reduces to the normal inverse matrix. For a general review and applications the reader is referred to, e.g., [3].

B. Appendix: Lateral-line Perception

In the following, we briefly enlighten a pretended paradox: Why can we use complex analysis to calculate the two-dimensional flow in the (x, y)-plane whereas we cannot calculate analytically rotational-symmetric settings, e.g. flow around bodies, even though this corresponds also to a setup with only two coordinates involved.

B.1 Plane potential flow

We define the potential flow as a setting where the flow field $\boldsymbol{v}$ can be calculated by means of the velocity potential $\Phi : \mathbb{R}^n \to \mathbb{R}$ through $\boldsymbol{v} = \nabla\Phi$. In these settings $\operatorname{rot} \boldsymbol{v} = 0$; the flow field is vortex free.

A special case of potential flow is the *plane* potential flow. Here, one of the flow field components $\boldsymbol{v} = (v_x, v_y, v_z)$ is zero, e.g. $v_z = 0$. In addition, the remaining two components are independent of the coordinate which corresponds to the flow field component that is zero, e.g., $\boldsymbol{v} = (v_x(x, y, t), v_y(x, y, t), 0)$. In case of plane potential flow, the following statements hold true for the potential Φ and the stream function Ψ; the latter is defined by means of $v_x = \partial\Psi/\partial y$ and $v_y = -\partial\Psi/\partial x$,

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} = 0 = \frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} \ . \tag{B.1}$$

That is, both the potential Φ and the stream function Ψ solve the plane Laplacian equation and both are connected through the Cauchy-Riemann differential equations,

$$\frac{\partial\Phi}{\partial x} = \frac{\partial\Psi}{\partial y} \qquad \text{and} \qquad \frac{\partial\Phi}{\partial y} = -\frac{\partial\Psi}{\partial x} \ . \tag{B.2}$$

For that reason, we can define a complex potential $F(z) := \Phi(x, y) + i\Psi(x, y)$ with $x + iy = z \in \mathbb{C}$ where we can derive the complex velocity $w(z)$,

$$\frac{dF}{dz} = w(z) = v_x(x, y) - iv_y(x, y) \ . \tag{B.3}$$

As we can see above, it is also composed of real valued functions, namely, v_x and v_y. In doing so, we can analytically solve many hydrodynamic problems such as the flow around profiles.

B.2 Rotational-symmetric potential flow

The complex description of hydrodynamic settings is quite powerful as it provides an analytical access to various hydrodynamic problems. It is obvious that only two-dimensional issues can be addressed with a one-dimensional complex function since the complex coordinate $z = x + iy = r\exp(i\phi)$ is composed of two real coordinates x, y or r, ϕ. Intuitively, one can think of extending the complex formalism to three-dimensions in case of symmetries that effectively reduce the three-dimensional problem at hand to a two-dimensional setup. However, it is not possible to do so as we will demonstrate below explicitly for rotational-symmetric flow.

B. Appendix: Lateral-line Perception

The velocity field for rotational-symmetric flow is given by $\boldsymbol{v} = (v_r(r,z,t), v_z(r,z,t), 0)$. The continuity equation $\operatorname{div} \boldsymbol{v} = 0$ now reads

$$\frac{1}{r}\frac{\partial(r v_r)}{\partial r} + \frac{\partial c_z}{\partial z} = 0 \ . \tag{B.4}$$

The continuity equation holds true for a stream function defined by $r\, v_r = -\partial\Psi/\partial z$ and $r\, v_z = \partial\Psi/\partial r$.

Since we are studying potential flow the flow field has to be vortex free, viz., $\operatorname{rot}\boldsymbol{v} = 0$. The two components $(\operatorname{rot}\boldsymbol{v})_z$ and $(\operatorname{rot}\boldsymbol{v})_r$ of the corresponding vorticity $\operatorname{rot}\boldsymbol{v}$ are obviously zero. According to symmetry reasons, however, the third component $(\operatorname{rot}\boldsymbol{v})_\phi$ is not obviously zero. That is, we have to claim

$$(\operatorname{rot}\boldsymbol{v})_\phi = \frac{\partial v_r}{\partial z} - \frac{\partial v_z}{\partial r} = -\left[\frac{1}{r}\frac{\partial^2\Psi}{\partial z^2} + \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial\Psi}{\partial r}\right)\right] = 0 \ . \tag{B.5}$$

If we combine the continuity equation (B.4) and $\operatorname{rot}\boldsymbol{v} = 0$ (B.5) similarly to (B.1) we can derive two equations for both the velocity potential Φ and the stream function Ψ,

$$\frac{\partial^2\Phi}{\partial r^2} + \frac{1}{r}\frac{\partial\Phi}{\partial r} + \frac{\partial^2\Phi}{\partial z^2} = 0 = \frac{\partial^2\Psi}{\partial r^2} - \frac{1}{r}\frac{\partial\Psi}{\partial r} + \frac{\partial^2\Psi}{\partial z^2} \ . \tag{B.6}$$

As we can see the left hand side and the right hand side of (B.6) are not of the same type. The stream function does not fulfill the rotational-symmetric Laplacian equation, i.e, $\Delta\Psi = 0$.

In addition, the stream function Ψ and the velocity potential Φ are not connected through the Cauchy-Riemann differential equations (B.2) but

$$\frac{\partial\Phi}{\partial r} = -\frac{1}{r}\frac{\partial\Psi}{\partial z} \qquad \text{and} \qquad \frac{\partial\Phi}{\partial z} = \frac{1}{r}\frac{\partial\Psi}{\partial r} \ . \tag{B.7}$$

Thus we cannot interpret the stream function Ψ and the velocity potential Φ as the real and the imaginary part of analytical function $F(z) = \Phi(x,y) + i\Psi(x,y)$, in concreto, of the complex velocity potential $F(z)$.

In summary, the stream function Ψ and the velocity potential Φ do not solve the same differential equation (B.6) – the Laplacian equation – and, in addition, they are not connected by the Cauchy-Riemann differential equations. Against intuition, we therefore cannot use complex analysis to describe rotational-symmetric flow and thus three-dimensional setups even in case of this special symmetry.

C. Appendix: iMT

C.1 Optimal combination of two modalities

The firing profile of a neuronal map can be interpreted as the likelihood of finding an object at a specific position; see Sec. 1.2 in the introduction. In other words, neuronal maps encode position estimators by means of their firing profiles. A modality, for instance, audition A or vision V, can be represented through Gaussian firing profiles

$$f^{A/V}(x) := S_{A/V} \exp\left[-(x-\mu_{A/V})^2/(2\sigma_{A/V}^2)\right] \tag{C.1}$$

with map position x and estimated position, i.e., mean $\mu_{A/V}$, standard deviation $\sigma_{A/V}$, and maximal amplitude $S_{A/V}$ of the activity profiles. To combine two such estimators to a multimodal estimate μ_M we have to calculate a weighted sum of the mean values μ_V and μ_A leading to

$$\mu_M = g_V \mu_V + g_A \mu_A \tag{C.2}$$

with the non-negative weights g_V and g_A fulfilling $g_V + g_A = 1$. The variance σ_M^2 of the combined estimator reads

$$\sigma_M^2 = g_V^2 \sigma_V^2 + (1-g_V)^2 \sigma_A^2 \ . \tag{C.3}$$

Next, we have to choose the weight g_V so as to minimize the variance σ_M^2 with respect to g_V; that is, $\partial\sigma_M/\partial g_V = 0$. Solving the latter equation leads to $g_V = 1 - g_A = \sigma_A^2/(\sigma_A^2+\sigma_V^2)$ and finally to

$$\sigma_M^2 \quad = \quad [\sigma_A^2/(\sigma_A^2+\sigma_V^2)]^2 \sigma_V^2 + ([\sigma_V^2/(\sigma_A^2+\sigma_V^2)]^2 \sigma_A^2 = \sigma_V^2 \sigma_A^2/(\sigma_A^2+\sigma_V^2) \ . \tag{C.4}$$

We can insert the above result into (C.2) and calculate the optimal combination scheme

$$\mu_M = \sigma_A^2/(\sigma_A^2+\sigma_V^2)\ \mu_V + \sigma_V^2/(\sigma_A^2+\sigma_V^2)\ \mu_A \ . \tag{C.5}$$

For mathematical details regarding the derivation see elsewhere [276].

The above formula is able to describe results of neuronal processes but without explaining how they are achieved. In almost all cases it is practically unknown how the neuronal system computes an optimal estimate. In the context of neuronal maps, however, a simple multiplication of the activity profiles of the involved maps can neuronally realize the above optimal combination scheme. The multimodal activity profile is then given by

$$\begin{aligned} f_M(x) \quad &= \quad \left\{S_A \exp\left[-(x-\mu_A)^2/(2\sigma_A^2)\right]\right\}\ \left\{S_V \exp\left[-(x-\mu_V)^2/(2\sigma_V^2)\right]\right\} \\ &= \quad \underbrace{S_A S_V \exp\left[-1/2\,(\mu_A-\mu_V)^2\,/\,\left(\sigma_V^2+\sigma_A^2\right)\right]}_{\text{scaling factor}} \end{aligned} \tag{C.6}$$

$$\exp\left\{-1/2\,\underbrace{\left[\left(\sigma_V^2+\sigma_A^2\right)/\left(\sigma_A^2\sigma_V^2\right)\right]}_{1/\sigma_M^2}\left[x-\underbrace{\left(\sigma_V^2\mu_A+\sigma_A^2\mu_V\right)/\left(\sigma_V^2+\sigma_A^2\right)}_{\mu_M}\right]^2\right\} \ . \tag{C.7}$$

Standard deviation σ_M and mean μ_M of the multimodal profile are identical to the optimal values we have derived above; cf. (C.4) and (C.5).

C.2 Optimal combination of three modalities

We extend the optimal combination scheme for two modalities as derived in Sec. C.1 to three modalities. The multimodal estimate μ_M of three modalities, say, audition A, vision V, and somatosensory S represented by their mean values μ_A, μ_V, and μ_S with their standard deviations σ_A, σ_V, and σ_S is given by

$$\mu_M = \frac{\sigma_V^2\sigma_S^2}{\sigma_A^2\sigma_V^2+\sigma_A^2\sigma_S^2+\sigma_V^2\sigma_S^2}\,\mu_A + \frac{\sigma_A^2\sigma_S^2}{\sigma_A^2\sigma_V^2+\sigma_A^2\sigma_S^2+\sigma_V^2\sigma_S^2}\,\mu_V + \frac{\sigma_A^2\sigma_V^2}{\sigma_A^2\sigma_V^2+\sigma_A^2\sigma_S^2+\sigma_V^2\sigma_S^2}\,\mu_S\ , \tag{C.8}$$

a convex combination of μ_A, μ_V, and μ_S. The combined maximum likelihood estimate μ_M features the smallest possible standard deviation σ_M given by

$$\sigma_M = \frac{\sigma_A\sigma_V\sigma_S}{\sqrt{\sigma_A^2\sigma_V^2+\sigma_A^2\sigma_S^2+\sigma_V^2\sigma_S^2}}\ . \tag{C.9}$$

The distance between the multimodal estimate and a unimodal one, for instance audition, is given by

$$t|\mu_M-\mu_A| = \sigma_A^2\,\frac{\sigma_S^2|\mu_A-\mu_V|+\sigma_V^2|\mu_A-\mu_S|}{\sqrt{\sigma_A^2\sigma_V^2+\sigma_A^2\sigma_S^2+\sigma_V^2\sigma_S^2}}\ . \tag{C.10}$$

For distances between the multimodal estimate and μ_V or μ_S one can interchange A with V or S.

C.3 Model Parameters

In Table C.1 we summarize the model parameters as used in our numerical experiments (see main text).

parameter	value
number of map neurons	$N = 100$
learning synapses	
minimal strength	$\boldsymbol{J}_{\min} = 0.0$
maximal strength	$\boldsymbol{J}_{\max} = 0.25$
maximal initial strength	$(\boldsymbol{J}_0)_{\max} = 0.25$
strength of teacher synapses	$J^T = -1$
postsynaptic response time	
to input spike	$\tau^I = 10\,\mathrm{ms}$
to teacher spike	$\tau^T = 25\,\mathrm{ms}$
input amplitude	$S_{A/V/S} = 50\,\mathrm{s}^{-1}$
teacher amplitude	$A^T = 100\,\mathrm{s}^{-1}$
teacher width	$\sigma^T = 0.025$
shift maps – teacher	$\Delta = 0.2$
learning trial length	$\mathcal{T} = 0.5\,\mathrm{s}$
simulation time step	$\Delta t = 0.5\,\mathrm{ms}$
learning parameter	$\eta = 3 \cdot 10^{-6}$
weight change	
upon input spike	$w^{in} = 1.5$
upon output spike	$w^{out} = -10.0$

Table C.1: *Model parameters as used in our numerical experiments (see main text). Shape and parameters of the learning window are taken from [78] with differing or additional parameters as indicated above.*

Bibliography

[1] Abbott, L.F. & Nelson, S.B. (2000) Synaptic plasticity: taming the beast, Nat Neurosci **3** 1178–1183.

[2] Alais, D. & Burr, D. (2004) The Ventriloquist effect results from near-optimal bimodal integration, Curr Biol **14** 257–262.

[3] Albert, A. (1972) *Regression and the Moore-Penrose pseudoinverse* (Academic, New York).

[4] Amemiya, F., Goris, R.C., Masuda, Y., Kishida, R., et al. (1995) The surface architecture of snake infrared receptor organs, Biomed Res-Tokyo **16** 411–421.

[5] Amemiya, F., Nakano, M., Goris, R.C., Kadota, T., et al. (1999) Ultrastructure of the crotaline snake infrared pit receptors: Sem confirmation of tem findings, Anat Rec **254** 107–115.

[6] Amemiya, F., Tatsuo, U., Goris, R.C., & Kusunoki, Y.T. (1996) Ultrastructure of the crotaline snake infrared pit receptors: Sem confirmation of tem findings, Anat Rec **246** 135–146.

[7] Anastasio, T.J., Patton, P.E., & Belkacem-Boussaid, K. (2000) Using bayes rule to model multisensory enhancement in the superior colliculus, Neural Comp **12** 1165–1187.

[8] Bakken, G.S. (2007) A critique of thermal modeling of snake infrared reception by Jones, Lynn and Stone (2001), J Theor Biol **244** 167–168.

[9] Bakken, G.S. & Krochmal, A.R. (2007) The imaging properties and sensitivity of the facial pits of pitvipers as determined by optical and heat-transfer analysis, J Exp Biol **210** 2801–2810.

[10] Bartels, M., Münz, H., & Claas, B. (1990) Representation of the lateral line and electrosensory systems in the midbrain of the axolotl, *Ambystoma mexicanum*, J Comp Physiol A **167** 347–356.

[11] Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., et al. (2004) Unraveling multisensory integration: patchy organization within human STS multisensory cortex, Nat Neurosci **7** 1190–1192.

[12] Ben-Israel, A. & Greville, T.N.E. (2003) *Generalized Inverses: Theory and Applications* (Springer, New York, NY), 2nd edition.

[13] Benson, D.L., Colman, D.R., & Huntley, G.W. (2001) Molecules, maps and synapse specificity, Nat Rev Neurosci **2** 899–909.

[14] Bergan, J.F. & Knudsen, E.I. (2009) Visual modulation of auditory responses in the owl inferior colliculus, J Neurophysiol **101** 2924–2933.

[15] Bi, G.Q. (2002) Spatiotemporal specicity of synaptic plasticity: cellular rules and mechanisms, Biol Cybern **87** 319–332.

[16] Bi, G.Q. & Poo, M.M. (1998) Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type, J Neurosci **18** 10 464–10 472.

[17] Bi, G.Q. & Poo, M.M. (2001) Synaptic modification by correlated activity: Hebb's postulate revisited, Annu Rev Neurosci **24** 139–166.

[18] Bialek, W., Rieke, F., de Ruyter van Steveninck, R.R., & Warland, D. (1991) Reading a neural code, Science **252** 1854–1857.

[19] Bisti, S. & Maffei, L. (1974) Behavioural contrast sensitivity of the cat in various visual meridians, J Physiol **241** 201–210.

[20] Blauert, J. (1999) *Spatial Hearing* (MIT Press, Cambridge, MA).

[21] Bleckmann, H. (1994) *Reception of hydrodynamic stimuli in aquatic and semiaquatic animals* (Fischer).

[22] Boll, S.F. (1979) Suppression of acoustic noise in speech using spectral subtraction, IEEE T Acoust Speech **27** 113–20.

[23] Brauchi, S., Orio, P., & Latorre, R. (2004) Clues to understanding cold sensation: Thermodynamics and electrophysiological analysis of cold receptor trpm8, Proc Natl Acad Sci USA **101** 15 494–15 499.

[24] Braun, H.A. & Hensel, H. (1980) Static and dynamic discharge patterns of bursting cold fibers related to hypothetical receptor mechanisms, Pflug Arch Eur J Phy **386** 1–9.
[25] Braun, H.A., Schäfer, K., & Wissing, H. (1984) Theorien und Modelle zum übertragungsverhalten thermosensitiver Rezeptoren, Funkt Biol Med **3** 26–36.
[26] Braun, H.B., Dewald, M., Schäfer, K., Voigt, K., et al. (1999) Low-dimensional dynamics in sensory biology 2: Facial cold receptors of the rat, J Comput Neurosci **7** 17–32.
[27] Braun, H.B., Huber, M.T., Dewald, M., Schäfer, K., et al. (1998) Computer simulations of neuronal signal transduction: The role of nonlinear dynamics and noise, Int J Bifurcat Chaos **8** 881–889.
[28] Braun, W., Eckhardt, B., Braun, H.A., & Huber, M. (2000) Phase-space structure of a thermoreceptor, Phys Rev E **62** 6352–6361.
[29] Bregman, A.S. (1990) *Auditory Scene Analysis* (MIT Press, Cambridge, MA).
[30] Bresciani, J.-P., Dammeier, F., & Ernst, M.O. (2008) Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events, Brain Res Bul **75(6)** 753 – 760.
[31] Bresciani, J.-P., Ernst, M.O., Drewing, K., Bouyer, G., et al. (2005) Feeling what you hear: auditory signals can modulate tactile taps perception, Exp Brain Res **162** 172–180.
[32] van Brunt, B. (2000) *The calculus of variations* (Springer, Heidelberg).
[33] Bullock, T.H. & Diecke, F.P.J. (1956) Properties of an infra-red receptor, J Physiol **134** 47–87.
[34] Buonomano, D.V. & Merzenich, M.M. (2004) Cortical plasticity: From synapses to maps, Annu Rev Neurosci **21** 149–186.
[35] Bürck, M., Friedel, P., Sichert, A.B., Vossen, C., et al. (2010) Optimality in mono- and multisensory map formation, Biol Cybern **103** 1–20.
[36] Bürck, M. & van Hemmen, J.L. (2007) Modeling the cochlear nucleus: A site for monaural echo suppression?, J Acoust Soc Am **122(4)** 2226–2235.
[37] Bürck, M. & van Hemmen, J.L. (2009) Neuronal identification of signal periodicity by balanced inhibition, Biol Cybern **100** 261 –270.
[38] Campbell, A.L., de Cock Bunning, T., Stone, M. O., Church, D., et al. (1999) Surface ultrastructure of pit organ, spectacle, and non pit organ epidermis of infrared imaging boid snakes: A scanning probe and scanning electron microscopy study, J Struct Biol **126** 105–120.
[39] von Campenhausen, C., Riess, I., & Weissert, R. (1981) Detection of stationary objects by the blind cave fish *Anoptichtys jordani*, J Comp Physiol A **143** 369–374.
[40] Carr, C.E., Maler, L., & Sas, E. (1982) Peripheral organization and cetral projections of the electrosensory nerves in gymnotiform fish, J Comp Neurol **211** 139–153.
[41] Carr, R.W., Pianova, S., McKemy, D.D., & Brock, J.A. (2009) Action potential initiation in the peripheral terminals of cold-sensitive neurons innervating the guinea-pig cornea, J Physiol **587** 1249–1264.
[42] Casseday, J. H. & Neff, W. D. (1973) Psychophysical studies of monkey vision: III. Spatial luminance contrast sensitivity tests of macaque and human observers, J Acoust Soc Am **54(2)** 365–372.
[43] Catania, K.C., Leitch, D.B., & Gauthier, D. (2009) Function of the appendages in tentacled snakes (*Erpeton tentaculatus*), J Exp Biol **213** 359–367.
[44] Claas, B. (1994) Removal of eyes in early larval stages alters the response of the clawed toad, *Xenopus laevis*, to surface waves, Physiol Beh **56** 423–428.
[45] Clegg, J.C. (1968) *Calculus of variations* (Oliver and Boyd, Edinburgh).
[46] Clifton, R.K. (1987) Breakdown of echo suppression in the precedence effect, J Acoust Soc Am **82(5)** 1834–1835.
[47] de Cock Buning, T. (1984) A theoretical approach to the heat sensitive pit organs of snakes, J Theor Biol **111** 509–529.
[48] de Cock Buning, T., Terashima, S.I., & Goris, R.C. (1981) Crotaline pit organs analyzed as

warm receptors, Cell Mol Neurobiol **1** 69–85.

[49] de Cock Buning, T., Terashima, S.I., & Goris, R.C. (1981) Python pit organs analyzed as warm receptors, Cell Mol Neurobiol **1** 271–278.

[50] Cohen, M.L. (1977) Measurement of the thermal properties of human skin. A review, J Invest Dermatol **69** 333–338.

[51] Cohen, Y.E. & Knudsen, E.I. (1999) Maps versus clusters: different representations of auditory space in the midbrain and forebrain, Trends Neurosci **22** 128 – 135.

[52] Coombs, S., Anderson, R., Braun, C.B., & Grosenbaugh, M. (2007) The hydrodynamic footprint of benthic, sedentary fish in undirectional flow, J Acoust Soc Am **122** 1227–1237.

[53] Copelli, M., Roque, A.C., Oliveira, R.F., & Kinouchi, O. (2002) Physics of psychophysics: Stevens and Weber-Fechner laws are transfer functions of excitable media, Phys Rev E **65** 60 901.

[54] Ćurčić-Blake, B. & van Netten, S.M. (2006) Source location encoding in the fish lateral line canal, J Exp Biol **209** 1548–1559.

[55] Dan, Y. & Poo, M.M. (2004) Spike timing-dependent plasticity of neural circuits, Neuron **44** 23–30.

[56] Davison, A.P. & Frégnac, Y. (2006) Learning cross-modal spatial transformations through spike timing-dependent plasticity, J Neurosci **26** 5604–5615.

[57] De Valois, R., Morgan, H., & Snodderly, D. (1974) Psychophysical studies of monkey vision: Iii. spatial luminance contrast sensitivity tests of macaque and human observers, Vision Res **14** 75–81.

[58] Dehaene, S. & Cohen, L. (2007) Cultural recycling of cortical maps, Neuron **56** 384–398.

[59] Deneve, S., Latham, P.E., & Pouget, A. (1999) Reading population codes: a neural implementation of ideal observers, Nat Neurosci **2(8)** 740–745.

[60] Deneve, S., Latham, P.E., & Pouget, A. (2001) Efficient computation and cue integration with noisy population codes, Nat Neurosci **4(8)** 826–831.

[61] Deneve, S. & Pouget, A. (2004) Bayesian multisensory integration and cross-modal spatial links, J Physiology-Paris **98** 249–258.

[62] Dräger, U.C. & Hubel, D.H. (1975) Responses to visual stimulation and relationship between visual, auditory, and somatosensory inputs in mouse superior colliculus, J Neurphysiol **38** 690–713.

[63] Ebert, J. & Westhoff, G. (2006) Behavioural examination of the infrared sensitivity of rattlesnakes (*Crotalus atrox*), J Comp Physiol A **192** 941–947.

[64] Eckertf, E.R.G. & Carlson, Walter O. (1961) Natural convection in an air layer enclosed between two vertical plates with different temperatures, Int J Heat Mass Transf **2** 106 – 120.

[65] Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., et al. (1988) Coherent oscillations: A mechanism of feature linking in the visual cortex?, Biol Cybern **60** 121–130.

[66] Ephraim, Y. & Malah, D. (2004) Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator, IEEE T Acoust Speech **32** 1109–21.

[67] Ernst, M.O. & Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion, Nature **415** 429–433.

[68] Ernst, M.O., Banks, M.S., & Bülthoff, H. (2000) Touch can change visual slant perception, Nat Neurosci **3** 69–73.

[69] Ernst, M.O. & Bülthoff, H. (2004) Merging the senses into a robust percept, Trends Cogn Sci **8** 162–169.

[70] Evans, L.C. (1998) *Partial Differential Equations*, volume 19 of *Graduate Series in Mathematics* (American Mathematical Society, Providence, RI).

[71] Faisal, A.A., Selen, L.P.J., & Wolpert, D.M. (2008) Noise in the nervous system, Nat Rev Neurosci **9** 292–303.

[72] Fiala, D., Lomas, K.J., & Stohrer, M. (1999) A computer model of human thermoregulation for a wide range of environmental conditions: the passive system, J Appl Physiol **87** 1957–1972.

[73] Flanagan, J.L. & Lummis, R.C. (1970) Signal processing to reduce multipath distortion in small rooms, J Acoust Soc Am **47(6)** 1475–81.

[74] Franosch, J.-M.P., Lingenheil, M., & van Hemmen, J.L. (2005) How a frog can learn what is where in the dark, Phys Rev Lett **95** 078 106.

[75] Franosch, J.-M.P., Sichert, A.B., Suttner, M.D., & van Hemmen, J.L. (2005) Estimating position and velocity of a submerged moving object by the clawed frog *Xenopus* and by fish—A cybernetic approach, Biol Cybern **93** 231–238.

[76] Franosch, J.-M.P., Sobotka, M.C., Elepfandt, A., & van Hemmen, J.L. (2003) Minimal model of prey localization through the lateral-line system, Phys Rev Lett **91** 158 101.

[77] Frens, M.A. & van Opstal, A.J. (1998) Visual-auditory interactions modulate saccade-related activity in monkey superior colliculus, Brain Res **46** 211–224.

[78] Friedel, P. & van Hemmen, J.L. (2008) Inhibition, not excitation, is the key to multimodal sensory integration, Biol Cybern **98** 597–618.

[79] Furuya, K. & Kataoka, A. (2007) Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction, IEEE T Audio Speech **15** 1579–91.

[80] Galan, R.F., Ermentrout, G.B., & Urban, N.N. (2007) Reliability and stochastic synchronization in type i vs. type ii neural oscillators, Neurocomputing **70** 2102–2106.

[81] Galan, R.F., Fourcaud-Trocme, N., Ermentrout, G.B., & Urban, N.N. (2006) Correlation-induced synchronization of oscillations in olfactory bulb neurons, J Neurosci **26** 3646–3655.

[82] Gancarz, G. & Grossberg, S. (1999) A neural model of saccadic eye movement control explains task-specific adaptation, Vision Res **39** 3123–3143.

[83] Gardner, J.L., Merriam, E.P., Movshon, J.A., & Heeger, D.J. (2008) Maps of visual space in human occipital cortex are retinotopic, not spatiotopic, J Neurosci **28** 3988–3999.

[84] Geisler, C.D. (1990) *From Sound to Synapse: Physiology of the Mammalian Ear* (Oxford University Press, Oxford).

[85] Gelfand, I.M. & Fomin, S.V. (1963) *Calculus of variations* (Prentice-Hall, Englewood Cliffs, NY).

[86] Gelfand, J.J., Pearson, J.C., Spence, C.D., & Sullivan, W.E. (1988) *Multisensor integration in biological systems* (IEEE Press, New York).

[87] Gerstner, W., Kempter, R., van Hemmen, J.L., & Wagner, H. (1996) A neuronal learning rule for sub-millisecond temporal coding, Nature **383** 76–78.

[88] Gerstner, W. & Kistler, W. (2002) *Spiking Neuron Models* (Cambridge University Press, Cambridge).

[89] Gerstner, W., Kreiter, A.K., Markram, H., & Herz, A.V.M. (1997) Neural codes: Firing rates and beyond, Proc Natl Acad Sci USA **94** 12 740 –12 741.

[90] Gillespie, B.W., Malvar, H.S., & Florencio, D.A.F. (2001) Speech dereverberation via maximum-kurtosis subband adaptive filltering, Proc IEEE Int Conf Acoust Speech Signal process **6** 3701–4.

[91] Gonzalez, R.C. & Woods, R.E. (2002) *Digital Image Processing* (Prentice Hall, Upper Saddle River, NJ), 2nd edition.

[92] Goris, R.C., Nakano, M., Atobe, Y., Kadota, T., et al. (2000) Nervous control of blood flow microkinetics in the infrared organs of pit vipers, Auton Neurosci-Basic **84** 98–106.

[93] Goris, R.C. & Nomoto, M. (1967) Infrared Reception in Oriental Crotaline Snake, Comp Biochem Physiol **23** 879–892.

[94] Goris, R.C. & Terashima, S.I. (1973) Central Response to Infra-red Stimulation of the Pit Receptors in a Crotaline Snake, Trimeresurus Flavoviridis, J Exp Biol **58** 59–76.

[95] Goulet, J., Engelmann, J., Chagnaud, B.P., Franosch, J.-M.P., et al. (2008) Object localization through the lateral line system of fish: Theory and experiment, J Comp Physiol A **194** 1–17.

[96] Grace, M.S., Woodward, O.M., Church, D.R., & Calisch, G. (2001) Prey targeting by the infrare-imaging snake *Python*: effects of experimental and congenital visual deprivation, Behav Brain Res **119** 23–31.

[97] Gracheva, E.O., Ingolia, N.T., Kelly, Y.M., Cordero-Morales, J.F., et al. (2010) Molecular basis of infrared detection by snakes, Nature **464** 1006–1012.
[98] Gray, C.M., König, P., Engel, A.K., & Singer, W. (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties, Nature **338** 334–337.
[99] Grossberg, S., Roberts, K., Aguilar, M., & Bullock, D. (1997) A neural model of multimodal adaptive saccadic eye movement control by superior colliculus, J Neurosci **24** 9706–9725.
[100] Grothe, B., Covey, E., & Casseday, J.H. (2001) Medial superior olive of the big brown bat: Neuronal responses to pure tones, amplitude modulations, and pulse trains, J Neurohysiol **86** 2219–2230.
[101] Grothe, B. & Neuweiler, G. (2000) The function of the medial superior olive in small mammals: temporal receptive fields in auditory analysis, J Comp Physiol A **186** 413–423.
[102] Gu, Y., Angelaki, D.E., & DeAngelis, G.C. (2008) Neural correlates of multisensory cue integration in macaque MSTd, Nat Neurosci **11** 1201–1210.
[103] Gutfreund, Y., Zheng, W., & Knudsen, E.I. (2002) Gated visual input to the central auditory system, Science **297** 1556–1559.
[104] Hafed, Z.M., Goffart, L., & Krauzlis, R.J. (2008) Superior colliculus inactivation causes stable offsets in eye position during tracking, J Neurosci **28** 8124–8137.
[105] Hänsel, E. & Schmidt, G. (editors) (2004) *Acoustic Echo and Noise Control – A Practical Approach* (Wiley, Hoboken).
[106] Harmening, W.M., Nikolay, P., Orlowski, J., & Wagner, H. (2009) Spatial contrast sensitivity and grating acuity of barn owls, J Vision **9(7)** 1–12.
[107] Harris, G.G. (1964) Considerations on the physics of sound production by fishes, in W.N. Tavolga (editor), *Marine Bio-Acoustics*, pp. 233–247 (Pergamon Press, Oxford).
[108] Harris, G.G., Flanagan, J.L., & Watson, B.J. (1963) Binaural interaction of a click with a click pair, J Acoust Soc Am **35** 672–678.
[109] Hartline, P.H., Kass, L., & Loop, M.S. (1978) Merging of modalities in the optic Tectum: Infrared and visual integration in rattlesnakes, Science **199** 1225–1229.
[110] Hasan, K.M., Salahuddin, S., & Khan, R.M. (2004) Reducing signal-bias from MAD estimated noise level for DCT speech enhancement, Signal Process **84** 151–62.
[111] Hassan, E.l.-S. (1985) Mathematical analysis of the stimulus for the lateral line organ, Biol Cybern **52** 23–36.
[112] Hebb, D.O. (1949) *The organization of behavior: a neuropsychological theory* (Wiley, New York).
[113] Helbig, H.B. & Ernst, M.O. (2007) Optimal integration of shape information from vision and touch, Exp Brain Res **179** 595–606.
[114] Helmholtz, H. (1879) *Die Thatsachen in der Wahrnehmung* (August Hirschwald, Berlin).
[115] van Hemmen, J.L. (2001) Theory of synaptic plasticity, in F Moss & S Gielen (editors), *Handbook of Biophysics*, pp. 771–823 (Elsevier, Amsterdam), 4th edition.
[116] van Hemmen, J.L. (2002) The map in your head: How does the brain represent the outside world?, Chem Phys Chem **3** 291–298.
[117] Hensel, H. (1975) Static and dynamic activity of warm receptors in Boa constrictor, Pflügers Archiv **353** 191–199.
[118] Hensel, H. & Wurster, R.D. (1970) Static properties of cold receptors in nasal area of cats, J Neurophysiol **33** 271–275.
[119] Herz, A.V.M., Gollisch, T., Machens, C.K., & Jaeger, D. (2006) Modeling Single-Neuron dynamics and Computations: A Balance of Detail and Abstraction, Science **314** 80–85.
[120] Hilis, J.M., Ernst, M.O., Banks, M.S., & Landy, M.S. (2002) Combining sensory information: Mandatory fusion within, but not between, senses, Science **298** 1627–1630.
[121] Hisajima, T., Kishida, R., Atobe, Y., Nakano, M., et al. (2002) Distribution of myelinated and unmyelinated nerve fibers and their possible role in blood flow control in crotaline snake

infrared receptor organs, J Comp Neurol **449** 319–329.
[122] Hötting, K., Rösler, F., & Röder, B. (2004) Altered auditory-tactile interactions in congenitally blind humans: An event-related potential study, Exp Brain Res **159** 370–381.
[123] Hubel, D.H. & Wiesel, T.N. (1977) Functional architecture of macaque monkey visual cortex, Proc Roy Soc Lond B **198** 1–59.
[124] Huber, M.T. & Braun, H.A. (2006) Stimulus-response curves of a neuronal model for noisy subthreshold oscillations and related spike generation, Phys Rev E **73** 0419 291–04192 910.
[125] Hughes, A. (1977) The topography of vision in mammals of contrasting life style: Comparative optics and retinal organization., in F. Crescitelli (editor), *Handbook of sensory physiology*, volume VII/5, chapter 8, pp. 613–756 (Springer, Berlin Heidelberg New York).
[126] Hürlimann, F., Kiper, D.C., & Carandini, M. (2002) Testing the bayesian model of perceived speed, Vision Res **42** 2253–2257.
[127] Hyde, P.S. & Knudsen, E.I. (2001) A topographic instructive signal guides the adjustment of the auditory space map in the optic tectum, J Neurosci **21** 8586–8593.
[128] Ito, M. (1989) Long-term depression, Annu Rev Neurosci **12** 85–102.
[129] Itti, L. & Koch, C. (2001) Computational modeling of visual attention, Nat Rev Neurosci **2(3)** 194–203.
[130] Järvilehto, M. (1985) The eye: Vision and perception, in G.A. Kerkut & L.I. Gilbert (editors), *Nervous System: Sensory*, volume 6 of *Comprehensive Insect Physiology, Biochemistry, and Pharmacology*, chapter FIND CHAPTER, pp. 355–429 (Pergamon Press, Oxford).
[131] Jazayeri, M. & Movshon, J.A. (2006) Optimal representation of sensory information by neural populations, Nat Neurosci **9** 690–696.
[132] Jeffress, L.A. (1948) A place theory of sound localization, J Comp Physiol Psychol **41** 35–39.
[133] Jiang, B., Treviño, M., & Kirkwood, A. (2007) Sequential development of long-term potentiation and depression in different layers of the mouse visual cortex, J Neurosci **27** 9648–9652.
[134] Jiang, W., Wallace, M.T., Jiang, H., Vaughan, J.W., et al. (2001) Two Cortical Areas Mediate Multisensory Integration in Superior Colliculus Neurons, J Neurophysiol **85** 506–522.
[135] Johnson, D.H. & Dudgeon, D.E. (1993) *Array Signal Processing: Concepts and Techniques* (Prentice-Hall, Upper Saddle River, NJ).
[136] Johnson, K.O., Hsiao, S.S., & Yoshioka, T. (2002) Neural coding and the basic law of psychophysics, Neuroscientist **8** 111–121.
[137] Jones, B.S., Lynn, W.F., & Stone, M.O. (2001) Thermal modeling of snake infrared reception: Evidence for limited detection range, J Theor Biol **209** 201–211.
[138] Jost, J. & Li-Jost, X. (1998) *Calculus of variations* (Cambridge University, Cambridge).
[139] Kaas, J.H. & Collins, C.E. (2004) The resurresction of multisensory cortex in primates: Connection patterns that integrate modalities, in G. Calvert, C. Spence, & B.E. Stein (editors), *The Handbook of Multisensory Processes*, chapter 17, pp. 285–293 (MIT Press, Cambridge, MA).
[140] Kadunce, D.C., Vaughan, J.W., Wallace, M.T., Benedek, G., et al. (1997) Mechanisms of Within- and Cross-Modality Suppression in the Superior Colliculus, J Neurophysiol **78** 2834–2847.
[141] Kailath, T. (1974) A view of three decades of linear filtering theory, IEEE T Inform Theory **20** 146–181.
[142] Kalmijn, A.J. (1988) Hydrodynamic and acoustic field detection, in J. Atema, R.R. Fay, A.N. Popper, & W.N. Tavolga (editors), *Sensory Biology of Aquatic Animals*, chapter 4, pp. 83–130 (Springer, New York, NY).
[143] Kandel, E.R., Schwartz, J.H., & Jessell, T.M. (2000) *Principles of neural science* (McGraw-Hill, New York), 4th (international) edition.
[144] Kay, S.M. (1993) *Fundamentals of Statistical Signal Processing* (Prentice Hall, Upper Saddle River, NJ).
[145] Kempter, R., Gerstner, W., & van Hemmen, J.L. (1999) Hebbian learning and spiking neurons,

Phys Rev E **59** 4498–4514.
[146] Kempter, R., Leibold, C., Wagner, H., & van Hemmen, J.L. (2001) Formation of temporal-feature maps by axonal propagation of synaptic learning, Proc Natl Acad Sci U S A **98** 4166–4171.
[147] King, A.J. (1999) Sensory experience and the formation of a computational map of auditory space in the brain, Bioessays **21** 900–911.
[148] King, A.J. (2009) Visual influences on auditory spatial learning, Phil Trans R Soc B **364** 331–339.
[149] King, A.J. & Hutchings, M.E. (1987) Spatial response properties of acoustically responsive neurons in the superior colliculus of the ferret: A map of auditory space, J Neurophysiol **57** 596–624.
[150] King, A.J., Hutchings, M.E., Moore, D.R., & Blakemore, C. (1988) Developmental plasticity in the visual and auditory representations in the mammalian superior colliculus, Nature **332** 73–76.
[151] King, A.J. & Palmer, A.R. (1983) Cells responsive to free-field auditory stimuli in guinea-pig superior colliculus: Distribution and response properties, J Physiol **342** 361–381.
[152] King, A.J. & Parsons, C.H. (1999) Improved auditory spatial acuity in visually deprived ferrets, Eur J Neurosci **11(11)** 3873–3883.
[153] Knierim, J.J. & van Essen, D.C. (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey, J Neurophysiol **67** 961–980.
[154] Knudsen, E.I. (1982) Auditory and visual maps of space in the optic tectum of the owl, J Neurosci **2** 1177–1194.
[155] Knudsen, E.I. (1999) Mechanisms of experience-dependent plasticity in the auditory localization pathway of the barn owl, J Comp Physiol A **185** 305–321.
[156] Knudsen, E.I. (2002) Instructed learning in the auditory localization pathway of the barn owl, Nature **417** 322–328.
[157] Knudsen, E.I., Blasdel, G.G., & Konishi, M. (1979) Sound localization by the barn owl (Tyto alba) measured with the search coil technique, J Comp Physiol A **133** 1–11.
[158] Knudsen, E.I. & Brainard, M.S. (1991) Visual instruction of the neural map of auditory space in the developing optic tectum, Science **253** 85–87.
[159] Knudsen, E.I. & Knudsen, P.F. (1985) Vision guides the adjustment of auditory localization in young barn owls, Science **230** 545–548.
[160] Knudsen, E.I. & Knudsen, P.F. (1989) Vision calibrates sound localization in developing barn owls, J Neurosci **9** 3306–3313.
[161] Knudsen, E.I., du Lac, S., & Esterly, S.D. (1987) Computational maps in the brain, Annu Rev Neurosci **10** 41–65.
[162] Kobayashi, S., Amemiya, F., Kishida, R., Goris, R.C., et al. (1995) Somatosensory and visual correlation in the optic tectum of a python, *Python regius*: a horseradish peroxidase and golgi study, Neurosci Res **22** 315–323.
[163] Kobayashi, S., Kishida, R., Goris, R.C., Yoshimoto, M., et al. (1992) Visual and infrared input to the same dendrite in the tectum opticum of the python, *Python regius*: electron-microscopic evidence, Brain Res **597** 350–352.
[164] Koch, C. (1999) *Biophysics of Computation* (Oxford University Press, Oxford).
[165] Körding, K. & Wolpert, D.M. (2004) Bayesian integration in sensorimotor learning, Nature **427** 244–247.
[166] Krauzlis, R.J., Basso, M.A., & Wurtz, R.H. (1997) Shared motor error for multiple eye movements, Science **276** 1693–1695.
[167] Krauzlis, R.J., Liston, D., & Carello, C.D. (2004) Target selection and the superior colliculus: Goals, choices and hypotheses, Vision Res **44** 1445–1451.
[168] Krochmal, A.R., Bakken, G.S., & LaDuc, T. (2004) Heat in evolutions kitchen: evolutionary perspectives on the functions and origin of the facial pit of pitvipers (viperidae: Crotalinae),

J Exp Biol **207** 4231–4238.
[169] Krueger, L.E. (1989) Reconciling Fechner and Stevens: Toward a unified psychophysical law, Behav Brain Sci **12** 251–320.
[170] Kuttruff, H. (1991) *Room Acoustics* (Elsevier Applied Science, London and New York), 3rd edition.
[171] Lamb, H. (1932) *Hydrodynamics* (Cambridge University Press), 6th ed. edition.
[172] Laming, D. (1997) *The Measurement of Sensation* (Oxford University, Oxford).
[173] Lebart, K., Boucher, J. M., & Denbigh, P. N. (2001) A new method based on spectral subtraction for speech dereverberation, Acta Acoustica **87(3)** 359–366.
[174] Leibold, C., Kempter, R., & van Hemmen, J.L. (2001) Temporal map formation in the barn owls brain, Phys Rev Lett **87** 248 101.
[175] Li, Z. (2002) A saliency map in primary visual cortex, Trends Cogn Sci **6(1)** 9–16.
[176] Liman, E.R. (2006) Thermal gating of TRP ion channels: Food for thought?, Sci Stke **326**.
[177] Linden, D.J. & Connor, J.A. (1995) Long-term synaptic depression, Annu Rev Neurosci **18** 319–357.
[178] Lingenheil, M. (2004) *Theorie der Beuteortung beim Krallenfrosch*, Master's thesis, Technische Universität München.
[179] Luksch, H. (2008) Sensorimotor integration: Optic tectum, in M. D. Binder, N. Hirokawa, U. Windhorst, & M. C. Hirsch (editors), *Encyclopedia of Neuroscience* (Springer, New York, NY).
[180] Luksch, H., Gauger, B., & Wagner, H. (2000) A candidate pathway for a visual instructional signal to the barn owl's auditory system, J Neurosci **20** RC70.
[181] Luo, L. & Flanagan, J.G. (2007) Development of continuous and discrete neural maps, Neuron **56** 284–300.
[182] Lynn, G.W. (1931) The structure and function of the facial pit of the pit vipers, Am J Anat **49** 97–133.
[183] Ma, W.J., Beck, J.M., Latham, P.E., & Pouget, A. (2006) Bayesian inference with probabilistic population codes, Nat Neurosci **9** 1432–1438.
[184] Mach, E. (1866) Über die physiologische Wirkung räumlich vertheilter Lichtreize, Sitzungsber Akad Wiss Wien II **54** 393–408.
[185] Madison, D.V., Malenka, R.C., & Nicoll, R.A. (1991) Mechanisms underlying long-term potentiation of synaptic transmission, Annu Rev Neurosci **14** 379–397.
[186] Magosso, E., Cuppinia, C., Serino, A., Di Pellegrino, G., et al. (2008) A theoretical study of multisensory integration in the superior colliculus by a neural network model, Neural Networks **21** 817–829.
[187] Manley, G.A., Köppl, C., & Konishi, M. (1988) A neural map of interaural intensity differences in the brainstem of the barn owl, J Neurosci **8** 2665–2676.
[188] Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs, Science **275** 213–215.
[189] McKemy, D.D., Neuhausser, W.M., & Julius, D. (2002) Identification of a cold receptor reveals a general role for TRP channels in thermosensation, Nature **416** 52–58.
[190] McLaughlin, T. & O'Leary, D.D.M (2005) Molecular gradients and development of retinotopic maps, Annu Rev Neurosci **28** 327–355.
[191] Meszler, R.M., Auker, C.R., & Carpenter, D.O (1981) Fine structure and organization of the infrared receptor relay, the lateral descending nucleus of the trigeminal nerve in pit vipers, J Comp Neurol **196** 571–584.
[192] Middlebrooks, J.C. & Knudsen, E.I. (1984) A neural code for auditory space in the cat's superior colliculus, J Neurosci **4** 2621–2634.
[193] Miller, K. (1970) Least squares methods for ill-posed problems with a prescribed bound, SIAM J Math Anal **1** 52–74.
[194] Mills, A.W. (1958) On the minimum audible angle, J Acoust Soc Am **30** 237.

[195] Miyoshi, M. & Kaneda, Y. (1988) Inverse Filtering of Room Acoustics, IEEE T Acoust Speech **36(2)** 145–152.

[196] Moiseenkova, V., Bell, B., Motamedi, M., Wozniak, E., et al. (2003) Wide-band spectral tuning of heat receptors in the pit organ of the copperhead snake (*Crotalinae*), Am J Physiol **284** 598–606.

[197] Molenaar, G.J. (1978) The sensory trigeminal system of a snake in the possession of infrared receptors: II. The central projections of the trigeminal nerve, J Comp Neurol **179** 137–152.

[198] Molenaar, G.J. (1992) Anatomy and physiology of infrared sensitivity of snakes, in C. Gans & P.S. Ulinski (editors), *Sensorimotor Integration*, volume 17 of *Biology of the Reptilia*, chapter 5, pp. 367–453 (University of Chicago Press, Chicago, IL).

[199] Molenaar, G.J. & Fizaan-Oostveen, J.L.F.P. (1980) Ascending projections from the lateral descending and common sensory trigeminal nuclei in python, J Comp Neurol **189** 555–572.

[200] Mooney, R.D., Klein, B.G., & Rhoades, R.W. (1987) Effects of altered visual input upon the development of the visual system and somatosensory representations in the hamster's superior colliculus, Neurosci **20** 537–555.

[201] Moore, E.H. (1920) On the reciprocal of the general algebraic matrix, B Am Math Soc **26** 394–395.

[202] Morgan, M.L., DeAngelis, G.C., & Angelaki, D.E. (2008) Multisensory integration in macaque visual cortex depends on cue reliability, Neuron **59** 662–673.

[203] Murphey, R.K. (1973) Mutual inhibition and the organization of a non-visual orientation in *Notonecta*, J Comp Physiol A **84** 31–40.

[204] Mysore, S.P. & Quartz, S.R. (2005) Modeling structural plasticity in the barn owl auditory localization system with a spike-time dependent hebbian learning rule, Proc IEEE Int Joint Conf Neural Networks **5** 2766–2771.

[205] Nakano, M., Atobe, Y., Goris, R.C., Yazama, F., et al. (2000) Ultrastructure of the capillary pericytes and the expression of smooth muscle α-actin and desmin in the snake infrared sensory organs, Anat Rec **260** 299–307.

[206] Neely, Stephen T. & Allen, Jont B. (1979) Invertibility of a room impulse response, J Acoust Soc Am **66(1)** 165–9.

[207] Nelson, D.A. (1998) Invited editorial on "penns 1948 paper revisited, J Appl Physiol **85** 2–3.

[208] Nelson, M.E. & Bower, J.M. (1990) Brain maps and parallel computers, Trends Neurosci **13** 403 – 408.

[209] Newman, E.A., Gruberg, E.R., & Hartline, P.H. (1980) The infrared trigemino-tectal pathway in the rattlesnake and in the python, J Comp Neurol **191** 465–477.

[210] Newman, E.A. & Hartline, P.H. (1981) Integration of visual and infrared information in bimodal neurons of the rattlesnake optic tectum, Science **213** 789–791.

[211] Newman, E.A. & Hartline, P.H. (1982) The infrared "vision" of snakes, Sci Am **246(3)** 98–107.

[212] Nilius, B., Talavera, K., Owsianik, G., Prenen, J., et al. (2005) Gating of TRP channels: a voltage connection?, J Physiol **12** 35–44.

[213] Norwich, K.H. & Wong, W. (1997) Unification of psychological phenomena: The complete form of Fechner's law, Perc Psychophys **59** 929–940.

[214] Olsen, J.F., Knudsen, E.I., & Esterly, S.D. (1989) Neural maps of interaural time and intensity differences in the optic tectum of the barn owl, J Neurosci **9** 2591–2605.

[215] Oppenheim, A.V., Schafer, R.W., & Stockham, T.G. (1968) Nonlinear filtering of multiplied and convolved signals, Proc IEEE **56(8)** 1264–91.

[216] van Opstal, A.J. & Munoz, D.P. (2004) Auditory-visual interactions subserving primate gaze orienting, in G. Calvert, C. Spence, & B.E. Stein (editors), *The Handbook of Multisensory Processes*, chapter 23, pp. 373–393 (MIT, Cambridge, MA).

[217] Oğuztöreli, M.N. & Caelli, T.M. (1985) An inverse problem in neural processing, Biol Cybern **53** 239–245.

[218] Panton, R.L. (2005) *Incompressible flow* (Wiley), 3rd edition.
[219] Pappas, T.C., Motamedi, M., & Christensen, B.N. (2004) Unique temperature-activated neurons from pit viper thermosensors, Am J Physiol Cell Physiol **287** 1219–1228.
[220] Patton, P.E., Belkacem-Boussaid, K., & Anastasio, T.J. (2002) Multimodality in the superior colliculus: an information theoretic analysis, Cognitive Brain Res **14** 10–19.
[221] Pecka, M., Zahn, T. P., Saunier-Rebori, B., Siveke, I., et al. (2007) Inhibiting the inhibition: a neuronal network for sound localization in reverberant environments., J Neurosci. **27(7)** 1782–1790.
[222] Pennes, H.H. (1948) Analysis of tissue and arterial blood temperatures in the resting human forearm, J Appl Physiol **1** 93–122.
[223] Penrose, R. (1955) A generalized inverse for matrices, Cambridge Philos Soc **51** 406–413.
[224] Perrault, T.J., Vaughan, J.W., Stein, B.E., & Wallace, M.T. (2005) Superior Colliculus Neurons Use Distinct Operational Modes in the Integration of Multisensory Stimuli, J Neurophysiol **93** 2575–2586.
[225] Pickles, J.O. (1988) *An Introduction to the Physiology of hearing* (Academic Press, London), 2nd edition.
[226] Planck, M. (1901) Ueber das gesetz der energieverteilung im normalspectrum, Ann Phys **309** 553–563.
[227] Pouget, A., Dayan, P., & Zemel, R.S. (2003) Inference and computation with population codes, Annu Rev Neurosci **26** 381–410.
[228] Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, B.P. (2007) *Numerical Recipes: The Art of Scientific Computing, 3rd ed.* (Cambridge University, Cambridge).
[229] Puetter, R.C., Gosnell, T.R.., & Amos Yahil, A. (2005) Digital image reconstruction: Deblurring and denoising, Annu Rev Astron Astrophys **43** 139–194.
[230] Putzar, L., Goerendt, I., Lange, K., Rösler, F., et al. (2007) Early visual deprivation impairs multisensory interactions in humans, Nat Neurosci **10** 1243–1245.
[231] Rauschecker, J.P. & Kniepert, U. (1994) Auditory localization behaviour in visually deprived cats, Eur J Neurosci **6(1)** 149–160.
[232] de Rivaz, Peter & Kingsbury, Nick (2001) Bayesian image deconvolution and denoising using complex wavelets, IEEE Image Proc **2** 273–276.
[233] Röder, B., Rösler, F., & Spence, C. (2004) Early vision impairs tactile perception in the blind, Curr Biol **14** 121–124.
[234] Roder, B., Teder-Salejarvi, W., Sterr, A., Rosler, F., et al. (1999) Improved auditory spatial tuning in blind humans, Nature **400** 162–166.
[235] Roper, P., Bressloff, P.C., & Longtin, A. (2000) A phase model of temperature-dependent mammalian cold receptor, Neural Comput **12** 1067–1093.
[236] Rosenfeld, D. (2002) New Approach to Gridding Using Regularization, Magn Reson Med **48** 193–202.
[237] Rowland, B.A., Quessy, S., Stanford, T.R., & Stein, B.E. (2007) Multisensory integration shortens physiological response latencies, J Neurosci **27** 5879–5884.
[238] Rowland, B.A., Stanford, T.R., & Stein, B.E. (2007) A bayesian model unifies multisensory spatial localization with the physiological properties of the superior colliculus, Exp Brain Res **180** 153–161.
[239] Rowland, B.A., Stanford, T.R., & Stein, B.E. (2007) A model of the neural mechanisms underlying multisensory integration in the superior colliculus, Perception **36(10)** 1431 – 1443.
[240] Rucci, M., Edelman, G.M., & Wray, J. (1999) Adaption of orienting behavior: From the barn owl to a robotic system, IEEE T Robot **15** 96–110.
[241] Rucci, M., Tononi, G., & Edelman, G.M. (1997) Registration of neural maps through value-dependent learning: Modeling the alignment of auditory and visual maps in the barn owl‘s optic tectum, J Neurosci **17** 334–352.
[242] Sarkar, T.K., Weiner, D.D., & Jain, V.K. (1981) Some mathematical Considerations in

Dealing with the Inverse Problem, IEEE T Antenn Propag **29** 373–379.
[243] Schäfer, K., Braun, H.A., & Hensel, H. (1982) Static and dynamic activity of cold receptors at various calcium levels, J Neurophysiol **47** 1017–1028.
[244] Schäfer, K., Braun, H.A., & Kürten, L. (1988) Analysis of cold and warm receptor activity in vampire bats and mice, Pflug Arch Eur J Phy **412** 188–194.
[245] Schreiner, C.E. & Langner, G. (1988) Periodicity coding in the inferior colliculus of the cat. II. Topograhical organization, J Neurophysiol **60** 1823–1840.
[246] Schroeder, D.M. & Loop, M.S. (1976) Trigeminal projections in snakes possessing infrared sensitivity, J Comp Neurol **169** 1–14.
[247] Sejnowski, T.J. (1995) Time for a new neural code?, Nature **376** 21–22.
[248] Seung, H.S. & Sompolinsky, H. (1993) Simple models for reading neuronal population codes, Proc Natl Acad Sci USA **90** 10 749–10 753.
[249] Shams, L., Kamitani, Y., & Shimojo, S. (2000) What you see is what you hear, Nature **408** 788.
[250] Shumway, C.A. (1989) Multiple electrosensory maps in the medulla of weakly electric gymnotiform fish. I. Physiological differences, J Neurosci **9** 4388–4399.
[251] Sichert, A.B., Bamler, R., & van Hemmen, J.L. (2009) Hydrodynamic object recognition: When multipoles count, Phys Rev Lett **102** 058 104.
[252] Sichert, A.B., Friedel, P., & van Hemmen, J.L. (2006) Snake's perspective on heat: Reconstruction of input using an imperfect detection system, Phys Rev Lett **97** 68 105.
[253] Sichert, A.B. & van Hemmen, J.L. (2010) How stimulus shape affects lateral-line perception: analytical approach to analyze natural stimuli characteristics, Biol Cybern **102** 177–180.
[254] Song, S., Miller, K.D., & Abbott, L.F. (2000) Competitive hebbian learning through spike-timing-dependent plasticity, Nat Neurosci **3** 919–926.
[255] Stanford, L.R. & Hartline, P.H. (1980) Spatial sharpening by second-order trigeminal neurons in crotaline infrared system, Brain Res **185** 115–123.
[256] Stanford, L.R., Schroeder, D.M., & Hartline, P.H. (1981) The ascending projection of the nucleus of the lateral descending trigeminal tract: A nucleus in the infrared system of the rattlesnake, *Crotalus viridis*, J Comp Neurol **201** 161–173.
[257] Stavenga, D.G. (2002) Reflections on colourful ommatidia of butterfly eyes, J Exp Biol **205** 1077–1085.
[258] Steenken, R., Colonius, H., Diederich, A., & Rach, S. (2008) Visual and auditory interaction in saccadic reaction time: Effects of auditory masker level, Brain Res **1220** 150 – 156.
[259] Stein, B.E. & Gaither, N.S. (1981) Sensory representations in reptilian optic tectum: Some comparisons with mammals, J Comp Neurol **202** 69–87.
[260] Stein, B.E., Jiang, W., & Stanford, T.R. (2004) Multisensory integration in single neurons of the midbrain, in G. Calvert, C. Spence, & B.E. Stein (editors), *The Handbook of Multisensory Processes*, chapter 15, pp. 243–264 (MIT, Cambridge, MA).
[261] Stein, B.E. & Meredith, M.A. (1993) *The Merging of the Senses* (MIT Press, Cambridge, MA).
[262] Stein, B.E. & Stanford, T.R. (2008) Multisensory integration: Current issues from the perspective of the single neuron, Nat Rev Neurosci **9** 255–266.
[263] Stryer, L. (1986) Cyclic gmp cascade of vision, Annual Review of Neuroscience **9** 87–119.
[264] Sullivan, W.E. & Konishi, M. (1986) Neural map of interaural phase difference in the owl's brainstem, Proc Natl Acad Sci USA **83** 8400–8404.
[265] Takahashi, T. & Konishi, M. (1986) Selectivity for interaural time difference in the owl's midbrain, J Neurosci **6** 3413–3422.
[266] Terashima, S.I., Goris, R.C., & Katsuki, Y. (1970) Structure of warm fiber terminals in the pit membrane of vipers, J Ultra Mol Struct R **31** 494–506.
[267] Thomson, R.F. (2001) *Das Gehirn* (Spektrum, Akad. Verl., Heidelberg, Berlin), 3rd edition.
[268] Tikhonov, A.N., Arsenin, V.Y., & John, F. (1977) *Solution of Ill-Posed Problems* (V.H.

Winston & Sons, Washington, DC), english translation of the original Russian text.
[269] Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V., & Yagola, A.G. (1995) *Numerical Methods for the Solution of Ill-Posed Problems* (Kluwer Academic Publishers, Dordrecht), english translation of the original Russian text.
[270] Truckenbrodt, E. (1999) *Fluidmechanik 2* (Springer), 4th edition.
[271] Udin, S. & Fawcett, J.W. (1988) Formation of topographic maps, Annu Rev Neurosci **11** 289–327.
[272] Unoki, M., Sakata, K., Furukawa, M., & Akagi, M. (2004) A speech dereverberation method based on the MTF concept in power envelope restoration, J Acoustical Science and Technology **25(4)** 243–254.
[273] Unoki, Masashi, Furukawa, Masakazu, & Sakata, Keigo anf Akagi, Masato (2004) An improved method based on the MTF concept for restoring the power envelope from a reverberant signal, J Acoustical Science and Technology **25(4)** 232–42.
[274] Ursino, M., Cuppinia, C., Magosso, E., Serino, A., et al. (2009) Multisensory integration in the superior colliculus: a neural network model, J Comput Neurosci **26(1)** 55–73.
[275] Voets, T., Droogmas, G., Wissenbach, U., Janssens, A., et al. (2004) The principle of temperature-dependent gating in cold- and heat-sensitive TRP channels, Nature **430** 748–754.
[276] van der Waerden, B.L. (1957) *Mathematische Statistik* (Springer, Berlin).
[277] Wallace, M.T., Carriere, B.N., Perrault Jr, T.J., Vaughan, J.W., et al. (2006) The development of cortical multisensory integration, J Neurosci **26(46)** 11 844–11 849.
[278] Wallace, M.T., Perrault Jr, T.J., Hairston, W.D., & Stein, B.E. (2004) Visual experience is necessary for the development of multisensory integration, J Neurosci **24** 9580–9584.
[279] Wallace, M.T. & Stein, B.E. (1997) Development of multisensory neurons and multisensory integration in cat superior colliculus, J Neurosci **17** 2429–2444.
[280] Wallace, M.T. & Stein, B.E. (2007) Early experience determines how the senses will interact, J Neurophysiol **97** 921–926.
[281] Wandell, B.A. (1995) *Foundations of Vision* (Sinauer Associates, Sunderland, MA).
[282] Wehr, Michael & Zador, Anthony M. (2003) Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex, Nature **426** 442–446.
[283] Wickesberg, R. E. & Oertel, D. (1990) Delayed, frequency-specific inhibition in the cochlear nuclei of mice: A mechanism for monaural echo suppression, J Neurosci **10** 1762–1768.
[284] Willshaw, D.J. & Malsburg, C. (1976) How patterned neural connections can be set up by self-organization, Proc Roy Soc Lond B **194** 431–445.
[285] Winkowski, D.E. & Knudsen, E.I. (2006) Top-down gain control of the auditory space map by gaze control circuitry in the barn owl, Nature **439** 336–339.
[286] Wissler, E.H. (1998) 'penns 1948 paper revisited, J Appl Physiol **85** 35–41.
[287] Witten, I.B., Knudsen, E.I., & Sompolinsky, H. (2008) A Hebbian learning rule mediates asymmetric plasticity in aligning sensory representations, J Neurophysiol **100** 1067–1079.
[288] Wu, Mingyang & Wang, DeLiang (2006) A two-stage algorithm for one-microphone reverberant speech enhancement, IEEE T Audio Speech **14** 774–84.
[289] Yang, Y., Chen, J., Engel, J., Pandya, S., et al. (2006) Distant touch hydrodynamic imaging with an artificial lateral line, Proc Natl Acad Sci USA **103** 18 891–18 895.
[290] Yegnanarayana, B., Avendano, C., Hermansky, H., & Satyanarayana Murthy, P. (1999) Speech enhancement using linear prediction residual, Speech Commun **28** 25–42.
[291] Yegnanarayana, B. & Satyanarayana Murthy, P. (2000) Enhancement of Reverberant Speech Using LP Residual, IEEE T Speech Audio P **8** 267–81.
[292] Yost, W.A. (1991) Auditory image perception and analysis: The basis for hearing, Hear Res **56(7)** 8–18.
[293] Zeil, J. & Hemmi, J.M. (2006) The visual ecology of fiddler crabs, J Comp Physiol A **192** 1–25.

[294] Zhang, L.L., Huizong, W.T., Holt, C.E., Harris, W.A., et al. (1998) A critical window for cooperation and competition among developing retinotectal synapses, Nature **395** 37–44.

[295] Zittlau, K.E., Claas, B., & Münz, H. (1988) Horseradish peroxidase study of tectal afferents in *Xenopus laevis* with special emphasis on their relationship to the lateral-line system, Brain Beh Evol **32** 208–219.