Auditory Information Processing in Systems with Internally Coupled Ears

Christine Voßen

This dissertation has been written in LaTeX using the *memoir* class. Typesetting was done with pdfLaTeX.

Funding has been provided by the Bernstein Center for Computational Neuroscience (BCCN) – Munich.

TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Theoretische Physik

Auditory Information Processing in Systems with Internally Coupled Ears

Christine Voßen

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. F. Simmel
Prüfer der Dissertation:	1. Univ.-Prof. Dr. J. L. van Hemmen
	2. Univ.-Prof. Dr. F. Bornemann

Die Dissertation wurde am 12.07.2010 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 27.07.2010 angenommen.

. . . es sind uns Dinge als außer uns befindliche Gegenstände unserer Sinne gegeben, allein von dem, was sie an sich selbst sein mögen, wissen wir nichts, sondern kennen nur ihre Erscheinungen, d. i. die Vorstellungen, die sie in uns wirken, indem sie unsere Sinne affizieren.

Immanuel Kant (1724 – 1804)
Prolegomena, 1783

Contents

A journey of a thousand miles must begin with a single step.

Lao Tzu

1. Introduction

This chapter introduces the fundamentals of sensory information processing, focusing on the processing of sound signals as the auditory system is one of the most widespread of the various sensory systems. The prominent role of the auditory system can be explained by the specific properties of sound perception, namely, hearing is omnidirectional and its processing is very fast. In comparison to light, sound waves have a much longer wavelength so they are not blocked by small objects. We can, for instance, hear something behind a tree but we cannot see it. Amongst other advantages audition therefore offers the possibility to react to approaching dangers that are not yet visible. To exploit the latter advantage, it is essential that an animal can *localize* a sound source, otherwise the prey could try to escape in the direction of the predator. To determine the direction of a sound stimulus, several steps are necessary in the animals studied in this thesis.

First, an object evokes an auditory stimulus. Section 1.1 describes how such a stimulus propagates through the surrounding medium and is amplified before arriving at the detectors of the animal.

Second, the arriving sound waves excite the *tympani*, a pair of thin membranes that are part of the mechanical auditory system. The anatomy of the auditory system is highly variable. The primary focus of this work is on internally coupled ears (ICE) in which one membrane can influence vibration of the other through internal cavities; see Sec. 1.2.

Third, tympanic vibrations are processed neuronally. The fundamental element of neuronal computation is the neuronal net, consisting of *neurons* as building blocks and the variable connections between them, called *synapses*. Section 1.3 reviews the general function and mathematical modeling of neurons and synapses.

Fourth, as result of neuronal computation, sensory stimuli give rise to neuronal representations of the stimuli, i.e., *neuronal maps*; see Sec. 1.4. Each neuron of the map represents a specific property, e.g., the stimulus at a specific point in space. Neighboring neurons respond to similar sensory inputs. Neuronal maps reconstruct the stimulus as well

as possible, i.e., *optimally* within the limits of processing. The precise calibration of the synapses required for stimulus reconstruction results from an experience-based learning process that takes into account inputs from all available sensory systems.

1.1 Sound stimuli

Sound waves are oscillations of pressure that are transmitted through the air. A sound stimulus is created when a movement compresses the surrounding air and creates traveling sound waves that arrive at the tympani of an animal. Depending on the spatial relationship between the two tympani and the sound source, the arriving sound waves differ in phase and amplitude between the tympani. These differences are denoted as interaural time differences (*ITD*s) and interaural amplitude differences (*IAD*s). A sound wave from a source at one side of the head, for instance, arrives at the facing, i.e., *ipsilateral* tympanum earlier and with higher amplitude than at the averted, i.e., *contralateral* tympanum. A sound wave from a source in front of the head, in contrast, arrives at both ears simultaneously and with the same amplitude. These two examples demonstrate that ITDs and IADs can encode the direction of a sound source. An evaluation of ITDs and IADs as performed by the auditory system and the consequent neuronal circuits could therefore decode the sound source direction.

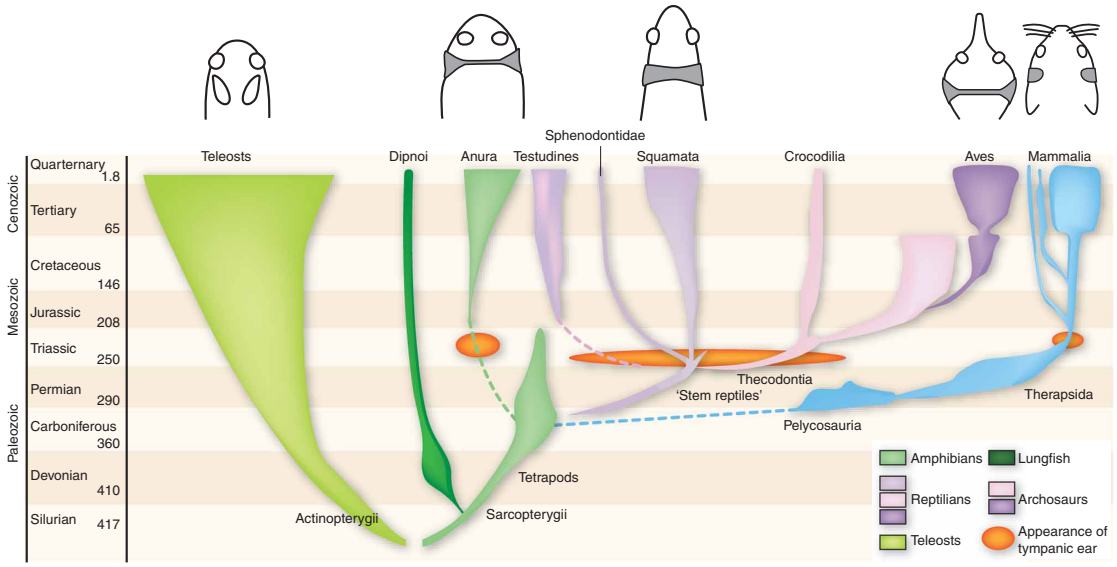

Figure 1.1: *Evolution of vertebrate ears. During the transition from water to land, tympanic middle ears capable of receiving airborne sound evolved separately among the ancestors of modern frogs, turtles, lizards, birds, crocodilians, and mammals. Diagrams at the top show cross-sections through different heads of these animals (middle ears - gray fill). Of interest are the auditory systems with internally coupled ears (ICE) as found in anura (frogs and toads), squamata (lizards and snakes), and aves (birds), where the tympani are coupled through differently shaped internal cavities; cf. as well Fig. 4.1. Mammals, in contrast, have independent ears, i.e., the tympani cannot influence each other. Figure due to Schnupp and Carr [193].*

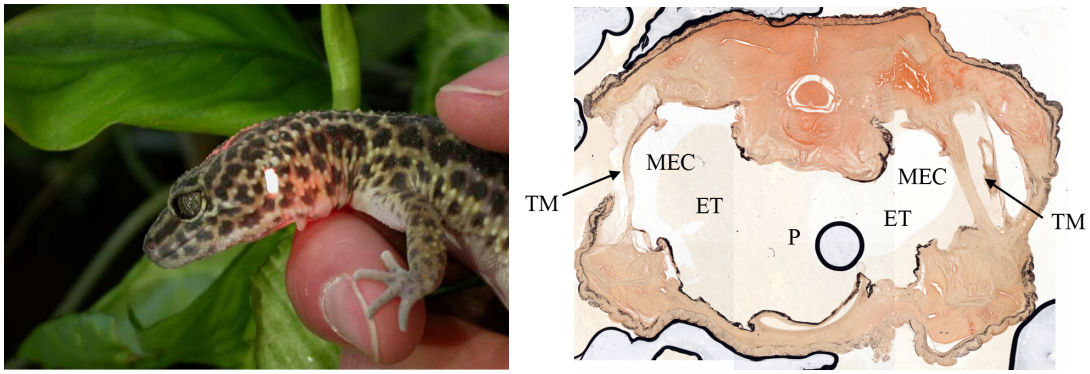

Figure 1.2: *Internally coupled ears in lizards. Left: A light source on the other side of the head illuminates the tympanic membrane from the back. Right: Intersection through a lizard's head. Large Eustachian tubes (ET) and the pharynx (P) connect the middle ear cavities (MEC) and the tympanic eardrums (TM) and allow the vibration of one tympanum to influence the vibration of the tympanum at the other side. Photographs due to Jakob Christensen-Dalsgaard and Catherine Carr.*

1.2 Mechanical processing of sound stimuli

1.2.1 Evolution of different auditory systems

To detect incoming sound waves, ancestors of modern frogs, turtles, lizards, birds, crocodilians, and mammals developed tympani, i.e., thin membranes that transmit sound from the air to the ossicles inside the middle ear; cf. Fig. 1.1. Cross-sections of heads of different animals with tympanic hearing show two fundamentally different constructions. Mammals possess tympani that are spatially separated and therefore acoustically independent of each other. In contrast, reptilia such as lizards, turtles, crocodiles, and birds have *internally coupled ears* (ICE) (for details, e.g., reviews [19, 24, 193]) in which tympanic membranes are connected through large Eustachian tubes as illustrated by Figs. 1.2 and 1.3. The evolutionary appearance of independent and internally coupled ears (Fig. 1.1) suggests that the latter are probably early tympanic ears. If this is indeed the case, the mammalian independent ears must be derived from internally coupled ears. This is reasonable since the enlarging mammalian brain could have grown into the internal cavities and disconnected the tympani (Manley, personal communication).

1.2.2 Internally coupled ears

Tympanic membranes of ICE are acoustically coupled in the sense that a signal arriving at one of the tympanic membranes can propagate through the internally interconnected cavities and influence the vibration of the other tympanic membrane. In this way, ICE translate incoming sound waves with specific ITDs and IADs into amplified vibrations of the two tympanic membranes. In turn, the tympanic vibrations differ through internal time differences ($iTDs$) and internal amplitude differences ($iADs$), being the result of both the external signal and the internal coupling. Interestingly, the internal coupling of the

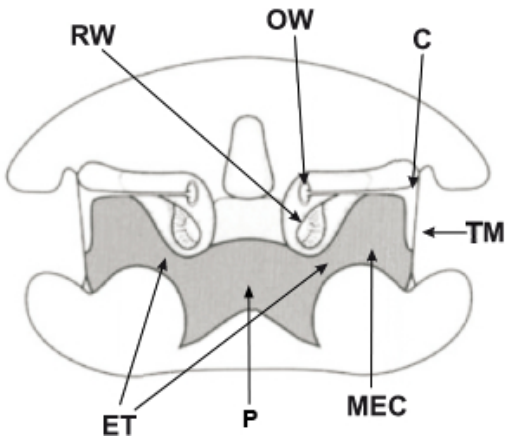

Figure 1.3: *Schematic cross-section through a lizard's head to illustrate the mechanical processing of auditory stimuli through internally coupled ears. Arriving sound waves excite the tympanic membranes (TM) as well as the air within the middle ear cavities (MEC), the Eustachian tubes (ET), and the pharynx (P). The resulting vibration of the tympanum evokes a movement of the attached middle ear, called* columella *(C). The lever construction transmits the vibration of the tympani to the oval window (OW), the membrane at the entrance of the fluid-filled cochlea. Through the vibration of the OW, the fluid in the cochlea is excited and gives rise to a local, frequency-dependent activation of the embedded basilar membrane and the underlying auditory nerve fibers. The round window (RW) is a membrane at the end of the cochlea that is needed to compensate the pressure within the fluid. Altogether, translation of tympanic vibrations into neuronal responses is similar for animals with internally coupled and independent ears despite several modifications; cf. Chap. 3. Figure taken from [27].*

tympanic membranes enhances iTDs in comparison to incoming ITDs by up to a factor of three [17]. In addition, internally coupled ears produce large iADs even for small interaural distances [27, 28]. Chap. 2 and [221] present a general model of internally coupled ears, the ICE model, that describes vibrations of the tympani and the realistic mouth cavity dependent on the characteristics of incoming sound stimuli. Once a sound wave has excited the tympanic membranes, the translation into neuronal responses is similar to mechanical sound processing as found in mammals. Tympanic vibrations are transmitted by the columella (the middle ear bone that is attached to the tympanum) to the oval window, a membrane at the entrance of the cochlea; see Fig. 1.3. Vibration of the oval window then results in vibration of the fluid within the cochlea and of the embedded basilar membrane. Due to its systematically varying stiffness, every part of the basilar membrane reacts only to a specific frequency. Auditory nerve fibers enervated by the movement of a restricted region of the basilar membrane therefore respond to a small range of frequencies and in this manner, the basilar membrane decomposes sound frequencies.

1.3 Neuronal processing of sound stimuli

As soon as sound waves are processed mechanically through the tympani and basilar membranes, processing continues on a neuronal level. The following section reviews important concepts of neuronal computation. For a detailed introduction to computational neuroscience see, e.g., Koch [125] or Izhikevich [97].

1.3.1 Building blocks of neuronal computation

In 1906, the Spanish anatomist Santiago Ramon y Cajal (1852-1934) received the Nobel Prize in Medicine "in recognition of his work on the structure of the nervous system". His

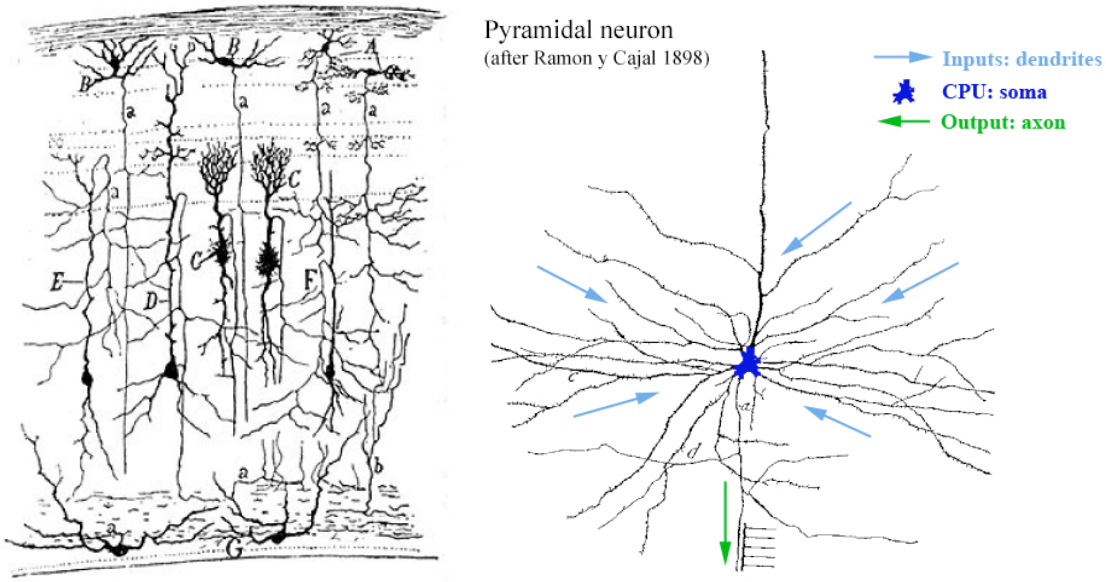

Figure 1.4: *Two of Ramon y Cajal's extraordinary drawings [236]. Left: Drawing of a section through the optic tectum of a sparrow showing individual neurons. Right: A pyramidal cell from the motor cortex extended by visualization of the three stages of neuronal information processing: collection of inputs through dendrites (light blue), processing of signals through the soma (blue), and transmission of output signals through the axon (green).*

work consisted of a systematic analysis of the neural tissue. Based on his revolutionary observations (cf. Fig. 1.4), Cajal showed that the nervous system consists of individual cellular elements [237]. The existence of these small structural and functional units had already been suggested in 1891 by Waldeyer-Hartz [220] who denoted them as *neurons*. Thanks to Cajal, scientists accepted that the neuronal tissue was not continuous but constructed from discrete elements that receive, process, and transmit signals. Accordingly, a neuron consists of three parts: dendrites to collect inputs from other neurons, the soma to process inputs, and the axon to transmit output signals to other neurons; cf. Fig. 1.4, right.

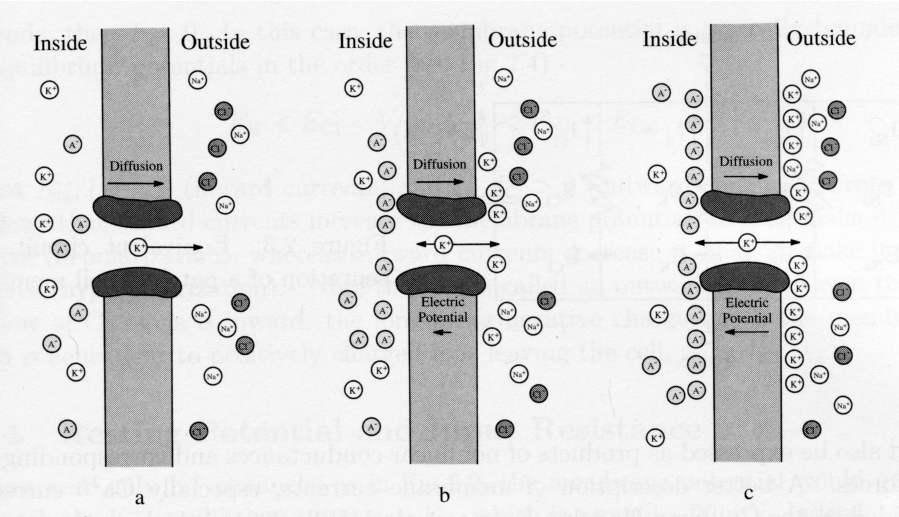

Figure 1.5: *Emergence of the ion-specific Nernst potential at the example of K+ ions. a) Following the concentration gradient, K+ ions diffuse into the extra-cellular domain. b) Since K+ ions are electrically charged, the change in their concentrations from the intra-cellular to extra-cellular domain creates an electric potential leading to an increasing force in the opposite direction. c) The potential in which electric and diffusive forces are balanced is the ion-specific Nernst potential denoted as E_K for K+ ions. Figure due to Izhikevich [97, p. 27].*

Starting with Cajal's individual neurons, researchers have steadily extended our knowledge of neuronal properties and function. Today, we know that the neuronal cell membrane is essential for processing neuronal, i.e., electric signals. The membrane separates ion solutions in the intra- and extra-cellular spaces. These solutions contain different concentrations of sodium (Na+), potassium (K+), chloride (Cl-), and calcium (Ca2+). K+ ions dominate the intra-cellular solution and Na+ ions the extra-cellular solution. Since the ions usually cannot pass through the membrane but instead assemble at its surface, the membrane acts as a capacitor with capacitance C_m. An electric potential appears across the membrane, called *membrane potential* $V(t)$. The ions can only cross the membrane through membrane-embedded ion-selective channels. Ion currents through these channels are determined by the membrane potential and the gradient of the ion concentrations between the intra- and extra-cellular spaces. As a consequence, ions move along an ion-specific reversal potential, the Nernst potential; see Fig. 1.5 for details. In the resting state, primarily K+ channels are open so that the resting membrane potential is dominated by the reversal potential of K+ ions $V_r \approx -75$ mV. The passive current through permanently open channels is called *leak* current.

In contrast, electric signals, called *action potentials* or *spikes*, are generated by voltage-gated Na+ and K+ channels that open and close actively with a characteristic timing. Through a rapid influx of Na+ ions, the membrane voltage is able to increase up to 30 mV for about a millisecond. Then a delayed outflow of K+ ions leads to a repolarization of the

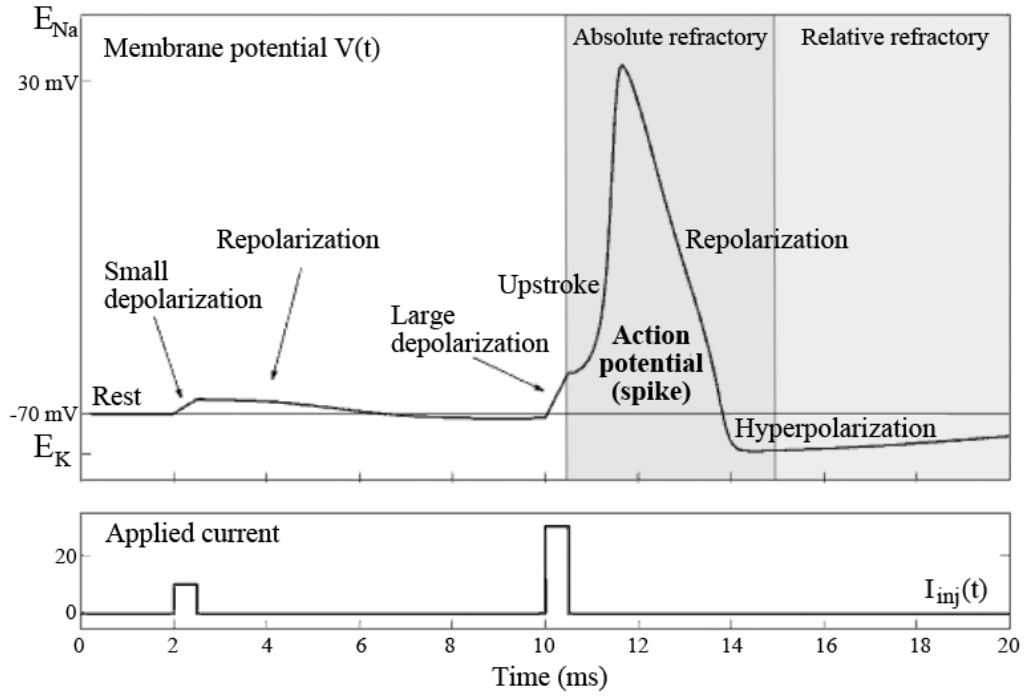

Figure 1.6: *Temporal development of an action potential (spike). The plot shows the membrane potential $V(t)$ (top) in response to an applied current (bottom). Without external input, the neuron remains in its equilibrium state where the membrane potential corresponds to the resting potential $V_r = -70$ mV. Na+ and active K+ channels are closed. The Na+ and K+ reversal potentials of $E_{Na} = 50$ mV and $E_K = -90$ mV are depicted. For a weak input current the membrane potential only slightly depolarizes and returns to its resting potential. In contrast, as soon as external input increases the membrane potential above a certain threshold, Na+ channels open. A large number of Na+ ions streams into the cell and depolarizes the membrane potential. As a consequence, the membrane potential quickly increases up to around 30 mV, an action potential is generated. At the same time more and more active K+ channels open so that the influx of K+ ions starts to decrease the membrane potential toward the reversal potential E_K. The membrane potential is therefore first repolarized, then even hyperpolarized, i.e., decreased below the membrane potential. During the phase of hyperpolarization another action potential can be generated but more excitatory current is needed in comparison to the resting state. This period is therefore denoted as* relative refractory *period. The period during the action potential where no spike generation is possible is called* absolute refractory *period. Figure modified from [97, p. 40].*

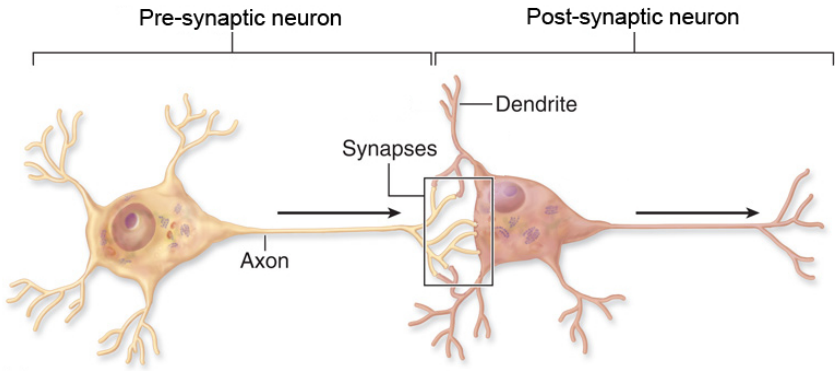

Figure 1.7: *Synapses connect pre- and post-synaptic neurons and allow action potentials from the pre-synaptic neuron to influence the membrane potential of the post-synaptic neuron with a synapse-specific efficiency.*

membrane potential. After a refractory period, the membrane potential reaches its original resting potential and the next action potential can be generated; for details see Fig. 1.6.

In the next step, an action potential generated in one specific neuron is transmitted to the connected neurons. Since neurons are individual cells, specialized structures, called *synapses* from Greek "syn-" together and "haptein" to clasp, are needed to pass information from a pre-synaptic to a post-synaptic cell; cf. Fig. 1.7. The majority of the synapses are *chemical* synapses that function through a release of neurotransmitter from the pre-synaptic cell into the synaptic cleft, a 20 nm thin gap between pre- and post-synaptic cell. These neurotransmitters can then diffuse through the cleft and reach neurotransmitter receptors of the post-synaptic cell. Each synapse is characterized by the synaptic strength, i.e., the efficiency of transmission of a pre-synaptic stimulus to the post-synaptic cell. Depending on their effect on the membrane potential, i.e., depolarizing (exciting) or hyperpolarizing (deactivating) a cell, synapses are called *excitatory* or *inhibitory*. Synaptic strengths indicate the degree to which neurons are connected.

To quantitatively describe neuronal behavior in response to incoming (synaptic) currents, neurons are described with mathematical models. These models are always compromises between detailed and generalized description. To gain intuitive *understanding* of neuronal behavior, the model used should include all processes needed to explain the observed effect and should leave out all others. The Hodgkin and Huxley model [89] is an example of a detailed neuron model that includes the dynamic behavior of voltage-depended Na+ and K+ channels. This model, however, is characterized by a large number of variables that are not able to be fully determined experimentally. Further simplification of the neuron model so that it only reproduces basic phenomenological and electric properties of a neuron is therefore reasonable. As a consequence, mathematically tractable neuron models such as the leaky integrate-and-fire and the Poisson model have been developed.

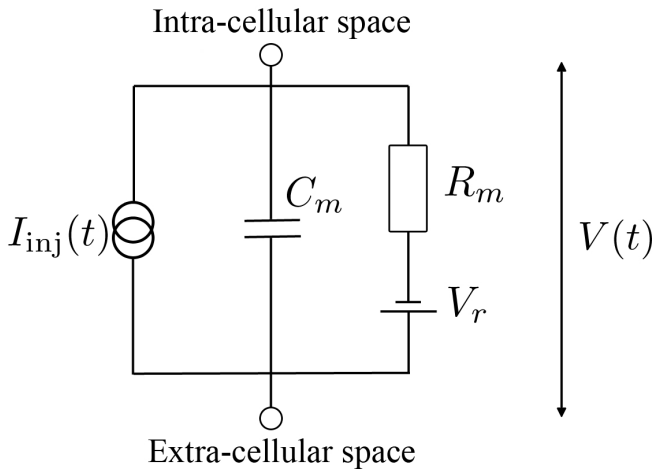

Figure 1.8: *Equivalent circuit for a leaky integrate-and-fire neuron. Behavior of the neuron is described by the membrane potential $V(t)$, that is, the potential between intra- and extra-cellular space. The membrane of a neuron can be characterized as capacitor with leak, i.e., a capacitor C_m in parallel with a resistance R_m and the resting potential V_r. The membrane potential is modified by the current $I_{\text{inj}}(t)$, i.e., synaptic currents or ions injected artificially through electrodes.*

1.3.2 Leaky integrate-and-fire neuron

The leaky integrate-and-fire neuron model reduces the neuron to a point neuron whose behavior is described by the membrane potential $V(t)$. In response to incoming currents, the membrane potential varies until it reaches a certain threshold $\overline{V}$ above which an action potential (spike) is generated, and the membrane potential is reset to the resting potential. Since the model focuses on the subthreshold membrane properties, it excludes the mechanisms responsible for generating the action potentials itself, i.e., the voltage-dependent sodium and potassium channels; for details see review [16]. The spike event is therefore only registered as a formal, discrete event when the membrane potential reaches its firing threshold. The exact form of the spike does not enter the model.

Subthreshold membrane properties have long been recognized. In 1907, Lapicque remarked that the membrane of a neuron can be characterized as capacitor with leak, i.e., a capacitor C_m in parallel with a resistance R_m; see [135] for Lapicque's original and [13, 14] for the translated and commented manuscript. Besides the Ohmic current over the membrane $[V_r - V(t)]/R_m$, the membrane potential can be modified by externally injected currents $I_{\text{inj}}(t)$, i.e., currents injected through synapses or intra-cellular electrodes. Kirchhoff's nodal rule for the equivalent circuit of a leaky integrate-and-fire neuron (cf. Fig. 1.8) requires that all currents flowing into a node are equal to the currents flowing out of the node. Application of the law yields the differential equation

$$C_m \dot{V}(t) = \left[V_r - V(t)\right]/R_m + I_{\text{inj}}(t) \ . \tag{1.1}$$

For a time-dependent input current I_{inj} the corresponding membrane potential is given by

$$V(t) = V_r + \exp\left(-\frac{t-t_0}{C_m R_m}\right)\left[\frac{1}{C_m}\int_{t_0}^{t} \mathrm{d}s\ \exp\left(\frac{s-t_0}{C_m R_m}\right) I_{\text{inj}}(s)\right] \tag{1.2}$$

which assumes that the neuron has fired and the membrane potential was reset to its resting potential V_r at time t_0. To prove that the above $V(t)$ fulfills Eq. (1.1), the initial condition for $t = t_0$ is first verified as $V(t_0) = V_r$. Second, the time derivative of $V(t)$ can be calculated as

$$\begin{aligned}
\dot{V}(t) &= -\frac{1}{C_m R_m}\exp\left(-\frac{t-t_0}{C_m R_m}\right)\left[\frac{1}{C_m}\int_{t_0}^{t} \mathrm{d}s\ \exp\left(\frac{s-t_0}{C_m R_m}\right) I_{\text{inj}}(s)\right] \\
&\quad + \exp\left(-\frac{t-t_0}{C_m R_m}\right)\left[\frac{1}{C_m}\exp\left(\frac{t-t_0}{C_m R_m}\right) I_{\text{inj}}(t)\right] \\
&= -\frac{1}{C_m R_m}\left[V(t) - V_r\right] + \frac{1}{C_m}\ I_{\text{inj}}(t)\ .
\end{aligned} \tag{1.3}$$

Multiplied with C_m, the resulting equation corresponds exactly to the differential equation (1.1). To illustrate Eq. (1.2), in the case of a constant current, the solution reduces to

$$V(t) = V_r + R_m I_{\text{inj}}\left[1 - \exp\left(-\frac{t-t_0}{C_m R_m}\right)\right]\ . \tag{1.4}$$

With time, the membrane potential approaches $V_r + R_m I_{\text{inj}}$ as the expression within the round brackets tends to 1. If this value lies below the threshold $\overline{V}$, the neuron does not spike at all, otherwise it spikes regularly with a constant interspike interval T given by the equation $V(t = T) = \overline{V}$. Taken together, the above calculations illustrate that a description of neuronal behavior through leaky integrate-and-fire neurons is possible but lengthy. For calculations including a large number of neurons in particular, it is advantageous to represent firing of each neuron stochastically.

1.3.3 Poisson neuron

The *Poisson neuron* reduces the firing of a neuron to a stochastic process, in other words, to an *inhomogeneous* Poisson process with a time-dependent rate function $\nu(t)$ (for a *homogeneous* Poisson process, this rate function is constant over time). An inhomogeneous Poisson process fulfills three aspects:

- disjoint intervals are independent,
- the probability of getting a single event at time t in an interval $[t, t + \Delta t)$ is given by $\nu(t)\Delta t$,
- the probability p_{mult} of getting multiple events follows $o(\Delta t)$, that is,

$$p_{\text{mult}}/\Delta t \to 0 \qquad \text{for} \qquad \Delta t \to 0\ . \tag{1.5}$$

For details of Poisson processes and Poisson neurons see [110] and Appendix B of [216]. The advantage of Poisson neurons lies in a compact mathematical description allowing exact solutions. For example, the probability that a neuron fires k times in an interval $[t_0, t)$ can be calculated explicitly. To this end, the interval is split up into N disjoint intervals of length $\Delta t = (t - t_0)/N$. First, the probability p_0 that a neuron remains silent over the entire interval $[t_0, t)$ is calculated as

$$p_0 = \prod_{l=0}^{N-1} [1 - \nu(t_l)\Delta t] \tag{1.6}$$

where t_l denotes $t_0 + l\Delta t$. Terms of higher order in Δt are neglected. To transform this probability into a more convenient form application of the equality $x = \exp[\ln(x)]$ leads to

$$p_0 = \exp\left\{\sum_{l=0}^{N-1} \ln[1 - \nu(t_l)\Delta t]\right\} . \tag{1.7}$$

For x small the logarithm $\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n+1} x^n/n$ can be approximated as $\ln(1+x) \approx x$. For a sufficiently small Δt, this approximation applies so that the probability p_0 transforms to

$$p_0 \approx \exp\left[-\sum_{l=0}^{N-1} \nu(t_l)\Delta t\right] \approx \exp\left[-\int_{t_0}^{t} \mathrm{d}t'\ \nu(t')\right] . \tag{1.8}$$

The probability p_0 of a neuron remaining silent consequently decreases exponentially with increasing length of the considered interval and increasing firing rate. Based on this probability p_0, the probability of getting k spikes in the interval $[t_0; t)$ is now calculated. For one specific realization of k spikes in N intervals, the times t_l^*, $l \in \mathbb{N}$, $0 \le l < k$ denote the k chosen succeeding spike intervals $t_l^* < t_{l+1}^*$; cf. Fig 1.9. The probability p_k^* of this specific realization can be calculated as the product of the probabilities of spikes in intervals t_l^* and the non-firing probability in the interspike intervals

$$p_k^* = \underbrace{\prod_{l=0}^{k-1} [\nu(t_l^*)\Delta t]}_{\text{firing probability at } t_l^*} \cdot \underbrace{\exp\left[-\int_{t_0}^{t_0^*} \mathrm{d}t'\ \nu(t')\right] \prod_{l=0}^{k-2} \exp\left[-\int_{t_l^*+\Delta t}^{t_{l+1}^*} \mathrm{d}t'\ \nu(t')\right] \exp\left[-\int_{t_{k-1}^*+\Delta t}^{t} \mathrm{d}t'\ \nu(t')\right]}_{\text{non-firing probability in interspike intervals according to Eq. (1.8)}} ; \tag{1.9}$$

see Fig. 1.9 for an illustration. As $\Delta t \to 0$ the above equation simplifies to

$$p_k^* = \prod_{l=0}^{k-1} [\nu(t_l^*)\Delta t]\ \exp\left[-\int_{t_0}^{t} \mathrm{d}t'\ \nu(t')\right] . \tag{1.10}$$

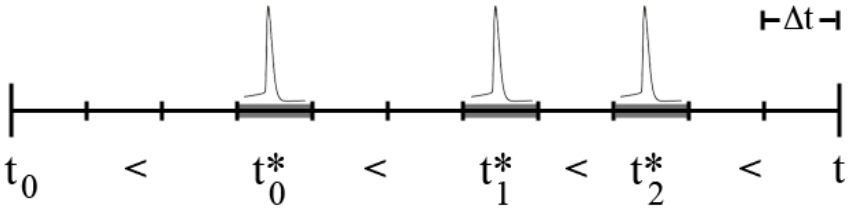

Figure 1.9: *Distribution of $k = 3$ spikes in the interval $[t_0; t)$. The interval is split up into N disjoint intervals of length $\Delta t = (t - t_0)/N$. The plot shows a specific realization of $k = 3$ spikes in the chronologically ordered intervals $\{t_0^*, t_1^*, t_2^*\}$.*

The equation gives the probability of getting k spikes at the specific chronologically ordered spike times $t_0^* < t_1^* < ... < t_{k-1}^*$. Since spikes are indistinguishable another arrangement, i.e., permutation, of the spike times corresponds to the same event with identical probability.

In a next step, the k spike times are randomly chosen from the N possible intervals. For reasons of mathematical convenience a random arrangement of spike times is allowed that is accounted for by devision through all additional permutations $k!$ of the same k spikes. Altogether the general probability for k spikes in N intervals corresponds to

$$p_k = \underbrace{\exp\left[-\int_{t_0}^{t} \mathrm{d}t'\ \nu(t')\right]}_{\text{non-firing probability}} \cdot \underbrace{\frac{1}{k!}\left[\sum_{n_1=0}^{N-1} \nu(t_{n_1})\Delta t\right]\left[\sum_{\substack{n_2=0 \\ n_2 \neq n_1}}^{N-1} \nu(t_{n_2})\Delta t\right] \ldots \left[\sum_{\substack{n_k=0 \\ n_k \neq n_1, \ldots n_{k-1}}}^{N-1} \nu(t_{n_k})\Delta t\right]}_{\text{firing probabilities of k spikes in different unordered intervals}} . \tag{1.11}$$

Again as $\Delta t \to 0$, the probability (1.11) can be written

$$p_k = \exp\left[-\int_{t_0}^{t} \mathrm{d}t'\ \nu(t')\right] \frac{1}{k!}\left[\int_{t_0}^{t} \mathrm{d}t'\ \nu(t')\right]^k . \tag{1.12}$$

Hence, fundamental stochastic properties yield a compact description of the firing probability to get k spikes within a given interval. Similar calculations involving Poisson neurons can lead to an intuitive understanding of complicated neuronal processes. This is essential since neurons are parts of larger arrays of neurons, e.g., of neuronal maps.

1.4 Neuronal representation of sound stimuli

1.4.1 Neuronal maps

To represent sensory stimuli within the brain, neurons do not function autonomously but are often grouped in well-ordered arrays called *neuronal maps* [124, 217]. The idea that a map-like architecture underlies certain aspects of sensory processing was suggested in 1879 by Helmholtz [86] who 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." That is, Helmholtz already recognized a topographic neuronal representation of a two-dimensional surface. A neuronal map is constructed from an array of neurons in which neighboring neurons respond to similar sensory stimuli [23, 43, 91, 114, 116, 119, 151, 202, 207]. In the auditory system, maps represent frequency, interaural time difference, interaural amplitude difference, and even amplitude modulation [62, 147, 167, 172, 194, 209].

Within given limitations, e.g., noise and number of available neurons, neuronal maps reconstruct a sensory stimulus as precisely as possible, i.e., *optimally*; see Chap. 5 for a general mathematical framework to calculate optimal connectivity. The neurophysiological conditions for such a precise topographic arrangement are an ordered arrangement of the connecting axons and a fine-tuned synaptic pattern. Accordingly, map calibration consists of two steps, a pre-wiring through growing axons and a fine-tuning through activity-based learning of the synapses.

1.4.2 Pre-wiring a map

For the pre-wiring of maps, molecular gradients and axon-axon interaction play a key role [140, 213]. When the brain develops, axons grow and connect various parts of the brain. Concentration gradients of specific molecules such as *ephrins* act as guidance cues in attracting or repelling axons. When the molecules and their receptors are distributed gradually along an array of neurons, growing axons arrange along the axis of the array and so form a coarse pre-wiring of a map.

In addition, growing axons tend to fill up available space smoothly because they are competing for one or more limiting molecules in the target. As a consequence, axons that grow from a topographically arranged structure, e.g., the retina or the cochlea interact with each other. The axons therefore maintain their structural arrangement and project it to the target area.
Throughout this thesis, the mechanisms of molecular gradients and axon-axon interaction are referred to implicitly as necessary tools for achieving a pre-wiring of the sensory maps. Beyond this, the work concentrates on the second stage of map formation, that is, on the fine-tuning of maps through activity-based learning.

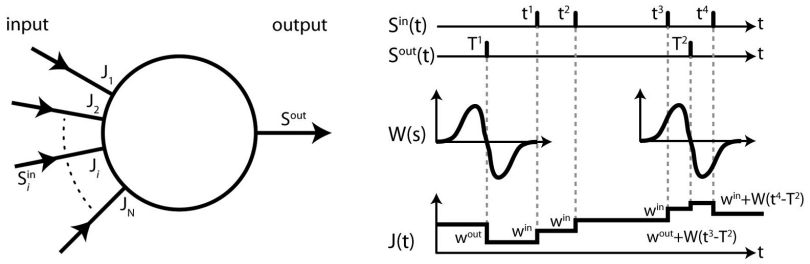

Figure 1.10: *Left: A neuron receives input from N neurons. The input spike trains S_i^{in} with $1 \leq i \leq N$ are transmitted through synapses with synaptic strength J_i to the soma. In reaction to the combined inputs, the neuron produces an output spike train S^{out}. Right: The incoming and outgoing spikes modify the efficiency of a synapse by values w^{in} and w^{out}. According to the timing between incoming spike t^i and outgoing spike T^j, the learning window $W(s)$ with $s = t^i - T^j$ additionally increases or decreases the synaptic strength. A synapse that transmits an input spike slightly before an output spike is likely to be important for the generation of the output spike and is enhanced. The opposite is the case for a synapse that transmits an input spike slightly after an output spike. Experimentally, this learning window has been measured in various forms, e.g., in rat hippocampus [240].*

1.4.3 Fine-tuning a map

The precise behavior of a neuronal map is determined by a fine-tuned connectivity, i.e., synaptic, pattern. An adult human brain contains from 10^{14} to $5 \cdot 10^{14}$ synapses. Each synapse has a specific strength that can be modified due to the firing activity of the neuron. The resulting amount of information is gigantic and cannot be encoded genetically but has to be learned during the development of an animal or human. As a consequence, fine-tuning of a map is based on experience, i.e., learning, that modifies the synaptic pattern of a map. The Canadian psychologist Donald Olding Hebb (1904 - 1985) was the first to understand how the activity of neurons contributes to the learning process. In his book "The organization of behavior" [79] Hebb wrote:

- "When one cell repeatedly assists in firing another, the axon of the first cell develops synaptic knobs (or enlarges them if they already exist) in contact with the soma of the second cell." (p. 63)
- "The general idea is an old one, that any two cells or systems of cells that are repeatedly active at the same time will tend to become 'associated', so that activity in one facilitates activity in the other." (p. 70)

Depending on the synchrony between firing of pre- and post-synaptic neurons, a synapse can undergo long-term potentiation (LTP), a long-lasting enhancement of the synapse, or the opposite, that is, long-term depression (LTD). Mathematically, both processes can be described in terms of spike-timing-dependent plasticity (STDP) [6, 7, 36, 66, 109, 149, 199, 216, 240]. Figure 1.10 illustrates the underlying mechanism. Here a neuron receives inputs, i.e., spike trains S_i^{in} from N input neurons via synapses with strengths J_i where $i \in \mathbb{N},\ 1 \leq i \leq N$. Modification of synaptic strengths in response to activity of the input

neurons is threefold. First, each incoming spike at time t^i changes the corresponding synapse by an amount w^{in}. Second, each output spike at time T^i changes all synapses by the amount w^{out}. Third, the precise timing between input and output spike changes the corresponding synapse via the *learning window* $W(s = t^i - T^i)$; see Fig. 1.10, right, middle row. For the last mechanism, synaptic strengths increase if a pre-synaptic firing precedes a spike, and decrease otherwise. Assuming the learning process to be much slower than the neuronal dynamics, synaptic changes can be described by the differential equations

$$\frac{d}{dt}J_i(t) \quad = \quad w^{in}\nu_i^{in}(t) + w^{out}\nu^{out}(t) + \int_{-\infty}^{\infty} \mathrm{d}s \; W(s)C_i(s,t) \tag{1.13}$$

with $\nu_i^{in} = \overline{< S_i^{in} >}$ and $\nu^{out} = \overline{< S^{out} >}$ the mean firing rates of input neurons and neuron considered, and

$$C_i(s,t) = \overline{< S_i^{in}(t+s)S^{out}(t) >} \tag{1.14}$$

the correlation term between the i-th input spike train at time $t + s$ and the output spike train at time t. Here $< f(t) >$ denotes the ensemble average of an arbitrary function $f(t)$ and $\overline{f(t)} := T^{-1} \int_t^{t+T} \mathrm{d}t' \; f(t')$ its time average in an interval T.

For the purpose of fine-tuning, STDP is applied to the synaptic pattern of a map that gets additional input from an already calibrated *teacher* map. The ensuing *supervised* learning can successfully explain map formation [52, 56]. In the first stage of brain development, however, all available maps show a bad resolution [223, 225]. One map nevertheless has to function as the teacher map to calibrate the other sensory maps. Given that the *multimodal* map as combination of all unimodal information from the available sensory systems is the most reliable and precise map, Chap. 6 suggests map formation based on the intrinsic multimodal teacher. As a consequence, map formation corresponds to a mutual improvement within *all* sensory maps.

Hear the other side. (Audi partem alteram.)
Saint Augustine (354 - 430 AD)

2. Modeling Internally Coupled Ears: The ICE model

Lizards, frogs, alligators, and many birds possess a specialized hearing mechanism: internally coupled ears (ICE) where the tympanic membranes connect through a large mouth cavity so that the vibrations of the tympanic membranes influence each other. This coupling enhances the phase differences and *creates* amplitude differences in the tympanic membrane vibrations. Both cues show strong directionality. This chapter presents a general model of internally coupled ears, the *ICE model*, that consists of two parts. On the one hand, a three-dimensional model of internally coupled ears with a simplified cylindrical mouth cavity allows for calculation of a complete vibration profile of the membranes. On the other hand, numerical simulations of the eigenfunctions in exemplary, realistically reconstructed mouth cavities further estimate the effects of the complex geometry. The model with the cylindrical mouth cavity additionally provides the opportunity to incorporate the effect of the asymmetrically attached extracolumella of the middle ear, which leads to the activation of higher membrane vibration modes. Incorporating this effect, the ICE model can explain measurements taken from the tympanic membrane of a living lizard, for example, data demonstrating an asymmetrical spatial pattern of membrane vibration. As the analytical calculations show, the internally coupled ears increase the directional response, appearing in large directional internal amplitude differences (iADs) and in large internal time differences (iTDs).

2.1 Introduction

In contrast to mammals, where the tympanic membranes are independent of each other, lizards possess *internally coupled ears* where the tympanic membranes connect through large Eustachian tubes, as illustrated by Fig. 2.1. Thus, a signal arriving at one of the tympanic membranes can propagate through the internally interconnected cavities and influence

the vibration of the other. Consequently, one tympanic membrane shows a directional response to incoming sound signals, a so-called pressure-gradient receiver characteristic as first described by Autrum [4] and later by Michelsen [150] for locusts. Similar internal coupling is present in many biological systems such as frogs [e.g., [29, 49, 103]], birds [e.g., [88] quails and [33] barn owls below 3 kHz], and lizards [in particular, [27, 28]]. Reviews by Carr et al. [19, 20] cover the evolutionary aspects of this coupling.

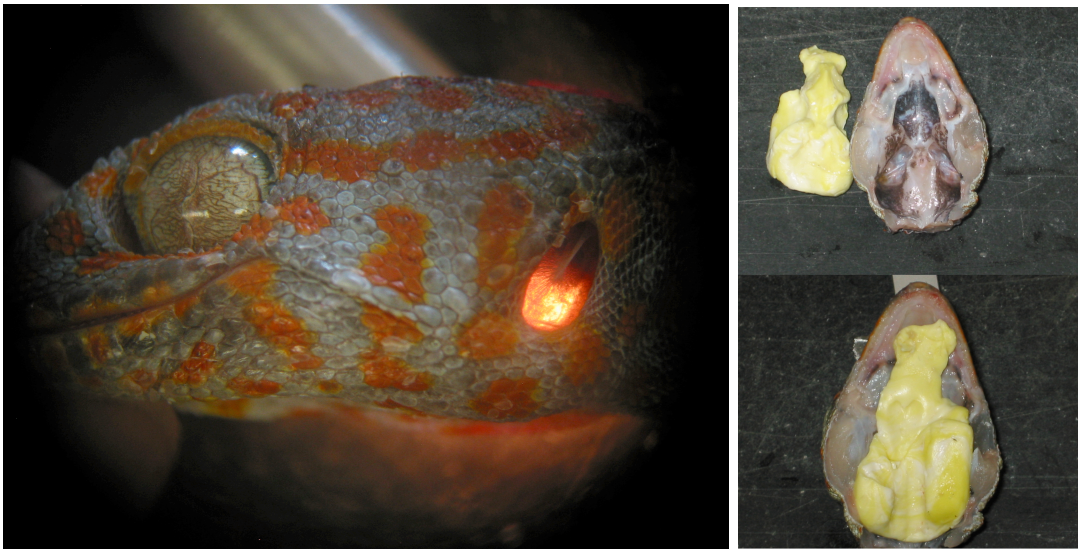

Figure 2.1: *Left: Picture of a Tokay gecko's head (snout to the left). A light source on the other side of the head illuminates the tympanic membrane from the back. The cartilaginous element attached to the membrane is part of the middle ear, called* extracolumella. *Right: Cast imprint of the mouth cavity of a gecko. Upper jaw of the lizard with the mouth pointing to the top of the picture. The figures illustrate the coupling of the tympanic membranes through large internal cavities giving rise to internally coupled ears (ICE). Photographs courtesy of Jakob Christensen-Dalsgaard.*

For ICE, phases and amplitudes of the tympanic membrane vibrations vary with the direction of the sound source, and they do so in a rather pronounced way. For animals of small head size amplitude differences between tympanic vibrations exclusively arise from ICE. This is because the small head results in negligible sound diffraction effects such that signals from the same sound source arrive at the two ears at slightly different times but with similar intensities. In contrast to systems having independent ears, the coupling of the tympanic membranes *creates* directional cues through the amplitudes of the tympanic membrane vibrations. ICE therefore translate the exterior phase shifted signals into tympanic membrane vibrations that vary both phase *and* amplitude with the direction of the sound source. As suggested by Jørgensen [103], a possible neuronal subtraction of the logarithmic vibration amplitudes from the two tympanic membranes, the internal amplitude difference (iAD), could further sharpen directionality.

Below is a three-dimensional model with a cylindrical mouth cavity and adjoined membranes, developed to elucidate the principal mechanism of internally coupled ears.

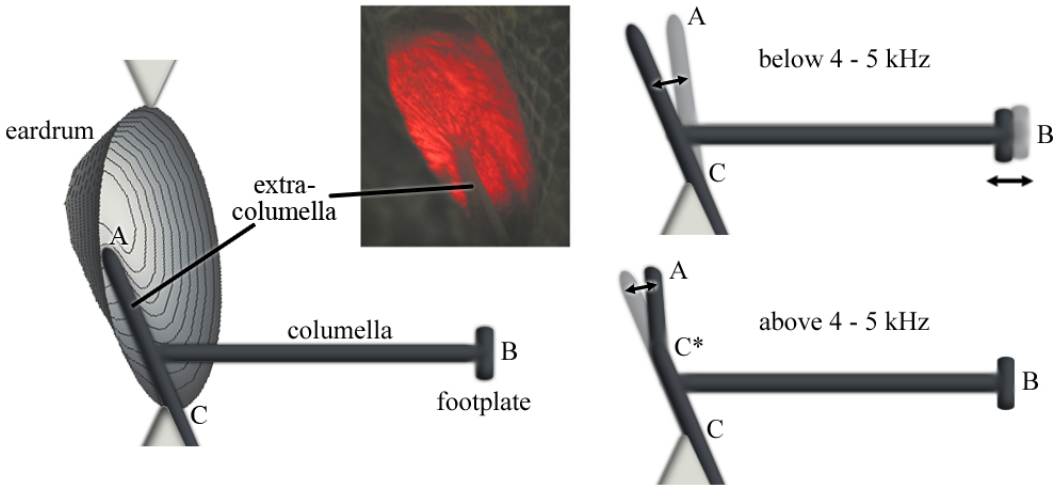

Figure 2.2: *Construction and functioning of the middle ear as reproduced from Manley [143]. Left: Schematic view of the tympanic membrane with the following middle ear. The cartilaginous element between A and C is attached to the tympanum and called* extracolumella. *The* columella *then directly connects via its footplate B to the oval window and transmits vibrations of the tympanic membrane to the cochlea. Top left: Photograph of the tympanic membrane with the attached extracolumella due to Jakob Christensen-Dalsgaard. Right top: For frequencies below* $4-5$ *kHz [143] the extracolumella between A and C is stiff and can only hatch around the point C. Thus, the displacement of the extracolumella and the attached tympanic membrane is linear in radial direction of the circular tympanum. The ICE model incorporates this restriction as additional boundary condition for the membrane displacement over the line of the extracolumella attachment; cf. Fig. 2.5. The footplate B reproduces the movement of the tip of the extracolumella A in an amplified manner. Right bottom: For frequencies above* $4-5$ *kHz (not considered in the following) the extracolumella starts to bend and a flection occurs at the point* C^* *in the columella-extracolumella link. The footplate B is no longer activated by the movement of the tip of the extracolumella A. As a consequence, the appearance of flection limits the hearing range of the animal.*

This first part of the ICE model is mathematically tractable and contains the most important aspect of ICE, viz., the coupling of the membranes through the internal mouth cavity. The vibrations of the membranes are calculated as a function of the force difference between the external pressure from the sound source and the pressure from the moving air in the internal mouth cavity. In the cavity itself, a wave equation with temporally varying boundary conditions at the eardrums describes the air movement. Consequently, the analytically soluble ICE model is the most fundamental, physically correct model for the three-dimensional structure of ICE. Furthermore, one can adapt this general model to any specific animal, in this case the lizard, and incorporate additional features.

One notable specification in the ICE model for lizards is the middle ear. Lizards have only one middle ear bone, the *columella*, that contacts the eardrum via a cartilaginous element, the *extracolumella*. As a consequence, the tympanic membrane cannot vibrate freely but is loaded by the accessory structure of the middle ear; see Fig. 2.2. The extracolumella moves as a completely stiff bar. Flection can only occur in the columella-

extracolumella attachment from frequencies above approximately $4 - 5$ kHz [143] and is not considered in the following. The stiffness of the extracolumella forces the membrane into a linear displacement over the line of the attachment. The solution therefore becomes a linear combination of eigenmodes for the free, unloaded membrane. As a consequence, higher modes emerge in addition to the fundamental mode. These modes cause a complex, asymmetric spatio-temporal pattern over the membrane as measured by Manley [144]. Given that the extracolumella is attached to the eardrum, the ICE model represents the vibration of the membrane, either to the right or to the left, as a linear combination of the external pressure excitation of the two tympanic membranes. The coefficients are, in this case, functions reflecting the internal response of the system to ipsilateral and contralateral local stimulation. For *Hemidactylus frenatus* and *Tokay gecko* amplitudes of the membrane vibrations, iADs, and iTDs are calculated and compared to experimental data.

A consistent, realistic evaluation of ICE further requires consideration of sound transmission variation due to the mouth cavity shape. The ICE model therefore further contains a numerical eigenfrequency analysis of the realistic mouth cavities. Three-dimensional scans of a cast imprint from the mouth cavity of *Hemidactylus frenatus* and a set of scans from slices through the head of *Tokay gecko* (cf. Fig. 2.1, right) allow a reconstruction of exemplary realistic mouth cavities. Numerical simulation programs can load the resulting meshes and calculate the eigenfunctions and the corresponding eigenfrequencies.

In previous work, lizard ear models consisted of a three impedance electrical circuit [27, 28], based on Fletcher [50, p. 164]. In the impedance model, abstract impedances detached from the geometric properties of the system represent both the tympanic membranes and the mouth cavity. Though electrical circuit models can explain the general pattern and magnitude of directionality, they can be seen as a truly crude approximation to the real acoustics of the ear, for example, by treating the cavity volume as a single impedance. Neglecting interesting phenomena, such electrical circuit models reduce a complicated system to a model with a small number of lumped parameters. For example, the attached extracolumella of the middle ear only influences the value of the membrane impedance; an average value for the whole membrane replaces the two-dimensional vibration profile of the membrane; the complex form of the mouth cavity is represented by its volume.

In contrast, the ICE model, as it contains a geometric representation of internally coupled ears, offers the possibility of analyzing the membrane vibration in *spatial* detail. This is crucial, for example, to understanding the effect of the attached extracolumella of the middle ear on membrane vibration; cf. the results below. In addition, the influence of the realistic mouth cavity can be estimated by calculating its lowest eigenfrequency. Furthermore, an "acoustical" model based on electrical circuits is valid only when the dimensions of the circuit elements are small as compared to the wavelength [159]. Given such a restricted point of view, a more general model will represent a broader range of systems far better and will also apply to larger animals having ICE.

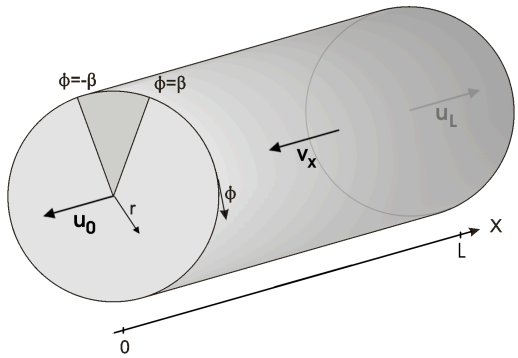

Figure 2.3: *Geometric representation of internally coupled ears as used for the first part of the ICE model. The internal cavity is represented by a cylinder using parameters* (x, r, ϕ) *with length* L *and radius* a *so that* $0 \leq x \leq L$ *and* $0 \leq r \leq a$*. Where* β *is estimated from anatomical data the circle segment* $-\beta \leq \phi \leq \beta$ *represents the contact surface between the membrane and the extracolumella. Arrows mark the orientations of the coordinate systems for the displacements* u_0 *and* u_L *of the moving membranes and for the velocity* v_x *of the moving air in the mouth cavity.*

2.2 Derivation of the ICE model with cylindrical mouth cavity

The equations derived in the following section describe the steady state solution for the tympanic membranes' vibrations, neglecting onset phenomena. The derivation step-by-step considers the different elements participating in the process. The cylindrical ICE model consists of elementary acoustic elements: membranes and an air filled tube as described in acoustic textbooks, such as Rschevkin [187, p. 107], Temkin [211, p. 247], and Fletcher [50, p. 73]. All used functions and parameters of the model are collected in Table 2.1.

As shown in Fig. 2.3, the model's geometry consists of a cylinder representing a simplified mouth cavity, using cylindrical coordinates (x, r, ϕ) as the parameters. The length of the cylinder is denoted by L and its radius by a. The two circular surfaces at $x = 0$ and $x = L$ represent the tympanic membranes. The segment $-\beta \leq \phi \leq \beta$ of the membrane models the extracolumella as a contact surface between the membrane and the attached extracolumella.

In the first step of sound processing, the external sound pressure excites the tympanic membranes. Consequently, the following Sec. 2.2.1 gives a formulation for the sound pressure arriving at the tympanic membranes from a given sound source. After arrival, the sound processing consists of two steps, viz., outside activation of the membranes and excitation of the air in the internal mouth cavity. Section 2.2.2 describes the fluctuating air in the internal cavity using a wave equation.
While equations describing a loaded force-driven membrane are presented in Sec. 2.2.3 here the derived equations account for a general force-driven membrane without specifying the eigenmodes. The latter are then calculated in a second step, though notably, the attached extracolumella requires additional modifications via boundary conditions.
Finally, Sec. 2.2.3 shows that adaptation of the boundary conditions leads to coupling of the

Table 2.1: *Functions and parameters used in the ICE model with cylindrical mouth cavity.*

θ	Horizontal angle of the source with respect to the head midline
ω	Angular excitation frequency of the incoming sound wave
$p_{ex}(\theta, \omega; t)$	Incoming sound wave
$\lvert p_{ex}\rvert$	Amplitude of the incoming sound wave
$\Psi(r, \phi; t)$	Driving force of the membrane
$u(r, \phi; t)$	Displacement function over the membrane, subscripts denote the ipsilateral ($x = 0$) and contralateral ($x = L$) membrane
c_M	Propagation velocity on the membrane
$f_{mn}(\phi, r)$	General eigenmode of the membrane displacement
$J_m(k_{mn}r)$	Bessel function of degree m with zero n
$-\beta \leq \phi \leq \beta$	Contact surface between tympanic membrane and extracolumella
$g_l(\phi, r)$	Eigenmode of the loaded membrane displacement
$\chi[(C_{mn})_{(m,n)}]$	Integrated quadratic error between a linear displacement and the function $\sum_{(m,n)} C_{mn} f_{mn}(r, \beta)$
(r, ϕ, x)	Cylinder coordinates
$p(x, r, \phi; t)$	Pressure function in the mouth cavity
$v_x(x, r, \phi; t)$	Velocity function in the mouth cavity
k	Wave number $k = \omega/c$ with the sound velocity $c = 343$ m/s
$G_{\text{ipsi}}(r, \phi)$	Ipsilateral filter
$G_{\text{contra}}(r, \phi)$	Contralateral filter

two domains. Taken together, the membrane vibration results from a linear combination of the external sound inputs with an ipsi- and a contralateral filter, defined as the response of an eardrum to contralateral and ipsilateral local stimulation (so only one sound input).

2.2.1 External sound input

Let p_{ex} be a harmonic external pressure wave with angular frequency ω and amplitude $|p_{ex}|$. The source is located at a horizontal angle θ measured from the central axis of the head. For small head size the sound signals arriving at the two tympani only differ in phase due to varying arrival times $kL/2\,[\sin(\theta) + \theta]$ with wave number $k = \omega/c$ and c as sound velocity in air; see Fletcher [50, p. 154] for details. Assuming a flat head form, the effect of sound traveling around the head vanishes and the phase difference for a lizard reduces to $kL\sin(\theta)$; cf. Fig. 2.4. Christensen-Dalsgaard [27, 28] experimentally verified that differences in amplitude and phase due to diffraction around the head and body of

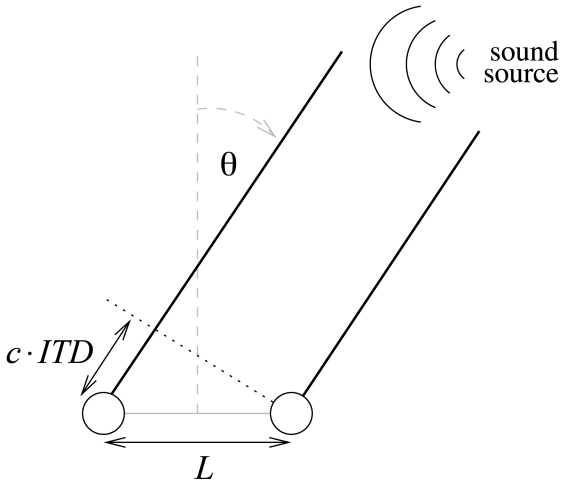

Figure 2.4: *Illustration of interaural time difference (ITD) as it arises from a sound source at angle θ with interaural distance between the tympani L and velocity of sound c. Depending on the angle of the sound source θ the distance between sound source and contralateral ear is longer than the distance between sound source and ipsilateral ear. The sound signal therefore reaches the contralateral ear later than the ipsilateral ear. The difference between the arrival times at the tympani is denoted by $ITD = L/c \sin(\theta)$.*

the lizard are negligible for the frequencies used here. The pressure functions at the two tympanic membranes ($x = 0$ and $x = L$) are therefore given by

$$\begin{aligned} p_{ex_0}(\theta, \omega; t) &= |p_{ex}|\ e^{ikL\sin(\theta)/2} e^{i\omega t}\ , \\ p_{ex_L}(\theta, \omega; t) &= |p_{ex}|\ e^{-ikL\sin(\theta)/2} e^{i\omega t} \end{aligned} \tag{2.1}$$

where the subscripts denote the corresponding membrane.

2.2.2 Internal cavity

The next section specifies the equations that describe the moving air in the internal cavity. Common acoustic models (cf. acoustic textbooks, such as Rschevkin [187, p. 107] and Temkin [211, p. 247]) assume that the air is moving due to pressure $p(x, r, \phi; t)$ described by the wave equation in cylindrical coordinates (cf. Fig. 2.3),

$$\frac{1}{c^2}\frac{\partial^2 p(x,r,\phi;t)}{\partial t^2} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\ \frac{\partial p(x,r,\phi;t)}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 p(x,r,\phi;t)}{\partial \phi^2} + \frac{\partial^2 p(x,r,\phi;t)}{\partial x^2} \tag{2.2}$$

with the sound velocity c as the propagation velocity. The above equation implies that there is no sound input through the mouth. That is, the mouth is closed, which is typical for a waiting animal. Adaptation of the *boundary conditions* (below) further incorporates the vibration of the two tympanic membranes.
To solve Eq. (2.2), a separation ansatz

$$p(x, r, \phi; t) = f(x)g(r)h(\phi)e^{i\omega t} \tag{2.3}$$

leads to

$$\begin{aligned} k^2 f(x)g(r)h(\phi) + f(x)h(\phi)\left[\frac{\partial^2 g(r)}{\partial r^2} + \frac{1}{r}\frac{\partial g(r)}{\partial r}\right] \\ + f(x)g(r)\frac{1}{r^2}\frac{\partial^2 h(\phi)}{\partial \phi^2} + g(r)h(\phi)\frac{\partial^2 f(x)}{\partial x^2} &= 0 \end{aligned} \tag{2.4}$$

with $k := \omega/c$. Separating the parameters yields the set of equations

$$\frac{d^2 f(x)}{dx^2} + k_x^2 f(x) = 0 \tag{2.5}$$

$$\frac{d^2 h(\phi)}{d\phi^2} + q^2 h(\phi) = 0 \tag{2.6}$$

$$\frac{\partial^2 g(r)}{\partial r^2} + \frac{1}{r}\frac{\partial g(r)}{\partial r} + \left[\underbrace{(k^2 - k_x^2)}_{=:k_{qs}^2} - \frac{q^2}{r^2}\right] g(r) = 0 \tag{2.7}$$

with separation constants q and k_x. The last equation is known as Bessel's differential equation [35, p. 313] that defines the Bessel functions $J_q(k_{qs}r)$ of the first kind, of order q, and with some constant k_{qs}. The solutions are then given by

$$f(x) = e^{\pm i k_x x}\ ,\ h(\phi) = e^{\pm iq\phi}\ ,\ \text{and}\ g(r) = J_q(k_{qs}r)\ . \tag{2.8}$$

In other words, the pressure p is given by a linear combination of waves propagating along the x-axis with wave number $k_x = \omega/c$ multiplied by functions of r and ϕ. Furthermore, the linear conservation of impulse for small displacements

$$-\nabla p = \rho\ \frac{\partial \mathbf{v}}{\partial t} \tag{2.9}$$

allows for the direct calculation of the particle velocity in the x direction,

$$\begin{aligned} v_x(x,r,\phi;t) = \sum_{q=0}^{\infty}\sum_{s=0}^{\infty} \quad & \left[-\frac{k_x}{\rho\omega}\left(A_{qs}^{+}e^{iq\phi} + A_{qs}^{-}e^{-iq\phi}\right)e^{ik_x x}\right. \\ & \left. + \frac{k_x}{\rho\omega}\left(B_{qs}^{+}e^{iq\phi} + B_{qs}^{-}e^{-iq\phi}\right)e^{-ik_x x}\right] J_q(k_{qs}r)e^{i\omega t}\ . \end{aligned} \tag{2.10}$$

The coefficients k_{qs}, A_{qs}^{+}, A_{qs}^{-}, B_{qs}^{+}, and B_{qs}^{-} must be adjusted in such a way that the vibration profiles of the internal air and the two membranes at $x = 0$ and $x = L$ are equal.

As the mouth cavity is a closed volume, there is no mass transport. The velocity is consequently purely oscillatory and mathematically possible additive constants [cf. Eq. (2.9)] are set to zero. The coupling of the internal cavity with the membranes necessitates an analysis of the vibrating membranes before adjusting the coefficients.

2.2.3 Vibration of the membrane

The aim of the following analysis is, given an external sound pressure p_{ex}, to find functions $u_0(r,\phi;t)$ and $u_L(r,\phi;t)$ for the displacement of the two tympanic membranes. The temporal derivatives $\dot{u}_{0/L}(r,\phi;t)$ then yield the vibration velocity; cf. Fig. 2.3. The driving force $\Psi(r,\phi;t)$ is the difference between the external pressure and the force exerted by the fluctuating air in the internal cavity. The membranes are further loaded by the extracolumella, which is an additional condition in calculating the membranes' vibrational eigenmodes. As discussed previously, the derivation first concentrates on the force-driven membrane before calculation of the eigenmodes in a second step.

Force-driven membrane Let $u(r,\phi;t)$ be the displacement of one of the membranes in x-direction. For an applied force $\Psi(r,\phi;t)$ the function $u(r,\phi;t)$ has to fulfill the two-dimensional damped wave equation in polar coordinates [50, p. 78],

$$-\frac{\partial^2 u(r,\phi;t)}{\partial t^2} - 2\alpha\frac{\partial u(r,\phi;t)}{\partial t} + c_M^2\left[\frac{\partial^2 u(r,\phi;t)}{\partial r^2} + \frac{1}{r}\frac{\partial u(r,\phi;t)}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u(r,\phi;t)}{\partial \phi^2}\right] = \frac{1}{\rho_m d}\Psi(r,\phi;t) \tag{2.11}$$

with α being the damping coefficient of the membrane, c_M the wave propagation velocity on the membrane, ρ_m the membrane density, and d its thickness.

Here the membrane is fixed at the boundary $r=a$, which means $u(a,\phi;t)=0$. The first step to solve (2.11) is the calculation of the homogeneous solution (that is, $\Psi=0$). Again a separation ansatz leads to equations of the form (2.6) and (2.7) so that the solution calculates as

$$u(r,\phi;t) = \sum_{m=0}^{\infty}\sum_{n=1}^{\infty} f_{mn}(r,\phi)e^{i\omega_{mn}t} \tag{2.12}$$

with $\omega_{mn} = k_{mn}c_M$ the resonance frequency of the eigenmodes

$$f_{mn}(r,\phi) \;=\; \left(M_{mn}^{+}e^{im\phi} + M_{mn}^{-}e^{-im\phi}\right)J_m(k_{mn}r)\,, \qquad m,n \;\in \mathbb{N}\,, \tag{2.13}$$

where J_m denotes a Bessel function of order m and k_{mn} guarantees that $J_m(k_{mn}a)=0$ is the n-th zero of J_m. In addition, the boundary conditions later determine the parameters M_{mn}^{+} and M_{mn}^{-}. The eigenmodes $f_{mn}(r,\phi)$ are orthogonal when integrated over the membranes,

$$\int_S f_{mn}(r,\phi)f_{ij}(r,\phi)\mathrm{d}S = \delta_{mi}\delta_{nj} \tag{2.14}$$

with $\int_S \mathrm{d}S$ being the integral over the surface of the membrane.

The next step consists of the calculation of the inhomogeneous solution to Eq. (2.11). In general, the force acting on the membranes consists of two parts. On one hand there is the (with respect to ϕ and r) constant external pressure force p_{ex} from the acoustic

excitement, and on the other hand the force from the oscillatory air in the internal cavity that depends on angular and radial coordinates ϕ and r. Whereas the external pressure is given by Eq. (2.1), the internal force results from the vibration velocity $v_x(r, \phi; t)$ shown in Eq. (2.10). But first, the boundary conditions help to simplify the following calculations. In general, the boundary conditions require that the velocities in the cylinder and of the vibrating membrane at $x = 0$ and $x = L$ are equal [211, p. 131], that is, the air in the cavity directly at the membrane moves with the same velocity as the membrane itself

$$\begin{aligned} \dot{u}_0(r, \phi; t) &= -v_x(0, r, \phi; t) \ , \\ \dot{u}_L(r, \phi; t) &= v_x(L, r, \phi; t) \ . \end{aligned} \tag{2.15}$$

The minus sign in the upper equation ensures that the directions defined for the displacement functions of the membranes u_0 and u_L are in line with the coordinate system of the cylinder; cf. Fig. 2.3. The functions $\dot{u}_{0/L}$ and v_x are linear combinations of orthogonal functions from the same set. Furthermore, only the external force from the sound source can activate movements of the internal cavity. From Eq. (2.15), the pressure function p in the internal cavity can only contain the vibration modes determined by the membranes. Thus, the pressure p has to be a linear combination of the modes of the membranes leading to

$$p(x, r, \phi; t) \quad = \quad \sum_{mn} \left(A_{mn} e^{ik_x x} + B_{mn} e^{-ik_x x} \right) f_{mn}(r, \phi) e^{i\omega t} \tag{2.16}$$

where the f_{mn} are the eigenmodes of the membranes as defined in Eq. (2.13) and ω is the angular frequency of the external sound source. The above function still satisfies Eq. (2.2). The applied force is then given by

$$\Psi(r, \phi; t) \quad = \quad \begin{cases} p_{ex_0} - p(0, r, \phi; t) & \text{for } x = 0 \ , \\ p_{ex_L} - p(L, r, \phi; t) & \text{for } x = L \ . \end{cases} \tag{2.17}$$

This force can be inserted into Eq. (2.11) leading for $x = 0$ to

$$\begin{aligned} -\frac{\partial^2 u(r, \phi; t)}{\partial t^2} - 2\alpha \frac{\partial u(r, \phi; t)}{\partial t} + c_M^2 \left[\frac{\partial^2 u(r, \phi; t)}{\partial r^2} + \frac{1}{r} \frac{\partial u(r, \phi; t)}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u(r, \phi; t)}{\partial \phi^2} \right] \\ = \frac{1}{\rho_m d} \left[p_{ex_0} - \sum_{mn} \left(A_{mn} + B_{mn} \right) f_{mn}(r, \phi) e^{i\omega t} \right] \ . \end{aligned} \tag{2.18}$$

For a solution of the form

$$u_0(r, \phi; t) = \sum_{mn}^{\infty} C_{mn} f_{mn}(r, \phi) e^{i\omega t} \tag{2.19}$$

with coefficients C_{mn} to determine, Eq. (2.18) then transforms to

$$\begin{aligned} \sum_{mn}^{\infty} C_{mn} \left[\omega^2 - 2i\alpha\omega - c_M^2 \left(k_{mn}^2 - \frac{m^2}{r^2} + \frac{m^2}{r^2} \right) \right] f_{mn}(r, \phi) e^{i\omega t} \\ = \frac{1}{\rho_m d} \left[p_{ex_0} - \sum_{mn} \left(A_{mn} + B_{mn} \right) f_{mn}(r, \phi) e^{i\omega t} \right] \ . \end{aligned} \tag{2.20}$$

Multiplying the equation by f_{mn} and integrating over the surface yields the solution

$$u_0(r,\phi;t) = \frac{1}{i\omega}\sum_{mn}\Omega_{mn}^{-1}\left[-p_{ex_0}\int_S f_{mn}(\phi,r)\mathrm{d}S + (A_{mn}+B_{mn})\right] f_{mn}(\phi,r)e^{i\omega t} \tag{2.21}$$

with the definition

$$\Omega_{mn}^{-1} := \frac{1}{\rho_m d}\frac{i\omega}{(\omega_{mn}^2-\omega^2)+2i\omega\alpha} \tag{2.22}$$

where $\omega_{mn} = c_M k_{mn}$. An analogue derivation for $x = L$ leads to the solution

$$u_L(r,\phi;t) = \frac{1}{i\omega}\sum_{mn}\Omega_{mn}^{-1}\left[-p_{ex_L}\int_S f_{mn}(\phi,r)\mathrm{d}S + \left(A_{mn}e^{ik_xL}+B_{mn}e^{-ik_xL}\right)\right] f_{mn}(\phi,r)e^{i\omega t}\,. \tag{2.23}$$

In the next section the boundary conditions determine the remaining coefficients A_{mn} and B_{mn}.

Adapting the velocity to the boundary conditions According to (2.15), the boundary conditions require that the velocities in the cylinder and of the vibrating membrane at $x = 0$ and $x = L$ be equal. Equation (2.15) with (2.21) or (2.23) on the left side and (2.9) and (2.16) on the right side leads for orthogonal functions f_{mn} and $x = 0$ to

$$-p_{ex_0}\int_S f_{mn}(\phi,r)\mathrm{d}S + A_{mn} + B_{mn} = \frac{\Omega_{mn}}{\rho c}(A_{mn}-B_{mn}) \tag{2.24}$$

with $k_x = \omega/c$ and for the other side, $x = L$, to

$$-p_{ex_L}\int_S f_{mn}(\phi,r)\mathrm{d}S + A_{mn}e^{ik_xL} + B_{mn}e^{-ik_xL} = -\frac{\Omega_{mn}}{\rho c}\left(A_{mn}e^{ik_xL}-B_{mn}e^{-ik_xL}\right)\,. \tag{2.25}$$

The two equations allow for a calculation of the coefficients A_{mn} and B_{mn} and a consequent representation of the solutions in the form

$$\begin{aligned}\dot{u}_0(r,\phi;t) &= G_{\mathrm{ipsi}}(r,\phi)\; p_{ex_0} + G_{\mathrm{contra}}(r,\phi)\; p_{ex_L}\;,\\ \dot{u}_L(r,\phi;t) &= G_{\mathrm{ipsi}}(r,\phi)\; p_{ex_L} + G_{\mathrm{contra}}(r,\phi)\; p_{ex_0}\;,\end{aligned} \tag{2.26}$$

with the ipsilateral filter

$$G_{\mathrm{ipsi}}(r,\phi) = -\sum_{mn}\int_S f_{mn}(\phi,r)\;\mathrm{d}S\;\left\{\frac{-\rho c i\cot(k_xL)+\Omega_{mn}}{\left[\Omega_{mn}-\rho c i\cot(k_xL)\right]^2+\rho^2c^2\sin^{-2}(k_xL)}\right\} f_{mn}(\phi,r) \tag{2.27}$$

and the contralateral filter

$$G_{\mathrm{contra}}(r,\phi) = \sum_{mn}\int_S f_{mn}(\phi,r)\;\mathrm{d}S\;\left\{\frac{-\rho c i[\sin(k_xL)]^{-1}}{\left[\Omega_{mn}-\rho c i\cot(k_xL)\right]^2+\rho^2c^2\sin^{-2}(k_xL)}\right\} f_{mn}(\phi,r)\;. \tag{2.28}$$

Loaded membrane In a final step, the model has to account for the fact that the tympanic membranes cannot vibrate freely but are *loaded by the extracolumella*; cf. Fig. 2.2. The model assumes the additional load to be attached to a segment of the circle S_C where $-\beta \leq \phi \leq \beta$; cf. Fig. 2.3. That is, the eigenmodes have to fulfill the homogeneous part of Eq. (2.11) with additional boundary conditions for $\phi = -\beta$ and $\phi = \beta$. The form of the eigenmodes f_{mn} therefore remains as defined in (2.13) but is using the reduced membrane area $S \setminus S_C$. As the attached load is rigid in the ϕ-direction, the velocity and the displacement at a certain radius are the same for $\phi = \beta$ and $\phi = -\beta$ simplifying the eigenmodes in (2.13) to

$$f_{mn}(r, \phi) \quad \rightarrow \quad C_{mn} \cos(m\phi) J_m(k_{mn} r) \ , \tag{2.29}$$

with $m, n \ \in \mathbb{N}$ and some coefficients C_{mn}. The choice of possible excitable Bessel functions reduces because the extracolumella, the cartilaginous element of the middle ear attached to the membrane, is stiff in the r direction; cf. Fig. 2.2, top right. This means that the displacement amplitude for $\phi = \pm\beta$ has to be linear with respect to r leading to

$$\sum_{m,n} C_{mn} \cos(m\beta) J_m(k_{mn} r) \quad = \quad \frac{a - r}{a} \ . \tag{2.30}$$

That is, the solution for the loaded membrane becomes a linear combination of eigenmodes of the unloaded membrane. Figure 2.5 illustrates how a sum of eigenmodes of the unloaded membrane can ensure a linear displacement of the membrane over the line of attachment with the extracolumella. At this step in the derivation, normalization limits the maximal amplitude (at $r = 0$) of the linear displacement to 1. However, the applied external force determines the maximal amplitude further below.
To solve an equation like (2.30) for a coefficient C_{ij}, usually one multiplies the equation by the function f_{ij} and integrates over the membrane. The coefficients could be calculated easily given the f_{mn} are orthogonal; cf. Eq. (2.14). Unfortunately, the given relation (2.30) is true only for the angles $\phi = \pm\beta$, which means that the functions f_{mn} are already evaluated at this angle and integration over the surface S is impossible. As a consequence, the coefficients C_{mn} should minimize the error function

$$\chi[(C_{mn})_{(m,n)}] := \int_0^r \left(\sum_{m,n} C_{mn} \cos(\pm m\beta) J_m(k_{mn} r) - \frac{a - r}{a} \right)^2 \mathrm{d}r \tag{2.31}$$

that measures the degree of similarity between the constructed test function and the linear function. Minimizing the error with respect to the coefficients C_{mn} leads to $\partial\chi / \partial C_{ij} = 0$, where (i, j) is one specific pair from the possible combinations (m, n). Interchanging the order of derivation and integration, the minimization condition leads to the linear system of equations

$$\begin{aligned} \sum_{m,n} C_{mn} \cos(m\beta) \cos(i\beta) \int_0^r J_m(k_{mn} r) J_i(k_{ij} r) \mathrm{d}r \\ = \frac{1}{a} \ \cos(i\beta) \int_0^r J_i(k_{ij} r) \, (a - r) \, \mathrm{d}r \end{aligned} \tag{2.32}$$

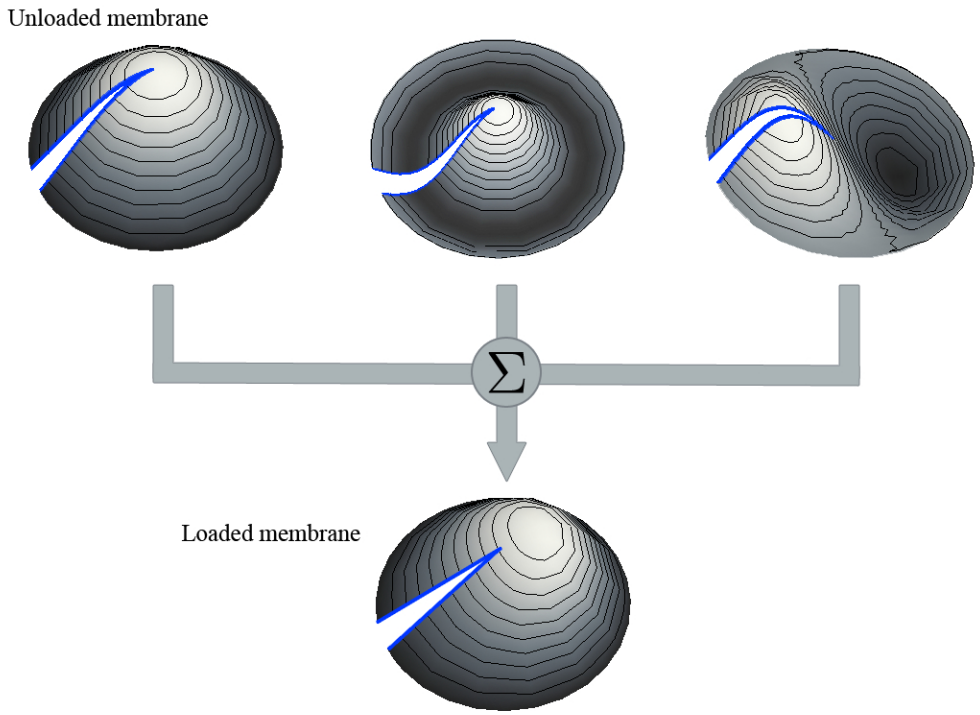

Figure 2.5: *Displacement of a loaded membrane (bottom) as linear combination of eigenmodes of the unloaded membrane (top); cf. Eq.* (2.33)*. The figure illustrates composition of the eigenmode* g_0 *from Table 2.2. The omitted segment (blue curves) represents the contact surface between membrane and extracolumella; cf. circle segment* $-\beta \leq \phi \leq \beta$ *in Fig. 2.3. An unloaded free membrane usually shows a symmetrical displacement corresponding to the fundamental mode (top left). The displacement of the fundamental mode is nonlinear in the radial direction. Below* $4 - 5$ *kHz the attached stiff extracolumella of the middle ear, however, forces the membrane into a linear displacement over the line of the attachment; cf. Fig. 2.2, top right. The model incorporates this restriction as additional boundary condition at the line of the attachment. As a consequence, the solution for the loaded membrane becomes a linear combination of eigenmodes for the free, unloaded membrane. Since higher modes do not have to be symmetrical (see, e.g., top right eigenmode) the membrane displacement can show a complex, asymmetric spatio-temporal pattern despite the linear displacement over the line of attachment with the extracolumella.*

for all (i, j) that occur in Eq. (2.29). This system is infinite, as is the number of eigenmodes, requiring restriction to a certain number of functions that suffice to approximate linear displacement. The limited set of eigenmodes that are taken into account is denoted by Z.

As an example, the calculations below use the set $Z = \{(m, n) \mid 0 \leq m \leq 5 \text{ and } 0 \leq n \leq 5\}$ of selected eigenfunctions. Higher modes result in a strongly fluctuating surface that has not been observed in experiments so far.

Table 2.2: *The loaded membrane: two possible eigenmodes g_0 and g_1 of the loaded membrane with their relative error χ when combining three eigenmodes of the free membrane to fulfill the condition of the attached extracolumella with an error $\chi \leq 0.05\%$. The three functions could be chosen out of a pool of 25 eigenmodes given by the set Z. The functions are orthogonalized by a Gram-Schmidt method. The eigenmode g_0 and its components are visualized in Fig. 2.5.*

n	g_n	χ [%]
0	$425.6\ J_0(k_{01}r) - 51.3\ J_0(k_{02}r)$ $-183.7\ \cos(\phi)\ J_1(k_{11}r)$	0.0008
1	$6.8\ J_0(k_{01}r) + 369.4\ J_0(k_{02}r)$ $-160.0\ \cos(\phi)\ J_1(k_{11}r) + 768.3\ \cos(2\phi)\ J_2(k_{21}r)$	0.0275

Furthermore, only a certain number of eigenfunctions from this set Z are necessary to build a linear function with minimal error. Likely the combination of two or three eigenfunctions is sufficient, thus in a first step different numbers of combined functions f_{mn} with $(m, n) \in Z$ are tested. The results show that three eigenmodes from the possible set Z already generate a linear profile with an error below 0.05%. The following calculations therefore center around combinations of three eigenmodes with indices Z_l with $l \in \mathbb{N}$ from the above set Z. Solving the resulting linear equation system (2.32) for the coefficients C_{mn} gives the eigenmodes

$$g_l(r, \phi) = a_l \sum_{(m,n)\in Z_l} C_{mn} \cos(m\phi) J_m(k_{mn} r) \tag{2.33}$$

with a scaling factor a_l. All chosen eigenmodes g_l must yield an error χ below 0.05%. Each of these functions is a combination of three Bessel functions. Two eigenmodes g_l can contain the same function f_{mn} (only a complete agreement of all three modes is excluded). Thus the new eigenmodes g_l are not necessarily orthogonal. Only subsequent application of an appropriate method such as Gram-Schmidt (described in textbooks, such as Cohen [32, p. 82] or Reed and Simon [179, p. 46]) leads to orthogonal eigenmodes. Taken together, this method produces a set of orthogonal eigenmodes that fulfill the membrane equation

and the additional boundary conditions induced by the attached extracolumella.
For the above set Z, Table 2.2 shows the two possible orthogonal eigenmodes g_l for a loaded membrane and the relative error $\chi \leq 0.05\%$. The eigenmode g_0 and its components are visualized in Fig. 2.5. In general, the eigenmodes correspond to the fundamental eigenmode f_{01} with modifications from the additional higher modes f_{02}, f_{11}, and f_{21}. The resonance frequencies of the eigenmodes g_l as needed for Eq. (2.21) are given by the resonance frequencies of the contributing modes f_{mn} depending on their excited fractions in the linear combination.

2.3 Numerical simulation of the eigenfunctions of realistic mouth cavities

As the complex shape of a realistic mouth cavity prevents an analytical description, the ICE model contains a numerical eigenfrequency analysis to estimate the eigenfrequencies of the internal volume. The eigenfrequency analysis is now performed in two steps. The first consists of the construction of simple geometries and the calculation of their eigenfrequencies with the simulation program COMSOL (see http://www.femlab.de/). In doing so, the influence of a narrowing or widening of a geometry can be estimated; see Fig. 2.13 for results.

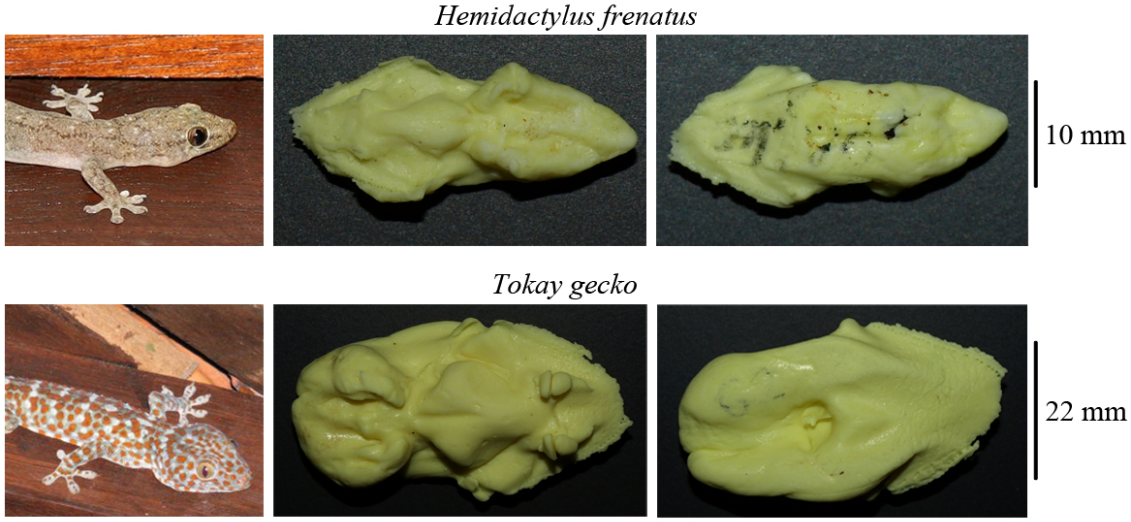

Figure 2.6: Hemidactylus frenatus *(top row) and* Tokay gecko *(bottom row) and casts of their mouth cavities produced by Jakob Christensen-Dalsgaard. As indicated on the right-hand side, the interaural distance is* 10 *mm for* Hemidactylus frenatus *and* 22 *mm for* Tokay gecko. *In addition to different interaural distances the cavities vary in shape and volume so that they cover a large range of possible realistic mouth cavities.*

In a second step, the eigenfrequencies of realistic mouth cavities for *Hemidactylus frenatus* and *Tokay gecko* (cf. Fig. 2.6) are calculated numerically. With different interaural

distances of 10 mm for *Hemidactylus frenatus* and 22 mm for *Tokay gecko* the analysis of the two lizards covers a large range of possible realistic mouth cavities. The prerequisite for a numerical eigenfrequency analysis is a closed three-dimensional mesh of the corresponding geometry. For *Hemidactylus frenatus* three-dimensional scans [154] of a cast imprint of the mouth cavity allow for a reconstruction of the realistic mouth cavity. Figure 2.7 illustrates the three major steps to construct the spatial mesh from the scanned cast of the mouth cavity. The first one (left) is alignment of the independently scanned profiles from the different sides of the geometry. In a second step (middle), the aligned profiles are combined in a closed geometry that permits, in a third and final step (right), the creation of a three-dimensional mesh of the mouth cavity interior.

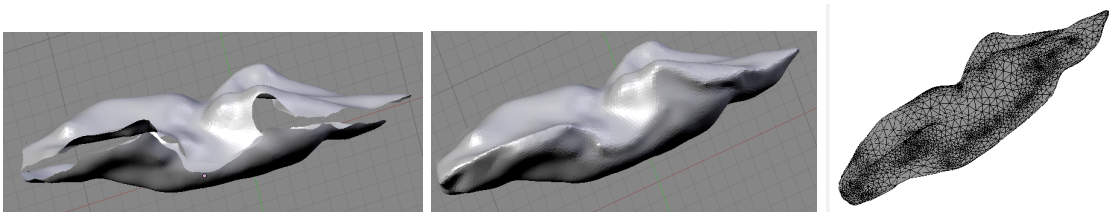

Figure 2.7: *Construction of a three-dimensional mesh from the scanned cast of the mouth cavity of* Hemidactylus frenatus. *Left: Alignment of the scanned profiles of the geometry from different sides. Middle: Construction of a closed geometry. Right: Construction of the three dimensional mesh. For the editing of the mesh the programs ICEM CFD and BLENDER were used. In a next step, the three-dimensional mesh of the realistic mouth cavity can be imported into the numerical simulation program COMSOL for eigenfrequency analysis; cf. Fig. 2.14.*

For *Tokay gecko*, spatial reconstructions of the mouth cavity have been received from Catherine Carr and Bruce Young. Figure 2.8 (left) shows the original received geometry as automatically reconstructed from a set of scanned slices of the lizard's head. The mesh, however, contains holes and disconnected or undefined elements which prevents a direct import into numerical simulation programs. In several post-processing steps the geometry therefore has to be closed, simplified and smoothened (Fig. 2.8, middle) to finally generate a closed three-dimensional mesh of the geometry (Fig. 2.8, right).

For the editing of both meshes the programs TGRID and ICEM CFD from ANSYS (see http://www.ansys.com/products/) and BLENDER (see http://www.blender.org/) were used. For both *Hemidactylus frenatus* and *Tokay gecko*, the closed meshes of the mouth cavities allow for an eigenfrequency analysis in the numerical simulation program COMSOL. Incorporation of the tympani to set up a complete *numerical* ICE model could be realized in a next future step. Numerical calculations of the eigenfrequencies for the isolated mouth cavities, however, allow for an analysis of the mouth cavity's effect on the behavior of the whole system of internally coupled ears.

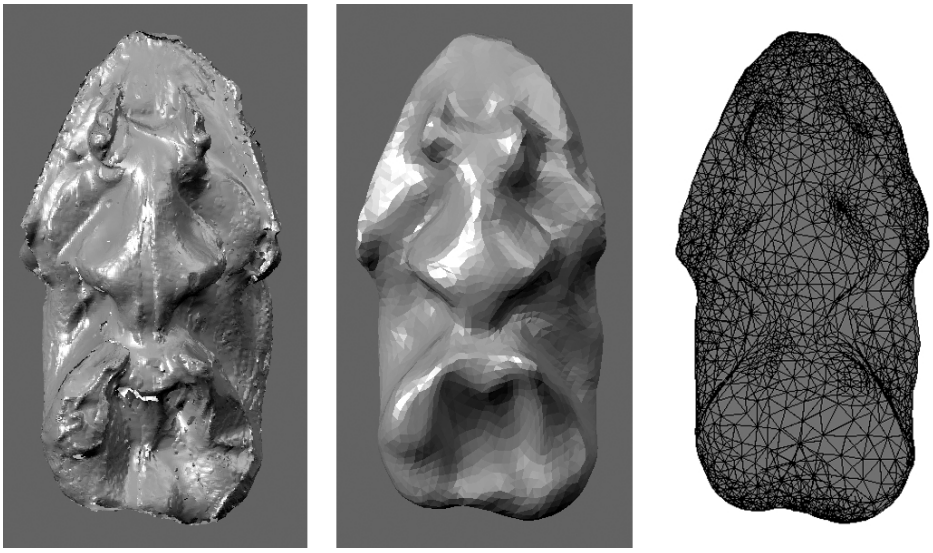

Figure 2.8: *Construction of a realistic mouth cavity of* Tokay gecko. *Left: Original mesh as automatically generated from scanned slices of the mouth cavity received from Catherine Carr and Bruce Young. Since the mesh at this stage contains holes and disconnected or undefined elements several post-processing steps are necessary. Middle: Closed geometry of the realistic mouth cavity as obtained from the original geometry using wrapping and smoothing algorithms from TGRID and BLENDER. Right: The obtained closed mesh of the realistic mouth cavity can be imported into the numerical simulation program COMSOL for an eigenfrequency analysis; cf. Fig. 2.19.*

2.4 Evaluation and results

Rating the significance of the ICE model necessitates a comparison of its predictions with measurements. To this end, the ICE model is evaluated using parameters from *Hemidactylus frenatus (gekkonid)* [28] as summarized in Table 2.3. All calculated results use the ICE model with a cylindrical mouth cavity.

The evaluation of the model for *Hemidactylus frenatus* proceeds in several steps. First, laser doppler vibrometry measurements of eardrum vibrations [28] are compared with the calculated data based on the ICE model. The analysis covers the vibration amplitude of the membrane, the gain of the interaural transmission pathway, the internal amplitude differences (iAD), and the internal time differences (iTD). In a second step, the eigenfunctions are numerically calculated for simple artificial geometries and for the realistic mouth cavity of *Hemidactylus frenatus*. Furthermore, the influence of the lowest eigenfrequency on iTDs and iADs is estimated.

To illustrate that the ICE model is a universal model for describing internally coupled ears and not only a specialized model for *Hemidactylus frenatus* the essential calculations described above are performed as well for *Tokay gecko*; cf. Sec. 2.4.3. Whereas *Hemidactylus frenatus* with an interaural distance of 10 mm is a rather small lizard, *Tokay gecko* is the second largest Gecko species. An additional confirmation of the ICE model for *Tokay gecko* therefore underlines that the ICE model can describe internally coupled ears for animals

Table 2.3: *Geometry parameters for* Hemidactylus frenatus *(gekkonid), obtained from Christensen-Dalsgaard [28].*

$L = 10$ mm	Length of the cylinder (interaural distance)
$a = 1.2$ mm	Radius of the tympanic membrane
$\alpha = 1000$ Hz$/(2 \cdot 1.2)$	Damping coefficient of the membrane
$\rho_m = 3.2$ mg/mm^3	Density of the membrane
$d = 10$ μm	Thickness of the membrane
$c = 343$ m/s	Velocity of sound

with largely varying interaural distance and mouth cavities.

The aim of the third step is to prove that the higher modes included in the cylindrical ICE model are responsible for the complex pattern of the vibration amplitude over the membrane as measured by Manley [144] for *Tokay gecko*. Therefore, calculation of a membrane profile of the vibration amplitude using the corresponding parameter set 2.4 allows for a comparison with the experimental data [144].

2.4.1 Directionality of the membrane vibration pattern

By means of the ICE model with a cylindrical mouth cavity, it is possible to estimate the dependence of the membrane vibration amplitude upon the specific frequency and direction of a sound source. To this end, it is necessary to evaluate the vibration amplitudes at a specified point on the membrane, which is usually the tip of the extracolumella. In the ICE model (cf. Fig. 2.3), the specific point is at the middle of the circular membrane, where $r = 0$.

The left side of Fig. 2.9 shows the logarithmic vibration amplitudes, dependent on the sound source direction and frequency as given by the ICE model for *Hemidactylus frenatus*. Therefore, one can directly compare the calculated (Fig. 2.9, left) and experimental (Fig. 2.9, right) data. The experimental data [28] consist of measured eardrum vibrations at the tip of the extracolumella of *Hemidactylus frenatus* using laser vibrometry. The plot shows vibration amplitude as a function of frequency (y-axis) and direction (x-axis). For every measured frequency, the mean square error between measured and calculated data averaged over the tested angular directions quantifies the goodness of the fit. The resulting deviation as illustrated in Fig. 2.10 lies around 1.3 dB re 1 mm/(s Pa). However, for frequencies above 5 kHz the error function systematically increases. In this frequency region, the experimental data show a systematic decline within the directional response pattern. In comparison, the calculated data predict a much stronger response. A possible explanation for this overestimation could be the occurrence of a limiting process at this frequency, such as bending of the extracolumella (cf. Fig. 2.2, right bottom) or an effect of the eigenfrequencies at the realistic mouth cavity of *Hemidactylus frenatus*. The ICE model does not consider a bending of the extracolumella, the inclusion of the latter could therefore be a future test application for the model. Numerical calculations below further detail the eigenfrequencies of the realistic mouth cavity.

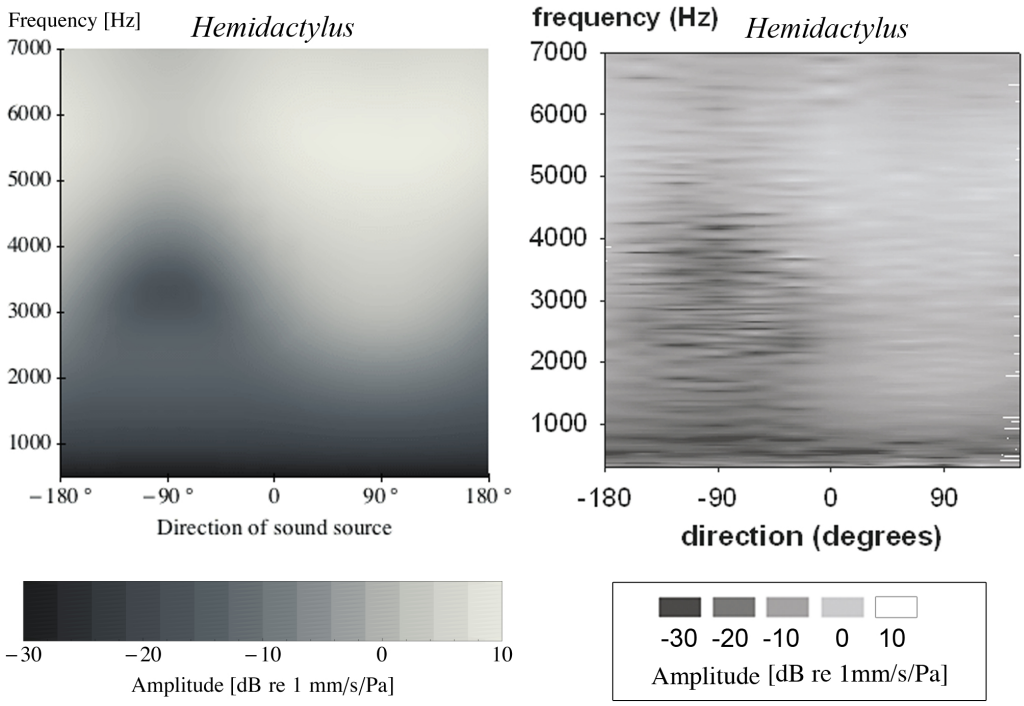

Figure 2.9: *Calculated (left) and experimental (right) amplitude of tympanic vibration for* Hemidactylus frenatus (gekkonid) *in dB re 1mm/(s Pa), i.e., the vibration amplitude with regard to a reference vibration velocity of 1 mm/s at 1 Pa. The amplitudes depend on the sound source direction (x-axis with directions varying from $-180°$ to $180°$; negative directions are contralateral, 0 is frontal, and positive directions are ipsilateral) and frequency (y-axis). The calculated results are based on the ICE model with a* cylindrical *mouth cavity. Measured eardrum vibration amplitudes for* Hemidactylus frenatus *are from Christensen-Dalsgaard [28]. The vibration amplitude pattern as calculated by the ICE model nicely reproduces experimental data. The model, however, overestimates the vibration amplitudes above 6 kHz and underestimates the response of the system for low frequencies. A quantitative comparison is presented in Fig. 2.10.*

In general, for both experimental and model data, at frequencies below 1 kHz the eardrum shows an identical vibration amplitude for all angles θ. Localization of a sound source based on the vibration amplitude is therefore not possible for these frequencies. In contrast, above 1 kHz the amplitude varies with angle θ. Consequently, for these frequencies the location of the sound source could be extracted from the amplitude variations. It is remarkable that, because of the special construction of ICE, these directional hints are already present at the level of the tympanic membranes.

Next, one can analyze the advantages of the transmission through the internal cavity in more detail by evaluating the transmission gain. Transmission gain is defined as the response ratio of eardrum vibration with contralateral and ipsilateral local stimulation (only one sound input). For local stimulation, one of the inputs is zero. Figure 2.11 shows

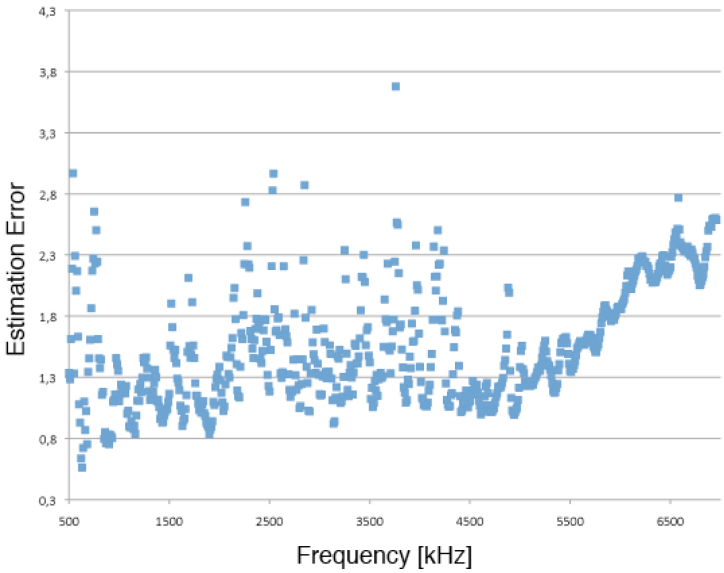

Figure 2.10: *Mean square error between measured and calculated amplitudes of tympanic vibrations for* Hemidactylus frenatus *averaged over the tested angular directions for every measured frequency. The error is around 1.3 dB re 1 mm/(s Pa); only for frequencies above 5 kHz the error function systematically increases indicating a possible attenuation of the biological data through lower eigenfrequencies of the realistic mouth cavity in comparison to the used cylindrical mouth cavity.*

the calculated data as light lines and the experimental results as black lines. The complex gain function consists of two parts, amplitude (Fig. 2.11, left) and phase (Fig. 2.11, right). Notably, experimental data from Christensen-Dalsgaard [28] present eardrum vibrations of *Hemidactylus frenatus* using a local sound source and compared the response to ipsi- and contralateral local stimulation. The qualitative form of those calculated gain functions over frequency is similar for data from the ICE model and experiments. The calculated data further show a smoother behavior and less extreme minima and maxima. This analytically soluble part of the ICE model, however, includes a cylindrical mouth cavity with a volume significantly smaller than the realistic one, thus additional modification is possible for the directional response that results from the eigenfrequencies of the realistically shaped mouth cavity.

For *Hemidactylus frenatus* the comparison of the data from the ICE model with the cylindrical mouth cavity and the experimental data [27] concludes with the evaluation of a possible neuronal processing based on the internal amplitude difference (iAD); see Fig. 2.12, top. The latter measures the vibration amplitude difference (in dB) between the input from the ipsi- and contralateral ear by subtracting the free field eardrum vibration (in dB) of the two eardrums. In fact, subtraction of dB values corresponds to division but is assumed to reflect the neuronal processing because, within a neuron's dynamic range,

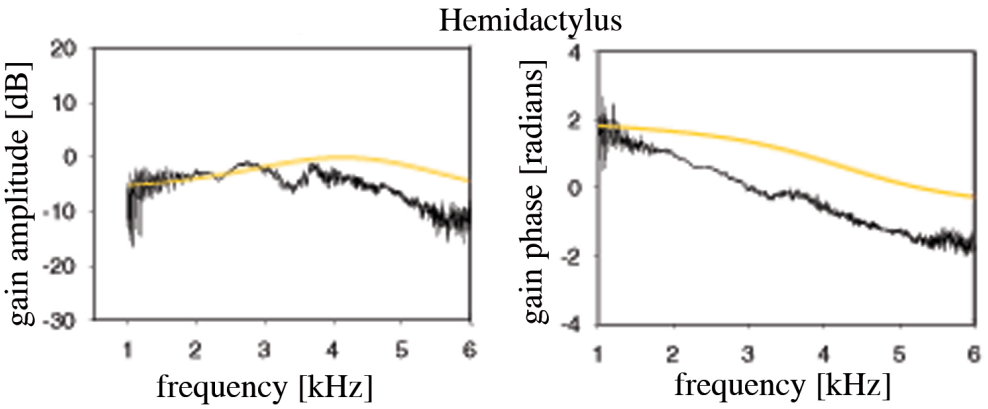

Figure 2.11: *Transmission gain amplitude (left side in dB) and phase (right side in radians) defined as the response ratio of eardrum vibration with contralateral and ipsilateral local stimulation (so only one sound input) for* Hemidactylus frenatus (gekkonid). *The transmission gain as calculated by the ICE model with a* cylindrical *mouth cavity is depicted as light line, the measured transmission gain [28] as black line. Calculated and experimentally measured transmission gains show the same tendencies and are in a good agreement. Only for frequencies above* 4 *kHz is there a gap between the calculated and the measured gain.*

the spike rate is a linear function of the sound level in dB. The iAD is therefore defined as

$$\text{iAD} \quad := \quad 20 \log \left(\frac{|\dot{u}_0|}{|\dot{u}_L|} \right) \ . \tag{2.34}$$

The processing function iAD corresponds to the IVAD function of Jørgensen et al. [103] and is measured in dB. As compared to experimental data from Christensen-Dalsgaard [28], the calculated function (positive values for angles exceeding zero, negative for those below zero) shows the same behavior. In both cases the response of the system is directional.

In addition to iADs, the vibrations of the tympani can differ in time as well leading to internal time differences (iTDs) defined as

$$\text{iTD} \quad := \quad \arg \left(\frac{\dot{u}_0}{\dot{u}_L} \right) / \omega \ . \tag{2.35}$$

Calculated iTDs are shown in Fig. 2.12, bottom left. For sound source directions $\pm\pi/2$ and resulting from the internally coupled ears, the maximal iTDs are around $\pm 86\ \mu s$ for frequencies below 2.5 kHz. These differences illustrate the enhancing effect of ICE, as a system with independent tympanic membranes and head size of *Hemidactylus frenatus* would, in comparison, produce maximal interaural time differences (ITDs) of $\pm 30\ \mu s$. The enhancement factor of maximal iTDs for internally coupled ears in comparison to maximal ITDs for independent ears in dependence upon sound frequency is illustrated in Fig. 2.12, bottom right. For frequencies below 3 kHz the internal coupling results in iTDs that are a factor 3 higher than ITDs for independent ears. Localization due to iTDs could therefore be possible at the signal level even for animals with a small interaural distance.

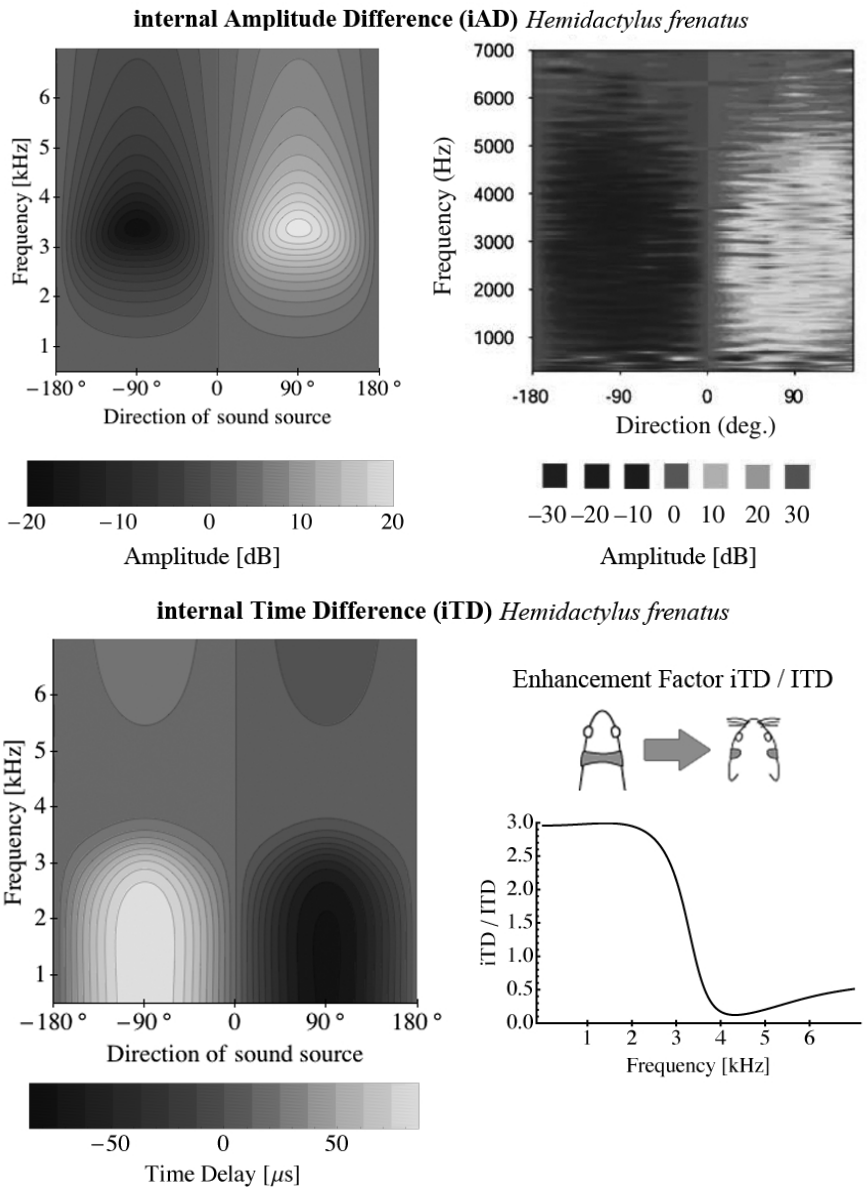

Figure 2.12: *Overview of possible localization cues for* Hemidactylus frenatus (gekkonid). *Top: Internal amplitude differences (iADs) in dB [cf. Eq.* (2.34)*] as a function of direction (x-axis, negative directions contralateral, 0 frontal, and positive directions ipsilateral) and frequency (y-axis) resulting from the ICE model with the cylindrical mouth cavity (left) and from experiments [28] (right). The obtained iADs systematically vary with direction of sound source for frequencies above* 1 *kHz. Calculated and experimental iADs [28] show a similar profile. The cylindrical ICE model, however, overestimates iADs for high frequencies and underestimates iADs for low frequencies. Bottom left: Internal time difference (iTDs) in μs [cf. Eq.* (2.35)*] as a function of direction (x-axis) and frequency (y-axis). Calculated iTDs vary with direction of sound source for low frequencies. Bottom right: Enhancement factor of maximal iTDs resulting from the ICE model with the cylindrical mouth cavity for animals having internally coupled ears in comparison to maximal* $\mathrm{ITD} = kL \sin(\theta = \pi/2)$ *of animals having independent ears in dependence upon different sound frequencies. For frequencies below* 3 *kHz the internal coupling leads to iTDs that are a factor 3 higher than ITDs for animals with independent ears. Taken together iADs and iTDs could deliver localization cues in* complementary *frequency ranges hinting at separated pathways for iTD and iAD processing; see Sec. 3.2 for details.*

As the vector strength (see Fig. 3.3 for details) decreases substantially for frequencies above 1 kHz [148], this frequency seems to be a natural limit for evaluation of the iTD cues.

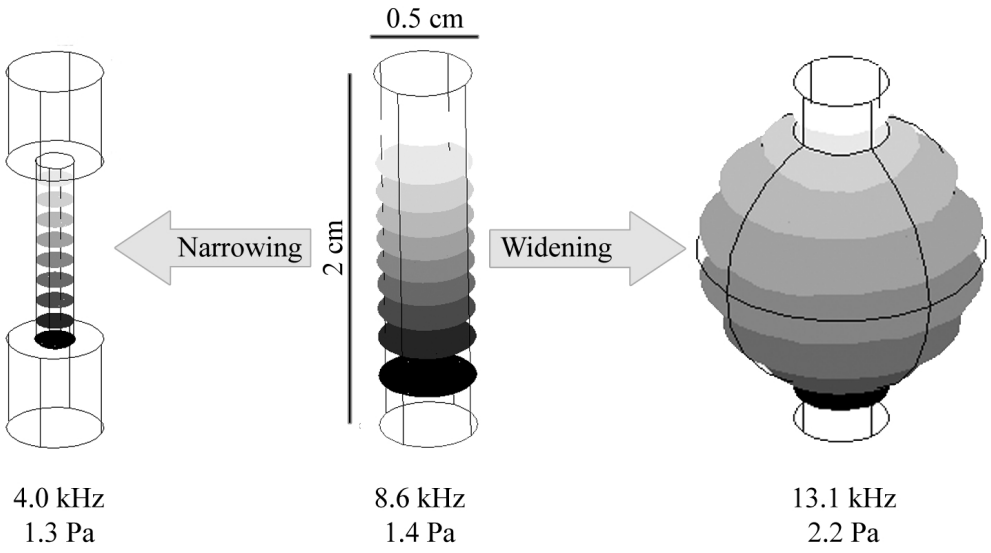

Figure 2.13: *Calculated first eigenmodes of simplified geometries with length 2 cm and top and bottom diameter of 0.5 cm. In comparison to the cylindrical geometry (middle) the two other geometries are either narrower (left) or broader (right). The plots show surfaces of equal pressure from the negative minimum (black) to the positive maximum (white). The absolute value of the extrema is equal and indicated below the geometries. The obtained lowest eigenfrequencies are 4.0 kHz (left), 8.6 kHz (middle), and 13.1 kHz (right) as indicated. A narrowing of the geometry therefore decreases, a widening increases the lowest eigenfrequency of the geometry.*

2.4.2 Eigenmodes of a realistic mouth cavity

The directional differences predicted by the cylindrical ICE model for *Hemidactylus frenatus* above 5 kHz contradict experimental data (see Fig. 2.10) and may result from the strongly simplified cylindrical model of the mouth cavity. Eigenfrequencies of the realistic mouth cavity could modify the response of the system. To test this hypothesis, two questions have to be addressed. First, what are the eigenfrequencies for complex geometries, in particular, for the realistic mouth cavity of *Hemidactylus frenatus*. Second, what is the influence of the lowest eigenfrequency on the response of the system, that is, on amplitudes of tympanic vibration, iTDs, and iADs.

To address the first question, numerical simulations of simplified geometries investigate the effect of a narrowing or widening of the geometry on its lowest eigenfrequency. The analyzed geometries are modifications of a cylinder of length 2 cm and radius 0.25 cm. Modification of this cylinder results in geometries that either widen or narrow in the

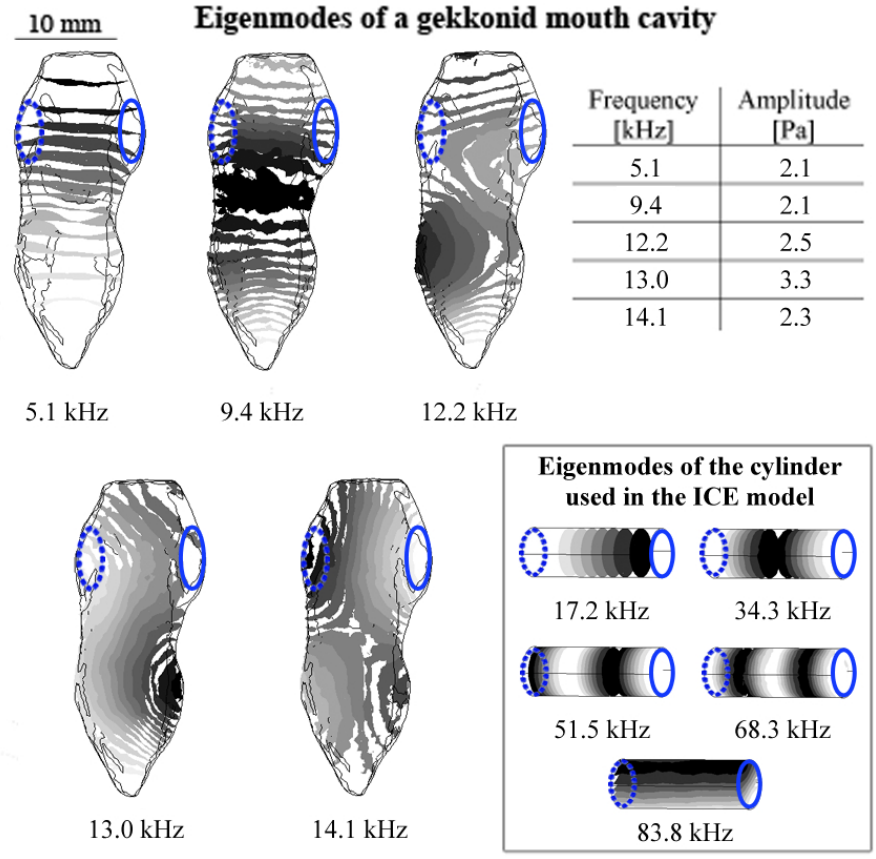

Figure 2.14: *Calculated eigenfunctions of the realistic mouth cavity (snout to bottom) for* Hemidactylus frenatus (gekkonid) *with the corresponding eigenfrequencies and their maximal amplitudes; see Fig. 2.7 (right) for the used three-dimensional mesh of the realistic mouth cavity. The estimated positions of the tympanic membranes are marked by blue circles. The plots show surfaces of equal pressure from the negative minimum (black) to the positive maximum (white). The absolute value of the extrema is equal and given in the table for each eigenfrequency. Shown in the bottom-right corner are the eigenfrequencies of the cylinder that corresponds to the space between the two eardrums used for calculations in the cylindrical ICE model; cf. Fig. 2.3. The eigenfrequencies of the cylinder lie high above the treated region and therefore have no effect on the system's response. In contrast, the eigenmodes of the realistic mouth cavity have eigenfrequencies starting at* 5.1 *kHz and may influence the directional response of the tympanic membranes.*

middle; cf. Fig. 2.13. Calculation of the eigenfrequencies reveals that a narrowing of the geometry decreases, a widening increases the lowest eigenfrequency of the geometry.

In the next step, numerical simulations based on the constructed three-dimensional mesh of the realistic mouth cavity of *Hemidactylus frenatus* (see Fig. 2.7, right) allow for the calculation and analysis of the corresponding eigenfrequencies. Figure 2.14 contrasts the calculated eigenfunctions of the realistic mouth cavity with the eigenfunctions of the corresponding cylinder used for the ICE model. In detail, the eigenfunction of the realistic mouth cavity with the lowest eigenfrequency, 5.1 kHz, shows a horizontal pattern. Intermediate eigenfunctions occur at around 12.2 kHz and 13.0 kHz with the highest pressure amplitudes, and weaker vertical eigenfunctions are around 14.1 kHz. In contrast, the lowest eigenfrequency for the corresponding cylinder used in calculating the ICE model lies at 17.2 kHz, therefore, outside the region of interest. Furthermore, strong intermediate eigenmodes are missing. To summarize, the eigenfrequencies of the realistic mouth cavity are lower than those of the corresponding cylinder. Therefore, they can modify the directional response, calculated using the cylindrical ICE model, in particular, in the frequency region above 5 kHz.

In the following, the second question is addressed, that is, how the lowest eigenfrequency influences the response of the system, in particular, amplitudes of the tympanic vibration, iADs, and iTDs. Here a re-calculation of tympanic vibrations with parameters as given in Table 2.3 and corresponding input ITDs but with varying length of the internal cylinder provides the possibility to estimate the influence of different lowest eigenfrequencies on the directional response of an ICE system. The cylinder varies between $L = 1$ cm, 2 cm, and 3 cm corresponding to lowest eigenfrequencies of 17.2 kHz, 8.6 kHz, and 5.7 kHz. Figure 2.15 summarizes the resulting amplitudes of the tympanic vibration for the three different interaural distances in dependence upon sound direction and frequency. With decreasing lowest eigenfrequency of the used cylinder the frequencies of maximal amplitudes of the tympanic vibration are shifted towards lower frequencies. As a consequence, vibration amplitudes for lower frequencies around 1 kHz increase. When sound frequency corresponds to the lowest eigenfrequency of the internal cavity the internal pressure reaches a local maximum. According to Eq. (2.17) the force acting on the tympani is therefore dominated by the pressure in the mouth cavity that is independent of the direction of the sound signal. The external inputs hardly influence the tympanic vibrations so that tympanic vibrations for the lowest eigenfrequency show almost identical amplitudes for all directions of sound source.

As illustrated in Fig. 2.16, iADs and iTDs are attenuated with decreasing lowest eigenfrequency of the mouth cavity. When the lowest eigenfrequency decreases from 17.2 kHz to 5.7 kHz, the maximal iTD decreases at the same time from 86 μs to around 50 μs, the maximal iAD from 15 dB to 8 dB. When sound frequency corresponds to the lowest eigenfrequency, tympanic vibrations do not vary anymore with direction of the sound source so that iADs and iTDs vanish. Localization based on iTDs or iADs therefore becomes impossible when reaching the lowest eigenfrequency. In particular, a rapid and complete decay of iADs has been measured experimentally for *Hemidactylus frenatus* at 5 kHz [28] which fits nicely to the calculated lowest eigenfrequency of the mouth cavity; cf. Fig. 2.14. A systematic calculation of the eigenfrequencies of several different mouth cavities together with the corresponding experimental data for iADs could verify whether

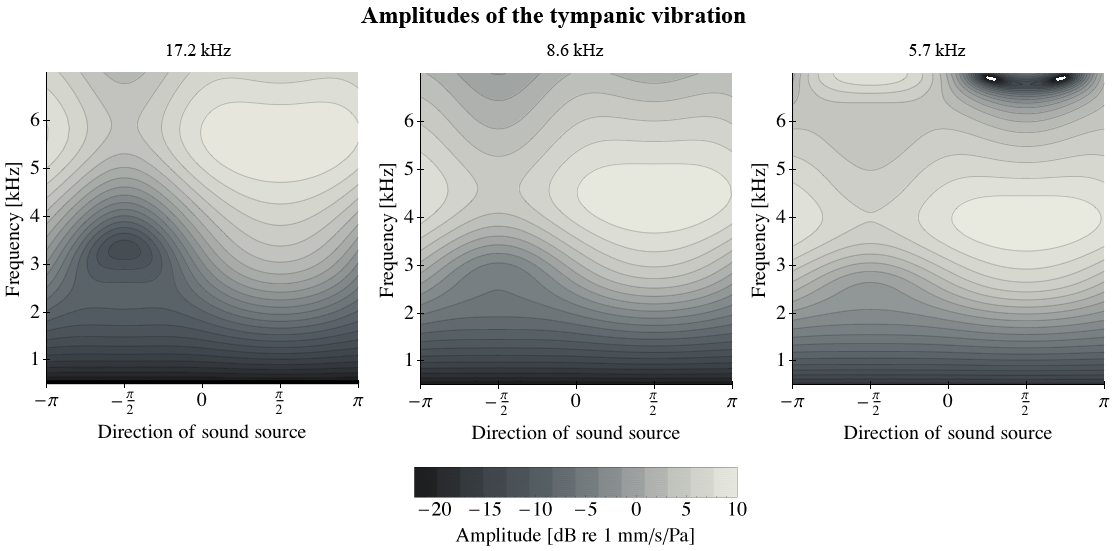

Figure 2.15: *Calculated amplitudes of the tympanic vibration in dependence upon the direction of sound source (x-axis) and frequency (y-axis) for varying lowest eigenfrequency of the mouth cavity. For the parameters of* Hemidactylus frenatus *as given in Table 2.3 and corresponding input ITDs only the length of the internal cylinder varies and consequently so does the eigenfrequency of the mouth cavity vary between 17.2 kHz (left column), 8.6 kHz (middle), and 5.7 kHz (right column). With decreasing lowest eigenfrequency (stepping to the right) the frequencies of the maximal amplitudes of the tympanic vibration are shifted towards lower frequencies. Vibration amplitudes for lower frequencies around* 1 *kHz increase. When sound frequency corresponds to the lowest eigenfrequency of the internal cavity the internal pressure reaches a local maximum. According to Eq.* (2.17) *the force acting on the tympani is therefore dominated by the pressure in the mouth cavity that is identical for every direction of the sound signal. The external inputs hardly influence the tympanic vibrations. For the lowest eigenfrequency tympanic vibrations therefore show almost identical amplitudes for all directions of sound source. In other words, directional hearing becomes impossible for the lowest, or whichever, eigenfrequency.*

the lowest eigenfrequency limits sound processing in the described way. For *Tokay gecko* this analysis is performed below.

In conclusion, for *Hemidactylus frenatus* the systematic increase within the estimation error between calculated and experimentally measured amplitudes of tympanic vibrations (cf. Fig. 2.10) could well result from a missing first eigenfrequency around 5 kHz. Without such an eigenfrequency, the ICE model with the cylindrical mouth cavity systematically overestimates iADs in the corresponding frequency region.

2.4.3 Generalization of the ICE model

The previous calculations are performed for *Hemidactylus frenatus* that is, with an interaural distance of 10 mm, one of the smallest lizards. To illustrate that the ICE model can generally describe tympanic vibrations of animals having internally coupled ears, in particular, all

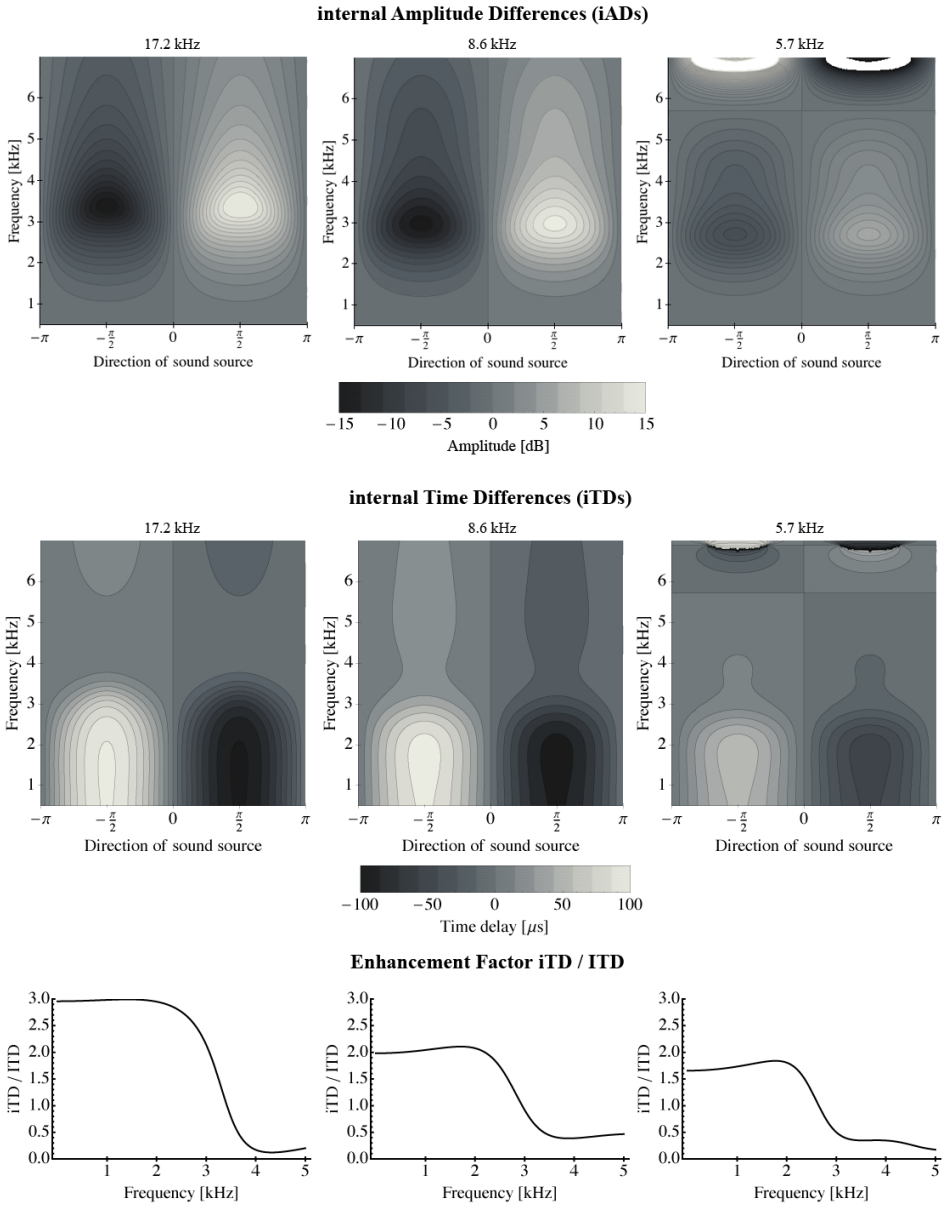

Figure 2.16: *Calculated iADs (top) [cf. Eq. (2.34)] and iTDs (middle) [cf. Eq. (2.35)] in dependence upon the direction of sound source (x-axis) and frequency (y-axis) and enhancement factor of maximal iTD for varying lowest eigenfrequency of the mouth cavity; for a description of the setup see caption of Fig. 2.15. The bottom row shows the enhancement factor of maximal iTDs resulting from the ICE model with the cylindrical mouth cavity for animals having internally coupled ears in comparison to maximal* $ITD = kL\sin(\theta = \pi/2)$ *of animals having independent ears in dependence upon different sound frequencies and lowest eigenfrequency of the mouth cavity. With decreasing lowest eigenfrequency (stepping to the right) iADs and iTDs are attenuated for frequencies below the lowest eigenfrequency. Frequencies of maximal iADs are further shifted towards lower frequencies. The enhancement factor of maximal iTDs in comparison to maximal ITDs decreases for decreasing lowest eigenfrequency, e.g., for a narrow mouth cavity; cf. Fig. 2.13. When sound frequency corresponds to the lowest eigenfrequency the amplitudes of the tympanic vibrations are almost identical for all directions of sound source; cf. Fig. 2.15. As a consequence, differences between tympanic vibrations, i.e., iADs and iTDs vanish completely for the lowest eigenfrequency. Localization based on iTDs or iADs therefore becomes impossible when reaching the lowest, or whichever, eigenfrequency.*

kinds of lizards, the model is evaluated as well for *Tokay gecko*, the second largest Gecko species. An agreement of experimental and analytical results for *Tokay gecko* then shows that the ICE model covers a large range of animals with different interaural distances having ICE. Chapter 4 further shows that the ICE model makes reasonable predictions for the best frequency, i.e., the frequency with the lowest hearing threshold, for different species of animals having ICE. Taken these two arguments together, the ICE model can be seen as universal model for describing internally coupled ears.

Table 2.4: *Geometry parameters for* Tokay gecko. *Length of the cylinder, i.e., interaural distance, is taken from [27], the other parameters are linearly scaled from the parameters of* Hemidactylus frenatus*; cf. Table 2.3. In particular, the interaural distance $L = 10$ mm of* Hemidactylus frenatus *is used for the scaling below.*

$L = 22$ mm	Length of the cylinder (interaural distance)
$a = 3.5$ mm $\cdot$ (22 mm/10 mm)	Radius of the tympanic membrane
$\alpha = 1000$ Hz$/(2 \cdot 1.2) \cdot$ (10 mm/22 mm)	Damping coefficient of the membrane
$\rho_m = 3.2$ mg/mm^3	Density of the membrane
$d = 10$ μm $\cdot$ (22 mm/10 mm)	Thickness of the membrane

The following comparison of calculated and experimental data analyzes vibration amplitudes of the tympanic membranes, iTDs, iADs, and the eigenfrequencies for the realistic mouth cavity of *Tokay gecko*. All calculations are performed with parameters as given in Table 2.4. The length of the cylinder, i.e., the interaural distance, is taken from [27], the other parameters are scaled from the parameters of *Hemidactylus frenatus* under the assumption that the geometry scales linearly with size. In detail, the radius a and the thickness d of the membrane increase linearly, whereas the eigenfrequencies of the tympani decrease linearly, with size; cf. [50, p. 74] for details.

Figure 2.17 contrasts calculated and experimentally measured amplitudes of tympanic vibration for *Tokay gecko* in dependence upon different directions and frequencies of sound. Both profiles show similar vibration amplitude patterns with identical maximum of 10 dB re 1 mm/s/Pa. The ICE model with the cylindrical mouth cavity, however, overestimates the directional response of the system above 2.5 kHz and underestimates the directional response for low frequencies. A lowest eigenfrequency of the realistic mouth cavity at round 3 kHz could provoke such a modification; cf. Fig. 2.15. At the same time, such a lowest eigenfrequency would modify as well internal amplitude differences (iADs) and internal time differences (iTDs) that are calculated in the following.

Figure 2.18 summarizes iADs and iTDs for *Tokay gecko* in dependence upon different directions and frequencies of sound. For frequencies above 1 kHz incoming ITDs are translated into directional iADs up to 15 dB. As already observed for *Hemidactylus frenatus*, the obtained iADs systematically vary with direction of sound source and reproduce experimental iADs [27]. For frequencies above 2.5 to 3 kHz, however, experimentally measured iADs vanish more quickly than the calculated iADs hinting at a modification due to a lowest eigenfrequency of the realistic mouth cavity around 3 kHz; cf. Fig. 2.16.

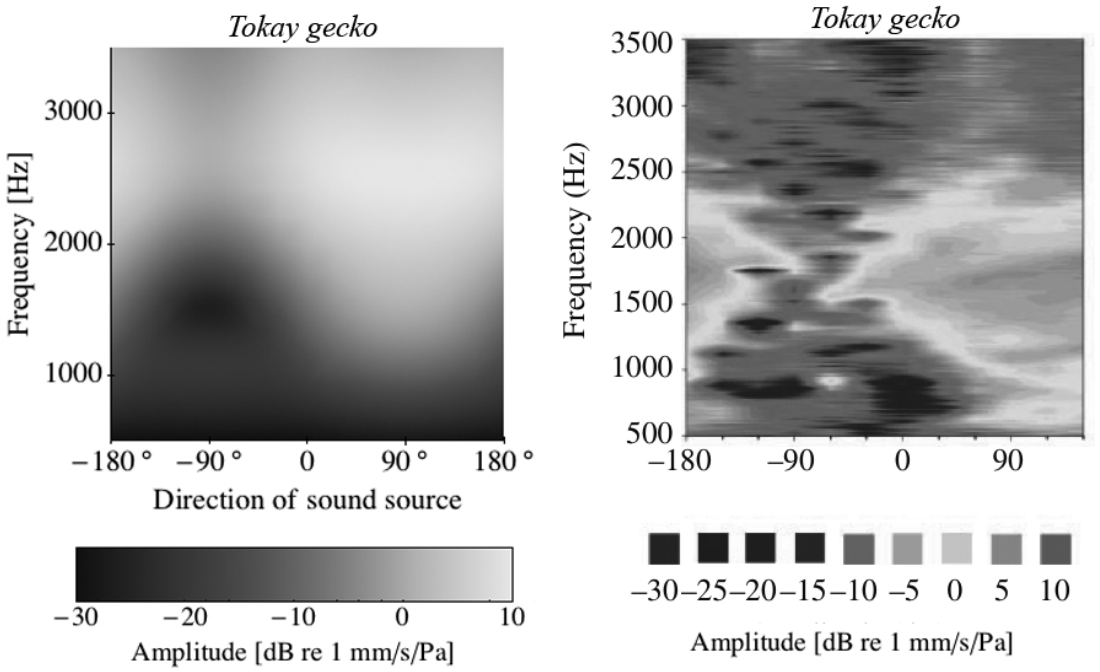

Figure 2.17: *Calculated (left) and experimental (right) amplitude of tympanic vibration for* Tokay gecko *in dB re 1mm/s/Pa, dependent on the sound source direction (x-axis with directions varying from* $-180°$ *to* $180°$*; negative directions are contralateral, 0 is frontal, and positive directions are ipsilateral) and frequency (y-axis). The calculated results are based on the ICE model with a* cylindrical *mouth cavity and parameters as given in Table 2.4. Measured eardrum vibration amplitudes for* Tokay gecko *are from Christensen-Dalsgaard [27]. The ICE model reproduces the measured vibration amplitude pattern but overestimates the directional response of the system above* 3 *kHz. In addition, calculated amplitudes of tympanic vibration are lower for frequencies below* 1 *kHz than experimentally measured vibration amplitudes. According to Fig. 2.15 a lowest eigenfrequency of the realistic mouth cavity around* 3 *kHz could provoke the observed modified response of the measured vibration amplitudes. A numerical calculation of the lowest eigenfrequency of the mouth cavity is therefore necessary.*

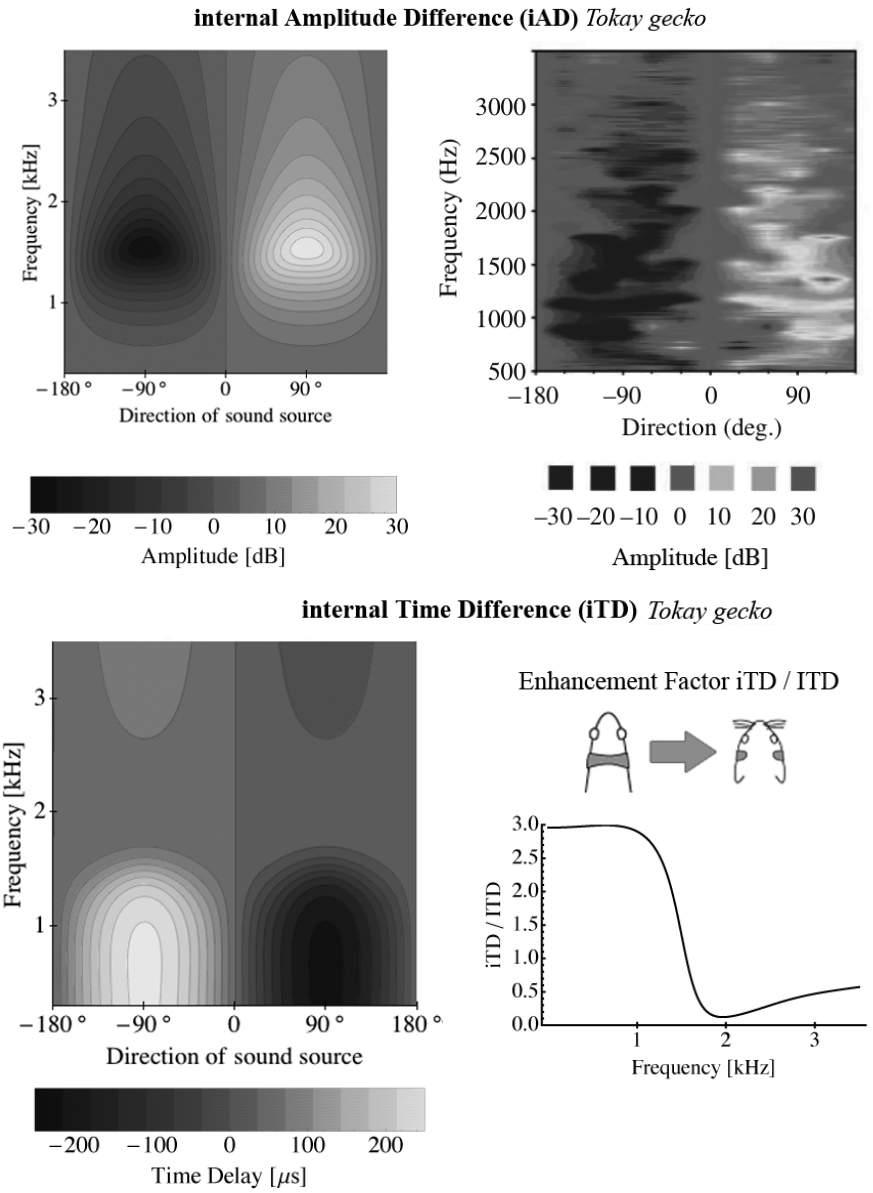

Figure 2.18: *Overview of possible localization cues for* Tokay gecko. *Top: Internal amplitude differences (iADs) in dB [cf. Eq.* (2.34)*] as a function of direction (x-axis, negative directions contralateral, 0 frontal, and positive directions ipsilateral) and frequency (y-axis) as resulting from the ICE model with the cylindrical mouth cavity (left) and from experiments [27] (right). The obtained iADs systematically vary with direction of sound source for frequencies above* 1 *kHz and reproduce experimental iADs [27]. Bottom left: Internal time difference (iTDs) in μs [cf. Eq.* (2.35)*] as a function of direction (x-axis) and frequency (y-axis). Calculated iTDs vary with direction of sound source for low frequencies. Bottom right: Enhancement factor of maximal iTDs resulting from the ICE model with the cylindrical mouth cavity for animals having internally coupled ears in comparison to maximal ITD* $= kL\sin(\theta = \pi/2)$ *of animals having independent ears in dependence upon different sound frequencies. For frequencies below* 1.5 *kHz the internal coupling produce time differences that are a factor of 3 higher than those for animals with independent ears. The additional enhancement could guarantee that an animal with a small head size could still localize sound signals by means of iTDs for frequencies below* 1 *kHz. Taken together iADs and iTDs could deliver localization cues in complementary frequency ranges.*

For low frequencies calculated iTDs are up to ± 225 μs for the large *Tokay gecko* and vary with direction of sound source. For frequencies below 1 kHz the internal coupling enhances iTDs in comparison to ITDs by a factor 3; cf. enhancement factor in Fig. 2.18, bottom right. Similar to the directional localization cues for *Hemidactylus frenatus* (cf. Fig. 2.18) iADs and iTDs could together deliver localization cues in complementary frequency ranges.

In a final step, the eigenfrequencies of the realistic mouth cavity of *Tokay gecko* (cf. Fig. 2.6, bottom) are calculated to account for a modification of tympanic vibrations due to the lowest eigenfrequency of the realistic mouth cavity (Figs. 2.15 and 2.16). Due to the complex shape of the cavity the eigenfrequencies can only be calculated numerically. For the eigenfrequency analysis the reconstructed three-dimensional mesh of the realistic mouth cavity (Fig. 2.8, right) is loaded into the simulation program COMSOL. Figure 2.19 illustrates the obtained eigenmodes and the corresponding eigenfrequencies. As already observed for *Hemidactylus frenatus* (cf. Fig. 2.14) the eigenfunction of the realistic mouth cavity with the lowest eigenfrequency, here 3.2 kHz, shows a horizontal pattern. Eigenfunctions of higher eigenfrequencies 5.9 kHz, 6.0 kHz, and 7.4 kHz with the highest pressure amplitudes show intermediate to vertical patterns. In comparison, the lowest eigenfrequency for the corresponding cylinder used in calculating the ICE model with the cylindrical mouth cavity lies at 7.8 kHz. The lowest eigenfrequency of the realistic mouth cavity could therefore modify the directional response as calculated using the cylindrical ICE model, in particular, in the frequency region around 3 kHz and below 1 kHz; cf. Figs. 2.15 and 2.16.

Taken results from *Hemidactylus frenatus* and *Tokay gecko* together, the ICE model can explain vibration patterns as observed in experiments. In this way, the ICE model presents itself as universal model to generally describe tympanic vibrations of animals having internally coupled ears.

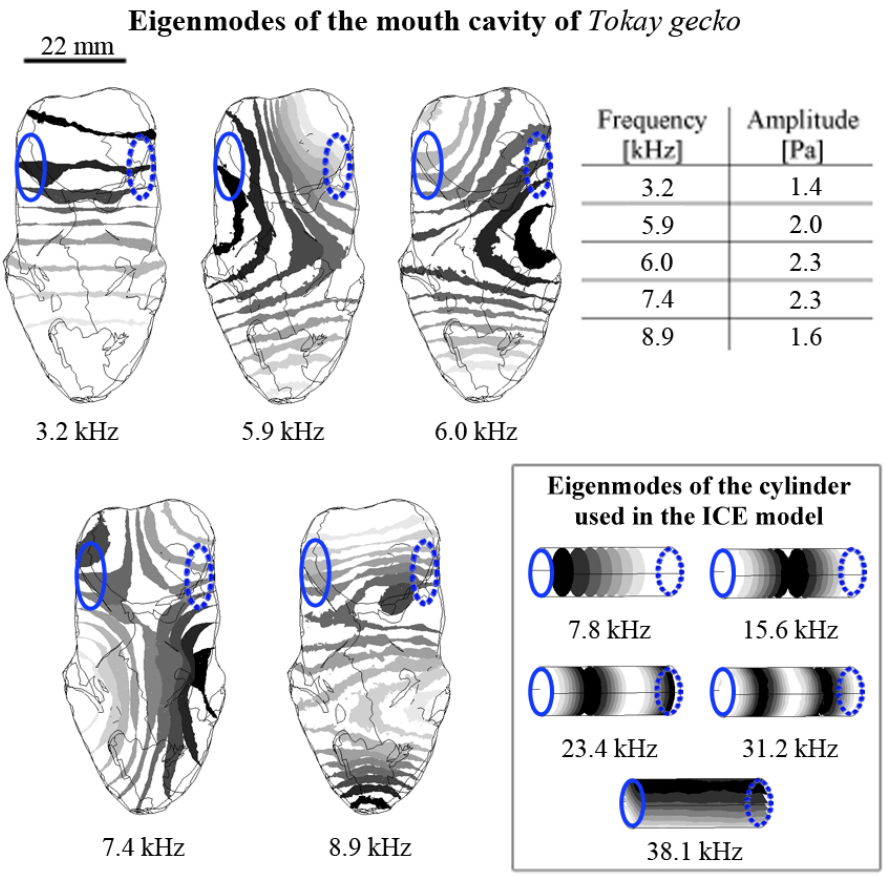

Figure 2.19: *Calculated eigenfunctions of the realistic mouth cavity for* Tokay gecko *with the corresponding eigenfrequencies and their maximal amplitudes; see Fig. 2.8 (right) for the used three-dimensional mesh of the realistic mouth cavity. The estimated positions of the tympanic membranes are marked by blue circles. The plots show surfaces of equal pressure from the negative minimum (black) to the positive maximum (white). The absolute value of the extrema is equal and given in the table for each eigenfrequency. The eigenmodes of the realistic mouth cavity have eigenfrequencies starting at* 3.2 *kHz. Similarly to* Hemidactylus frenatus, *the realistic eigenfrequencies lie below the lowest eigenfrequency of the cylinder used by the ICE model. According to Figs. 2.15 and 2.16, the lowest eigenfrequency of the realistic mouth cavity can modify the directional response of the system around* 3.2 *kHz.*

2.4.4 Spatial vibration pattern of the membrane

The final step of the evaluation consists of the analysis of the spatial pattern in the membrane with respect to the exciting frequency. The tympanic vibration pattern was measured experimentally by Manley [144] for a *Tokay gecko* with the strongest vibration response around 1 kHz. Manley determined the vibration amplitude for eight locations on the membrane and measured a complex pattern (see Fig. 2.20, right) that does not correspond to a symmetrical eigenmode with only one maximum. Between 4 kHz and 8 kHz, the pattern shows two maxima, hinting that higher modes are likely involved.

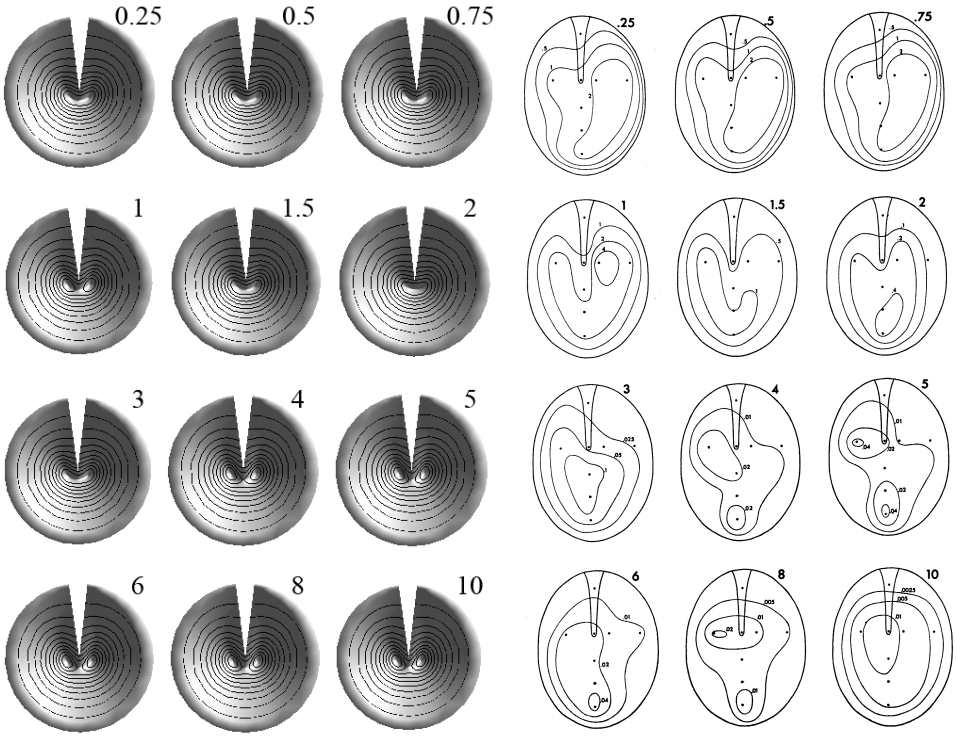

Figure 2.20: *Calculated (left) and experimental (right) vibration amplitude pattern over the membrane, dependent on sound frequency from* 0.5 *to* 10 *kHz for* Tokay gecko. *Experimental data have been taken from Manley [144]. The calculated data result from the ICE model with a* cylindrical *mouth cavity. In both cases, a similar complex vibrational pattern varies with frequency. The inclusion of higher eigenmodes in the ICE model results in complex amplitude patterns over the membrane that are in good agreement with experimental data.*

In the ICE model with a cylindrical mouth cavity, higher modes are necessary to satisfy the boundary conditions of the membrane resulting from the attached extracolumella of the middle ear. Figure 2.20 (left) illustrates the vibration amplitude pattern over the membrane, calculated using Eq. (2.26) for different frequencies with the parameters in Table 2.4 for *Tokay gecko*. The omitted segment represents the attached extracolumella. The comparison of the calculated data with the experimental data is qualitative, as the parameters used are only estimated values. The aim of the calculation is to show that

reproduction of a complex vibrational pattern over the membrane is possible. As measured experimentally (cf. Fig. 2.20, right), the asymmetrical pattern varies with frequency. Due to the inclusion of higher modes into the ICE model the spatial vibration pattern of the tympanum has a complex shape that clearly differs from a symmetrical pattern. Depending on the frequency the calculated profile can either show a pattern with only one maximum or with two neighboring maxima. That is, the model results correspond qualitatively with experimental results such that the vibration pattern is no longer rotation symmetric and contains more than one maximum. Consequently, the appearance of higher modes for the tympanic membrane could influence its vibration pattern.

2.5 Conclusion

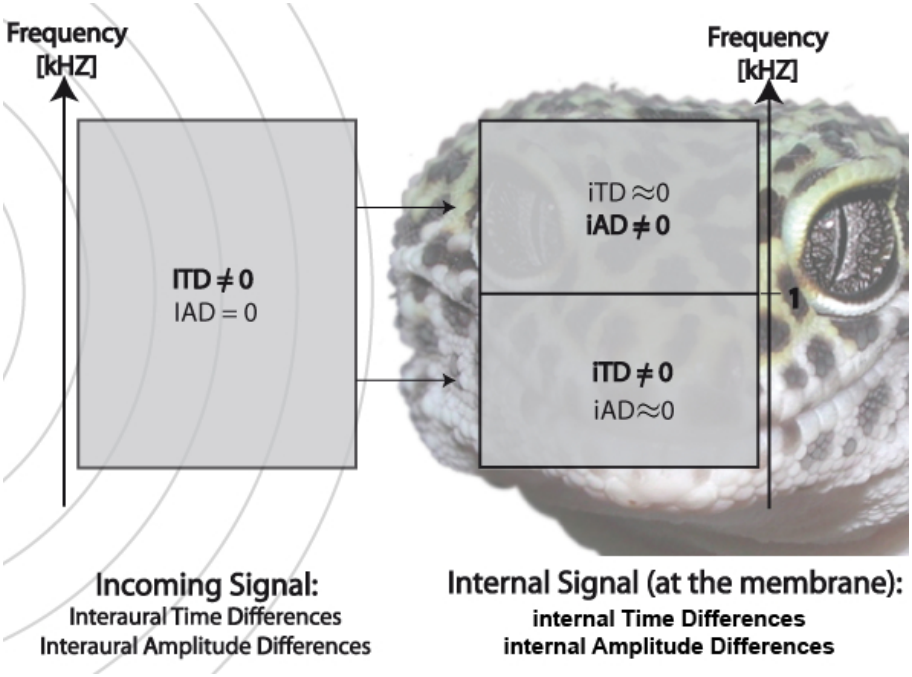

Figure 2.21: *Overview of the two processing pathways as resulting from the ICE model; cf. Fig. 2.12 and [221]. This figure assumes a small interaural distance so that amplitude differences between the sound waves at the two ears can be neglected. Incoming interaural time differences (ITDs) are processed by the vibrating membranes coupled through the mouth cavity and, hence, giving rise to internally coupled ears (ICE). For low frequencies ITDs are translated into internal time differences (iTDs), for high frequencies into internal amplitude differences (iADs).*

In summary, this chapter has presented a general model of internally coupled ears, the ICE model, that consists of a cylindrical, analytically soluble model to calculate the vibration pattern of two loaded membranes coupled by the internal cavity and a numerical eigenfrequency analysis of the realistic mouth cavities. To include the effect of the attached

extracolumella of the middle ear, as found in lizards, higher modes of the membrane vibration are part of the solution. The calculated results together with the numerical calculations agree well with experimental data. Evaluating membrane vibration at the tip of the extracolumella for different angles of the sound source and frequencies, the model can largely reproduce the directional responses of ICE systems as found in *Hemidactylus frenatus* and *Tokay gecko*.

For *Hemidactylus frenatus* the ICE system creates iTDs up to ± 86 μs for low frequencies which reflects an enhancement of the incoming ITDs due to the internal coupling by a factor 3. Incoming ITDs are translated into directional iADs up to 20 dB for high frequencies. For the larger *Tokay gecko* calculated iTDs are up to ± 225 μs for low frequencies, and calculated iADs are up to 15 dB for high frequencies. Taken together, iADs and iTDs could deliver localization cues in complementary frequency ranges for both animals. Figure 2.21 summarizes the frequency-dependent response of an ICE system to incoming ITDs.

To further analyze the effect of the internal mouth cavity, three-dimensional meshes of the realistic mouth cavities of *Hemidactylus frenatus* and *Tokay gecko* are reconstructed from scanned data. The eigenfrequencies for the internal cavities are then calculated numerically. For *Hemidactylus frenatus* the eigenmodes start at 5 kHz (see Fig. 2.14), for the larger *Tokay gecko* at 3 kHz; see Fig. 2.19. In both cases the lowest eigenfrequency may well provoke a modified response of the realistic ICE system in comparison to the ICE model based on a simplified cylindrical mouth cavity. In concrete terms, the lowest eigenfrequency could limit the range of directional hearing in such a way that iADs and iTDs vanish for sound signals around this frequency. For *Hemidactylus frenatus* and *Tokay gecko* experimentally measured iADs [27, 28] confirm this hypothesis by vanishing around the calculated lowest eigenfrequency of the corresponding realistic mouth cavity. A systematic calculation of eigenfrequencies for realistic mouth cavities in combination with measured iADs could finally confirm the hypothesis and would be an interesting topic for further study.

The results of Sec. 2.4.1 could be obtained by means of a simple model including three impedances for the two membranes and the cavity [27, 28]. However, the new ICE model explains much more of the experimental data and can be generalized for all kinds of species. A new, interesting feature of the model is the activation of higher modes. Compared to the fundamental eigenmode that is rotationally symmetric with one maximum at the middle of the circular membrane, higher vibration modes exhibit a pattern with a number of maxima. Including higher modes in the solution therefore breaks the rotational symmetry and reproduces the complex vibration patterns observed experimentally.

The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' (I found it!) but 'That's funny ...'

Isaac Asimov (1920 - 1992)

3. Neuronal processing of iTDs and iADs

3.1 Introduction

The neuronal processing ensuing from iTDs and iADs is exemplarily analyzed by means of data from *Hemidactylus frenatus*. For the larger *Tokay gecko* maximal iADs are in the same range and maximal iTDs are even larger than those of *Hemidactylus frenatus*. As a consequence, the neuronal processing has to cope with the smaller species as the more difficult case. For *Hemidactylus frenatus*, analytical results of the ICE model based on a simplified cylindrical mouth cavity (cf. Fig. 2.12) and experimental data [27, 28] show that the tympani of lizards respond to sound signals in a manner dependent on the horizontal direction of the sound source. More precisely, internal time differences (iTDs) and internal amplitude differences (iADs), as they appear at the level of the tympanic membranes, vary systematically with the angle of the sound source. In the next step of sound processing, the iTD and iAD sound localization cues have to be extracted from the vibrations of the tympani.

Vibrations of the tympanic membranes are first transmitted by the middle ear (through the columella) to the cochlea and to the embedded basilar membrane; cf. Fig. 1.3. Due to the systematically varying stiffness, the basilar membrane is tonotopically organized in such a way that each of its hair cells is maximally excited by stimulation at one specific characteristic frequency (CF). Hair cells along a small region of the basilar membrane in turn enervate local auditory nerve fibers. As a consequence, the frequency-specific tonotopic representation is also preserved at the level of neuronal processing. Accordingly, the simulations below have been performed for one specific frequency.

The mechanical transduction process from tympani to the hair cells of the basilar

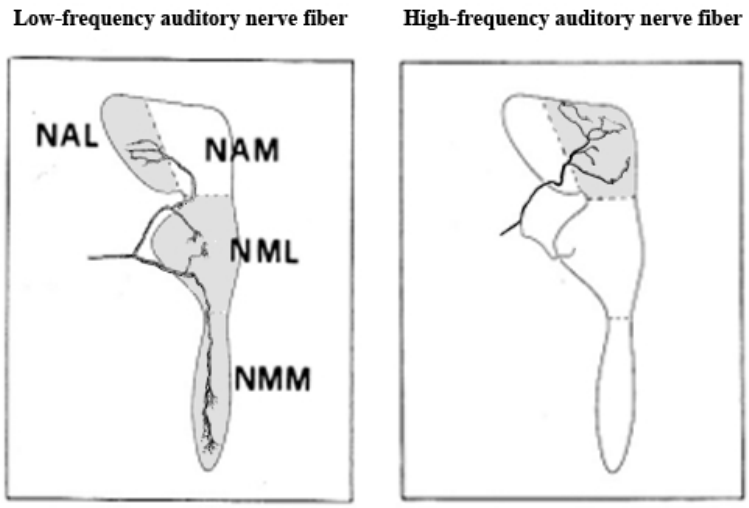

Figure 3.1: *Different neuronal projections from low- and high-frequency auditory nerve fibers to the first nuclei the nerve fibers encounter in the auditory brain stem. Figure modified from Szpir [208] for* Alligator lizard. *Left: Low-frequency fibers project to the lateral nucleus angularis (NAL) and the lateral and medial nucleus magnocellularis (NML and NMM). Right: In contrast, high-frequency fibers project almost exclusively to the medial nucleus angularis (NAM). The different projection pathways of low- and high-frequency nerve fibers suggest an* independent *processing of low- and high-frequency signals.*

membrane to the auditory nerve is simplified in such a way that the response of the auditory nerve is proportional to the corresponding tympanic vibration. This simplified model might not be the case in the real world, the assumption, however, ensures that the iTD and iAD properties are conserved. That is, if mechanical processing of the membrane vibration reduces amplitude or shifts the phase of the signal, the modifications apply at both sides in the same manner. Modifications made along the path cancel each other out, and iADs and iTDs remain unchanged by the specific transduction from tympanic vibration to the response of the auditory nerve. Low-frequency auditory nerves are therefore assumed to reproduce the temporal profile of the membrane vibrations. Spike rates of high-frequency auditory nerves are assumed to be linear functions of the tympanic vibration level in dB. In the next steps of neuronal processing, low and high frequencies, i.e., iTDs and iADs, are assumed to be treated separately.

3.2 Separated pathways for iTD and iAD processing

Results of the analytical calculations for *Hemidactylus frenatus* (gekkonid) as illustrated in Figs. 2.21 and 2.12 show fundamental differences between iADs and iTDs for frequencies below and above 1 kHz. For the low frequency range, there are iTDs available ranging up to 86 μs (Fig. 2.12, bottom left), whereas intensity differences are negligible. In contrast, for high frequencies, Fig. 2.12 (top) illustrates that there are iADs up to 20 dB between the vibrations of the two tympanic membranes. Consequently, separated processing pathways

for iTDs in the low frequency and for iADs in the high frequency regions are probable. For mammals, such a "duplex theory of sound localization" has been suggested more than 100 years ago by Thompson [212] and Rayleigh [178]. They stated that phase and amplitude cues are used for sound localization in complementary frequency ranges.

For lizards, the hypothesis of two separated processing pathways for iTDs and iADs is supported by a number of experimental studies. Manley [148], for instance, showed that auditory nerve fibers cannot reliably represent, i.e., they do not phase-lock to, periodic signals above 1 kHz; see Fig. 3.3 and next section for a detailed explanation. Thus, neuronal processing of iTDs becomes impossible.

Furthermore, the basilar membrane of lizards, which follows the tympanic membrane in the mechanical sound processing chain, differs from the mammalian basilar membrane. It is divided into two segments: the apical segment and the basal segment [127, 145]. Whereas hair cells on the apical segment react best to low frequencies, hair cells on the basal segment are best tuned to frequencies above 1 kHz. This separation of low and high frequencies is conserved within the neuronal projections from the hair cells to different brain areas; see Fig. 3.1 and [208]. Based on these data, low and high frequencies, i.e., iTDs and iADs, are likely processed independently in the brain.

In contrast to the above measurements, the subsequent neuronal processing pathways of iTDs and iADs, i.e., the involved brain areas and their connections, are not experimentally determined for lizards. Therefore, the parallel existence of two widespread processing strategies has been analyzed suggesting a Jeffress-like model for low-frequency signals and excitatory-inhibitory (EI) processing for high-frequency signals.

3.3 Processing of iTDs

The most popular and strikingly simple model for neuronal localization through iTDs is the Jeffress model [99] as suggested by Lloyd Jeffress in 1948; see, e.g., [73, 104] for further information. Figure 3.2 illustrates the Jeffress-like model that is equivalent to the Jeffress model (see [57]) and used for iTD processing in lizards. The Jeffress-like model consists of the following three elements.

- **Phase-locked and phase-shifted inputs from the two sides**
  Vibrations of the tympani are translated into *phase-locked* neuronal spike patterns of the auditory nerve. That means that spikes always occur at a specific phase of the incoming wave, e.g., at the maximum; see Fig. 3.3 (left) for illustration. This translation of the sound waves into spikes is not precise but shows a certain variation around a perfect synchronization. For the first time, Goldberg and Brown [69] have quantified synchronization by computing a cycle histogram of spikes plotted relative to their phase within the cycle. The length of the average vector, called *vector strength*, corresponds to the degree of synchronization. As a convex combination, the vector strength lies in the interval $[0, 1]$. A vector strength of 1 represents perfect synchronization since all spikes have to point at exactly the same direction in the cycle histogram. For a vector strength of 0 all spikes are randomly distributed within the cycle histogram, there is no synchronization at all. Figure 3.3 (right) shows vector strengths for low- and high-frequency fibers of the lizard *Tiliqua rugosa* in dependence upon sound frequency.

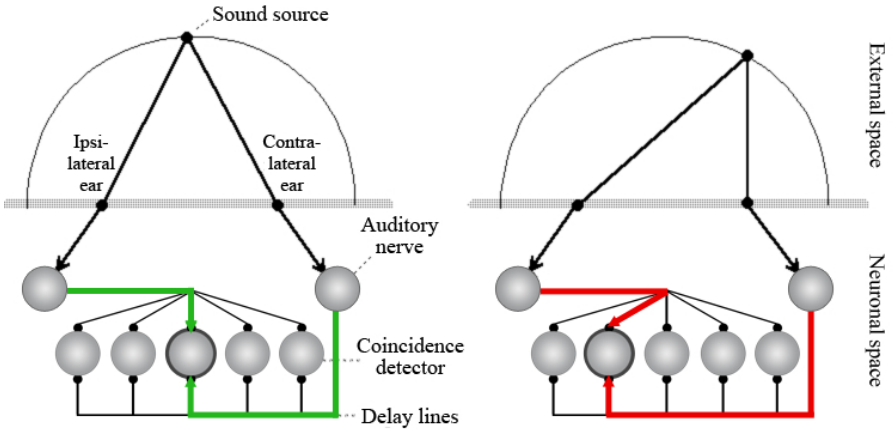

Figure 3.2: *Visualization of sound localization by means of interaural time differences through a Jeffress-like model illustrated for two different positions of a sound source; figure modified from [126]. Input signals from the ipsilateral ear arrive at all map neurons (coincidence detectors) simultaneously whereas input signals from the contralateral ear arrive at each map neurons with a certain delay that systematically varies over the array of map neurons. Left: Sound signals from a source directly in front of the head arrive at the two tympani at the same time. The neuron that fires the most is the one with delay lines that show an equal delay for the ipsi- and contralateral side. Right: Sound waves from a sound source at the contralateral side of the head arrive earlier at the contralateral than at the ipsilateral tympanum. The neuron that fires the most is the one with delay lines that compensate for this time shift, that is, the delay of the ipsilateral axon is longer than the delay of the contralateral axon.*

For a specific frequency ω smaller than 1 kHz, the firing rate f_{ipsi} of the ipsilateral auditory nerve has to reflect both timing information of the tympanum and coding precision of the auditory nerve fiber. According to Friedel et al. [57] the firing rate can therefore be described by a sum of Gaussian profiles with maxima at $\omega t = 2\pi n$ for $n \in \mathbb{N}_0$ leading to

$$f_{\text{ipsi}}(t) = A \exp\left\{ \left[\underbrace{\min\left(|\omega t - 2\pi n| , n \in \mathbb{N}_0 \right)}_{\in [0,\pi]} \right]^2 / \left(2\sigma^2\right) \right\} \tag{3.1}$$

with standard deviation σ and A as maximal amplitude. The value of σ can be determined from measured vector strengths (Fig. 3.3 B) as the absolute value of the ratio between first and zero-th Fourier coefficient; see [57] for details. For $\sigma \to 0$, the firing rate converges to a sum of delta functions with frequency ω. Small values of σ therefore correspond to a very precise neuronal representation of the incoming wave, and large values result in a blurred representation.
Given that the inputs are phase-locked they also have to be phase-shifted. This is the case since the firing rate f_{contra} of the contralateral auditory nerve is delayed by the internal time difference (iTD) with respect to the ipsilateral firing rate f_{ipsi}. The

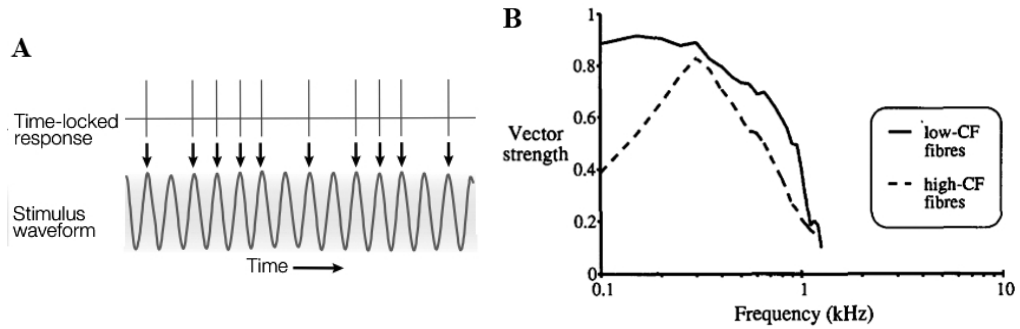

Figure 3.3: *(A) Illustration of phase-locking. The neuronal firing pattern of a phase-locking neuron (top) reproduces the form of the incoming sound wave (bottom). That is, a neuron always fires at one specific phase of the sound wave (e.g., at its maxima). Figure taken from Grothe [73]. (B) The quality of phase-locking is measured by the vector strength. A vector strength of* 1 *corresponds to perfect synchronization between incoming wave and neuronal spike pattern; a vector strength of* 0 *denotes completely uncorrelated signals. The plot shows averaged vector strengths of auditory-nerve fibers with low and high characteristic frequencies (CFs) for the lizard* Tiliqua rugosa *[148]. Whereas low-frequency fibers show a vector strength near 1 for frequencies below 1 kHz, high-frequency nerves are not tuned to phase-lock to incoming signals with a high reliability. The animal can therefore exploit iTDs for sound localization only* below *about 1 kHz. Figure taken from [148].*

iTDs result from the ITDs of the incoming sound signal and their processing by the internally coupled ears and depend on the angle θ of the sound source. The firing rate f_{contra} of the contralateral auditory nerve is therefore given by Eq. (3.1) with a modified arriving time

$$f_{\text{contra}}(t) = A \exp\left\{ \left[\underbrace{\min\left(\left| \omega t + \omega \text{ iTD} - 2\pi n \right|, n \in \mathbb{Z} \right)}_{\in [0,\pi]} \right]^2 \Big/ \left(2\sigma^2\right) \right\} . \tag{3.2}$$

Taken together, firing profiles from ipsi- and contralateral auditory nerves provide phase-locked and phase-shifted inputs.

- **An array of coincidence detection map neurons**
  In the next step, the phase-locked and phase-shifted signals from the ipsi- and contralateral auditory pathways arrive at an array of coincidence detection neurons referred to as map neurons. Each neuron fires maximally when spikes from the two sides temporally coincide, i.e., arrive simultaneously.

- **Systematically varying delay lines**
  The map neurons are connected to left and right input neurons by axons that transmit input signals with a certain delay. For a Jeffress-like model this delay is identical for all axons from the ipsilateral side whereas the delays of axons from the contralateral side vary systematically over the array of map neurons. That is, a specific spike from the contralateral side arrives at each map neuron at a slightly different time compared

to the neighboring map neurons. The coincidence detection map neurons then react most to sound input with the specific iTD that compensates exactly for the difference between the delay times of the axons from the two sides. Given that iTDs correspond to different angles of the sound source, the array of iTD-detecting neurons corresponds to a spatial map. Figure 3.2 illustrates the Jeffress-like model for two examples of sound sources at different horizontal directions. A sound source directly in front of the listener maximally excites the neuron in the middle of the array, that is, the neuron connected by axons with an identical delay to the ipsi- and contralateral side. In contrast, a sound source at the right side of the listener maximally excites a neuron at the left edge of the array.

The following numerical simulations of the Jeffress-like model refrain from specifying the exact brain areas and auditory pathways (possible projections through intermediate brain areas). Instead, the model includes right and left input neurons as well as map neurons in an abstract way. With regard to Fig. 3.1 (left), the input neurons might be neurons of the nucleus magnocellularis, the map neurons of the superior olive or the nucleus laminaris. In any case, input neurons are modeled as stochastically firing *Poisson* neurons (see Sec. 1.3.3 for details) with firing rate functions defined by Eqs. (3.1) and (3.2). Their number is estimated by about 30, which corresponds to the measured number of low-frequency auditory nerve fibers [146].

The map neurons are modeled as *leaky integrate-and-fire* neurons; see Sec. 1.3.2 and [67] for details. Each map neuron is directly connected through axons to the input neurons. The total current I_n^{tot} arriving at the n-th map neuron is given by the sum of ipsi- and contralateral currents. According to the assumption of the two processing pathways for low frequencies these currents only differ in phase, whereas the temporal course and the strength of the currents are identical. Based on these considerations, the ipsilateral current is defined to be $I(t)$. The contralateral input then corresponds to the ipsilateral current, except it is shifted in time by two effects. On the one hand, the vibrating tympani generate iTDs that depend on the angle of the sound source θ. On the other hand, the neuronal response is delayed through the map neuron-specific axonal delay Δt_n. Taken together, the contralateral current is the given by $I[t + \Delta t_n + \mathrm{iTD}]$ so that the total current arriving at the n-th map neuron equals

$$I_n^{tot}(t) = I(t) + I(t + \Delta t_n + \mathrm{iTD}) \ . \tag{3.3}$$

As a coincidence detector, the n-th map neuron fires maximally for input from a preferred iTD_n that exactly compensates for its neuronal delay, i.e., $\Delta t_n = -\mathrm{iTD}_n$. Another $\mathrm{iTD}_m \neq \mathrm{iTD}_n$ does not perfectly compensate for the neuronal delay. Therefore, the more the input iTD_m differs from a neuron's specific axonal delay, the more the firing rate of that neuron decreases. Thus within an array of coincidence detection neurons with varying delay lines neurons with preferred iTDs corresponding to the incoming iTD fire most, and other neurons fire in an attenuated manner depending on how much their preferred iTD differs from the incoming iTD.

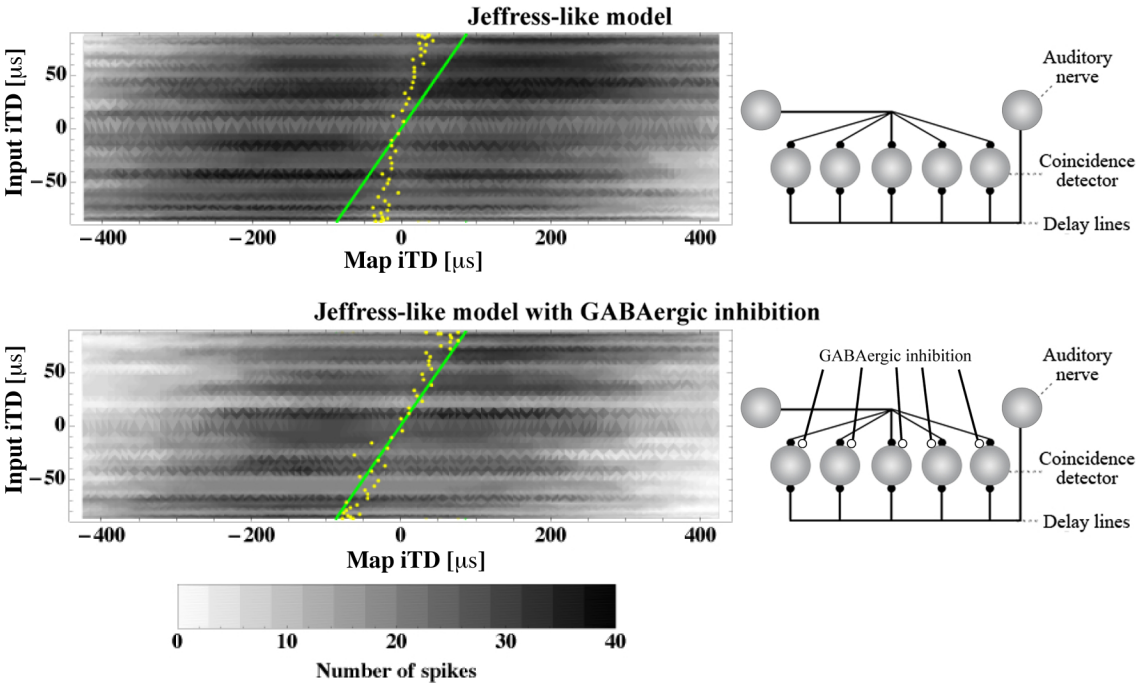

Figure 3.4: *Firing pattern of iTD map neurons (left) in dependence upon incoming iTDs (green line) for a Jeffress-like models without and with additional GABAergic inhibition (right). The intensity of the gray color indicates the number of the spikes from the map neurons as denoted. Yellow dots denote the estimated* $\widehat{\text{iTD}}$ *as resulting from a rate-weighted mean of the firing profile of the iTD map; cf. Eq.* (3.4)*. Top: Simulation results of a Jeffress-like model (right) with parameters given in Table 3.1, in particular, with a membrane constant* $\tau_m = 500$ *µs and a synaptic strength* $J = 0.015$*. The estimation error for the firing profile is calculated to be* $E = 39$ *µs over a range of* $[-86\ \mu s, 86\ \mu s]$ *for the angles* $[-\pi/2, \pi/2]$*. The horizontal angle of misestimation is then approximately* $39\ \mu s/86\ \mu s \cdot(\pi/2) \approx 41°$*. Bottom: Simulation results of a Jeffress-like model with additional GABAergic inhibition (right) as found in birds [58, 96, 238]. Additional GABAergic inhibition leads to a reduced membrane constant* $\tau_m = 250$ *µs [37]. The synaptic strength is set to* $J = 0.021$ *to ensure sufficient firing activity of the map neurons. With additional GABAergic inhibition the firing profile of the iTD map sharpens and the misestimation angle reduces to* $\approx 19°$*.*

Given the firing activity of all map neurons, the system can calculate the estimated $\widehat{\text{iTD}}$ as a rate-weighted mean of the map

$$\widehat{\text{iTD}} = \left(\sum_i \nu_i \ \text{iTD}_i\right) \Big/ \sum_i \nu_i \tag{3.4}$$

where every map neuron "votes" with its firing rate ν_i for its encoded iTD_i. Such an evaluation corresponds to an evaluation of the population vector; for details see, e.g., [65, 206, 218]. The advantage of a such "democratic" evaluation is that information, i.e., firing rates, of all map neurons is considered, in contrast to a determination of the estimated $\widehat{\text{iTD}}$ by the maximum of the iTD map. For broad firing profiles without a clearly

defined maximum, only a calculation of the rate-weighted mean is reasonable at all. To quantify the quality of the estimates the root mean square error E can be calculated. That means for M different inputs iTD_i with $1 \leq i \leq M$, and corresponding neuronal estimates $\widehat{\mathrm{iTD}}_i$ the error is defined as

$$E := \sqrt{\left[\sum_{i=1}^{M} \left(\mathrm{iTD}_i - \widehat{\mathrm{iTD}}_i\right)^2\right] / M} \ . \tag{3.5}$$

The value of the error E then gives the mean mismatch between the input iTD and the estimated $\widehat{\mathrm{iTD}}$.

Numerical results of simulations performed with parameters as given in Table 3.1 are illustrated in Fig. 3.4, left. The upper plot shows the firing profiles of the iTD map and the estimated $\widehat{\mathrm{iTD}}$s [cf. Eq. (3.4)] in dependence upon different input iTDs for a Jeffress-like construction; Fig. 3.4, top right. Given that iTDs for *Hemidactylus frenatus* are very small in comparison to neuronal time scales, e.g., the length of an excitatory postsynaptic current (EPSC), the membrane constant, or the refractory period of typical neurons, the firing profile of the iTD map is very broad. To account for the slow decay of the firing profile, the neuronal map has to represent a large range of iTDs that by far overshoot the physically accessible range. The phenomenon has been measured, e.g., in alligators [24], where the range of the neuronal iTD map of 1000 μs extremely exceeds the physically accessible range. In the simulations the overshoot is fixed at 5, which means that map neurons cover an iTD range from $[-5\ \mathrm{iTD}_{\max}, 5\ \mathrm{iTD}_{\max}]$, where $\mathrm{iTD}_{\max}$ is the maximal iTD within the natural physically accessible range, i.e., $\mathrm{iTD}_{\max} = 86$ μs for *Hemidactylus frenatus*; cf. Fig. 2.12, bottom left.

The estimation error for the firing profile as shown in Fig. 3.4 (top left) is calculated to be $E = 39$ μs over a range of $[-86\ \mu s, 86\ \mu s]$ for the angles $[-\pi/2, \pi/2]$. The horizontal angle of misestimation is then approximately $39\ \mu\mathrm{s}/86\ \mu\mathrm{s} \cdot (\pi/2) \approx 41°$. This huge estimation error results from the fact that the firing activity between neighboring map neurons is very similar; cf. Fig. 3.4, top left. The Gaussian profile almost resembles a constant function. The rate-weighted mean, i.e., the estimated $\widehat{\mathrm{iTD}}$ [cf. Eq. (3.4)], for a constant function corresponds to the center of the map. In a similar way, the estimated $\widehat{\mathrm{iTD}}$s (Fig. 3.4 top left, yellow dots) for the calculated almost flat firing profile of the iTD map shift toward the center of the map and away from the input iTDs (Fig. 3.4 top left, green line). The neuronal system therefore systematically underestimates the input iTDs.

A possibility of how the map firing profile could be sharpened is an additional GABAergic inhibition of the iTD map neurons as found in birds [58, 96, 238]. Here iTD-processing neurons of nucleus laminaris (NL) receive long-lasting depolarization from neurons of the superior olivary nucleus (SON). Since inhibition seems to be present in neuronal processing of lizards as well (Catherine Carr and Jakob Christensen-Dalsgaard, personal communication) the effect of such an inhibition on iTD processing is incorporated into the presented model and evaluated in the following. The resulting circuit is illustrated in Fig. 3.4, bottom right. Experimental data [37, 58, 156, 238] unraveled that GABAergic inhibition can increase the precision of time information processing. More precisely, GABAergic inhibition decreases the membrane resistance R_m of the map neurons, which has two effects as neatly visible

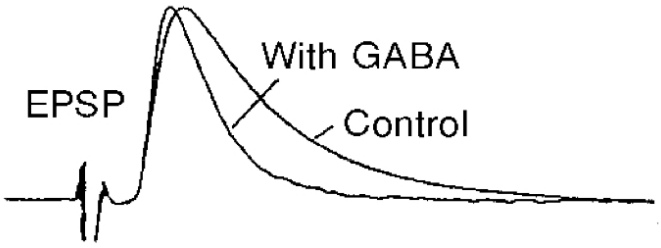

Figure 3.5: *Temporal course of an excitatory post-synaptic potential (EPSP) of nucleus laminaris (NL) neurons in chicks without (control) and with GABAergic inhibition (with GABA) due to Funabiki et al. [58]. In birds coincidence detecting NL neurons receive a long-lasting depolarization from neurons of the superior olivary nucleus (SON) [58, 96, 238] that speeds up the rise and decay times of excitatory post-synaptic potentials of NL neurons. GABAergic inhibition therefore sharpens the firing profiles of the iTD-processing NL neurons and may serve to increase coincidence detection.*

in Eq. (1.4) for a leaky integrate-and-fire neuron and a constant input current. First, more excitatory current I_{inj} is needed to increase the membrane potential above its firing threshold. This effect is called *gain control* [73] and prevents monaural summation at high sound intensities. Thus, the coincidence detection neurons remain within an appropriate working range. Given that simulated neurons are already modeled within their working range, the inclusion of an additional gain control does not alter the results of the presented Jeffress-like model.

The second effect, however, is more promising. That is, GABAergic inhibition reduces the membrane time constant $\tau_m := R_m C_m$ and consequently speeds up the rise and decay times of excitatory post-synaptic potentials; cf. Fig. 3.5. For Eq. (1.4) the exponential function tends to 0 more quickly when τ_m decreases. In concrete terms, Dasika et al. [37] showed that the membrane constant of the map halfens with additional inhibition so that map neurons can still react to smaller time delays.

For the numerical simulations inhibitory neurons are not explicitly modeled since their exact firing activity is smeared out by the long-lasting character of the inhibition. As a consequence, the effect of GABAergic inhibition is incorporated as reduced membrane constant $\tau_m = 250\ \mu$s of the map neurons. The Jeffress-like model therefore remains the same only that the parameters of the map neurons are adapted.

For a coincidence detector neuron the width of the excitatory post-synaptic potential determines the size of the time window within that two spikes are received as simultaneous. As a consequence, the neuron fires less when τ_m decreases. To nevertheless ensure a sufficient firing of the map neurons synaptic strengths are increased for a Jeffress-like model with GABAergic inhibition. All used parameters are summarized in Table 3.1 and, if different from the ones previously used, are specially marked.

Figure 3.4 (bottom left) shows the obtained firing profile of the iTD map in dependence upon different input iTDs, including the effect of GABAergic inhibition. The estimated $\widehat{\text{iTDs}}$ [cf. Eq. (3.4)] nicely reproduce input iTDs. The estimation error reduces to $E = 18\ \mu$s, and thus the misestimation angle is calculated as $18\ \mu\text{s}/86\ \mu\text{s} \cdot(\pi/2) \approx 19°$. Similar resolutions are found in birds, such as 23° for great tits [118], 27° for budgerigars [169], and

Parameter			Value
maximal iTD (*Hemidactylus frenatus*)	$\mathrm{iTD}_{\max}$	=	86 μs
auditory nerve firing rate	A	=	300 Hz
auditory nerve vector strength	VS	=	0.95
number of input neurons	N_{in}	=	30
number of map neurons	N_{map}	=	100
ITD range of map neurons	$\mathrm{iTD}_{\mathrm{map}}$	$\in$	$[-430\ \mu\mathrm{s},\ 430\ \mu\mathrm{s}]$
synaptic strength	J	=	0.015
with inhibition	J	=	0.021
post-synaptic current width	τ_{EPSC}	=	250 μs
membrane constant of map neuron	τ_m	=	500 μs
with inhibition [37]	τ_m	=	250 μs
map neuron refraction time	τ_{refr}	=	1000 μs
map neuron capacitance	C	=	1
map neuron resting potential	V_o	=	0
map neuron threshold	V_{thresh}	=	1

Table 3.1: *Parameters used in computer simulations of a Jeffress-like model. If not marked otherwise, all parameters are taken from Friedel et al. [57]. Parameters that change as a consequence of an assumed additional GABAergic inhibition are marked and written in blue. With decreasing membrane constant the map neuron can distinguish more reliable between non-coinciding input spikes and fires less. To ensure a sufficient firing of the map neurons despite additional inhibition the synaptic strength J is increased.*

29° for canaries [169]. The proposed Jeffress-like model with GABAergic inhibition therefore provides reasonable results for iTD processing. Exact time constants for *Hemidactylus frenatus* and the existence of a long-lasting depolarization through GABAergic inhibition, however, have to be experimentally verified.

3.4 Processing of iADs

Similar to neuronal processing of iTDs in lizards, neuronal processing of iADs is not yet experimentally clarified. A neuronal construction to estimate amplitude differences, however, are EI neurons, that is, neurons that receive inhibitory input from the ipsilateral side and excitatory input from the contralateral side. Experiments [59, 69, 155, 173, 191] show that EI neurons are sensitive to amplitude variation. In the following, firing activity of EI neurons are simulated in response to inputs differing in amplitude between the two sides by the tympanic iAD as calculated from the ICE model based on a simplified cylindrical mouth cavity; cf. Chap. 2. As illustrated by Fig. 3.6 the model consists of two stages. First, tympanic vibrations excite neurons of the nucleus angularis (NA); cf.

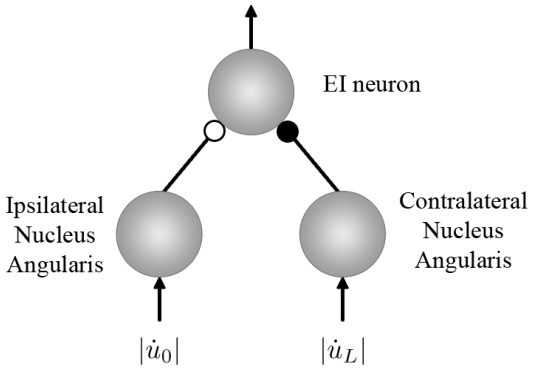

Figure 3.6: *Suggested neuronal circuit to process iADs. For both the experimental and the natural stimulus Eq.* (2.26) *of the ICE model allows for the calculation of the tympanic vibration amplitudes* $|\dot{u}_0|$ *and* $|\dot{u}_L|$ *in dB. The latter are used as input firing rates for the Poisson neurons of the ipsi- and contralateral nucleus angularis (NA). In the next step, the ipsilateral NA projects via inhibitory synapses (white sphere) and the contralateral NA projects via excitatory synapses (black sphere) to EI neurons that are modeled as leaky integrate-and-fire neurons; see Sec. 1.3.2 for details.*

high-frequency projection in Fig. 3.1. In the second step, the contralateral NA projects via excitatory synapses, and the ipsilateral NA projects via inhibitory synapses to EI neurons. Given that neuronal projections from the NA remain to be elucidated, the exact location of these EI neurons and the exact projecting pathway from the NA to these EI neurons, e.g., through intermediate neurons, are not specified. In mammals, EI neurons are topographically organized within the inferior colliculus (IC) [59, 69, 173]. In chicks, EI neurons are found within the dorsal lateral lemniscal nucleus (LLD) [191].

3.4.1 Experimental and natural stimulus

To estimate the stimulus exciting the auditory nerves, amplitudes of the tympanic vibrations are calculated. More concretely, Eq. (2.26) of the ICE model applies for incoming pressures p_{ex_0} and p_{ex_L}. For the latter two different stimuli are possible.

- **Experimental stimulus**
  To reproduce experimental setups the incoming pressure is varied in intensity between the tympani, IAD $\neq 0$, but not in phase, ITD $= 0$. Hence, pressure functions at the two tympanic membranes ($x = 0$ and $x = L$) are given by

$$\begin{aligned} p_{ex_0}(\omega, \mathrm{IAD}; t) &= |p_{ex}| \exp\left(\frac{\mathrm{IAD}}{20}\right) \exp\left(i\omega t\right), \\ p_{ex_L}(\omega; t) &= |p_{ex}| \exp\left(i\omega t\right) \end{aligned} \tag{3.6}$$

  where $\mathrm{IAD} = 20 \log\left(|p_{ex_L}/p_{ex_0}|\right)$ induces the exponential form of the amplitude modification. The input IADs, however, do not necessarily enter the neuronal sound processing circuits. In contrast, internally coupled ears first modify the stimulus at the level of the tympanic vibrations, resulting in iADs $\neq$ IADs.

- **Natural stimulus**
  For an animal of small head size such as *Hemidactylus frenatus* (gekkonid) shielding effects due to the head can be neglected. That is, natural stimuli only differ in phase between the two sides, ITD $\neq 0$, whereas intensity differences can be neglected, IAD $= 0$. Similarly to Eq. (2.1) incoming pressures p_{ex_0} and p_{ex_L} are then given by

$$\begin{aligned} p_{ex_0}(\omega, \mathrm{ITD}; t) &= |p_{ex}| \exp\left[i\omega(t + \mathrm{ITD})\right], \\ p_{ex_L}(\omega; t) &= |p_{ex}| \exp\left(i\omega t\right) \ . \end{aligned} \tag{3.7}$$

  The reason why such a stimulus is not used in experiments is obvious, there are no amplitude differences in the stimulation, so why should an experimentalist who looks for neuronal circuits for amplitude processing expect a reaction? The answer lies within the properties of ICE. Given external ITDs, internally coupled ears generate iADs up to 20 dB between the vibrations of the two tympani for frequencies above 1 kHz; cf. Fig. 2.12, top. In other words, neuronal circuits are not characterized by the properties of the input but by the properties of the tympani. Thus neurons can only be **i**TD- or **i**AD-sensitive. Given that simultaneous measurements of the tympanic vibrations are necessary to estimate the input to the neuronal system, experiments with the natural stimulus are, at a first glance, more complicated than experiments with the experimental stimulus. For the latter, the experimentalists do not have to be aware of the modification due to the internally coupled ears. Neglecting the generation of iADs from ITDs could nevertheless lead to wrong interpretations of experimental results, e.g., wrong classification of iAD-sensitive neurons as ITD-sensitive neurons.

The following simulation of NA neurons distinguishes between the experimental and the natural stimulus. Measured data are only available for homologous brain areas in chicks in response to an experimental stimulus [191].

3.4.2 Simulation of nucleus angularis (NA)

In a first step, simulation of NA neurons requires an estimation of how these neurons react to tympanic vibrations of varying intensity. Köppl and Manley have measured responses of the auditory nerve for the experimental stimulus in the bobtail lizard *Tiliqua rugosa*; see [128] and Fig. 3.7. According to this study, auditory nerves possess a dynamic range in which the firing rates vary linearly with sound pressure level in dB. Exploiting these experimental results, the model concentrates on the dynamic range of the NA neurons that are consequently simulated as Poisson neurons; see Sec. 1.3.3. Their input rate functions are assumed to be proportional to the amplitudes of the tympanic vibrations $|\dot{u}_0|$ and $|\dot{u}_L|$ in dB as calculated by Eq. (2.26) of the ICE model.

Simulations of contra- and ipsilateral NA neurons in response to the experimental and the natural stimulus of 1 kHz lead to firing profiles as illustrated in Fig. 3.8. The obtained firing profiles for the experimental stimulus (left) nicely reproduce experimental data from the contralateral auditory nerve of chicks in response to varying IADs; cf. Fig 3.9 (top left) by Sato et al. [191]. Given that internally coupled ears have been experimentally proven in several birds [17, 33, 88, 183], in particular in chicks [95], the assumptions made for the modeling input and the firing of NA neurons seem reasonable for an ICE system.

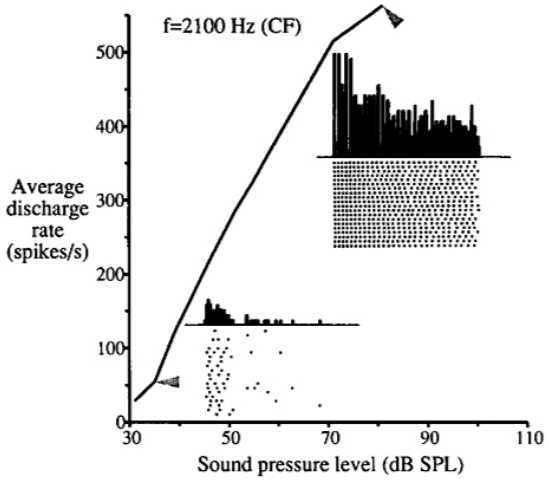

Figure 3.7: *Discharge rates of auditory nerve fibers of* Tiliqua rugosa *with a characteristic frequency (CF) of* 2.1 *kHz dependent on incoming sound pressure level. Within the dynamic range from 40-80 dB SPL, the firing of the auditory nerve varies almost linearly with the stimulus strength in dB. Figure due to Köppl and Manley [128].*

The firing activity of NA neurons for the natural stimulus of ITDs $\in [-28\ \mu\text{s}, 28\ \mu\text{s}]$ are presented in Fig. 3.8, right. Here firing activities vary monotonically with input ITDs. The firing activity of ipsilateral NA neurons is maximal for positive ITDs and minimal around $-20\ \mu$s where it hardly varies. The firing activity of contralateral NA neurons is exactly the opposite with maximal values at negative ITDs.

3.4.3 Simulation of EI neurons

In a second step, neurons of the ipsilateral NA project via inhibitory synapses and neurons of the contralateral NA project via excitatory synapses to EI neurons; cf. Fig. 3.6. EI neurons are modeled as leaky integrate-and-fire neurons; see Sec. 1.3.2. Due to the assumption of two processing pathways, high frequency inputs only contain amplitude and not timing information. The firing rates of NA neurons are therefore assumed to be constant over time. Equation (1.4) determines the membrane potential of a leaky integrate-and-fire neuron in the case of a constant input current I_{inj} as

$$V(t) \quad = \quad V_r + R_m I_{\text{inj}} \left[1 - \exp\left(-\frac{t}{C_m R_m} \right) \right] \tag{3.8}$$

with $t_0 = 0$. The injected current I_{inj} is the sum of excitatory currents from the contralateral NA and inhibitory currents from the ipsilateral NA. Given that neurons of the NA fire with a probability ν_{ipsi} and ν_{contra}, the injected current can be expressed as

$$I_{\text{inj}} \quad = \quad J_{\text{inh}}\ \nu_{\text{ipsi}} + J_{\text{exc}}\ \nu_{\text{contra}} \tag{3.9}$$

with $J_{\text{inh}} \leq 0$ and $J_{\text{exc}} \geq 0$ being the strengths of inhibitory and excitatory synaptic connections, respectively. With $J_{\text{inh}} = -J_{\text{exc}}$ and ν_{ipsi} and ν_{contra} being proportional to the vibration amplitude of the tympani, the injected current I_{inj} is proportional to the iAD.

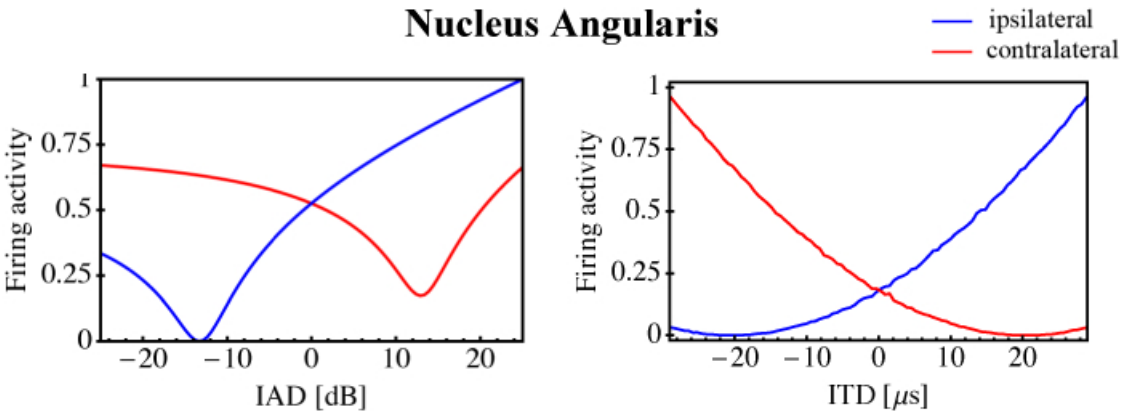

Figure 3.8: *Simulated firing profiles of neurons from the ipsilateral (blue) and contralateral (red) nucleus angularis (NA) (cf. Fig. 3.6) in response to tympanic vibration amplitudes $|\dot{u}_0|$ and $|\dot{u}_L|$ in dB as calculated from Eq. (2.26) of the ICE model. Simulations are performed with parameters given in Table 3.2. Left: Simulations performed with the experimental stimulus, i.e., varying input IADs. Here the firing rates of ipsilateral and contralateral NA neurons have either a clear minimum or maximum and cannot represent the input IADs in an unambiguous way. Right: Simulations performed with the natural stimulus, i.e., varying input ITDs. The firing rates of ipsilateral and contralateral NA neurons vary monotonically with the input ITDs and can therefore unambiguously represent the stimulus.*

As analyzed in Sec. 1.3.2, the neuron will not spike at all if $V_r + R_m I_{\text{inj}}$ lies below the firing threshold $\overline{V}$. Otherwise, the spike rate is proportional to $1/T$, with T being the constant interspike interval. The latter is given by Eq. (3.8) and $V(t = T) = \overline{V}$, such that

$$\overline{V} \;=\; V_r + R_m I_{\text{inj}} \left[1 - \exp\left(-\frac{T}{C_m R_m}\right)\right] \;. \tag{3.10}$$

Given that the interspike interval is limited by the refractory period T_{refrac} as the minimum interval, the interspike interval can be calculated to be

$$T \;=\; \begin{cases} -C_m R_m \ln\left(1 - \frac{\overline{V}-V_r}{R_m I_{\text{inj}}}\right) & \text{for } T \geq T_{\text{refrac}} \\ T_{\text{refrac}} & \text{for } T < T_{\text{refrac}} \end{cases} \;. \tag{3.11}$$

This formula is simplified by the prerequisite for a firing neuron, that is, $V_r + R_m I_{\text{inj}} > \overline{V}$. As a consequence, the fraction $(\overline{V} - V_r)/(R_m I_{\text{inj}}) < 1$. For $(\overline{V} - V_r)/(R_m I_{\text{inj}}) \ll 1$ the logarithm can be approximated by $\ln(1 + x) \approx x$ which leads to

$$T \;=\; \begin{cases} C_m \left(\overline{V} - V_r\right) \;/\; I_{\text{inj}} & \text{for } T \geq T_{\text{refrac}} \\ T_{\text{refrac}} & \text{for } T < T_{\text{refrac}} \end{cases} \;. \tag{3.12}$$

Altogether the neuron shows three different firing behaviors. Either it remains silent ($V_r + R_m I_{\text{inj}} < \overline{V}$), fires constantly at its maximal rate ($T < T_{\text{refrac}}$), or shows a firing rate $1/T$ that varies with $I_{\text{inj}} = J_{\text{inh}}\ \nu_{\text{ipsi}} + J_{\text{exc}}\ \nu_{\text{contra}}$.

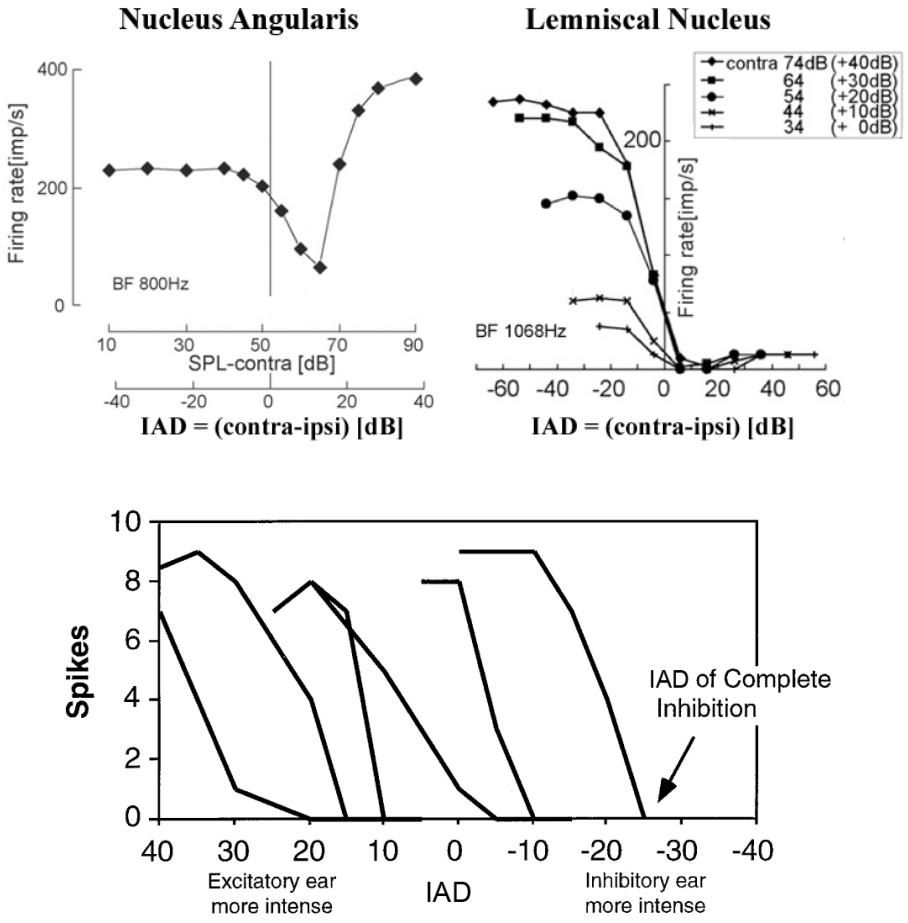

Figure 3.9: *Top: IAD processing in chicks [191]. Auditory nerves first project to the nucleus angularis (NA). The contralateral NA then projects via excitatory synapses and the ipsilateral NA via inhibitory synapses to the dorsal lateral lemniscal nucleus (LLD). Top left: Measured firing rate in impulses per second [imp/s] of intensity-sensitive neurons of the contralateral NA with best frequency BF = 800 Hz in response to the experimental stimulus, i.e., varying IADs. Top right: Responses of LLD units with BF = 1068 Hz to binaural sound with varying IAD. Rate-intensity curves are measured in a single LLD unit at five different contralateral sound levels above the threshold of this unit (34 dB SPL), as indicated, while the ipsilateral sound level was varied. The measured responses of NA and LLD neurons to the experimental stimulus nicely correspond to calculated data for the experimental stimulus, as illustrated in Figs. 3.8 (left) and 3.10 (left). Bottom: IAD processing in bats [170]: measured firing rates of intensity-sensitive neurons in response to varying IADs. When the input from the inhibitory ear gets more intense the EI neuron stops firing at a neuron-specific IAD denoted as* IAD of complete inhibition. *Within the array of EI neurons, IADs of complete inhibition systematically vary over the whole physically accessible range of possible IADs. For a specific input IAD neurons with IADs of complete inhibition above the considered IAD fire, whereas the other neurons remain silent. Thus, the IAD of complete inhibition of the neuron at the edge of this step function could encode the estimated* $\widehat{\text{IAD}}$. *Since the IADs systematically vary with the direction of the sound source (cf. head-related transfer function [141]) the bat can use IADs to localize a sound source. It is therefore fair to say that the neurons represent a neuronal map of estimated* $\widehat{\text{IAD}}$*s and hence a neuronal map of the horizontal direction of the sound source.*

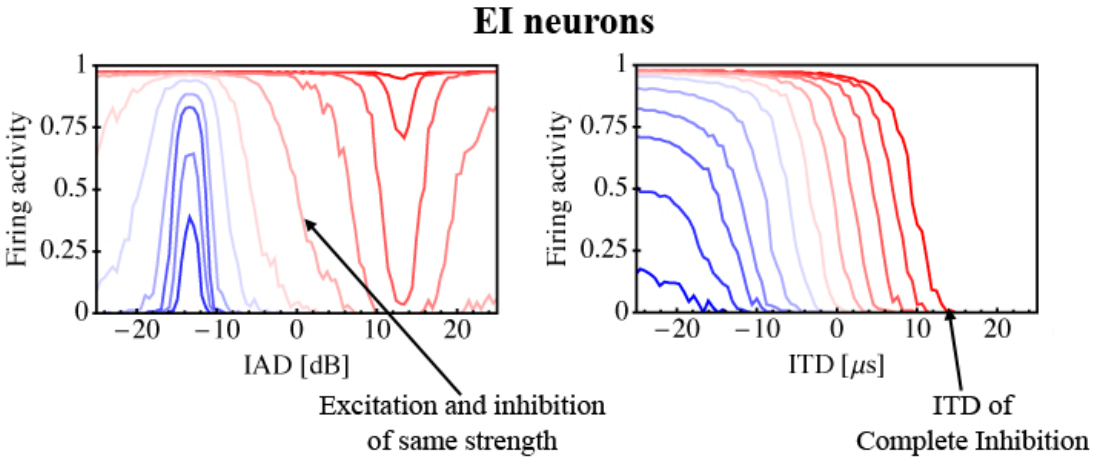

Figure 3.10: *Simulated firing profiles of EI neurons receiving inhibitory inputs from the ipsilateral side and excitatory inputs from the contralateral side; cf. Fig. 3.6. Simulations are performed with parameters given in Table 3.2. Strengths of excitatory and inhibitory synapses vary systematically between 0 and 1. Red indicates more excitation than inhibition, and blue represents the opposite. The intensive red line, for instance, corresponds to strong excitatory synapses near 1 and almost ineffective inhibitory synapses with strength almost* 0. *Left: Simulated firing activity of EI neurons in response to the experimental stimulus. An arrow marks the firing profile for an EI neuron for which excitatory and inhibitory synapses are assumed to have the same strength, i.e.,* $J_{\text{inh}} = -J_{\text{exc}}$. *This firing profile nicely fits experimental data for EI neurons, as shown in Fig. 3.9, top right. The other profiles, however, either have a maximum at* -15 *dB or a minimum at* 15 *dB indicating that the presented EI neurons are not optimally adapted to the stimulus. Right: Simulated firing profiles of EI neurons in response to the natural stimulus.* ITDs of complete inhibition *vary systematically with strength of inhibition and cover large parts of the physically accessible range. Only high ITDs above* 15 *μs do not induce firing activity of the shown EI neurons. EI neurons from the other side of the brain, however, respond to the corresponding negative ITDs. Firing profiles of both ipsi- and contralateral EI populations therefore provide an encoding strategy for sound source localization (viz., direction) that is similar to the one found, e.g., in bats; see Fig. 3.9 (bottom) and Park et al. [170].*

As before the numerical simulations of EI neurons distinguish between the experimental and the natural stimulus. The input rates of the NA neurons are proportional to the tympanic vibration amplitudes $|\dot{u}_0|$ and $|\dot{u}_L|$ in dB; cf. Eq. (2.26). Firing rates of EI neurons as a result of simulations with the experimental stimulus are shown in Fig. 3.10, left. Here inhibition and excitation systematically vary in strengths. An arrow marks the firing profile of the EI neuron where excitatory and inhibitory synapses have the same strength $J_{\text{inh}} = -J_{\text{exc}}$. Firing profiles with similar strength of inhibition and excitation vary monotonically with input IADs and show the above described three different firing behaviors. For negative input IADs the excitatory input dominates the input currents, and the EI neuron fires at maximal rate. Around 0 dB the firing rate decreases with input IADs in a linear manner; cf. Eq. (3.12). As soon as input IADs reach a certain threshold, called *IAD of complete inhibition*, the inhibitory input is so strong that the membrane potential remains below its firing threshold, and the neuron stops firing. Firing profiles obtained as a result of the numerical simulations for $J_{\text{inh}} \approx -J_{\text{exc}}$ have been measured in the dorsal lateral lemniscal nucleus of chicks; cf. Fig. 3.9 lemniscal nucleus, top right.

Firing profiles of EI neurons with different strengths of inhibition and excitation (Fig. 3.10, left), however, do not monotonically vary with input IADs. Instead the profiles either have a maximum at −15 dB or a minimum at 15 dB. The occurring ambiguity could indicate that the used experimental stimulus does not correspond to the stimulus the map is tuned for. If this hypothesis holds true firing profiles of EI neurons should rather be adapted to the natural stimulus.

Figure 3.10 (right) shows the according firing profiles of EI neurons in response to the natural stimulus with varying strengths of inhibition. Here the firing activities of EI neurons are perfectly adapted to the natural stimulus in such a way that they monotonically vary with input ITD, i.e., horizontal direction of sound source. In addition, *ITDs of complete inhibition* systematically vary with the strength of inhibition and cover large parts of the physically accessible range. The ITDs of complete inhibition of the EI neurons could therefore encode the horizontal direction of sound. For a whole array of EI neurons with varying strength of inhibitory and excitatory synapses the map firing profile in response to a specific input ITD looks like a step function. Neurons with ITDs of complete inhibition above the considered ITD fire, whereas other neurons remain silent. Thus, the ITD of complete inhibition of the neuron at the edge of this step function could encode the estimated $\widehat{\text{ITD}}$ and hence the horizontal direction of the sound source. A similar encoding principle based on IADs has been measured in bats [51, 170]; cf. Fig 3.9, bottom.

3.5 Conclusion

Based on the results of the ICE model with a simplified cylindrical mouth cavity for *Hemidactylus frenatus* (gekkonid) and experimental data, two separated pathways for low and high frequencies are suggested, a Jeffress-like model without and with GABAergic inhibition for processing of iTDs and EI neurons for processing of iADs.

Concerning the processing of iTDs, the simulation of an exclusively excitatory Jeffress-like model results in a horizontal estimation error of $\approx 41°$. An additional GABAergic inhibition could improve the angle of misestimation to $\approx 19°$. Both values are in good

Parameter	Value
experimental stimulus: maximal input IAD	$\mathrm{IAD}_{\max} = 25$ dB
natural stimulus: maximal input ITD	$\mathrm{ITD}_{\max} = 28\ \mu$s
excitatory synaptic strength	$J_{\mathrm{exc}} \in [0;1]$
inhibitory synaptic strength	$J_{\mathrm{inh}} \in [-1;0]$
post-synaptic current width	$\tau_{\mathrm{EPSC}} = 500\ \mu$s
map neuron time constant	$\tau_{\mathrm{relax}} = 500\ \mu$s
map neuron refraction time	$\tau_{\mathrm{refr}} = 1$ ms
map neuron capacitance	$C = 1$
map neuron resting potential	$V_o = 0$
map neuron threshold	$V_{\mathrm{thresh}} = 1$

Table 3.2: *Parameters used for simulations of EI neurons for iAD processing. For the simulations the experimental stimulus (varying input IADs) and the natural stimulus (varying input ITDs) are used as denoted.*

agreement with resolutions found in birds. Exact time constants for *Hemidactylus frenatus* and the existence of a long-lasting depolarization through GABAergic inhibition, however, have to be experimentally verified.

Concerning the processing of iADs, the neuronal activity of the NA and the ensuing EI neurons have been calculated. Here two different stimuli were specified. The experimental stimulus varies the input IADs, whereas the natural stimulus varies the input ITDs. The obtained firing rates of NA and EI neurons in response to the experimental stimulus are in perfect agreement with the measured firing rates of corresponding brain areas in the chick.

Firing rates of EI neurons in response to the natural stimulus with varying strengths of inhibition provide systematically varying ITDs of complete inhibition and hence a possible encoding strategy of horizontal sound source direction within an array of EI neurons.

Taken together, the results of the two processing pathways of iTDs and iADs are reasonable and fit experimental results from closely related birds. Given that auditory processing could, however, differ between lizards and chicks, neurophysiological measurements from the lizard's auditory systems are needed to finally verify the suggested processing of iTDs and iADs. In particular, the involved brain areas and pathways have to be clarified.

The smaller the lizard, the larger its hope to become a crocodile.

Proverb from Abyssinia

4. Auditory Sensitivity and Internally Coupled Ears

The vertebrate auditory system exhibits large variations in structure and mechanics. In contrast to the acoustically independent ears of mammals, many other vertebrates have morphologically and functionally coupled tympanic membranes, called *internally coupled ears* (ICE). Here general relations between a vertebrate's head size, i.e., interaural distance, and its best frequency, viz., the frequency with the lowest hearing threshold, are explored. Time and amplitude differences between the membrane vibrations determine hearing abilities of animals with internally coupled ears. The corresponding equations then deliver an inverse relation $f_{\text{best}} \propto L^{-1}$ between interaural distance L and best frequency f_{best}, which is confirmed by experimental data. Experimental best frequencies of animals with independent ears provide an inverse dependence with an exponent around -0.5 that cannot be explained by analysis of the corresponding acoustics. Apart from different dependences upon interaural distance, calculations and experimental measurements show that animals with internally coupled ears have in general lower best frequencies than animals of the same interaural distance with independent ears. Internally coupled ears thus present themselves as a distinct auditory mechanism that assures an animal's capabilities of hearing low frequencies despite a small head size.

4.1 Introduction

Internally coupled ears are probably the least studied among the various hearing systems, despite being common in both invertebrate and vertebrate taxa. By means of the direct coupling of the left and right middle ear cavities, the vibration of each tympanic membrane shows a directional response to incoming sound signals, the so-called pressure-gradient receiver characteristic, as first described by Autrum [4] for locusts. This means that phase as well as amplitude of a single tympanic membrane vibration varies with respect to that

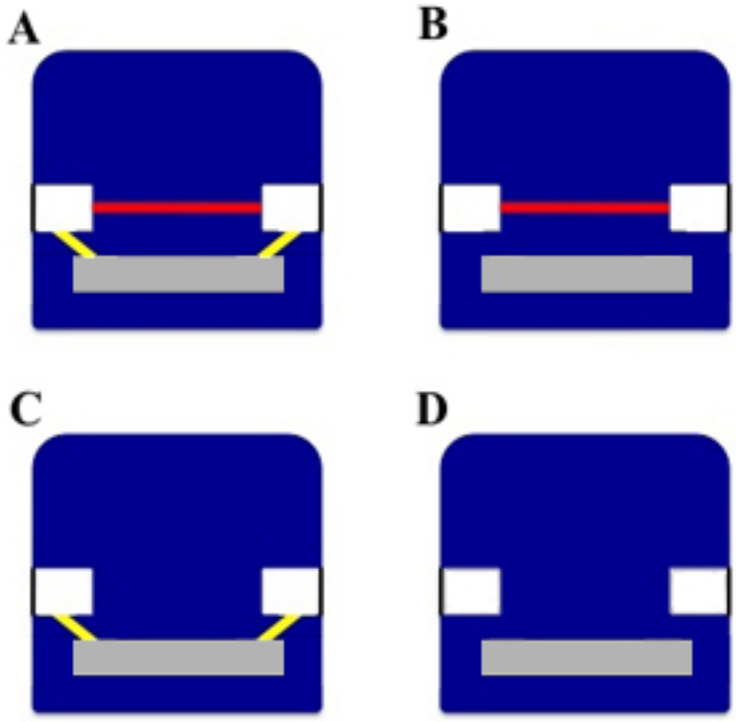

Figure 4.1: *Illustration of anatomical variations within the auditory system. The plots show cross-sections through the head at the level of the ear of animals with ICE (A and B) and with independent ears (C and D). Black (tympanic membrane), white (middle ear cavity), yellow (Eustachian tube), gray (pharynx), red (direct linkage). A Directly linked middle ear cavities with Eustachian tubes, e.g., lizards, frogs, alligators, and birds. For lizards all cavities are merged to one huge mouth cavity so that the middle ear cavities are* directly *linked through Eustachian tubes and pharynx; cf. Fig. 1.2, right. B Directly linked middle ear cavities without Eustachian tubes, e.g., chameleons. C Indirect connection of the middle ear cavities through Eustachian tubes and pharynx, e.g., mammals. D No connection between the middle ear cavities, e.g., a number of lizards and snakes.*

of the other membrane depending on the direction of the sound source. The expression "pressure-gradient receiver" refers to the fact that such a system is not sensitive to the pressure amplitude of a sound wave at one specific point but rather to the direction of the sound wave already at the stage of the tympani. More precisely, Autrum [4] noted that such a receiver reacts to the pressure *difference* between the tympani. To avoid misunderstandings concerning technical terms, throughout the thesis cavity-coupled tympanic membranes are referred to as *internally coupled ears* (ICE). Descriptions and measurements of ICE are available from a number of different animals including crickets [150], frogs [29, 49, 103], barn owls and other birds [17, 33, 88], and lizards [27, 28]. Evolutionary aspects have been covered in several reviews [19, 20].

Simplistically the ears of the tetrapodal (or non-piscine) vertebrates can be classified into one of the conditions (Fig. 4.1), with a variety of anatomical variation in each condition.

The first auditory condition (Fig. 4.1 A) in which there is a direct connection between the right and left middle ear cavities, as well as an indirect connection via Eustachian tubes linking the middle ear cavities to the pharynx, occurs in some frogs, lizards, alligators, and birds, and has frequently been interpreted as an auditory specialization. For lizards direct and indirect connection are identical since all cavities are merged to one huge mouth cavity. The middle ear cavities are therefore directly linked through Eustachian tubes and pharynx; cf. Fig. 1.2, right. The rarest arrangement (Fig. 4.1 B) is for the middle ear cavities to be directly linked, but for the Eustachian tubes to be either absent or represented by a single opening between the pharynx and the connection between the right and left middle ear cavity. In this condition, best represented by the Chameleons [231], the pharynx does not play any role in establishing a pressure difference. A more common configuration (Fig. 4.1 C) occurs when the middle ear cavities are only indirectly linked by paired Eustachian tubes opening into the pharynx. This condition typifies the mammalian ear but is also frequently encountered in reptiles. In the fourth configuration (Fig. 4.1 D) the contralateral middle ear cavities are anatomically isolated from one another, and maybe greatly reduced or even lost. This configuration is encountered, for instance, in a number of lizards and snakes.

Using experimental measurements of directional and dependent vibrations of the two tympani most animals group as having either independent ears or ICE. In general, conditions A and B from Fig. 4.1 are found in animals with ICE, while conditions C and D are found in animals with independent ears. The question of how direct the coupling between the middle ear cavities has to be in order for the ears to form ICE remains unsolved. The following analysis explores the relationship between head size, more exactly interaural distance, and peak auditory sensitivity since head size has a direct impact on the morphometrics of the coupling(s) that establish the ICE.

The inverse relation between body size and frequency of peak sensitivity both among and within species was recently reviewed [71]. Previous studies have used a variety of metrics for animal size including body mass, head width, and functional head width (interaural distance divided by the speed of sound in the surrounding medium). Nevertheless, the relationship between peak sensitivity and size in animals (excluding subterranean species) with independent ears [80, 81, 82, 83] is clearly different from the relationship reported in animals with ICE as described, e.g., by Werner [229, 230] and Gleich [68]. The primary purpose of the present chapter is to explore the biophysical basis of this difference and, by doing so, explore aspects of the functional performance of ICE.

4.2 Theoretical description of membrane vibrations

In general, sound processing is based on differences between the vibrations of the tympanic membranes. In the following equations for time and amplitude differences between the membrane vibrations both for independent and internally coupled ears are derived. The best frequencies correspond to the frequencies where these differences become maximal. This is reasonable since the best frequency is determined by the frequency with the lowest hearing threshold; cf. Fig. 4.2.

Membrane vibrations are responses to sound waves of a certain frequency ω from a

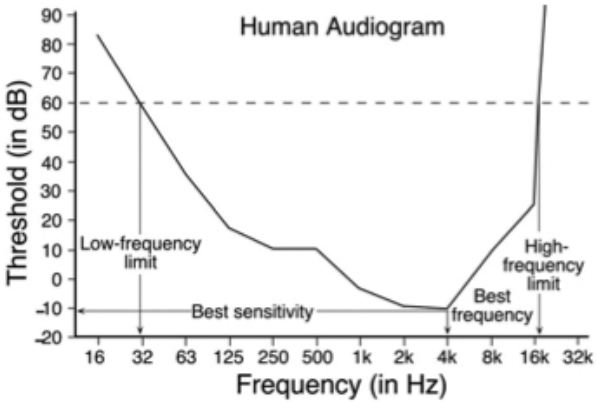

Figure 4.2: *Visualization of best frequency at the example of human hearing [81]. The audiogram shows hearing thresholds in dependence upon the frequency of the sound signal. The best frequency is defined as frequency with the lowest hearing threshold. In the present example the best frequency corresponds to* 4 *kHz.*

specific sound source at distance D and angle θ. Because of the different positions of the two tympani with respect to the sound source the sound pressure waves $p_{ex_{0/L}}$ arrive at different times and with different amplitudes at the two tympani. The pressure at a point with angle θ on a spherical head can be calculated [see, e.g., Kuhn [133]] to be

$$p_{ex_{0/L}}(\theta,\omega;t) = i\rho c/(4\pi L^2)e^{-i\omega t}\sum_n P_n[\sin(\pm\theta)]h_n\left(\frac{\omega}{c}\,D\right)/h_n'\left(\frac{\omega}{c}\,L\right) \tag{4.1}$$

where L is the interaural distance, c the velocity of sound, and ρ the air density. Furthermore, P_n is the Legendre polynomial of order n and h_n is the spherical Hankel function of order n. Thus sound arrives at the contralateral tympanum later and more alleviated than at the ipsilateral tympanum. For $\omega/c\ L$ less than one, the difference between the two arriving pressure functions reduces to a phase shift. Experimental data from vertebrates with ICE (see Table 4.1) fulfill this assumption. Accordingly, the simplification generally applies for internally coupled ears and for independent ears in the low-frequency range. In this case the differences between the input pressures are sufficiently described by the interaural time difference

$$\mathrm{ITD} = (L/c)\sin(\theta)\ ; \tag{4.2}$$

for details see Sec. 2.2.1. In doing so the effect of sound traveling *around* the head is neglected.

For independent ears and higher frequencies, Eq. (4.1) cannot be simplified. Interaural amplitude differences (IADs) are calculated by the full solution to be

$$\mathrm{IAD} \quad := \quad 20\log\left[|p_{ex_0}(\theta,\omega;t)|/|p_{ex_L}(\theta,\omega;t)|\right]\ . \tag{4.3}$$

For a sound source directly in front of the head, the angle θ becomes zero, and consequently so do the corresponding ITD and IAD values.

4.2.1 Membrane vibration differences of independent ears

For independent ears the two membranes process incoming sound independently. The membrane vibrations $\dot{u}_{0/L}$ are therefore solutions of a forced membrane equation [50, 221]. That is,

$$\dot{u}_{0/L}(r;t) \;=\; \Omega^{-1}\left[p_{ex_{0/L}}\int_S f_0(r)\mathrm{d}S\right] f_0(r)e^{i\omega t} \tag{4.4}$$

with the definition

$$\Omega^{-1} := \frac{1}{\rho_m d}\frac{i\omega}{\left(\omega_0^2-\omega^2\right)+2i\omega\alpha}\ . \tag{4.5}$$

The parameters are the following

r,t	solution parameters: radial position on the membrane and time,
S	surface of the membrane,
a	radius of the membrane,
d	thickness of the membrane,
ρ_m	density of the membrane,
ω_0	first eigenfrequency of the membrane, and
α	damping coefficient of the membrane.

The fundamental vibration mode of the membrane, a normalized Bessel function [35, p. 313] of first kind of order 0, is denoted by $f_0(r) := [J_1(k_1a)]^2\, J_0(k_1r)$ with the k_1 defined by $J_0(k_1a)$=0. To determine the specific parameters for different systems, values for the gecko *Hemidactylus frenatus* (gekkonid) [221] are used and linearly scaled with the system as a whole. Radius a and thickness d of the membrane therefore increase linearly. In contrast, the eigenfrequencies of the tympani ω_{mn} decrease linearly with head size, i.e., interaural distance and radius. The inverse relation between ω_{mn} and radius is illustrated by Eq. (2.7) that is valid for the radial function $g(r)$ of internal cavity and membrane. When the geometry is scaled by γ the radius r in Eq. (2.7) is replaced by $r=\gamma r'$ leading to

$$\frac{\partial^2 g(\gamma r')}{\gamma^2\ \partial r'^2}+\frac{1}{\gamma r'}\frac{\partial g(\gamma r')}{\gamma\ \partial r'}+\left[k_{qs}^2-\frac{q^2}{\gamma^2\ r'^2}\right]g(\gamma r') \;=\; 0\ . \tag{4.6}$$

Multiplying the equation by γ^2 leads to

$$\frac{\partial^2 g(\gamma r')}{\partial r'^2}+\frac{1}{r'}\frac{\partial g(\gamma r')}{\partial r'}+\left[\underbrace{\gamma^2\ k_{qs}^2}_{:=(k'_{qs})^2}-\frac{q^2}{r'^2}\right]g(\gamma r') \;=\; 0\ . \tag{4.7}$$

The resulting scaled solutions are again Bessel functions [35, p. 313] but with modified k'_{qs}. Since the eigenfrequencies of the original system are given by $\omega_{mn}=k_{qs}c_M$ with c_M the propagation velocity on the membrane [50, p. 74] the eigenfrequencies of the scaled membrane are given by $\omega_{mn}=k_{qs}c_M=k'_{qs}/\gamma\ c_M=\omega'_{mn}/\gamma$. In particular, the eigenfrequency of the fundamental mode scales as $\omega_0=\omega'_0/\gamma$ for $r=\gamma r'$.

The time shift between the membrane vibrations for independent ears $\mathrm{ITD}_{\mathrm{ind}}$ can then be calculated as a difference between the arguments, denoted as arg, of the complex vibrations of the tympani leading to

$$\begin{aligned} \mathrm{ITD}_{\mathrm{ind}} &= \{\arg[\dot{u}_0(r;t)] - \arg[\dot{u}_L(r;t)]\}/\omega \\ &= \arg[\dot{u}_0(r;t)/\dot{u}_L(r;t)]/\omega \\ &= \arg[p_{ex_0}/p_{ex_L}]/\omega \ . \end{aligned} \tag{4.8}$$

The $\mathrm{ITD}_{\mathrm{ind}}$ of independent ears consequently reproduces the time shift of the exciting external pressure waves; cf. Eq. (4.2). As a consequence of neuronal phase-locking, which only occurs if the frequency is low enough, ITDs are only used in the low-frequency range so that application of the simplification (4.2) leads to

$$\mathrm{ITD}_{\mathrm{ind}} = L/c \, \sin(\theta) \ . \tag{4.9}$$

The difference between the vibration amplitudes of the membranes, i.e., the $\mathrm{IAD}_{\mathrm{ind}}$, can be calculated as

$$\begin{aligned} \mathrm{IAD}_{\mathrm{ind}} &= 20 \log \left[|\dot{u}_0(r;t)/\dot{u}_L(r;t)| \right] \\ &= 20 \log \left(|p_{ex_0}/p_{ex_L})| \right) \ . \end{aligned} \tag{4.10}$$

Again the vibration amplitudes of the membranes of independent ears reflect the IADs of the incoming sound wave; cf. Eq. (4.3).

4.2.2 Membrane vibration differences of internally coupled ears

Tympani of internally coupled ears do not vibrate independently but influence each other. That is, the external stimuli *and* the internal coupling determine the membrane response. Differences between time and amplitude of membrane vibrations (iTD and iAD) can be described by the ICE model presented in Chap. 2 and [221]. For reasons of simplicity the problem is assumed to be rotationally symmetrical, i.e., a possible asymmetry within the attachment of the middle ear is neglected. By a **reduced ICE** model [see Eqs. (2.26)-(2.28)] membrane vibrations can be calculated to be

$$\begin{aligned} \dot{u}_0(r;t) &= G_{\mathrm{ipsi}}(r;t) \ p_{ex_0} + G_{\mathrm{contra}}(r;t) \ p_{ex_L} \ , \\ \dot{u}_L(r;t) &= G_{\mathrm{ipsi}}(r;t) \ p_{ex_L} + G_{\mathrm{contra}}(r;t) \ p_{ex_0} \ , \end{aligned} \tag{4.11}$$

with the ipsilateral filter

$$\begin{aligned} &G_{\mathrm{ipsi}}(r;t) \\ &= -\int_S f_0(r) \ \mathrm{d}S \ \left\{ \frac{-\rho c i \cot(\omega/cL) + \Omega}{\left[\Omega - \rho c i \cot(\omega/cL)\right]^2 + \rho^2 c^2 \sin^{-2}(\omega/cL)} \right\} f_0(r) \end{aligned} \tag{4.12}$$

and the contralateral filter

$$\begin{aligned} &G_{\mathrm{contra}}(r,\omega;t) \\ &= \int_S f_0(r) \ \mathrm{d}S \ \left\{ \frac{-\rho c i [\sin(\omega/cL)]^{-1}}{\left[\Omega - \rho c i \cot(\omega/cL)\right]^2 + \rho^2 c^2 \sin^{-2}(\omega/cL)} \right\} f_0(r) \end{aligned} \tag{4.13}$$

where ρ denotes density of air in the internal cavity. The other parameters have been introduced in the previous section.

The internal time difference (iTD) for internally coupled ears is then given by

$$\begin{aligned} &\text{iTD} \\ &:= \arg\left[\dot{u}_0(r;t)/\dot{u}_L(r;t)\right]/\omega \\ &= \arg\left\{\frac{[\rho ci\cot(\omega/cL)-\Omega]\ p_{ex_0}-\rho ci[\sin(\omega/cL)]^{-1}\ p_{ex_L}}{[\rho ci\cot(\omega/cL)-\Omega]\ p_{ex_L}-\rho ci[\sin(\omega/cL)]^{-1}\ p_{ex_0}}\right\}/\omega \end{aligned} \tag{4.14}$$

where the first equality is a definition and the second follows from (4.11) to (4.13). Furthermore, the internal amplitude difference (iAD) for internally coupled ears is in a similar way given by

$$\begin{aligned} &\text{iAD} \\ &:= 20\log\left[|\dot{u}_0(r;t)/\dot{u}_L(r;t)|\right] \\ &= 20\log\left\{\left|\frac{[\rho ci\cot(\omega/cL)-\Omega]\ p_{ex_0}-\rho ci[\sin(\omega/cL)]^{-1}\ p_{ex_L}}{[\rho ci\cot(\omega/cL)-\Omega]\ p_{ex_L}-\rho ci[\sin(\omega/cL)]^{-1}\ p_{ex_0}}\right|\right\}\ . \end{aligned} \tag{4.15}$$

For internally coupled ears iTDs and iADs are responses of the whole acoustical system and not simply a direct reflection of the vibrations of the incoming ITDs and IADs.

4.2.3 Empirical analysis

In the next step, for every interaural distance L best frequencies both for independent and internally coupled ears are estimated by maximizing amplitude differences between the tympanic vibrations.

Since best frequencies for animals with independent ears are above the region where ITDs are used, Eq. (4.10) determines the frequency where calculated IADs become maximal for every interaural distance L; see Fig. 4.3, top. This figure also presents literature data for interaural distance and best frequency for a number of species with independent ears; see Table 4.1. Linear regression analysis of the literature data reveals a slope of around -0.5 with R^2 value of 0.4; see Table 4.2. The theoretical curve cannot reproduce the experimentally found regression function, rather it consistently underestimates the best frequency. Since the presented model only includes the acoustical system and its processing through the tympani, experimental data for animals with independent ears could reflect substantial middle ear and neuronal processing, which could also explain the greater level of variation (lower R^2 value) in this data set.

Despite vanishing IADs within the input function, internally coupled ears create iADs for higher frequencies as illustrated in Fig. 2.12. The animal can therefore use iTDs and iADs depending on the frequency of the incoming sound; cf. Fig. 2.21. Similar to animals with independent ears, for every interaural distance L the frequency where iADs (4.15) become maximal provides an estimation of the best frequency, viz., the frequency with the lowest hearing threshold (cf. Fig. 4.2), of the system. Figure 4.3 (bottom) shows the resulting theoretical best frequencies together with literature data for interaural distance

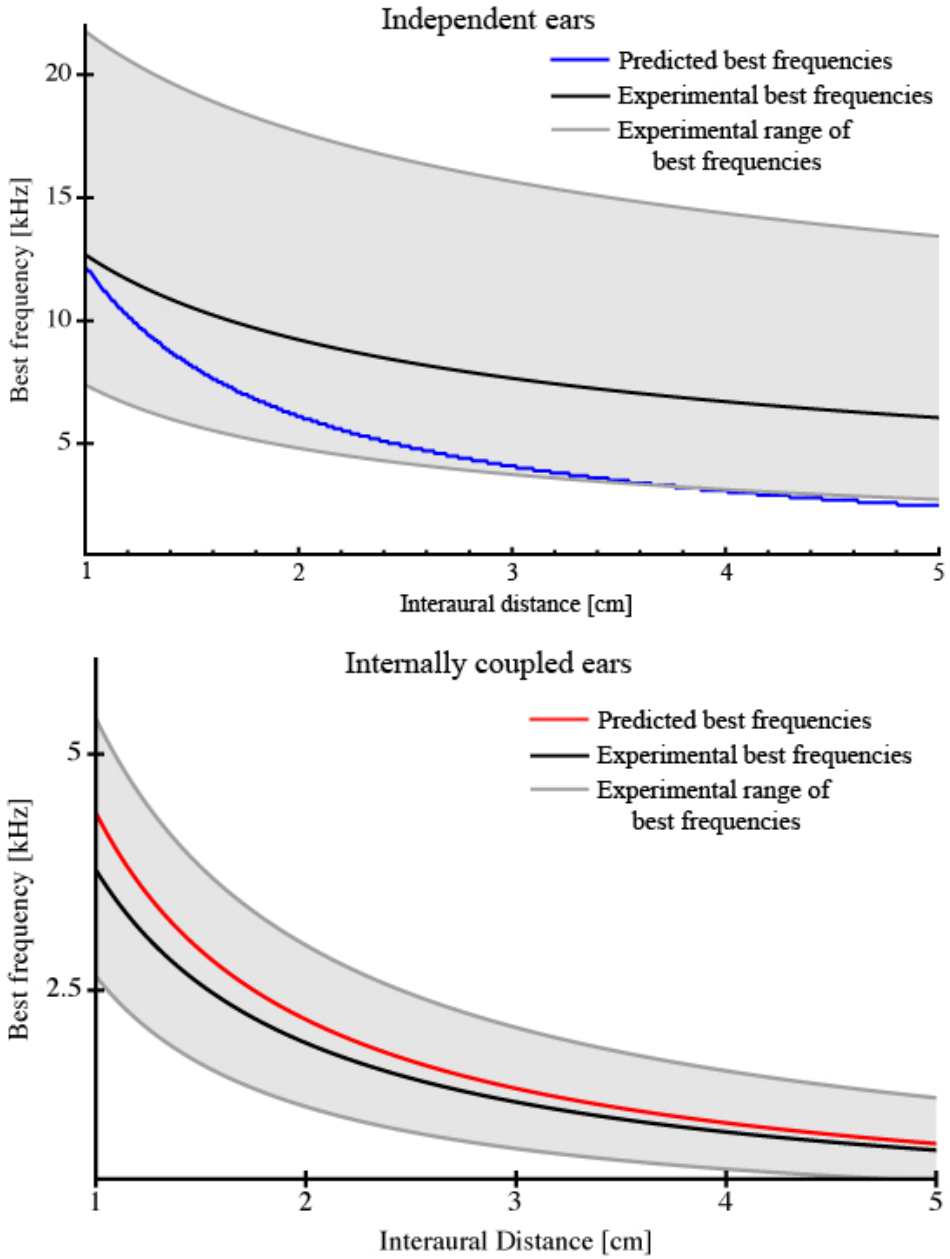

Figure 4.3: *Best frequencies [kHz] (vertical axis) as a function of the interaural distance [cm] (horizontal axis) for animals with independent ears (top) and internally coupled ears (bottom). The regression functions as calculated from experimentally measured best frequencies are shown in black; cf. Table 4.2. The gray curves and areas show possible statistical variations of the curve resulting from different slopes; cf. Table 4.2. Top: For independent ears the blue curve shows best frequencies as resulting from maximizing IADs* (4.10)*. The mismatch between predicted and experimental data probably results from middle ear and neuronal processing that are not included in the presented purely acoustical model. Bottom: For internally coupled ears the red overlaying curves show best frequencies as resulting from maximizing iADs* (4.15)*. Predicted best frequencies nicely fit to the experimentally obtained range of best frequencies.*

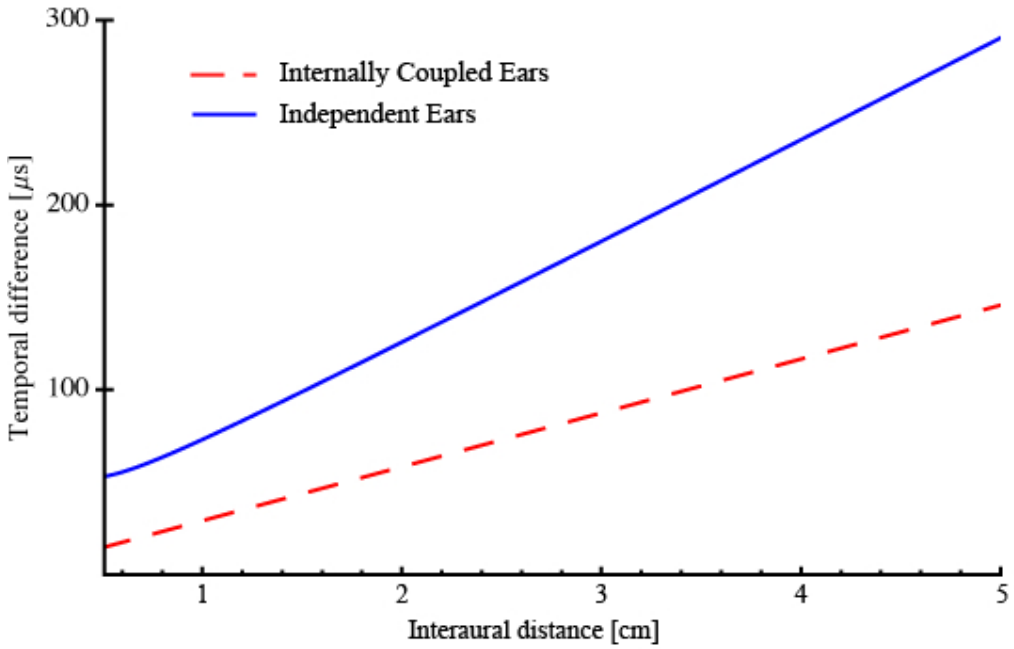

Figure 4.4: *Maximal temporal differences [μs] (vertical axis) as a function of the interaural distance [cm] (horizontal axis) for animals with independent ears (dashed red line) and internally coupled ears (solid blue line). Values are calculated by Eq.* (4.9) *for ITDs of independent ears and Eq.* (4.14) *for iTDs of internally coupled ears. Depending on the interaural distance the coupling of the tympani enhances time differences starting from a factor of around* 3.5 *for an interaural distance of* 0.5 *cm to a factor of around* 2 *for a head size of* 5 *cm. The internal coupling of the tympani therefore effectively causes a magnification of interaural distance.*

and best frequency for a number of species with ICE; see Table 4.1. Linear regression analysis of the literature data reveals a slope of around -1 with R^2 value of 0.8. Figure 4.3 (bottom) illustrates the good fit between the theoretical data and this experimental data set.

Comparing data of internally coupled and independent ears, the calculated regressions for the two literature data sets, and for the two theoretical curves, are significantly different (cf. Table 4.2), reflecting the underlying different biophysics of these two hearing systems. In Fig. 4.3 the horizontal axes are the same, whereas the vertical axes are different. These different vertical axes may obscure one of the key differences between these two auditory systems; animals with internally coupled ears have lower best frequencies than animals with independent ears, often by a factor of $2-4$. The system of internally coupled ears may ensure an animal's capabilities of hearing low frequencies despite a small head size.

The presented theoretical model of the tympanic vibrations predicts an enhancement of the time differences between the membrane vibrations for a system with ICE in comparison to independent ears. Figure 4.4 compares maximal ITDs for independent ears and iTDs for internally coupled ears. Depending on the interaural distance, the enhancement factor lies between 2 and 3.5. The internal coupling of the tympani therefore effectively causes a

magnification of interaural distance, so that best frequencies of systems with internally coupled ears are consistently, and significantly, below those of similar-sized organisms with independent ears; see Table 4.1. Calford [17] has documented these time differences experimentally and reported an enhancement of the same magnitude predicted by the presented biophysical analysis.

4.3 Conclusion

In principle, internally coupled ears (ICE) translate an external signal arriving with interaural time differences (ITDs) and interaural amplitude differences (IADs) at both ears into coupled vibrations of the two tympanic membranes with internal time differences (iTDs) and internal amplitude differences (iADs), the latter two being the result of both the external signal and the internal coupling. Because of the coupling of the tympanic membranes the iTDs are, according to analytical calculations and experiments [17], enhanced in comparison to the incoming ITDs by a factor depending on interaural distance, of two to three for low frequencies; see Fig. 4.4.

For animals with ICE there is both a theoretically and empirically significant inverse relation between functional head size and best frequency, reflecting the limitations resulting from the wavelength of the signal. The slope of this relation is nearly twice as steep in animals with internally coupled ears (-1) as compared to animals with independent ears (-0.5), a difference that is significant. The different regression functions suggest a different sound processing for animals with independent and internally coupled ears. Animals with internally coupled ears have enhanced capabilities to hear low frequencies.

4.A Methods: Empirical analysis

For the empirical analysis interaural distance L and best frequency f_{best} for a number of different animals are collected. Table 4.1 summarizes experimental data [3, 12, 17, 27, 44, 48, 87, 162, 192]. The animals presented have either independent or internally coupled ears, as indicated. The sound propagation velocity used for the calculation of the functional head size L/c is $c = 343$ m/s when transmission is through the air at 20° C and $c = 1483$ m/s when the transmission is through water at 20° C.
To cope for the expected inverse relation $\exp(a)\, L^m = f_{\text{best}}$ between interaural distance L and best frequency f_{best} the equation is linearized in the form

$$\log f_{\text{best}} \quad = \quad m \log L + a \tag{4.16}$$

with slope m and the intercept parameter a. Since the measured data points are blurred by noise they fulfill a given relation only within a certain error range. To find optimal estimates for the slope and the intercept a minimization of the expectation value $\langle \log(f_{\text{best}}) - m \log(l) + a \rangle$ with respect to m and a is necessary. Equivalently, minimization of the sum of squared residuals $\sum_{i=1}^{N} [\log(f_{\text{best}_i}) - m \log L_i - a]^2$ determines m and a. This method is called a least-square linear regression analysis [235]. A measure for the quality of the fit is the R^2 of the regression. Let $\widehat{f_{\text{best}_i}}$ be the estimated value of f_{best_i} as computed

through a fit by means of (4.16) and the sample average by $\overline{f}_{\text{best}} = N^{-1} \sum_{i=1}^{N} f_{\text{best}_i}$. The R^2 of the regression, sometimes called coefficient of determination, is defined as

$$R^2 := \frac{\sum_{i=1}^{N} \left(\widehat{f}_{\text{best}_i} - \overline{f}_{\text{best}} \right)^2}{\sum_{i=1}^{N} \left(f_{\text{best}_i} - \overline{f}_{\text{best}} \right)^2} \in [0; 1]. \tag{4.17}$$

The R^2 measures the ratio of the explained variation compared to the total variation, i.e., the fraction of the sample variation in f_{best} that is explained by L. The R^2 equals 1 for a perfect and 0 for a poor fit. An additional t-test on the estimation results further evaluates a hypothetical relation between two variables, that is, whether the regression coefficients are likely to be zero. The t-test therefore analyzes whether there is a statistically significant relation between the explaining and the explained variable. The result of the t-test, called p-value, gives the likelihood that the coefficients m and a are equal to zero. It is common to define a relation to be statistically "significant" for p-values of 0.05 or 0.01, corresponding to a 5% or 1% chance of an outcome like the observed one without a correlation between the variables. For details on regression analysis see, e.g., Wooldridge [235].

Table 4.1: *Overview of different animals with their interaural distance L in [m], their functional head size L/c in [μs] with the speed of sound c in the surrounding medium, and their best frequencies in [kHz] taken from papers as indicated in the 4th column. The animals have independent or internally coupled ears, as indicated in the 5th column. For the porpoise and the alligator in water the sound propagation velocity is taken to be 1483 m/s. Otherwise the sound propagation velocity in air is 343 m/s.*

animal	interaural distance [m]	functional head size [μs]	best frequency [kHz]	internal coupling
budgerigars	0.01	43.7	2.86 [12]	yes
zebra finch (owl)	0.011	32.1	4.0 [48]	yes
skink	0.013	37.9	3.2 [27]	yes
agamid	0.013	37.9	3.19 [27]	yes
pigeon	0.015	58.3	2.0 [17]	yes
grass frog	0.019	55.4	1.1 [48]	yes
kestrel (owl)	0.02	58.3	1.5 [17]	yes
chick	0.021	61.2	1.5 [192]	yes
gekkonid	0.022	64.1	1.82 [27]	yes
alligator (water)	0.025	167.9	0.8 [87]	yes
tree frog	0.03	87.5	0.9 [48]	yes
leopard frog	0.03	87.5	1.1 [162]	yes
iguanid	0.032	93.3	1.95 [27]	yes
bullfrog	0.075	218.7	0.6 [48]	yes
cotton rat	0.015	42.3	8.0 [48]	no
mouse	0.018	52.5	15.0 [48]	no
gerbil	0.03	87.5	4.0 [48]	no
rat	0.05	145.8	8.0 [48]	no
rabbit	0.095	277.0	10.0 [48]	no
chinchilla	0.1	291.5	2.0 [48]	no
porpoise (water)	0.12	80.9	8.0 [3]	no
cat	0.15	437.3	8.0 [48]	no
dog	0.20	568.5	3.0 [48]	no
sheep	0.2	583.1	10.0 [48]	no
human	0.21	612.2	1.5 [48]	no
cow	0.295	860.1	8.0 [48]	no
horse	0.3	874.6	2.0 [48]	no
elephant	1.3	3790.1	1.0 [48]	no

Table 4.2: *Regression analysis for the best frequency in dependence upon the functional head size (i.e., interaural distance L divided by propagation velocity of sound c) for animals with independent and internally coupled ears as given in Table 4.1. Slope m and the intercept parameter a with a given confidential rate and standard errors have been calculated by minimizing the expectation value $\langle \log(f_{\text{best}}) - m \log(l) + a \rangle$. The calculations were performed only for internally coupled ears (left column), for animals with independent ears (column in the middle), and for the two groups together (right column). The dummy variable C equals 0 for independent ears and 1 for internally coupled ears. It indicates to which degree the two populations differ. The value R^2 is the fraction of the sample variation in f_{best} that is explained by the functional head size L/c; cf. Eq.* (4.17)*. That is $R^2 = 1$ would be a perfect, $R^2 = 0$ a poor fit. The p-value further gives the likelihood that the coefficients m and a are equal to zero, i.e., whether there is a statistically significant relation between the explaining and the explained variable. P-values of 0.05 or 0.01 indicate statistically "significance", corresponding to a 5% or 1% chance of an outcome like the observed one without a correlation between the variables. The results show a statistically significant inverse relation between functional head size and best frequency for animals with internally coupled ears. In contrast best frequencies of animals with independent ear show an inverse dependence upon the functional head size with an exponent around -0.5. The high values of R^2 and the low p-values underline the goodness and reliability of the fit. The difference between the two groups is statistically significant.*

Variables	Internally Coupled Ears	Independent Ears	both
m	-0.97***	-0.46**	-0.52***
	(0.16)	(0.16)	(0.11)
C			-1.8***
			(0.3)
a	11.5***	11.0***	11.4***
	(0.7)	(0.9)	(0.7)
R^2	0.8	0.4	0.7

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

All our knowledge has its origins in our perceptions.

Leonardo da Vinci (1452 - 1519)

5. Optimality in mono- and multisensory map formation

In the struggle for survival in a complex and dynamic environment, nature has developed a multitude of sophisticated sensory systems. To exploit the information provided by these sensory systems, higher vertebrates reconstruct the spatio-temporal environment from each of the sensory systems they have at their disposal. That is, for each modality the animal computes a neuronal representation of the outside world, a monosensory neuronal map. The here presented universal framework allows for the calculation of the specific layout of the involved neuronal network by means of a general mathematical principle, viz., *stochastic optimality*. A step-by-step tutorial illustrates how to apply the theoretical framework to concrete situations. That is, given a known physical signal transmission and rudimental knowledge of the detection process, the approach allows to predict neuronal properties of biological systems. Finally, information from different sensory modalities has to be integrated so as to gain a unified perception of reality for further processing, e.g., for the creation of distinct motor commands. Concepts of multimodal interaction and the evolvement of a multimodal space by alignment of monosensory maps are briefly discussed.

5.1 Introduction

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, for instance, allows for directional motor commands is an essential concept for surviving. A neuronal representation of the external world is denoted as neuronal *map*. Depending on the map processing information from one or many sensory systems, the map is called *uni-* or

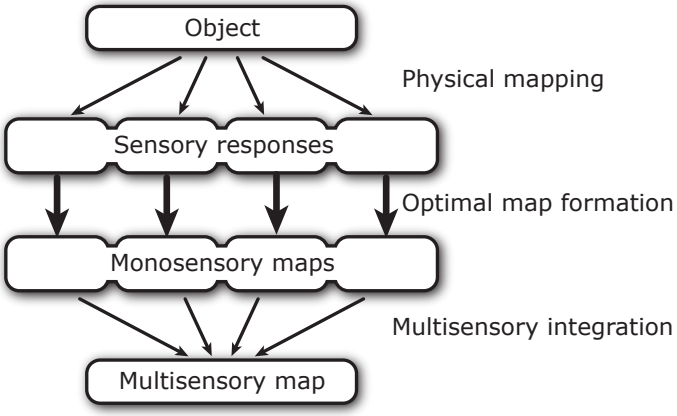

Figure 5.1: *The three steps of sensory processing leading to the formation of a unified multisensory map. An object in the outside world generates physical input signals, which can be detected by different sensory systems. To form a map, the physical mapping must be "inverted" in some suitable way. After monosensory map formation, the distinct maps are combined into a unified multisensory map.*

multimodal. The advantageous concept of a neuronal map is discussed in detail in the next section.

The processing of sensory information, from its generation to multimodal map formation, can be subdivided into the "golden three" of sensory processing: physical mapping, optimal map formation, and multimodal integration; see Fig. 5.1.

Physical Mapping. An object in the outside world reveals its presence by generating different signals that are transmitted along distinct physical pathways. A running animal may, for instance, generate sound and a changing visual image as it moves, as well as vibrations and an infrared profile. In concrete terms, given any signal that varies as a function of spatial position and time, it is possible to calculate the time-dependent response of the receptor neurons. That is, the physical mapping of the signal onto the neuronal detector response can be described by a set of *transfer functions*, indicated by the upper arrows in Fig. 5.1. The responses of the sensory systems then represent particular physical quantities such as sound, light intensity, volatile molecules, or heat originating from the object.

Optimal map formation. From the sensory responses an observer needs to reconstruct a map that represents the spatio-temporal stimulus. For some sensory systems a pre-stage map already exists inherently, e.g., for the visual system on the retina. For other systems, such as the auditory system [9, 21, 22, 111, 136], spatial information is not readily available and a neuronal map must be constructed more explicitly. In either case, the map has to represent the environment as accurately as possible, that is, *optimally*. The task of the brain is to obtain an optimal reconstruction of the signal; cf. Fig. 5.1, middle arrows. The key to success is the choice of the right neuronal connections between the sensory systems and the corresponding maps. That is, the synaptic connections have to be adjusted in

such a way that the network "inverts" the physical mapping of the signal to the sensory response [168, 210, 241]. The firing activity of the map neurons then accurately represents the spatio-temporal signal.

Multimodal integration. In a final step of sensory processing the monosensory maps merge into a single unambiguous multisensory map. Here two difficulties arise. First, the successful fusion of unimodal maps requires proper map alignment as observed, for example, in the *superior colliculus* (SC)[1] [18, 112, 204]; see Sec. 5.2. Second, the monosensory maps should be combined optimally [74, 158] to maximize the quality of the integrated map in comparison to that of the contributing maps. Besides this "integration of information" the multimodal map together with the aligned unimodal maps allows for a new concept in multimodal processing, the so-called "pooling" of information, an efficient way to identify and characterize an object by its sensory properties.

The first processing step, physical mapping, is a purely physical description of the signal generation and detection process. For internally coupled ears, the ICE model describes the process of physical mapping of auditory stimuli; cf. Chap. 2 and [221]. This chapter therefore focuses on optimal map formation and touches multimodal integration superficially; a detailed model of multimodal map alignment is presented in Chap. 6. Section 5.2 provides a review of the concept of a neuronal map. Section 5.3 discusses a general framework that describes how a neuronal map can emerge from a given sensory input in a *stochastically optimal* way [160]. After introducing in Sec. 5.4 a "recipe" how the presented framework can be applied to realistic situations, Sec. 5.5 treats the integration of monosensory into multisensory maps and reviews the current literature from the perspective of maps. Basic concepts such as "integration" and "pooling" of information are presented. The final section addresses the question of how a common sensory space can develop at all.

5.2 Fundamental concept of neuronal maps

A major role in sensory processing is reserved for *maps* [124, 217]. A neuronal map is a neuronal representation of the external world realized by a topographically arranged array of neurons. Neighboring map neurons respond to similar sensory stimuli.

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") [107, 227]. Here neighboring neurons respond to visual input from neighboring points in space. Such *spatial* maps have been discovered in various sensory systems in many groups of vertebrates [23, 43, 91, 114, 116, 119, 151, 202, 207].

One might argue that neuronal maps exist simply because their neuronal architecture only 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.

[1] The SC is called *optic tectum* in non-mammals. Here "SC" simply refers to either the optic tectum or the superior colliculus, depending on the context.

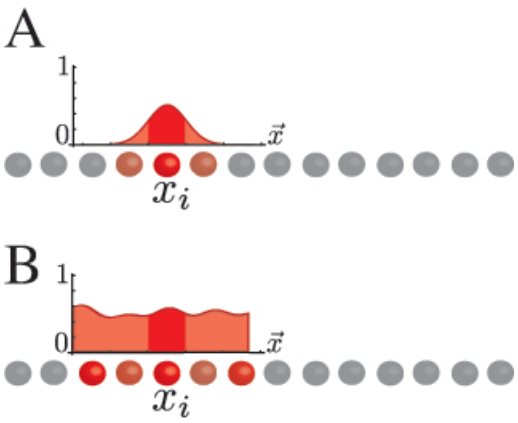

Figure 5.2: *The firing profile of a map encodes the likelihood of finding an object at a certain position in sensory space. Focusing on the firing rate of only a* single *neuron such as neuron x_i and ignoring its neighbors prevents a faithful perception of reality. A faithful perception can only be achieved if the activity of the whole neuronal map is taken into account, that is, just compare the 'neighborhood' of x_i with the rest. Consequently, even though the nominal value of the firing rate of neuron x_i is identical in case A and B, the represented physical reality differs significantly in both cases.*

This argument, however, does not hold for every sensory map, in particular, not for auditory maps of interaural time and amplitude differences [24, 147, 167, 209]. In this case it is certainly *not* straightforward to build a map; cf. Chap. 3.

A key question [217] therefore is: What is the *function* of a neuronal map? That is, why choose a map structure for neuronal processing? One argument is that, in contrast to *arbitrary* population coding, neuronal maps ensure a topographic neuronal organization. This organization then underlies the neuronal processing and provides an efficient representation of a continuously varying input signal. For instance, it allows for the interpretation of a firing pattern on a spatial map as the *likelihood* to find a sensory object at a certain position [41, 98, 174, 195, 217]. As illustrated by Fig. 5.2, in the map perspective one needs to consider the activity of the complete map in order to retrieve meaningful information from the firing rate of a single neuron.

Nelson and Bower [163] suggested a justification of maps based on their function by comparing computational principals of the brain and parallel computers. Accordingly, they distinguished three types of maps.

- *Continuous maps* are topographically arranged neuronal arrays that represent a continuous parameter. Interaction within the map mainly takes place between neighboring neurons. An exemplary continuous map is found in the somatosensory cortex of rats [228]. Here neurons are locally interconnected through shortrange axonal arborization of stellate cells. Continuous maps are used for spatial filtering and local feature extraction. The local connectivity ensures a balancing of the load.

- *Scattered or discrete maps* are characterized by a lack of systematic structure, i.e.,

non-topographic arrangement of the neurons. The used definition of a map therefore excludes scattered maps. A neuronal example is the olfactory cortex of rats [76] since the olfactory inputs do not provide a topographic order. Consequently an extensive network of fibers interconnects all cortical regions.

- *Patchy maps* are intermediate between continuous and scattered maps. The interaction within the map has both a local and a global component. The somatosensory map within the cerebellar cortex of rats [196] constitutes a neuronal example of a patchy map. The neuronal structure includes long-distance parallel fiber connections as well as strong local influence through shortrange axons. The task of this brain area is the analysis of local sensory information within a more global sensory context.

In summary, structure and function of a map are highly correlated. For each computational unit the load-balance should be optimized with respect to the given task. Within a neuronal context the limited capacity of the supply network requires such a load-balancing since oxygen and glucose are transported to the cells via supply networks of small capillaries. A high, local increase in tissue metabolic activity leads to an undersupply of the neurons. Brain areas that perform, e.g., local feature extraction are therefore organized in continuous maps with strong inter-neighbor connections to distribute computational load equally.

The real computational power of neuronal maps, however, can only be appreciated when the *interplay* of several maps is considered. Here the SC is a beautiful, well-studied example of a collection of different maps of sensory systems that provide spatial information in a map-like form [18, 204]. The SC contains multisensory, predominantly monosensory, as well as motor maps, i.e., motor neurons organized in a map-like structure. All neuronal maps are mutually aligned to gain a unified multisensory representation of sensory space [18, 112, 204]. The combined sensory information can then, for instance, activate motor maps and generate directional motor responses [138, 203, 219]. Direct evidence for this hypothesis has recently been found in eye tracking experiments [77].

Moreover, external *objects* can be *identified* by their position encoded through the firing pattern in a neuronal map. More concretely, the position serves as appropriate and necessary information for defining a sensory object. When combining different sensory systems, the spatial information is needed to bind information associated with the same sensory object into one single multimodal percept for further processing.

5.3 Mathematical model

The mathematical model for optimal stimulus reconstruction derived in the following is based on the initial division of sensory processing into three major steps (see Fig. 5.1) but focuses on optimal map formation. That is, stimulus characteristics have to be extracted at the best from the sensory response as described in Sec. 5.1. Mathematically the *inverse transfer function* is needed that can perform an optimal reconstruction of a particular stimulus from the sensory response. This inverse transfer function can then be translated into a neuronal connectivity pattern.

The derivation below is based on two reasonable simplifications. First, all sensory maps are assumed purely *monosensory* which reflects that many spatial maps are clearly dominated by a single sensory modality [226].

The second assumption is 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. Nonlinear relations between the stimulus and the detector responses, e.g., a logarithmic response [34, 102, 132, 134, 166], can in principle be treated with the presented model as well (see Appendix 5.A for details) but are excluded in the following.

5.3.1 Definition of the problem

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

In the next step, 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 $1 \leq i \leq N$ can be a complete sensory organ, such as the left ear, or a part of a detector array such as a specific interval of best frequencies in the cochlea. In principle, the detector combines information from past signals within the whole sensory space. The response is therefore described by

$$r_i(t) = \int_{\text{all space}} \mathrm{d}\mathbf{x} \int_{-\infty}^{t} \mathrm{d}\tau \; s^{\mathbf{x}}(\tau) h_i^{\mathbf{x}}(t-\tau) \tag{5.1}$$

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. Transfer functions within the auditory system, 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, $h_i^{\mathbf{x}}(t) = 0$ for large values of $|\mathbf{x}|$ and t. This property reflects the intuition that events occurring far away or long ago do not influence the state of a sensor. Moreover, since any detector can only react to temporal-causal, i.e., past signals, it is $h_i^{\mathbf{x}}(t) = 0$ for $t < 0$. The response function (5.1) with adapted integration limits then transforms to a convolution [see (5.3) and box for definition and further information] with respect to time,

$$\begin{aligned} r_i(t) &= \int \mathrm{d}\mathbf{x} \int_{-\infty}^{\infty} \mathrm{d}\tau \; s^{\mathbf{x}}(\tau) h_i^{\mathbf{x}}(t-\tau) \\ &=: \int \mathrm{d}\mathbf{x} \; (s^{\mathbf{x}} \star h_i^{\mathbf{x}})(t) \; . \end{aligned} \tag{5.2}$$

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

Convolution ($\star$) and autocorrelation ($\circ$)
for arbitrary functions $a(t)$ and $b(t)$ are defined by

$$(a\star b)(t) \;:=\; \int_{-\infty}^{\infty} \mathrm{d}\tau\; a(t-\tau)b(\tau) \tag{5.3}$$

and

$$(a\circ b)(t) \;:=\; \int_{-\infty}^{\infty} \mathrm{d}\tau\; a(t+\tau)b(\tau)\;. \tag{5.4}$$

Useful properties are

$$a\star b \;=\; b\star a\;, \tag{5.5}$$
$$a\star(b\star c) \;=\; (a\star b)\star c\;, \tag{5.6}$$
$$(a\circ b)(t) \;=\; (b\circ a)(-t)\;. \tag{5.7}$$

For a Fourier transformation $\mathcal{F}$, in particular, it is

$$\mathcal{F}(a\star b) \;=\; \mathcal{F}(a)\mathcal{F}(b)\;, \tag{5.8}$$
$$\mathcal{F}(a\circ b) \;=\; \overline{\mathcal{F}(a)}\mathcal{F}(b)\;. \tag{5.9}$$

Compositions of convolution and autocorrelation can be calculated as

$$\begin{aligned}(a\star b\star c)(t) &= \int_{-\infty}^{\infty} \mathrm{d}\tau\; (a\star b)(t-\tau)c(\tau)\\ &= \int_{-\infty}^{\infty} \mathrm{d}\tau\;\mathrm{d}\tau'\; a(t-\tau-\tau')b(\tau')c(\tau)\;,\end{aligned} \tag{5.10}$$

$$\begin{aligned}[(a\star b)\circ c]\,(t) &= \int_{-\infty}^{\infty} \mathrm{d}\tau\; (a\star b)(t+\tau)c(\tau)\\ &= \int_{-\infty}^{\infty} \mathrm{d}\tau\;\mathrm{d}\tau'\; a(t+\tau-\tau')b(\tau')c(\tau)\;.\end{aligned} \tag{5.11}$$

First, information may get lost during the transfer from an object of the external world to the detecting sensory system. Second, noise influences all steps in the detection and reconstruction process [47]. Third, limitations of the neuronal hardware, for instance, the limited dynamic range of receptors, constrain possible solutions; see Sec. 5.3.5 for details.

The mathematical model incorporates these three restrictive factors as additional noise terms. Accordingly, a term describing background noise $\xi^{\mathbf{x}}(t)$ must be added to the signal. Transfer function and sensory response are hampered by additional noise terms $\eta_i^{\mathbf{x}}(t)$ and $\chi_i(t)$. Consequently (5.2) is modified for a biological system so as to read

$$r_i(t) = \int \mathrm{d}\mathbf{x}\;[(s^{\mathbf{x}}+\xi^{\mathbf{x}})\star(h_i^{\mathbf{x}}+\eta_i^{\mathbf{x}})](t)+\chi_i(t)\;. \tag{5.12}$$

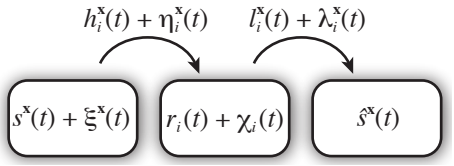

Figure 5.3: *Illustration of functions and entities involved in the process of optimal stimulus reconstruction.* Physical mapping*: signal $s^{\mathbf{x}}(t)$ with background noise $\xi^{\mathbf{x}}(t)$ is mapped onto a noisy receptor response $r_i(t) + \chi_i(t)$ through the noisy transfer function $h_i^{\mathbf{x}}(t) + \eta_i^{\mathbf{x}}(t)$.* Optimal map formation*: the application of the (possibly noisy) inverse transfer function $l_i^{\mathbf{x}}(t) + \lambda_i^{\mathbf{x}}(t)$ provides an estimate $\hat{s}^{\mathbf{x}}(t)$ of the original 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 5.1: *Functions and error terms describing detection and processing of sensory information.*

To reconstruct an estimated signal from the detector responses $r_i(t)$, the above transformation must be "inverted" in some appropriate way. The time-dependent inverse transfer function between detector i and the map at position $\mathbf{x}$ is calculated as $l_i^{\mathbf{x}}(t)$. Application of $l_i^{\mathbf{x}}(t)$ to the receptor responses at i leads to the estimate $\hat{s}^{\mathbf{x}}(t)$ of the original signal $s^{\mathbf{x}}(t)$ given by

$$\hat{s}^{\mathbf{x}}(t) = \sum_i [r_i \star (l_i^{\mathbf{x}} + \lambda_i^{\mathbf{x}})](t) \tag{5.13}$$

where the term $\lambda_i^{\mathbf{x}}(t)$ represents the noise due to the concrete realization of the theoretical inverse transfer function. In contrast to elsewhere [168, 175] the present model is non-iterative so that the neuronal realization results in a purely feedforward network structure as described in Sec. 5.3.5.

Figure 5.3 summarizes the whole mathematical procedure of sensory information processing, corresponding to the first two steps of Fig. 5.1. All the relevant terms are summarized in Table 5.1. The next section indicates how to calculate inverse transfer functions $l_i^{\mathbf{x}}(t)$ that enable optimal signal reconstruction.

5.3.2 Optimal reconstruction

Sensory system are tuned 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 are denoted 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. An optimal reconstruction therefore requires to minimize the *expectation value* of the squared difference between signal and reconstruction.

This is possible because all quantities and functions (cf. Figs. 5.1 and 5.3) involved in both the process of physical mapping (see Appendix 5.B) and the neuronal process of optimal map formation (cf. Sec. 5.3.5) are *self-averaging*. The mathematical definition of self-averaging allows for a description in terms of expectation values.

The consequent derivation of the inverse transfer functions $l_i^{\mathbf{x}}(t)$ that enable optimal signal reconstruction for a class of typical signals requires the minimization of the expectation value of the squared error between estimated and real signal

$$\begin{aligned} E\left[\mathbf{l}^{\mathbf{x}}(t), t\right] &:= \left\langle \int_{t-T}^{t} \mathrm{d}t' \int \mathrm{d}\mathbf{x} \;\; \left[s^{\mathbf{x}}(t') - \hat{s}^{\mathbf{x}}(t')\right]^2 \right\rangle \\ &= \int_{t-T}^{t} \mathrm{d}t' \int \mathrm{d}\mathbf{x} \;\; \left\langle \left[s^{\mathbf{x}}(t') - \hat{s}^{\mathbf{x}}(t')\right]^2 \right\rangle \end{aligned} \tag{5.14}$$

with $\mathbf{l}^{\mathbf{x}}(t)$ the vector of $l_i^{\mathbf{x}}(t)$. Here the brackets $\langle . \rangle$ denote the expectation value with respect to the different types of noise. The typical processing time is denoted by T.

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; see van der Waerden [222].

Minimizing the mean squared error means to minimize both the variance of the estimate and the systematic shift between estimate and expectation value of the signal. For further details see van der Waerden [222]. In addition, a quadratic form of the error term has been proven to be a reasonable and practical choice in many physical optimization problems [152]. 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 [101, 108]; see Sec. 5.3.4.

Mathematically, the error (5.14) 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 [31, 63, 105, 215]. For the present situation the first variation with respect to every inverse transfer function $l_j^{\mathbf{x}}(t')$ is to vanish. That is,

$$\frac{\partial \left\langle \left[s^{\mathbf{x}}(t') - \hat{s}^{\mathbf{x}}(t')\right]^2 \right\rangle}{\partial\, l_j^{\mathbf{x}}(t')} = 0 \qquad \text{for every } j. \tag{5.15}$$

Solving Eq. (5.15) requires to substitute (5.13) for the estimate $\hat{s}^{\mathbf{x}}(t)$ and replace $r_i(t)$ by its description (5.12). Expanding the square produces expectation values of products consisting of varying combinations of noise and signal terms. Here all noise terms as well as the signal itself are assumed to be stochastically independent of each other so that the expectation of a product of independent terms factorizes, for instance,

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

For a product consisting of the same kind of term, the autocorrelation of an arbitrary quantity $f^{\mathbf{x}}(t)$ is given by

$$\left\langle f^{\mathbf{x}}(t) f^{\mathbf{x}'}(t') \right\rangle = \delta(\mathbf{x}-\mathbf{x}')\delta(t-t')(\mu_f^2+\sigma_f^2) \tag{5.16}$$

with μ_f the mean and σ_f the variance of the quantity $f^{\mathbf{x}}(t)$. That is, the values for different spatio-temporal positions are in a first step completely uncorrelated, a kind of worst-case scenario.
Given that the means of all noise terms μ_f vanish the autocorrelations for the noise terms reduce to

$$\left\langle \xi^{\mathbf{x}}(t)\xi^{\mathbf{x}'}(t') \right\rangle = \delta(\mathbf{x}-\mathbf{x}')\delta(t-t')\ \sigma_\xi^2\ , \tag{5.17a}$$

$$\left\langle \chi_i(t)\chi_j(t') \right\rangle = \delta_{ij}\delta(t-t')\ \sigma_\chi^2\ , \tag{5.17b}$$

$$\left\langle \eta_i^{\mathbf{x}}(t)\eta_j^{\mathbf{x}'}(t') \right\rangle = \delta_{ij}\delta(\mathbf{x}-\mathbf{x}')\delta(t-t')\ \sigma_\eta^2 \quad \text{with} \quad |\mathbf{x}| < x^{\max} \ \text{ and } \ 0 < t < t^{\max}\ . \tag{5.17c}$$

The final equation accounts for the fact that the noise $\eta_i^{\mathbf{x}}(t)$ vanishes for large values of t and $|\mathbf{x}|$ in the same way as it does for the transfer function $h_i^{\mathbf{x}}(t)$.

The autocorrelation (5.16) 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 others equal to zero. In the following, choosing the absolute signal strength μ_S of the signal as the appropriate quantity and therefore taking σ_S zero reduces Eq. (5.16) to

$$\left\langle s^{\mathbf{x}}(t) s^{\mathbf{x}'}(t') \right\rangle = \delta(\mathbf{x}-\mathbf{x}')\delta(t-t')\ \mu_s^2\ . \tag{5.18}$$

Whereas (5.17) incorporates reasonable assumptions for all noise terms, the correlation (5.18) 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 moment to the next. A Gaussian correlation term

$$\left\langle s^{\mathbf{x}}(t) s^{\mathbf{x}'}(t') \right\rangle = A \exp\left[-\left|\mathbf{x}-\mathbf{x}'\right|^2/(2\sigma_{s_x}^2)\right] \exp\left[-\left|t-t'\right|^2/(2\sigma_{s_t}^2)\right]\ , \tag{5.19}$$

for instance, can take into account correlations between neighboring points in space and time. Here σ_{s_x} and σ_{s_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; see Appendix 5.D for details. For reasons of clarity, relation (5.18) is used in the following.

To solve (5.15) the application of the correlations (5.17) and (5.18) leads to

$$l_j^{\mathbf{x}}(t)\left[\sigma_\chi^2 + (\mu_s^2+\sigma_\xi^2)\int_{\substack{|\mathbf{y}|<y^{\max}\\ 0<\tau<t^{\max}}} \mathrm{d}\mathbf{y}\mathrm{d}\tau\ \sigma_\eta^2\right] + (\mu_s^2+\sigma_\xi^2)\sum_i \int \mathrm{d}\mathbf{y}\ \left[(h_i^{\mathbf{y}} \star l_i^{\mathbf{x}}) \circ h_j^{\mathbf{y}}\right](-t) = \mu_s^2 h_j^{\mathbf{x}}(-t)\ ; \tag{5.20}$$

for details on the calculations see Appendix 5.C. The open circle $\circ$ denotes the autocorrelation integral as defined in (5.4). To simplify (5.20) two noise measures are introduced. The parameter τ represents an inverse signal-to-noise ratio and is defined by

$$\tau^2 := \frac{\sigma_\xi^2}{\mu_s^2} \,. \tag{5.21}$$

For processing of sensory signals it is reasonable to assume a small value of τ. The second noise parameter σ defined by

$$\sigma^2 := \frac{\sigma_\chi^2}{\mu_s^2} + \int_{\substack{|\mathbf{y}|<y^{\max} \\ 0<\tau<t^{\max}}} \mathrm{d}\mathbf{y}\mathrm{d}\tau \; \frac{\sigma_\eta^2(\mu_s^2+\sigma_\xi^2)}{\mu_s^2} \tag{5.22}$$

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, the value of σ cannot be assumed to be small and has to be adjusted according to the situation at hand.

To further simplify (5.20) the equation is translated to Fourier space, where convolution (5.3) and autocorrelation (5.4) become ordinary multiplications combined with complex conjugations; cf. Eqs. (5.8) and (5.9). Denoting Fourier transforms by capital letters and the complex conjugation by an overline, Eq. (5.20) with the introduced noise measures (5.21) and (5.22) simplifies to

$$\sum_i L_i^{\mathbf{x}} \left[\sigma^2\delta_{ij} + (1+\tau^2)\int \mathrm{d}\mathbf{y}\; H_i^{\mathbf{y}}\overline{H_j^{\mathbf{y}}}\right] = \overline{H_j^{\mathbf{x}}} \,. \tag{5.23}$$

Equation (5.23) is the main result of the presented derivation. In principle, it allows to calculate the inverse transfer functions $L_i^{\mathbf{x}}$ for optimal signal reconstruction. A calculation of the second variation confirms that the obtained inverse transformation indeed minimizes the error; see Appendix 5.C, in particular, Eq. (5.48), for details.

5.3.3 Matrix notation

To rewrite (5.23) in a more convenient notation "matrices" $\mathcal{H}$ and $\mathcal{L}$ are introduced by

$$\mathcal{H}_{[ix]} = H_i^{\mathbf{x}} \qquad \text{and} \qquad \mathcal{L}_{[xi]} = L_i^{\mathbf{x}} \,. \tag{5.24}$$

The notation illustrates 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. $\mathcal{H}$ and $\mathcal{L}$ are therefore 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, leads to a true matrix formulation.

In addition, the covariance matrix $\mathcal{C}(\mathbf{R})$ of the receptor response $\mathbf{R}$ as described, e.g., in [101, 108], is calculated to be

$$\begin{aligned} \mathcal{C}(\mathbf{R}) &:= \left\langle (\mathbf{R}-\langle\mathbf{R}\rangle)\overline{(\mathbf{R}-\langle\mathbf{R}\rangle)}^T \right\rangle \\ &= \mu_s^2\left(\sigma^2\mathbb{I} + \tau^2\overline{\mathcal{H}}\mathcal{H}^T\right) \end{aligned} \tag{5.25}$$

where the superscript T denotes the matrix transpose and $\mathbb{I}$ the identity matrix. Equation (5.23) now simplifies to

$$\mathcal{M}\mathcal{L}^T = \overline{\mathcal{H}} \qquad \text{with the 'model matrix'} \qquad \mathcal{M} := \mu_s^{-2}\mathcal{C} + \overline{\mathcal{H}}\mathcal{H}^T \ . \tag{5.26}$$

Given that $\mathcal{M}$ is an invertible matrix the solution for $\mathcal{L}$ turns out to be

$$\mathcal{L} = \left(\mathcal{M}^{-1}\overline{\mathcal{H}}\right)^T = \overline{\mathcal{H}}^T \left[\left(\mu_s^{-2}\mathcal{C} + \mathcal{H}\overline{\mathcal{H}}^T\right)^{-1}\right]^T \ . \tag{5.27}$$

This equation gives a unique solution for the optimal reconstruction for any given set of transfer functions and noise constants (σ, τ). The estimated signal can be calculated from the measured response vector $\mathbf{R}$ using (5.13) in matrix form as

$$\hat{\mathbf{S}} = \mathcal{L} \cdot \mathbf{R} \ . \tag{5.28}$$

5.3.4 Relation to the maximum-likelihood approach

The challenge of signal reconstruction has a long tradition, and, accordingly, the above formalism should incorporate methods that have been established in this field. The following discussion, in particular, relates the presented model to methods based on the *maximum-likelihood* analysis [101, 108]. Within the maximum-likelihood scheme one computes the stimulus that is the most likely one *given* a set of detector responses $\mathbf{R}$. Experiments have shown that optimal or near-optimal stimulus combinations can indeed describe several phenomena of sensory processing [1, 45, 84, 93, 129, 158]. A method of optimal stimulus combination like the maximum-likelihood approach is therefore highly relevant to neuronal information processing and ought to be included in the present model.

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

$$\mathbf{R} = \mathcal{H}\mathbf{S} + \boldsymbol{\chi} \tag{5.29}$$

with $\boldsymbol{\chi}$ representing the noise. The noise follows a Gaussian distribution with zero mean and the standard deviation σ_χ. The maximum-likelihood method minimizes the noise χ. That is, based on the fundamental definitions of Bayesian statistics, it maximizes the conditional probability density function

$$p(\mathbf{R}|\mathbf{S}) \propto \exp\left[-\frac{1}{2\sigma_\chi^2}(\mathbf{R} - \mathcal{H}\mathbf{S})^T(\overline{\mathbf{R} - \mathcal{H}\mathbf{S}})\right] \tag{5.30}$$

with respect to the signal $\mathbf{S}$. This leads to a linear system of equations

$$\mathbf{S} = \underbrace{\left(\overline{\mathcal{H}}^T\mathcal{H}\right)^{-1}\overline{\mathcal{H}}^T}_{=:\mathcal{L}_{ML}}\mathbf{R} \ . \tag{5.31}$$

Using on the other side the above assumptions for the presented model, viz., $\sigma_s = \infty$, $\eta = 0$, and $\xi = 0$, Eq. (5.26) reduces to

$$\left(\overline{\mathcal{H}}\mathcal{H}^T\right)\mathcal{L}^T \ = \ \overline{\mathcal{H}} \ . \tag{5.32}$$

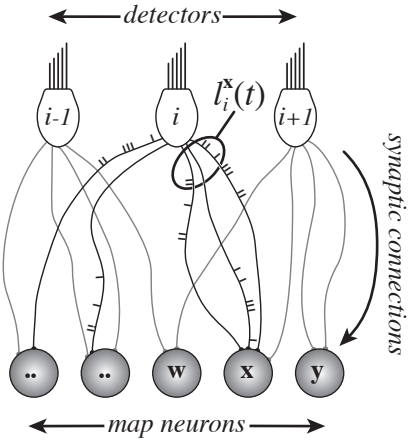

Figure 5.4: *Neuronal realization of optimal stimulus reconstruction. The sensory response of the detector neurons (here hair cells labeled by i) to an external sensory object is projected onto map neurons. The synaptic connections have to be adjusted in such a way that the network "inverts" the physical mapping of the signal to the detector response. That is, the synaptic connections have to represent the inverse transfer functions $l_i^{\mathbf{x}}(t)$. To this end, each sensor connects to several map neurons. Spatial processing is then governed by the topographic structure of the network, that is, which detector i is connected to which map neuron (encoding the location $\mathbf{x}$). Temporal processing is determined by the distribution of delays within the set of connections. Together the inverse transfer functions $l_i^{\mathbf{x}}(t)$ can be reliably represented in the neuronal network.*

To test whether the two filters are equal, $\mathcal{L}_{ML}$ is inserted into (5.32). Application of the transposition rules shows that with the used assumptions it is $\mathcal{L} = \mathcal{L}_{\mathcal{ML}}$. Under the made assumptions the two strategies are therefore identical; for details see [101, 108, 182, 190].

5.3.5 Neuronal realization of the model

The general mathematical algorithm of optimal stimulus reconstruction is now translated into a concrete neuronal context. One therefore has to verify first whether the assumptions made in Sec. 5.3.2 are fulfilled in neuronal processing. That is, neuronal quantities and functions of optimal map formation have to be *self-averaging*. As mentioned before, quantities and functions within the physical mapping process are indeed self-averaging; see Appendix 5.B. For the neuronal side this assumption is as well fulfilled since on the one hand firing of neurons is correlated with the self-averaging input, and neuronal noise can be described by a stochastic process, e.g., a Gaussian one; as discussed in a minute. The mathematical 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 [109]. In conclusion, the conditions needed to exploit the mathematical framework

(cf. Sec. 5.3) are fulfilled.

Consequently, a translation of the inverse transfer functions $l_i^{\mathbf{x}}(t)$ into neuronal hardware is possible. In a neuronal architecture, the actual processing is performed by the synaptic connections between detectors and neurons. The mathematical framework already takes into account the discrete character of detectors and maps 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 $\lambda_i^{\mathbf{x}}$. Spatial processing is then governed by the topographic structure of the network, that is, which detector is connected to which neuron. Temporal processing is determined by the distribution of delays within the set of connections. That is, a number of dendrites with appropriate delays are chosen to sample the time course of $l_i^{\mathbf{x}}(t)$ in a sufficient way [54]. The response of the map neuron is shown to be robust with respect to the sampling method of the temporal delays [137].

Figure 5.4 shows an example of the whole neuronal setup. As illustrated a feedforward network of excitatory and inhibitory connections can implement the presented unified framework and can form a unimodal map from arbitrary input [54, 137, 198]. The question, however, remains how such a connectivity pattern is established in a real biological system. Franosch et al. [52] and Friedel et al. [56] have shown that correct synaptic connections can be *learned* from a teacher such as the visual system. Non-visual maps can develop by means of (supervised) STDP; for details see Sec. 5.5.2. By means of the present mathematical framework the learned connectivity pattern can be compared with the optimal one as given by Eqs. (5.23) and (5.27).

A meaningful comparison of the mathematically optimal network architecture with an actual biological setup, though, may not be straightforward. In real biological systems, error minimization as in Eq. (5.14) – that is, realizing the optimal connectivity – may not be possible because of neuronal limitations. The mathematical framework can nevertheless cope for the limited neuronal accuracy by reducing the error only below a certain error threshold, which may even vary in space. A biological example of non-uniform stimulus reconstruction are visual fields being sampled with different spatial and spectral resolutions [92, 200, 239]. Such a focus on specific spatio-temporal domains can mathematically be realized by introducing a positive weighting function into the integral of Eq. (5.14). 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.

Taken together, the formalism of optimal map formation is capable to deliver concrete neuronal connectivity pattern, just as illustrated in Fig. 5.4. As a consequence, the mathematical framework can make direct forecasts of the firing activity of concrete maps.

5.4 A recipe of making maps

In the previous section optimal connectivity pattern between sensory system and map was calculated [Fig. 5.3 and Eq. (5.23)] and realized neuronally [cf. Fig. 5.4].

To bring to life the mathematical framework of Sec. 5.3, an easy step-by-step "recipe" provides the optimal connectivity in a realistic biological setup:

- First, the transfer function $h_i^{\mathbf{x}}(t)$ is specified that determines the response of the detector i to a stimulus pulse that occurred t time units ago at position $\mathbf{x}$.
- Next, the Fourier transform $H_i^{\mathbf{x}}$ of the transfer function $h_i^{\mathbf{x}}(t)$ is calculated.
- Values of τ and σ are chosen in a suitable way. 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 [53, 54, 198].
- The entries of the model matrix M_{ij} are calculated by Eq. (5.26). Inversion of the matrix leads to $\mathcal{M}^{-1}$.
- The inverted matrix $\mathcal{M}^{-1}$ is multiplied by the matrix $\overline{\mathcal{H}}$ so as to find the input connection strengths $\mathcal{L}^T$.
- Finally, transposition and inverse Fourier transformation of $\mathcal{L}^T$ provide the inverse transfer functions $l_i^{\mathbf{x}}(t)$.

5.5 Multimodality

In deriving the mathematical framework in Sec. 5.3 all sensory maps are purely monosensory. In spite of this assumption, animals and humans perceive their environment through *several*, multimodal, sensory systems. To fully access the complete information of all monosensory maps, their information has to be combined. Physiological and behavioral experiments show that the monosensory perceptions are not independent but mutually interact with each other [1, 11, 46, 197]. A combination leads to the formation of *multi*sensory maps, i.e., maps receiving input from more than one sensory system [106, 158].

A number of brain areas, such as the midbrain in mammals, more precisely the SC [204], or even higher brain areas, such as the Anterior Ectosylvian Sulcus (AES) [25], contain clearly distinct monosensory as well as multisensory neurons. Since the perceptual and behavioral role of higher brain areas remains vague the well-studied SC is used as an example for multimodal interaction in the following.

The SC features a layered organization of spatial maps from all sensory systems that dispose of topographic, map-like information (such as vision but not olfaction) [204]. All these maps, uni- or multisensory, are mutually aligned [18, 112, 204] and thus provide a common *reference system* of sensory space.

5.5.1 Multimodal interaction

In general, there are two categories of multimodal interaction: *integration* and *pooling* of information.

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 the unimodal maps, features increased information reliability and saliency and an improved sensitivity in both space and time [77, 130, 131, 184]. For example, if visual

and auditory sensory system 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 artifact. At the same time, the integrated signal is stronger and allows for faster reactions (e.g., "escape!"). In some cases an integrated signal is even *optimal* [74, 158]. Despite increased reliability of the integrated map, its individual input streams cannot be distinguished anymore. That is, the information of 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, for instance, vision or audition, remains unresolved.

Chapter 6 provides a detailed description of multimodal integration. Further neuronal models based on statistical methods have been presented elsewhere [40, 41, 42]. Concrete theoretical models of multimodal integration within the SC have been developed as well [2, 142, 171, 185, 186, 214].

Pooling. Within a common reference system all monosensory as well as multisensory maps are present in parallel. Since all maps are aligned a simultaneous access to the diverse information, thus, signal characteristics, is possible. Consequently an object at one specific position in space-time can be identified and characterized by this location in order to select, e.g., 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 object, in other words, a possible prey. If it is visible and cold 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 [164, 165]. These prominent examples of pooling in the SC could enable target selection and ensure appropriate motor commands in a complex environment.

In summary, integration of information allows for a reliable spatial determination of an object, the key task of *object formation*. 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.

5.5.2 Development of multisensory space

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* [120, 205, 217]. The following section and Chap. 6 discuss the question of how a common multimodal space can evolve at all.

An obvious strategy would be the determination of one dominant modality as reference for all other modalities [120, 122]. Modifications of this reference map would then automatically lead to modifications of all other maps. 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 [157], cat [224, 226], clawed frog [30], ferret [115], barn owl [123], and snakes [70]. Psychophysical experiments with congenitally blind and normally sighted humans have shown that visual input early in life is necessary for multimodal interaction to occur [90, 176, 180]. Consequently vision seems to be the dominant guiding modality, i.e., the "teacher", for non-visual modalities.

A plausible argument supporting this idea 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 [213, 232]. Together with the subsequent development of deeper layers in the visual cortex (for mice, see [100]) the intrinsic topography of the retina could step-by-step dictate the organization and alignment of higher visual and, potentially, also multimodal maps.

An example where the alignment has been studied in detail, both experimentally and theoretically, is audio-visual integration within the SC of the barn owl. Here experiments [94, 122] have shown that the auditory map follows systematic changes within the visual input. The general mechanism facilitating such an alignment of maps is supervised spike-timing-dependent plasticity (STDP) [6, 7, 36, 66, 109, 149, 199, 216, 240]. The precise nature of the teaching signal has not been clarified experimentally but *selective neuronal disinhibition*, or gating, seems to play a key role [75, 233]. Theoretical studies have confirmed that excitatory and inhibitory teaching input can account for proper map formation and thus development of multimodal space [38, 56]. It is, however, only by *inhibitory* teaching input that an already aligned map can be re-aligned later on [56].

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. It has been shown both theoretically and experimentally that, although imprecise, a map of azimuthal sound location can be learned *without* any visual input [111, 122] though admittedly on a genetically determined substrate. In addition, non-visual modalities can influence each other as well, e.g., audition can influence haptics [11]. Moreover, somatosensory maps already sharpen in a postnatal phase when only auditory but no visual neurons are present [223, 225]. Behavioral and psychophysical studies show that visual perception can even be influenced by other modalities such as haptics [46] or audition [55, 197, 201]. More importantly, vision itself can improve, respectively sharpen as found in the visual system of young cats [223, 225].

Taken together, the presented experimental and theoretical findings put into question the current picture of vision-guided map alignment [113]. Wallace and Stein [225] 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. A concept of integrated multimodal teaching that is based on these considerations is presented in Chap. 6.

5.6 Discussion

In summary, sensory processing is based on three general concepts, denoted as the 'golden three' of sensory processing: physical mapping, optimal map formation, and multimodal integration; cf. Fig. 5.1. Given transfer functions to describe how a signal stimulates a detector, the formulated mathematical framework is able to quantify how the detector response is processed so as to lead to a "reconstruction" of the original signal. In the context of neuronal information processing the framework extends the established maximum-likelihood method by linking its parameters to easily accessible experimental quantities. The mathematical principle of *stochastic optimality* leads to a discrete and optimal representation of the outside world – a map.

Most importantly, the mathematical setup can be translated into neuronal architecture. That is, a discretization in space-time of the mathematical model allows for a derivation of synaptic connection patterns between detector and map neurons. To illustrate the relation to real biological settings, a step-by-step recipe offers the possibility of applying the mathematical framework to concrete biological situations. The generality of the method of optimal map formation can now be tested to model and analyze experimental results. In particular, the measurement of internal connections, for instance, in the SC, and firing profiles for specific sensory systems would provide a possibility to experimentally access the inverse transfer functions as defined in Sec. 5.3.1.

Based on the understanding of monosensory map formation, multimodal interaction and the development of multisensory space have been discussed. Here the concept of neuronal maps as compared to single neuron effects can deliver new perspectives on *multimodal* interaction, viz., *integration* and *pooling* of information. While integration of information allows for a reliable spatial localization of an object, pooling of information assures the access to the details of an object. Pooling of information could be the reason as to why multiple maps are found in the SC instead of a single multimodal map.

Finally, the importance of proper map alignment for multimodal interaction was emphasized. Here STDP learning algorithms with an inhibitory teacher signal can account for both initial map formation and even subsequent re-alignment of maps. Further experimental studies on inhibitory teacher input, e.g., within the SC, are nevertheless needed to clarify the precise role of inhibition in the alignment process. In addition, such experiments could answer the crucial question of which sensory systems determine the formation of multimodal space.

In other words, a deep understanding of how multimodal interaction is realized and established at an anatomical level requires more experimental and theoretical studies. For example, through which concrete mechanism could a collection of aligned maps allow for the pooling of information? Does such a mechanism also include feature selection in a common sensory space? How does multimodal interaction of maps contribute to the formation of a common sensory space? And to what extent does such a finding contradict the current picture of vision-guided map alignment? Finally, to what extent does the mathematical framework, that has been substantiated so far only by findings in the SC, apply to other areas of the brain, such as the AES, a well-defined multisensory cortical area observed in cats.

Some of these questions are addressed in the following chapter where a concept of

multimodal teaching is presented.

Appendix

5.A Nonlinearities in information processing

The presented model assumes a linear relation between stimulus and detector response. For a number of sensory systems, however, non-linearities appear during sensory processing. First, the transfer function h can be a non-linear function $\tilde{h}$. Second, the neuronal detector response can be nonlinear, typically logarithmic [34, 102, 132, 134, 166]. In case of a nonlinear transfer function and a logarithmic response the detector responses $\tilde{r}_i(t)$ have to be rewritten from (5.1) as

$$\tilde{r}_i(t) = \log\left[\int_{\text{all space}} \mathrm{d}\mathbf{x} \int_{-\infty}^{t} \mathrm{d}\tau \; s^{\mathbf{x}}(\tau)\tilde{h}_i^{\mathbf{x}}(t-\tau)\right] . \tag{5.33}$$

To nevertheless apply the presented model an incorporation of an additional computational step can in a first step cancel the logarithm. In a biological system this can be realized, e.g., by neurons with exponential firing behavior. Assuming such a neuronal step the detector response $\tilde{r}_i(t)$ remains with a nonlinear $\tilde{h}$ as compared to (5.1). A linearization of the nonlinear transfer function can be achieved by a redefinition of the signal $s \to \tilde{s}$. That is, appropriate characteristics of the stimulus are identified that are linearly related to r. For example, instead of looking at the heat distribution $T(\mathbf{x}, t)$ the intensity distribution of the corresponding radiation $\sim T^4(\mathbf{x}, t)$ can be considered due to the Stefan-Boltzmann law. Together, the incorporation of an additional computational step and the *reasonable* redefinition of detector response and signal allow for an optimal stimulus reconstruction by means of the presented linear framework.

5.B Self-averaging

Why can physical but noisy input quantities be expected to be self-averaging? To understand this valuable property a detector receives input signals f_i with $1 \leq i \leq N$ as a sum $\eta_N \sum_{i=1}^{N} f_i$ where η_N is a scaling factor. For the sake of convenience η_N is set to $1/N$ to get a decent scaling behavior as N increases. Moreover, f_i are stochastic random variables with mean a_i and finite variance. Finally, if the $1 \leq i \leq N$ represent, for example, different positions in space, in biological reality the stochastic correlation between positions that are far apart is small. Hence $f_i = a_i + \phi_i$ where the noise terms ϕ_i (cf. Fig. 5.3) are taken as *independent* random variables that by construction all have zero mean. For inputs of the form

$$\sum_{i=1}^{N} f_i = \left(\sum_{i=1}^{N} a_i\right) + \left(\sum_{i=1}^{N} \phi_i\right) =: A_N + \Phi_N \tag{5.34}$$

the number A_N is deterministic. Regarding Φ_N the strong law of large numbers (for details see [216], Appendix A) applies, so as to conclude that, as $N \to \infty$, it is $\eta_N \, \Phi_N \to 0$

independent of the *specific* realization of the $\{\phi_i\}$. The latter circumstance is exactly what is needed in practical work since one never knows the realization until it is all over. The strong law of large numbers guarantees that Φ_N vanishes as N becomes large.

The only, minor, drawback of all this is twofold. First, in reality the ϕ_i are never perfectly independent. Nevertheless, as long as correlations fall off fast enough as the distance $|i-j|$ becomes large, the strong law of large numbers still holds. Second, in practical work N is and remains finite. Then the central limit theorem as described in Appendix A of [216] shows that for arbitrary independent ϕ_i provided the second moment $\langle \phi_i^2 \rangle$ is finite and N large

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i \tag{5.35}$$

has a *Gaussian* distribution with mean zero. The standard deviation gives information about the width of the Gaussian. The same holds true for weakly dependent ϕ_i as found in biophysical reality.

5.C Remaining derivation steps leading to (5.23)

The following section elaborates some steps skipped in the derivation of (5.23). Here ideas due to the calclulus of variations [105] apply. Equation (5.15) is used as starting condition to minimize the expectation value of the quadratic error with respect to the optimal inverse transfer functions $l_j^{\mathbf{x}}(t)$. This leads to

$$\begin{aligned} \frac{\partial \left\langle [s^{\mathbf{x}}(t') - \hat{s}^{\mathbf{x}}(t')]^2 \right\rangle}{\partial\, l_j^{\mathbf{x}}(t')} &= 0 \\ \Leftrightarrow \qquad \left\langle \left[s^{\mathbf{x}}(t') - \hat{s}^{\mathbf{x}}(t')\right] \frac{\partial\, \hat{s}^{\mathbf{x}}(t')}{\partial\, l_j^{\mathbf{x}}(t')} \right\rangle &= 0 \qquad \text{for every } j. \end{aligned} \tag{5.36}$$

To solve (5.36), the estimate $\hat{s}^{\mathbf{x}}(t)$ is expanded using Eqs. (5.12) and (5.13) giving

$$\begin{aligned} \hat{s}^{\mathbf{x}} = \sum_i \Big\{ \chi_i \star l_i^{\mathbf{x}} + \chi_i \star \lambda_i^{\mathbf{x}} + \int \mathrm{d}\mathbf{y} \, \big[& s^{\mathbf{y}} \star h_i^{\mathbf{y}} \star l_i^{\mathbf{x}} + s^{\mathbf{y}} \star h_i^{\mathbf{y}} \star \lambda_i^{\mathbf{x}} + s^{\mathbf{y}} \star \eta_i^{\mathbf{y}} \star l_i^{\mathbf{x}} + s^{\mathbf{y}} \star \eta_i^{\mathbf{y}} \star \lambda_i^{\mathbf{x}} \\ & + \xi^{\mathbf{y}} \star h_i^{\mathbf{y}} \star l_i^{\mathbf{x}} + \xi^{\mathbf{y}} \star h_i^{\mathbf{y}} \star \lambda_i^{\mathbf{x}} + \xi^{\mathbf{y}} \star \eta_i^{\mathbf{y}} \star l_i^{\mathbf{x}} + \xi^{\mathbf{y}} \star \eta_i^{\mathbf{y}} \star \lambda_i^{\mathbf{x}} \big] \Big\} \, . \end{aligned} \tag{5.37}$$

Variation of $\hat{s}$ leads to

$$\frac{\partial\, \hat{s}^{\mathbf{x}}(t')}{\partial\, l_j^{\mathbf{x}}(t')} = \Big[\chi_j + \int \mathrm{d}\mathbf{y} \, \left(s^{\mathbf{y}} \star h_j^{\mathbf{y}} + s^{\mathbf{y}} \star \eta_j^{\mathbf{y}} + \xi^{\mathbf{y}} \star h_j^{\mathbf{y}} + \xi^{\mathbf{y}} \star \eta_j^{\mathbf{y}} \right) \Big](0) \, . \tag{5.38}$$

As before, all noise terms as well as the expectation of the input are stochastically independent of each other. All noise terms have zero mean. With these assumptions, the expectation values $\langle s\, \partial\hat{s}/\partial l \rangle$ and $\langle \hat{s}\, \partial\hat{s}/\partial l \rangle$ from (5.36) can be written

$$\left\langle s^{\mathbf{x}}(t') \frac{\partial\, \hat{s}^{\mathbf{x}}(t')}{\partial\, l_j^{\mathbf{x}}(t')} \right\rangle = \int \mathrm{d}\mathbf{y} \, \left\langle s^{\mathbf{x}}(t') \, (s^{\mathbf{y}} \star h_j^{\mathbf{y}})(0) \right\rangle \tag{5.39}$$

and combining (5.37) and (5.38)

$$\begin{aligned}\left\langle \hat{s}^{\mathbf{x}}(t')\frac{\partial\,\hat{s}^{\mathbf{x}}(t')}{\partial\,l_j^{\mathbf{x}}(t')}\right\rangle = \sum_i \Big\{ \left\langle (\chi_i \star l_i^{\mathbf{x}})(t')\chi_j(0)\right\rangle + \int \mathrm{d}\mathbf{y}\mathrm{d}\mathbf{y}' \Big[\left\langle (s^{\mathbf{y}} \star h_i^{\mathbf{y}} \star l_i^{\mathbf{x}})(t')(s^{\mathbf{y}'} \star h_j^{\mathbf{y}'})(0)\right\rangle \\ + \left\langle (s^{\mathbf{y}} \star \eta_i^{\mathbf{y}} \star l_i^{\mathbf{x}})(t')(s^{\mathbf{y}'} \star \eta_j^{\mathbf{y}'})(0)\right\rangle + \left\langle (\xi^{\mathbf{y}} \star h_i^{\mathbf{y}} \star l_i^{\mathbf{x}})(t')(\xi^{\mathbf{y}'} \star h_j^{\mathbf{y}'})(0)\right\rangle \\ + \left\langle (\xi^{\mathbf{y}} \star \eta_i^{\mathbf{y}} \star l_i^{\mathbf{x}})(t')(\xi^{\mathbf{y}'} \star \eta_j^{\mathbf{y}'})(0)\right\rangle \Big] \Big\}\,.\end{aligned} \tag{5.40}$$

To illustrate the calculations, which simplify (5.39) and (5.40), two isolated terms from (5.40) are considered as concrete example. The other terms are treated in a similar way. First consider

$$\begin{aligned}\sum_i \int \mathrm{d}\mathbf{y}\mathrm{d}\mathbf{y}' \left\langle (s^{\mathbf{y}} \star h_i^{\mathbf{y}} \star l_i^{\mathbf{x}})(t')(s^{\mathbf{y}'} \star h_j^{\mathbf{y}'})(0)\right\rangle \\ = \sum_i \int \mathrm{d}\mathbf{y}\mathrm{d}\mathbf{y}'\mathrm{d}\tau\mathrm{d}\tau'\mathrm{d}\tau'' \left\langle s^{\mathbf{y}}(t'-\tau-\tau')h_i^{\mathbf{y}}(\tau')l_i^{\mathbf{x}}(\tau)s^{\mathbf{y}'}(-\tau'')h_j^{\mathbf{y}'}(\tau'')\right\rangle\end{aligned} \tag{5.41}$$

where the definition of the convolution (5.3) and the composition rule (5.10) are applied. Exploiting the autocorrelation for the signal (5.18) this expression becomes

$$\begin{aligned}&\mu_s^2 \sum_i \int \mathrm{d}\mathbf{y}\mathrm{d}\mathbf{y}'\mathrm{d}\tau\mathrm{d}\tau'\mathrm{d}\tau''\delta(\mathbf{y}-\mathbf{y}')\delta(t'-\tau-\tau'+\tau'')h_i^{\mathbf{y}}(\tau')l_i^{\mathbf{x}}(\tau)h_j^{\mathbf{y}'}(\tau'') \\ &\quad = \mu_s^2 \sum_i \int \mathrm{d}\mathbf{y}\mathrm{d}\tau\mathrm{d}\tau'' h_i^{\mathbf{y}}(t'-\tau+\tau'')l_i^{\mathbf{x}}(\tau)h_j^{\mathbf{y}}(\tau'') \\ &\quad = \mu_s^2 \sum_i \int \mathrm{d}\mathbf{y}[(h_i^{\mathbf{y}} \star l_i^{\mathbf{x}}) \circ h_j^{\mathbf{y}}](t')\end{aligned} \tag{5.42}$$

where the last step results from the composition rule of convolution and autocorrelation (5.11). Focusing on the third term in the right-hand side of (5.40) leads in a similar way to

$$\begin{aligned}\sum_i \int \mathrm{d}\mathbf{y}\mathrm{d}\mathbf{y}' \left\langle (s^{\mathbf{y}} \star \eta_i^{\mathbf{y}} \star l_i^{\mathbf{x}})(t')(s^{\mathbf{y}'} \star \eta_j^{\mathbf{y}'})(0)\right\rangle \\ = \sum_i \int \mathrm{d}\mathbf{y}\mathrm{d}\mathbf{y}'\mathrm{d}\tau\mathrm{d}\tau'\mathrm{d}\tau'' \left\langle s^{\mathbf{y}}(t'-\tau-\tau')\eta_i^{\mathbf{y}}(\tau')l_i^{\mathbf{x}}(\tau)s^{\mathbf{y}'}(-\tau'')\eta_j^{\mathbf{y}'}(\tau'')\right\rangle\end{aligned} \tag{5.43}$$

which simplifies with (5.17) and (5.18) to

$$\begin{aligned}\mu_s^2\sigma_\eta^2\sum_i\int_{\substack{|\mathbf{y}|<y^{\max}\\0<\tau'<t^{\max}}}\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{y}'\mathrm{d}\tau\mathrm{d}\tau'\mathrm{d}\tau''\delta(\mathbf{y}-\mathbf{y}')\delta(t'-\tau-\tau'+\tau'')\\ \delta_{ij}\delta(\mathbf{y}-\mathbf{y}')\delta(\tau'-\tau'')l_i^{\mathbf{x}}(\tau)\\ =\mu_s^2\sigma_\eta^2\int_{\substack{|\mathbf{y}|<y^{\max}\\0<\tau'<t^{\max}}}\mathrm{d}\mathbf{y}\mathrm{d}\tau\mathrm{d}\tau'\delta(-\tau+t')l_j^{\mathbf{x}}(\tau)\\ =\mu_s^2\sigma_\eta^2\int_{\substack{|\mathbf{y}|<y^{\max}\\0<\tau'<t^{\max}}}\mathrm{d}\mathbf{y}\mathrm{d}\tau' l_j^{\mathbf{x}}(t')\ .\end{aligned}\tag{5.44}$$

Taken together, the final expressions for the expectation values become

$$\left\langle s^{\mathbf{x}}(t')\frac{\partial\,\hat{s}^{\mathbf{x}}(t')}{\partial\,l_j^{\mathbf{x}}(t')}\right\rangle=\mu_s^2h_j^{\mathbf{x}}(-t')\tag{5.45}$$

and

$$\begin{aligned}\left\langle\hat{s}^{\mathbf{x}}(t')\frac{\partial\,\hat{s}^{\mathbf{x}}(t')}{\partial\,l_j^{\mathbf{x}}(t')}\right\rangle=\sigma_\chi^2\ l_j^{\mathbf{x}}(t')+\sigma_\eta^2(\mu_s^2+\sigma_\xi^2)\int_{\substack{|\mathbf{y}|<y^{\max}\\0<\tau<t^{\max}}}\mathrm{d}\mathbf{y}\mathrm{d}\tau\ l_j^{\mathbf{x}}(t')\\ +(\mu_s^2+\sigma_\xi^2)\sum_i\int\mathrm{d}\mathbf{y}\ \left[(h_i^{\mathbf{y}}\star l_i^{\mathbf{x}})\circ h_j^{\mathbf{y}}\right](t')\ .\end{aligned}\tag{5.46}$$

Equation (5.36) therefore transforms to

$$\begin{aligned}l_j^{\mathbf{x}}(t)\left[\sigma_\chi^2+(\mu_s^2+\sigma_\xi^2)\int_{\substack{|\mathbf{y}|<y^{\max}\\0<\tau<t^{\max}}}\mathrm{d}\mathbf{y}\mathrm{d}\tau\ \sigma_\eta^2\right]\\ +(\mu_s^2+\sigma_\xi^2)\sum_i\int\mathrm{d}\mathbf{y}\ \left[(h_i^{\mathbf{y}}\star l_i^{\mathbf{x}})\circ h_j^{\mathbf{y}}\right](-t)=\mu_s^2h_j^{\mathbf{x}}(-t)\ .\end{aligned}\tag{5.47}$$

Inserting the parameter σ and τ and applying a Fourier transformation finally leads to Eq. (5.23).

To test whether the obtained extremum is a minimum, the second variation is calculated, which reads

$$\begin{aligned}\frac{\partial^2\left\langle\left[s^{\mathbf{x}}(t')-\hat{s}^{\mathbf{x}}(t')\right]^2\right\rangle}{\left(\partial\,l_j^{\mathbf{x}}(t')\right)^2}&=2\ \frac{\partial}{\partial\,l_j^{\mathbf{x}}(t')}\left[\left\langle s^{\mathbf{x}}(t')\frac{\partial\,\hat{s}^{\mathbf{x}}(t')}{\partial\,l_j^{\mathbf{x}}(t')}\right\rangle-\left\langle\hat{s}^{\mathbf{x}}(t')\frac{\partial\,\hat{s}^{\mathbf{x}}(t')}{\partial\,l_j^{\mathbf{x}}(t')}\right\rangle\right]\\ &=2\ \left[\sigma_\chi^2+\sigma_\eta^2(\mu_s^2+\sigma_\xi^2)\int_{\substack{|\mathbf{y}|<y^{\max}\\0<\tau<t^{\max}}}\mathrm{d}\mathbf{y}\mathrm{d}\tau\ +(\mu_s^2+\sigma_\xi^2)\int\mathrm{d}\mathbf{y}\int\mathrm{d}\tau\ \left[h_j^{\mathbf{y}}(\tau)\right]^2\right].\end{aligned}\tag{5.48}$$

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

5.D Gaussian blurred signal

This subsection presents an equation equivalent to (5.23) but for a Gaussian blurred signal. As in (5.19), a realistic signal would fulfill some kind of Gaussian relation for the expectation value

$$\left\langle s^{\mathbf{x}}(t)s^{\mathbf{x}'}(t')\right\rangle = A\exp\left(-\frac{|\mathbf{x}-\mathbf{x}'|^2}{2\sigma_{s_x}^2}\right)\exp\left(-\frac{|t-t'|^2}{2\sigma_{s_t}^2}\right)\ . \tag{5.49}$$

For this case a derivation of an equation like (5.23) is possible analogously to Appendix 5.C. Since for the signal the Gaussian correlations, however, replace the delta functions, e.g., in (5.42) and (5.44), 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, the analogue to (5.23) is derived as

$$\begin{aligned}
&\int \mathrm{d}\epsilon\ \mathrm{d}\epsilon_t\ A\exp\left(-\frac{|\epsilon|^2}{2\sigma_{s_x}^2}\right)\exp\left(-\frac{|\epsilon_t|^2}{2\sigma_{s_t}^2}\right)h_j^{\mathbf{x}+\epsilon}(t''+\epsilon_t)\\
&\quad= \sigma_\chi^2\ l_j^{\mathbf{x}}(t'') + \sigma_\eta^2\sigma_\xi^2\int_{\substack{|\mathbf{y}|<y^{\max}\\0<\tau<t^{\max}}}\mathrm{d}\mathbf{y}\mathrm{d}\tau\ l_j^{\mathbf{x}}(t'') + \sigma_\xi^2\sum_i\int\mathrm{d}\mathbf{y}\ \left[(h_i^{\mathbf{y}}\star l_i^{\mathbf{x}})\circ h_j^{\mathbf{y}}\right](t'')\\
&\qquad+\sigma_\eta^2\int_{\substack{|\mathbf{y}|<y^{\max}\\0<\tau<t^{\max}}}\mathrm{d}\mathbf{y}\mathrm{d}\epsilon\mathrm{d}\tau' A\exp\left(-\frac{{\epsilon_t}^2}{2\sigma_{s_t}^2}\right)l_j^{\mathbf{x}}(t''+\epsilon_t)\\
&\qquad+\sum_i\int\mathrm{d}\mathbf{y}\mathrm{d}\epsilon\mathrm{d}\epsilon_t A\exp\left(-\frac{|\epsilon|^2}{2\sigma_{s_x}^2}\right)\exp\left(-\frac{\epsilon_t^2}{2\sigma_{s_t}^2}\right)\left[(h_i^{\mathbf{y}}\star l_i^{\mathbf{x}})\circ h_j^{\mathbf{y}+\epsilon}\right](t''+\epsilon_t)\ .
\end{aligned} \tag{5.50}$$

The effect of the additional spatio-temporal integrals as compared to (5.23) is a smoothening of the final reconstruction. Not only is the value at a specific point in space and time $(\mathbf{y}, t'')$ taken into account but neighboring points in a nearly area surrounding it are included as well.

Was die Abbildung der Raumverhältnisse betrifft, so geschieht eine solche allerdings an den peripherischen Nervenenden im Auge und an der tastenden Haut in einem gewissen Grade, aber doch nur in beschränkter Weise, da das Auge nur perspectivische Flächenabbildungen giebt, die Hand die objective Fläche auf der ihr möglichst congruent gestalteten Körperoberfläche abbildet. Ein directes Bild einer nach drei Dimensionen ausgedehnten Raumgrösse giebt weder das Auge noch die Hand. Erst durch die ***Vergleichung*** *der Bilder beider Augen, oder durch Bewegung des Körpers, beziehlich der Hand, kommt die Vorstellung von Körpern zu Stande.*

Helmholtz [85], p.445

6. Multimodal map formation: Calibration of neuronal maps through integrated Multimodal Teaching (iMT)

For neuronal processing of sensory information and motor commands, spatially aligned uni- and multimodal maps play a key role. The precise alignment of maps, however, is not present at birth but has to be *learned* during the development of an animal. Here the unifying concept of *integrated Multimodal Teaching* (iMT) is introduced that is based on supervised spike-timing-dependent plasticity (sSTDP) and multimodal integration to calibrate inputs from different sensory modalities. Analytical calculations and numerical experiments demonstrate that a multimodal teacher can ensure proper *intrinsic* map formation and alignment. New interpretations of existing experiments are presented and the dominance of vision in guiding map alignment is demonstrated to be a natural consequence of the exceptional precision of the visual system. Finally, new experimental setups that can shed light on and clarify intrinsic development of an aligned multisensory space are suggested.

6.1 Introduction

The 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

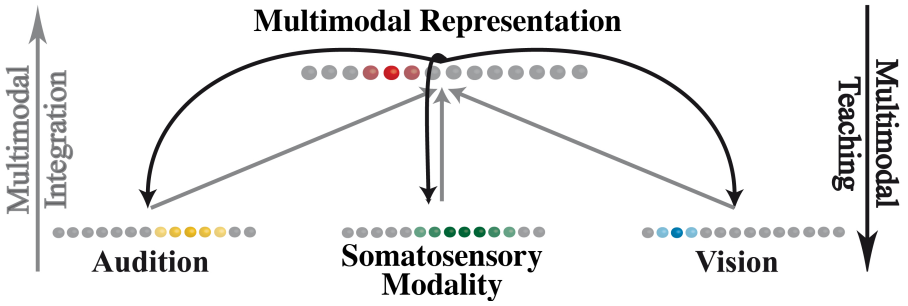

Figure 6.1: *Illustration of integrated Multimodal Teaching (iMT) as fundamental concept for unimodal map formation and alignment. All unimodal modalities are integrated into an integrated multimodal map (gray arrows). The activity of the multimodal map then induces an inhibitory teaching signal (black arrows) that leads to an adaptation of the unimodal maps by means of supervised spike-timing-dependent plasticity (sSTDP). The detailed model setup employed in this chapter is explained in Sec. 6.2.2.*

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 [213, 226].

The process of sensory calibration is commonly seen as vision-guided; cf. King's review [113]. 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 [157], cat [224, 226], clawed frog [30], ferret [115], barn owl [123], and snakes [70], support the idea of vision as teaching modality. However, vision-guided map calibration is not as evident as it seems at a first glance. First, sensory maps, although imprecise, can form without any visual input [111, 122, 213]. Experiments on visually deprived animals and humans have shown that despite a missing visual system, the auditory and somatosensory systems can develop normally; see [117, 177, 181] for details. Second, behavioral [10, 46, 55, 197, 201] and physiological [158] studies have shown that vision is not static but can be influenced by other modalities. Third, and most importantly, Wallace and Stein [223, 225] have pointed out that non-visual sensory maps in the superior colliculus (SC) already sharpen at times where *no* visual neurons are present. Furthermore, they have observed that all modalities, including vision, develop in a *temporally coinciding* way. But if vision acts 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 only predominant guiding modality. That is, what is a coherent alternative to vision as a teaching modality?

For the following, two findings serve as starting points. First, the formation of different modalities begins at the time of appearance of multimodal integration [225]. Second, the motor, i.e., a multimodal map in the superior colliculus (SC) and in its non-mammalian homolog, the optic tectum (OT), has been suggested to provide a teaching signal for auditory space alignment; cf. [139] and Fig. 6.2.

Therefore, an *integrated Multimodal Teaching* concept (iMT) (cf. Fig. 6.1) is suggested that can unify contradictory findings as to how different modalities can be aligned with respect to each other as well as explain the well-proven dominance of the visual system.

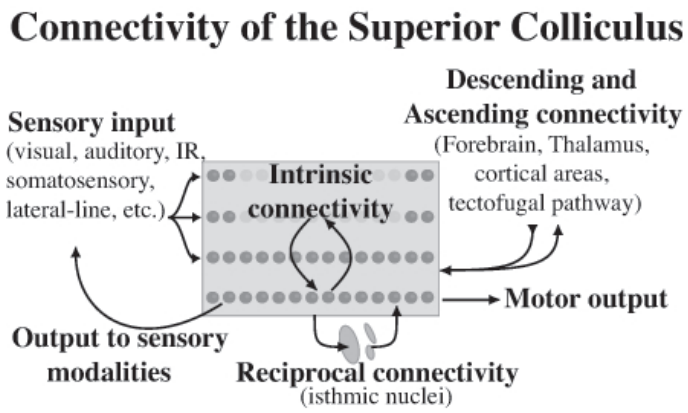

Figure 6.2: *Schematic drawing of the SC and its connectivity. The SC is a layered arrangement of aligned uni- and multisensory maps that are intrinsically connected. Inputs to the SC come from all sensory systems that provide information in a map-like structure. The SC produces output for motor responses and partially to influence the alignment process of associated sensory systems [139]. Altogether the SC is a beautiful example to study unimodal map formation and alignment.*

The introduction of an integrated multimodal teacher requires a careful discussion of its intrinsic characteristics and implications on a conceptual level. Here the focus lies 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 supervised STDP. The question of dynamic adaptation is omitted 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 [61]. Attempts to model this issue can be reviewed in [60, 72, 188, 189]. Within the SC/OT, however, multisensory and predominantly monosensory layers are mutually aligned to gain a unified multisensory representation of sensory space [18, 112, 204]. The combined sensory information can then be used to generate directional responses in the SC/OT motor map [130, 131, 138, 203, 219]. In other words, at this stage all sensory maps share a common representation system. Similar to other theoretical models [56, 234], the suggested iMT concept concentrates on the question of map calibration appearing in the context of the very common sensory space.

Besides the derivation of the general concept for iMT, applications of the model to concrete examples are presented. In particular, the experiments of Knudsen et al. [123] are reanalyzed in detail and contrasted with findings of Wallace and Stein [223, 225] in the light of the iMT concept. Finally, the concluding section gives experimentally testable predictions of the model.

6.2 The integrated multimodal teacher

The iMT concept structurally consists of a stack of unimodal maps, each connected with the multimodal map; cf. Fig. 6.1. In a first step, different unisensory modalities are merged into the multimodal map. Such a multimodal integration of unimodal maps was found within the experimentally well described SC/OT [204, 225]. 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. The teacher input therefore does not originate from a single modality, for example, vision, but from a common integrated multimodal map.

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 hardly any information and are randomly distributed or coarsely pre-wired; see Sec. 1.4. The connections have to evolve during the calibration process by means of LTP and LTD of the synapses (in mathematical terms described by STDP). This case is denoted as map *formation*.

The second class requires a collection of already pre-wired maps (see Sec. 1.4 and [213]) and corresponds to an *alignment* of maps, for instance, during growth or in shifting experiments [122, 123]. The stimulus position is 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.

A careful analysis of the consequences of applying an integrated Multimodal Teaching (iMT) signal requires an understanding of map formation and alignment from a procedural point of view. In general, the iMT process (as shown in Fig. 6.3) constitutes a closed feedback loop that consists of three repeating steps. First, during *multisensory 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 are inherently connected by means of three important questions: How do unisensory maps determine iMT? How do the characteristics of iMT influence map adaptation? And how does iMT calibrate different unimodal maps? Taken together, these three questions provide a structure for the analysis of the iMT concept and guide through the following sections.

6.2.1 How do unisensory maps determine iMT?

To start the analysis, the essential terms have to be clarified. A sensory modality is referred to as its neuronal representation on a spatial map [15, 140, 213, 217]. 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 [41, 98, 174, 195, 217]. Reflecting the statistics of the firing profile, the mean μ corresponds to the estimated position and the deviation σ describes the estimator's

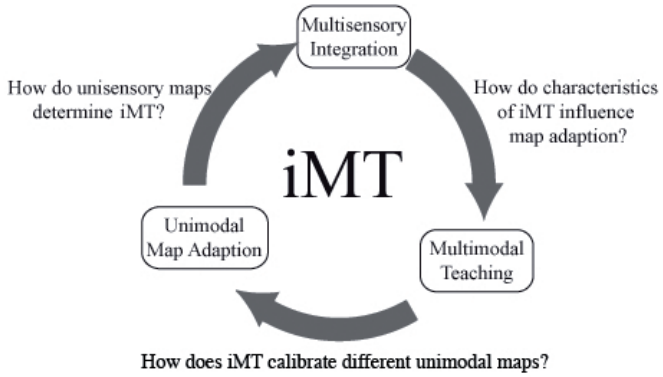

Figure 6.3: *Main procedural aspects of iMT. The multimodal map as combination of different unimodal maps shows certain characteristics. What are these characteristics and how do they influence map calibration by means of supervised spike-timing-dependent plasticity (sSTDP) where the multimodal signal serves as teaching signal? To close the feedback loop, the question has to be answered how different unimodal maps adapt to a common teaching signal. Taken together, these three questions provide a structure for the analysis of the iMT as presented in this chapter.*

accuracy. The integration of different sensory maps is therefore equivalent to a combination of position estimators [1, 45, 74, 84, 93, 129, 158].

The exact neuronal mechanism as to how different modalities are integrated into a multimodal map [41, 42] is still largely an open question. The family of Bayesian combination schemes nevertheless mathematically describes the result of optimal multimodal integration. For the sake of simplicity, the following analysis is reduced to the maximum-likelihood combination of two modalities, namely, *vision* V and *audition* A. The combination of three modalities is described in Appendix 6.B. The visual and auditory maps are characterized by their means, i.e., the position estimates, μ_V (vision) and μ_A (audition). Furthermore, the accuracy of the position estimates is given by σ_V (vision) and σ_A (audition) as illustrated by Fig. 6.4.

The multimodal position estimate μ_M as a maximum-likelihood combination from temporally coinciding visual and auditory inputs is given by [101, 108]

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

As a consequence of optimal integration, the multimodal estimator features the smallest possible standard deviation σ_M as given by

$$\sigma_M \quad = \quad \sigma_A \sigma_V \ / \ \sqrt{\sigma_A^2 + \sigma_V^2} \ . \tag{6.2}$$

For a detailed derivation of the two above equations, please see Appendix 6.A.

The distance between the multimodal position μ_M and the unimodal positions $\mu_{A/V}$ by means of (6.1) are computed to

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

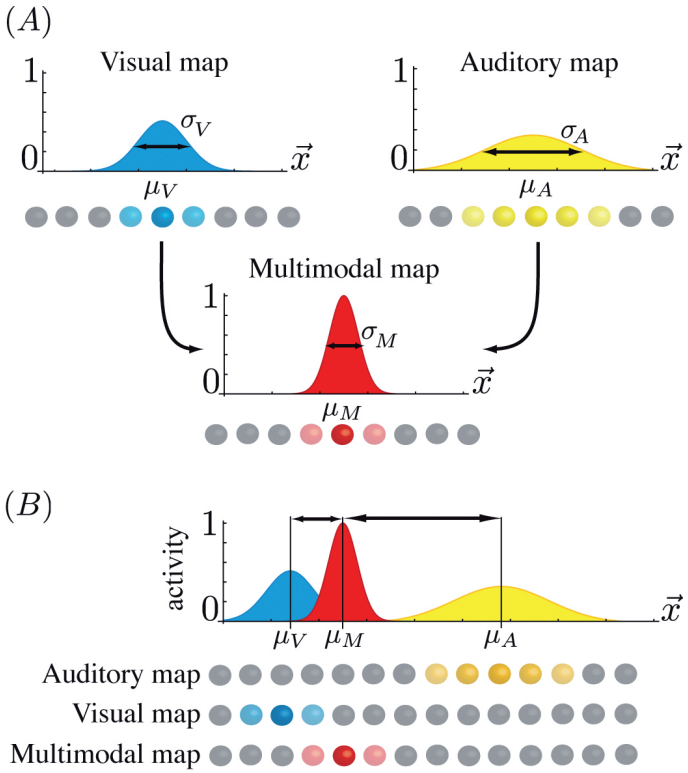

Figure 6.4: *Multimodal integration of two unisensory maps. The firing profiles of neuronal maps for audition, vision, and the multimodality can be interpreted as position estimators of the stimulus and are therefore assumed to be Gaussian functions with mean* $\mu_{A/V/M}$ *and standard deviation* $\sigma_{A/V/M}$*. The multimodal map is integrated by means of a maximum-likelihood method; cf. Eqs.* (6.1) *and* (6.2)*.* (*A*) *Accordingly, the multimodal map shows the smallest standard deviation* $\sigma_M < \sigma_V < \sigma_A$ *and is thus the most precise of all maps; cf.* (6.2)*.* (*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)*.*

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 and in Table 6.1.

In the next step the integrated multimodal map provides 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 effects of synaptic plasticity (mathematically described by STDP), as discussed in Sec. 6.2.2. This section focuses on fundamental characteristics of the multimodal map that follow directly from *optimal* multimodal integration. The unimodal maps that form the multimodal map are presumed to be 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. These two effects lead to a variable multimodal map with time-dependent $\mu_M(t)$ and $\sigma_M(t)$ [see Eqs. (6.1) and (6.2)], in particular, during the calibration process of the unimodal maps.

Figure 6.5 illustrates two varying unimodal maps together with their combined multimodal map; see Eqs. (6.1) to (6.3). One of the unimodal maps changes in precision (A) or position (B) while the other remains constant. First, a sensory map that becomes more precise (Fig. 6.5 A, blue profile) "attracts" the position of the multimodal map, i.e., the learning signal. Sensory maps that increase in accuracy get more positionally stable during sSTDP learning processes. Second, a shifting sensory map (Fig. 6.5 B, blue profile) induces a shift of the multisensory map toward the stationary sensory map (yellow profile). The upshot in the context of map alignment is that a more static sensory map attracts the teaching signal and will therefore hardly adapt at all.

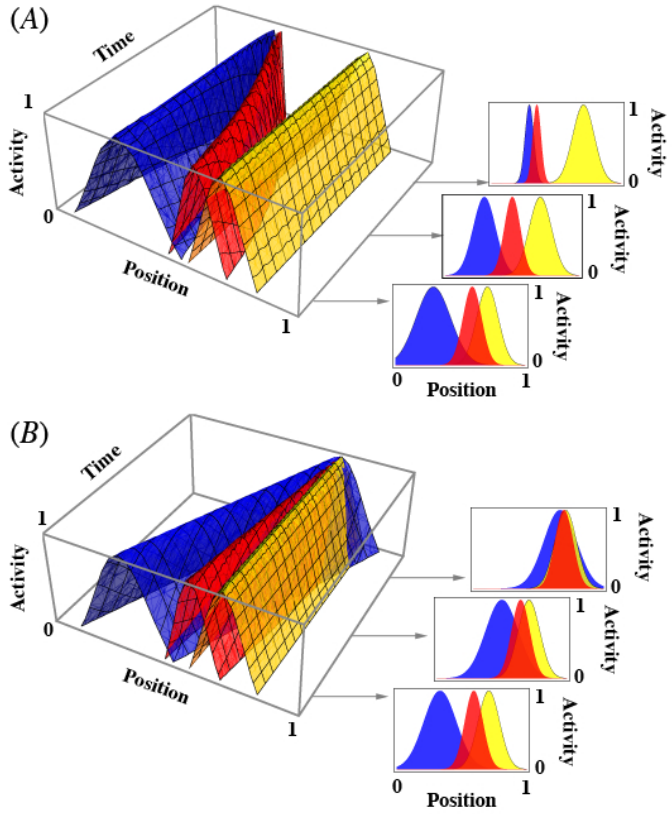

Figure 6.5: *Characteristics of an integrated multimodal map. During the iMT learning process the different modalities align with the teaching signal ($\mu_{A/V}$ changes) and their neuronal representations become more precise ($\sigma_{A/V}$ decreases). Given that the multimodal map is a combination of the unimodal sensory maps [cf. Eqs.* (6.1) *to* (6.3)*], its precision and position are determined by the unisensory maps and consequently change as well. Here one modality (blue profile) exemplarily varies in precision (A) and position (B) whereas the other modality (yellow profile) remains unchanged. The resulting multimodal teacher is depicted as red profile. (A) The modality (blue profile) that gets more precise attracts [see Eq.* (6.3)*] and sharpens [see Eq.* (6.2)*] the multimodal teacher (red profile). (B) 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 decreases as well; cf. Eq.* (6.3).

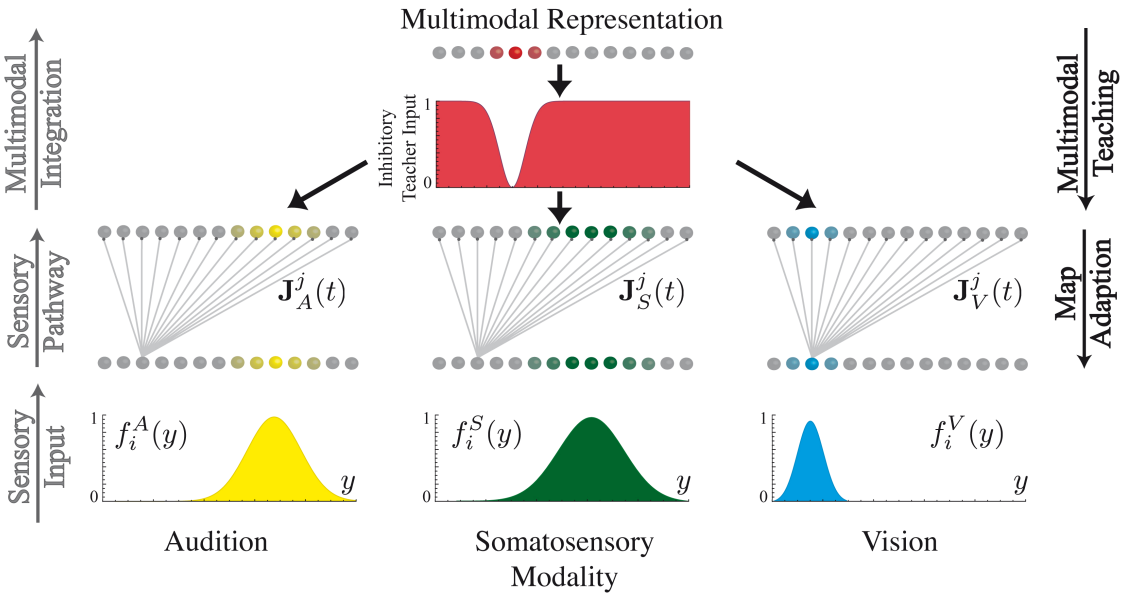

Figure 6.6: *Model architecture. A multimodal object such as a water drop hitting the skin produces sensory input for audition, somatosensory modality, and vision. The input is modeled by means of a Gaussian* $f_i^{A/S/V}(y)$ *so as to stimulate neuron* i *of the corresponding input layer. Neurons of the input layer project through all-to-all synaptic connections* $\mathbf{J}^j_{A/S/V}(t)$ *to neuron* j *of the map layer. These synaptic connections are learned by means of sSTDP to align the output of the map layer to the inhibitory teacher input.*

In summary, the analysis of the optimally integrated multimodal map has unraveled three essential aspects. First, the multimodal map is always more precise than any unimodal map; cf. Eq. (6.2). Second, the optimal multisensory teaching signal is located more closely to the position of the more precise modality; see Eq. (6.3). Third, on the basis of an adaption of the monosensory maps through sSTDP learning rules, Eqs. (6.1) and (6.2) describe how position $\mu_M(t)$ and precision $\sigma_M(t)$ of the multimodal teacher vary for given $\mu_{A/V}(t)$ and $\sigma_{A/V}(t)$ of the input maps.

6.2.2 How do iMT characteristics influence map adaptation?

In the previous section, 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 and consequently the teacher vary during map calibration. This section therefore discusses how teachers with different accuracies and teachers that are shifting during learning 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 the model consists of the three modalities audition A, the somatosensory modality S, and vision V that adapt to the multimodal inhibitory teacher modality T via STDP learning rules; see Fig. 6.6. The fundamental principles of inhibition-mediated STDP map alignment have been described elaborately by Friedel and van Hemmen [56]. All sensory maps are

one-dimensional maps of N Poisson neurons [109]. Each modality consists of an input map with a Gaussian firing profile

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

for $1 \leq i \leq N$, 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 $\mathbf{J}^j_{A/S/V}(t)$ 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.

The simulations incorporate a teacher that shows the essential characteristics 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 neurons, and σ_T being the width of the teacher influence. The process where a teacher suppresses map activity at locations different from the real object's position y, is called *selective disinhibition* [56, 94].

For numerical experiments on map alignment a shift Δ is introduced in the sensory position of the modality, e.g., through prisms in the Knudsen [122, 123] experiment on the barn owl. As explained in Sec. 6.3.1, a shift of vision induces at the same time 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 the multimodal teacher. In mathematical terms, $y + \Delta_{A/S/V}$ replaces y in Eq. (6.5) where the shift depends on the modality according to Eq. (6.3). Therefore, the biological experiments and the numerical experiments agree.

The changes within the synaptic connectivity pattern during the alignment process can be described and explained by STDP [56, 109]. The *map estimation error* E as a least-square error quantifies changes within the synaptic pattern. It measures how well a map can reproduce spatial information of M systematically varied incoming signals,

$$E = \sqrt{\sum_{i=1}^{M} \left(\overline{x}_i - x_i^T\right)^2 / M} \tag{6.6}$$

where $\overline{x}_i$ 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 the numerical experiments are given in the Table 6.2.

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 [140, 213]. Furthermore, Wallace and Stein [225] 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. To cope for the worst situation, it is necessary to show that learning still works or at least starts with a highly inaccurate teacher. 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, the teacher accuracy is modeled by an Gaussian distributed random shift Δ_T added to the correct teacher position. As a result of the numerical experiments with different teacher shifts Δ_T, the unimodal maps adapt more slowly and with a larger map estimation error to a less accurate teacher signal; cf. Fig. 6.7. Within a given time window, an inaccurate teacher therefore slows down but does not prevent the map alignment process. In conclusion, the iMT concept does not depend crucially on the concrete realization of the multimodal integration.

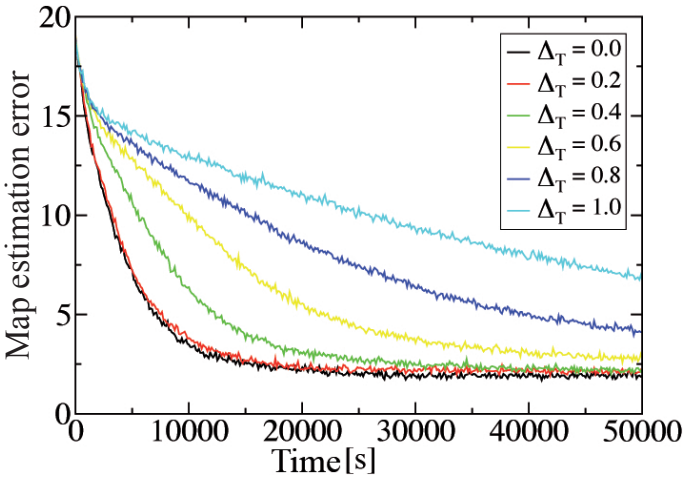

Figure 6.7: *Influence of teacher precision Δ_T on map alignment. The plot presents the map estimation error E [see Eq. (6.6)] for map alignment with a multimodal teacher of different precisions. The map estimation error drops faster and to a lower level with a more precise teacher (lower Δ_T). The precision of the teacher therefore influences the temporal development of map alignment but does not prevent it.*

Multimodal teacher with dynamically changing position

Similar to Fig. 6.5, the position of the integrated multimodal teacher continuously shifts in space during the alignment process. As a consequence, the influence of a shifting teacher, more concretely, its shifting velocity, on map alignment has to be analyzed. Figure 6.8 shows the synaptic pattern of sensory maps for the time step $t = 2000$ s. Synapses are modified by means of teaching signals shifting with different velocities. The dynamic of the

teacher essentially determines to which degree a map adapts 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; cf. Eq. (6.4). A less precise modality can still adapt to a teacher with a given velocity, whereas a precise map cannot follow anymore; cf. Fig. 6.8, change from top to bottom.

In conclusion, a low teacher accuracy does not prevent alignment of different maps

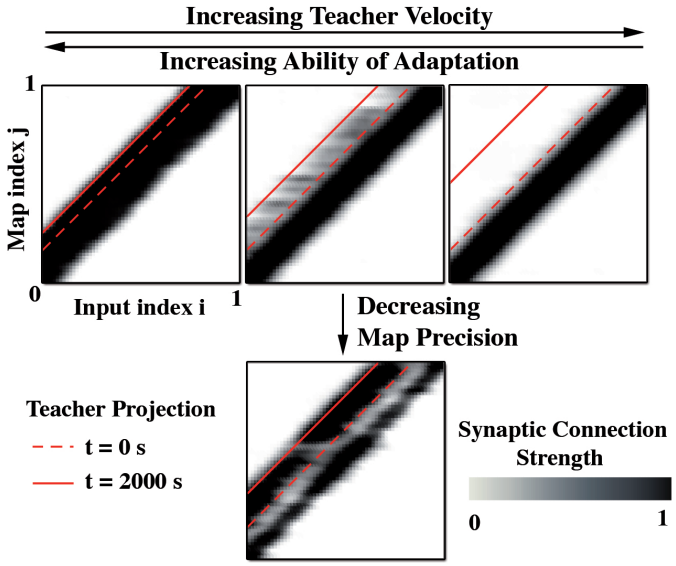

Figure 6.8: *Influence of dynamically changing teacher position on map alignment. Simulations performed with input standard deviation* $\sigma = 0.03$ *(top line),* $\sigma = 0.05$ *(bottom), and teacher standard deviation* $\sigma_T = 0.02$*. For details regarding the parameters, see Appendix 6.C. The teacher signal, that is, the inverse Gaussian profile, is shifted in space. The red lines indicate the mean position of the teacher. In the 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. By stepping to the right, the teacher velocity doubles. The synaptic pattern of the sensory maps are initially located around the diagonal and are modified by means of supervised spike-timing-dependent plasticity (sSTDP). The box at the top left shows the synaptic pattern at time* 2000 *s caused by a slowly shifting teacher. With a certain delay, the synaptic pattern of the map has adapted to the teacher map. With increasing teacher velocity, that is, by stepping to the right, the map adapts much less. For the synaptic pattern in the top right box the teacher velocity has reached a critical level 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 box at top middle to the box at bottom middle. Here the synaptic pattern of the 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.*

by means of sSTDP but influences the temporal development and, to some degree, the accuracy of the final state. 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.

6.2.3 How does iMT calibrate different unimodal maps?

The third and last step of the iMT analysis (see Fig. 6.3) addresses the question of how the essential characteristics of the different modalities, in particular, the different accuracies of the sensory systems, influence map formation and alignment. Results of the corresponding simulations are shown in Fig. 6.9. Here map formation and alignment by means of the multimodal teacher map are 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. 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.

6.3 Applications of the iMT concept

After detailed analysis of multimodal teaching, the experiments presented in the introduction are reviewed in the light of the iMT concept. In doing so, the aforementioned experiments that can be explained by a vision-guided concept of map calibration and those that cannot are treated separately. The iMT concept offers explanations for both.

In the following a group of experiments is analyzed that motivated the idea of vision as the dominant guiding modality: the prism glass experiments of Knudsen and coworkers [94, 122]. These experiments are explained while denying a vision-guided map alignment. A re-interpretation of experimental results from Stein et al. [223, 225] further illustrates how the iMT concept could work.

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. [94, 122]. 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. The visual map, in contrast, remains constant during the whole alignment process.

Can the iMT concept reproduce this *dominance of vision* within the learning process? To answer this question the alignment of maps with different variances and distances [cf. Eq. (6.3)] to the teacher map has been simulated. 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. The iMT concept now serves to reconsider synaptic plasticity during a prism shift experiment.

- **Initial state** (Fig. 6.10, top)
  Auditory and visual maps are shifted with respect to each other because of an artificial

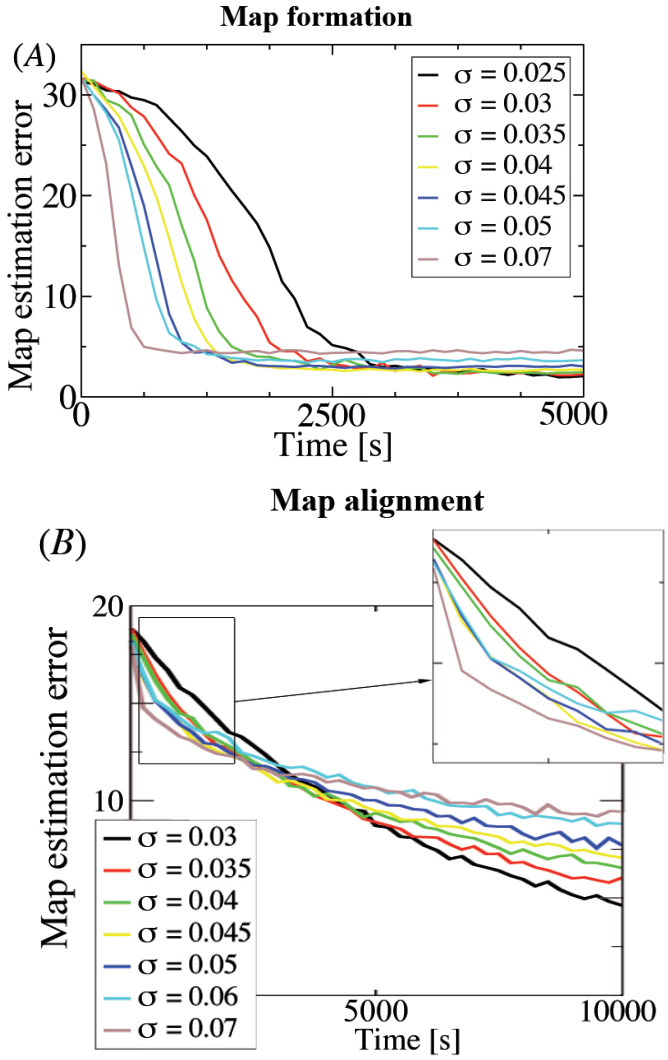

Figure 6.9: *Influence of varying input precision σ on map formation (A) and alignment (B). Both plots show the map estimation error E as given by Eq.* (6.6) *between the unimodal maps and the teacher. The input maps differ in precision (input standard deviation σ) as indicated. The graphs show the map estimation error E for map formation (A) and for map alignment (B). The boxed region in the bottom plot is enlarged to clearly visualize which map adapts faster. Both plots show that maps with a high precision adapt more slowly to the teacher input than low-precision maps but regain higher precision after the map calibration.*

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 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

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 [39, 153], barn owl [78, 121], and cat [8, 26]. 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. The resolutions are assumed to be the inverse of these values. By means of the variances the two unimodal maps are combined to a multimodal teacher and the distance of each map to the multimodal map [cf. Eq.* (6.3)*] is calculated and normalized by the distance between the two modalities. The visual map as more precise map is nearer to the multimodal map than the less accurate auditory map.*

multimodal teacher map is mainly determined by the characteristics of the visual map, that is $\sigma_M \approx \sigma_V$ and $\mu_M \approx \mu_V$; cf. Eqs. (6.1) and (6.2). Equation (6.3) illustrates that the shift between vision as the most precise sensory system and the multimodal teacher map is smaller than for any other system.

- **Intermediate state** (Fig. 6.10, middle)
  Synaptic connections to the auditory map start changing to compensate for the shift between audition and the multimodal teacher. Based on the previous studies as presented in Secs. 6.2.2 and 6.2.3 and because of its higher precision, the visual map adapts much more slowly than the auditory map. Due to the adaptation of the auditory map the teacher, as a combination of the two modalities, nevertheless 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. 6.5 B.

- **Final state** (Fig. 6.10, bottom)
  The two maps have been realigned. The position of the visual map has hardly changed, whereas the auditory map has shifted almost the whole distance induced by the prisms at the beginning of the experiment.

In summary, the iMT concept can reproduce the different steps of the shifting experiments of Knudsen and others [94, 122]. Here the observed dominance of the visual system only reflects the more general concept of a *dominance of accuracy* following from the iMT concept.

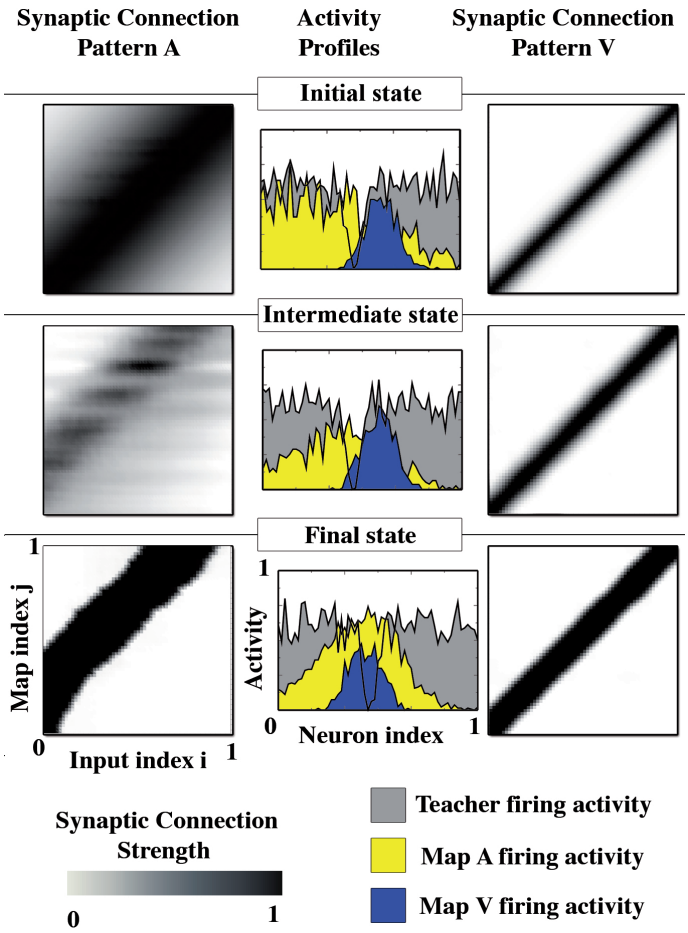

Figure 6.10: *Numerical realignment experiments. The plot illustrates the shifting process for the auditory map A with variance* $\sigma_A = 0.07$ *and the visual map V with* $\sigma_V = 0.03$. *The distances between teacher and unimodal map positions are calculated by Eq.* (6.3) *for the corresponding 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 *(see color code in the bottom left corner). The activity profiles of the auditory map A, the visual map V, and the teacher map are shown in the middle column; for details of the activity profiles see Sec. 6.2.2. 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. In summary, the iMT concept can therefore reproduce experiments* pro-*vision-guided map formation.*

6.3.2 Experiments *contra*-vision-guided map formation

The introduction has reviewed some experiments that question a vision-guided learning process. The analysis of the iMT 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 [46, 55, 158, 197, 201].

Furthermore, the iMT concept should be able to reproduce formation of unimodal maps in visually deprived animals. Studies [117, 177, 181] show that blind humans, cats, and ferrets can localize sound as precisely as individuals with normal vision. Measurements of Wallace and Stein [223, 225] support the critical view of vision-dominated learning by showing that sensory maps already sharpen at a time 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 the 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. 6.7 and [111, 122, 213]. Furthermore, in the intrinsic learning process all maps develop and improve in parallel, as reported in experiments such as [223, 225]. Here the reproduction of unimodal map formation in a visually deprived animal allows for studying whether somatosensory and auditory maps can develop *without vision*, i.e., with a very imprecise teacher.

- **Initial state** (Fig. 6.11, top)
  Somatosensory and auditory maps are only coarsely pre-wired, and the visual map does not exist at all. The multimodal map as a combination of somatosensory and auditory maps is very imprecise; cf. Eq. (6.2).

- **Intermediate state** (Fig. 6.11, middle)
  Even with a very imprecise teacher map the auditory and somatosensory maps start to improve so that their variances decrease. The earlier calculations (see Fig. 6.7) illustrate that the learning process with an imprecise teacher starts very slowly, corresponding to the topmost curve in the plot. 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. 6.7), the learning process accelerates.

- **Final state** (Fig. 6.11, bottom)
  The two maps are aligned and have reached their best-possible resolutions.

Together, an interpretation of both experiments *pro*-vision-guided map formation and *contra*-vision-guided map formation is possible within the iMT concept. As a consequence, iMT is a successful concept for calibrating different unimodal maps that is in accordance with known experimental results of map formation and alignment.

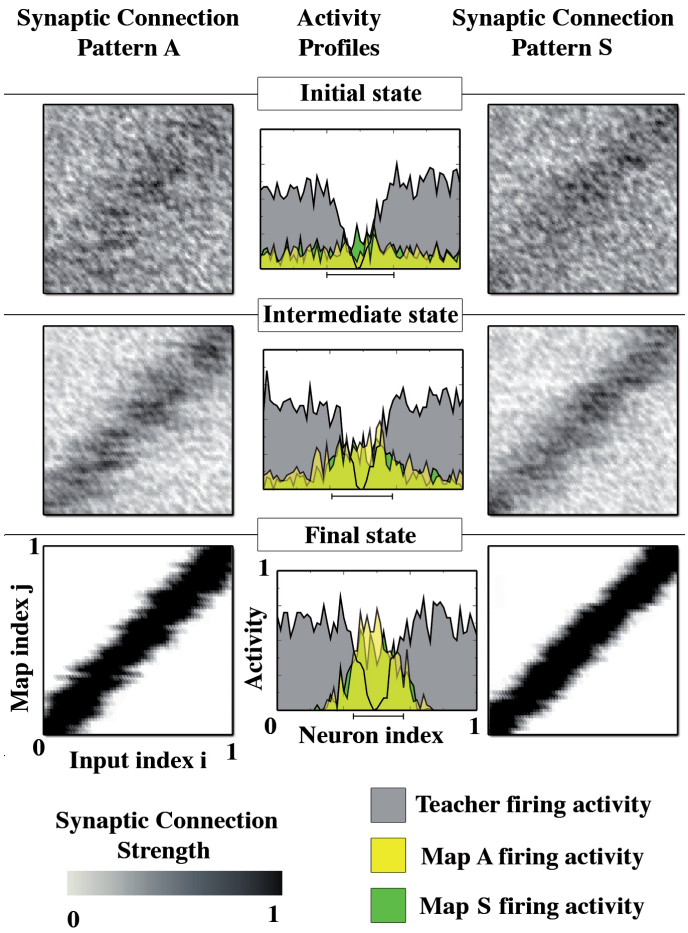

Figure 6.11: *Numerical map formation experiments with an improving teacher. Activity profiles of the auditory map A ($\sigma_A = 0.03$), the somatosensory map S ($\sigma_S = 0.045$), and the teacher map are shown in the middle column. Colors of the profiles are specified in the bottom right corner. 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; see color code in the bottom left corner. The initial synaptic pattern is only coarsely pre-wired. Despite the initially imprecise teacher, both maps develop properly. The iMT concept can therefore reproduce experiments* contra-*vision-guided map formation.*

6.4 Discussion

In summary, the present chapter has introduced the concept of integrated Multimodal Teaching. The application of iMT to map formation and alignment leads to four fundamentally new ideas. First, 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. Forth, the more precise a sensory system is, the more slowly it adapts to teaching signals. This natural consequence, which is denoted 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 simulations of multimodal teacher characteristics for map calibration on the level of spiking neurons and synaptic plasticity, namely, sSTDP, are presented. 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 [5]. Taken together, the iMT concept therefore connects two fundamental neurophysiological phenomena: inhibition-mediated map calibration and multimodal integration.

Previous theoretical studies have focused on one or parts of these aspects. Gelfand and Mysore [64, 161] 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, vision as teacher signal nevertheless has to be questioned. Stein and Wallace [223, 225] therefore suggested a connection between multimodality and map formation in cats. The latest experimental studies on barn owls from Bergan and Knudsen [5] 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. [234] 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 neuronal representation 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. These conditions are both quite implausible in the context of map alignment in ICX and OT. In concrete terms, experimental data from Bergan and Knudsen [5] suggest that teaching signals to the ICX are gated by inhibition. In addition, Kempter et al. [109] demonstrated that the assumptions underlying rate-based Hebbian learning are not necessarily valid, in particular, not in the auditory system of the barn owl.

Including these experimental results, Friedel et al. [56] 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 map alignment.

The present iMT model combines the fundamental aspects of the models from Witten

et al. [234] and Friedel et al. [56]: 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 at a first step. In a second step, the multimodal map provides a teaching signal for map calibration to all modalities. Therefore, the iMT model can explain the dominant role of vision in map alignment, plasticity in non-visual modalities without vision, plasticity within the visual modality guided by other modalities, plasticity of non-visual modalities guided by other non-visual modalities, and any combination of these settings.

Testable predictions of the iMT concept
Given that the iMT concept can reproduce both experiments supporting and questioning vision-dominated learning, suggestions regarding how the iMT concept can be experimentally verified are needed. To distinguish vision-dominated teaching from iMT, experimental setups have to bypass the dominance of vision. To this end, they 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 the teaching modality. A third possibility would be to deactivate the multimodal integration during the learning process. Given iMT, deactivation of the multimodal map(s) corresponds 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 should be observed on the learning process.

6.A Optimal combination of two modalities

The firing profile of a neuronal map can be interpreted as the likelihood to find an object at a specific position. 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

$$f^{A/V}(x) := S_{A/V} \exp\left[-(x-\mu_{A/V})^2/(2\sigma^2_{A/V})\right] \tag{6.7}$$

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. For the sake of simplicity, only one-dimensional maps are considered here.

To combine two such estimators to a common estimate μ_M a weighted sum of the mean values μ_V and μ_A is calculated as

$$\mu_M = g_V \mu_V + g_A \mu_A \tag{6.8}$$

with the weights g_V and g_A fulfilling $g_V + g_A = 1$. Replacing g_A by $g_V - 1$ the variance σ_M^2 of the combined estimator is then given as

$$\sigma_M^2 = g_V^2 \sigma_V^2 + (1 - g_V)^2 \sigma_A^2 \ . \tag{6.9}$$

In the next step the weight g_V has to minimize the variance σ_M^2, 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 therefore to

$$\begin{aligned} \sigma_M^2 &= [\sigma_A^2/(\sigma_A^2 + \sigma_V^2)]^2 \sigma_V^2 + [\sigma_V^2/(\sigma_A^2 + \sigma_V^2)]^2 \sigma_A^2 \\ &= \sigma_V^2 \sigma_A^2/(\sigma_A^2 + \sigma_V^2) \end{aligned} \tag{6.10}$$

and

$$\mu_M = \sigma_A^2/(\sigma_A^2 + \sigma_V^2)\ \mu_V + \sigma_V^2/(\sigma_A^2 + \sigma_V^2)\ \mu_A \ . \tag{6.11}$$

For mathematical fundamentals of the derivation see [222].

In the context of neuronal maps, a multiplication of the activity profiles of two maps realizes the above optimal combination scheme of the corresponding neuronal estimators $\mu_{A/V}$. The multimodal activity profile is then given by

$$\begin{aligned} f_M(x) &= \left\{ S_A \exp\left[-(x-\mu_A)^2/(2\sigma_A^2)\right] \right\} \ \left\{ S_V \exp\left[-(x-\mu_V)^2/(2\sigma_V^2)\right] \right\} \\ &= \underbrace{S_A S_V \exp\left[-1/2\,(\mu_A - \mu_V)^2 \,/\, \left(\sigma_V^2 + \sigma_A^2\right)\right]}_{\text{scaling factor}} \\ &\quad \exp\left\{ -1/2 \underbrace{\left[\left(\sigma_V^2 + \sigma_A^2\right) / \left(\sigma_A^2 \sigma_V^2\right)\right]}_{1/\sigma_M^2} \left[x - \underbrace{\left(\sigma_V^2 \mu_A + \sigma_A^2 \mu_V\right) / \left(\sigma_V^2 + \sigma_A^2\right)}_{\mu_M} \right]^2 \right\} \ . \end{aligned} \tag{6.12}$$

Standard deviation σ_M and mean μ_M of the multiplied multimodal profile are identical to the optimal values derived above; cf. (6.10) and (6.11).

6.B Optimal combination of three modalities

The optimal combination scheme for two modalities as derived in Appendix 6.A is now extended to three modalities. The multimodal estimate μ_M of three modalities, say, audition A, vision V, and the somatosensory modality S represented by their mean values μ_A, μ_V, and μ_S and standard deviation σ_A, σ_V, and σ_S is given by

$$\begin{aligned} \mu_M &= \frac{\sigma_V^2 \sigma_S^2}{\sigma_A^2 \sigma_V^2 + \sigma_A^2 \sigma_S^2 + \sigma_V^2 \sigma_S^2}\ \mu_A \\ &\quad + \frac{\sigma_A^2 \sigma_S^2}{\sigma_A^2 \sigma_V^2 + \sigma_A^2 \sigma_S^2 + \sigma_V^2 \sigma_S^2}\ \mu_V \\ &\quad + \frac{\sigma_A^2 \sigma_V^2}{\sigma_A^2 \sigma_V^2 + \sigma_A^2 \sigma_S^2 + \sigma_V^2 \sigma_S^2}\ \mu_S \ , \end{aligned} \tag{6.13}$$

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

$$\sigma_M = \frac{\sigma_A \sigma_V \sigma_S}{\sqrt{\sigma_A^2 \sigma_V^2 + \sigma_A^2 \sigma_S^2 + \sigma_V^2 \sigma_S^2}} . \tag{6.14}$$

The distance between the multimodal estimate (6.13) and, for instance, the auditory estimate A is given by

$$|\mu_M - \mu_A| = \sigma_A^2 \, \frac{\sigma_S^2 |\mu_A - \mu_V| + \sigma_V^2 |\mu_A - \mu_S|}{\sqrt{\sigma_A^2 \sigma_V^2 + \sigma_A^2 \sigma_S^2 + \sigma_V^2 \sigma_S^2}} . \tag{6.15}$$

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

6.C Model Parameters

Table 6.2 summarizes the model parameters as used in the numerical experiments; see Sec. 6.2.2.

parameter	value
number of map neurons	$N = 100$
learning synapses	
minimal strength	$\mathbf{J}_{\min} = 0.0$
maximal strength	$\mathbf{J}_{\max} = 0.25$
maximal initial strength	$(\mathbf{J}_0)_{\max} = 0.25$
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 6.2: *Model parameters as used in the numerical experiments. Shape and parameters of the learning window are taken from [56] with differing or additional parameters as indicated above.*

7. Frequently used abbreviations and functions

abbreviation / function	description
ICE	internally coupled ears
ITD	interaural time difference
IAD	interaural amplitude difference
iTD	internal time difference (at the level of the tympanic vibrations)
iAD	internal time difference (at the level of the tympanic vibrations)
CF	characteristic frequency
(s)STDP	(supervised) spike-timing-dependent plasticity
LTP	long-term potentiation
LTD	long-term depression
NA	nucleus angularis
NM	nucleus magnocellularis
SON	superior olivary nucleus
NL	nucleus laminaris
LLD	dorsal lateral lemniscal nucleus
EI neuron	neuron with excitatory and inhibitory input
SC	superior colliculus
OT	optic tectum
iMT	integrated multimodal teaching

Bibliography

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

[2] Anastasio, T. J., Patton, P. E., & Belkacem-Boussaid, K. (2000) Using Bayes' rule to model multisensory enhancement in the superior colliculus, Neural Comput **12** 1165–1187.

[3] Andersen, S. (1970) Auditory sensitivity of the harbour porpoise, *Phocoena phocoena*, In Investigations on Cetacea **2** 255–259.

[4] Autrum, H. (1940) Über Lautäußerungen und Schallwahrnehmung bei Anthropoden II., J Comp Physiol A **28** 326–352.

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

[6] 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.

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

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

[9] Brand, A., Behrend, O., Marquardt, T., McAlpine, D., et al. (2002) Precise inhibition is essential for microsecond interaural time difference coding, Nature **417** 543–547.

[10] 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.

[11] 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.

[12] Brittan-Powell, E. F. & Dooling, R. J. (2002) Auditory brainstem responses in adult budgerigars (*Melopsittacus undulatus*), J Acoust Soc Am **112(3)** 999–1008.

[13] Brunel, N. & van Rossum, M. C. W. (2007) Lapicques 1907 paper: from frogs to integrate-and-fire, Biol Cybern **97(5)** 337–339.

[14] Brunel, N. & van Rossum, M. C. W. (2007) Quantitative investigations of electrical nerve excitation treated as polarization, Biol Cybern **97(5)** 341–349.

Bibliography

[15] Buerck, M., Friedel, P., Sichert, A. B., Vossen, C., et al. (2009) Optimality in mono- and multisensory map formation, Biol Cybern **103(1)** 1–20.

[16] Burkitt, A. N. (2006) A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input, Biol Cybern **95** 1–19.

[17] Calford, M. B. (1988) Constraints on the coding of sound frequency imposed by the avian interaural canal, J Comp Psychol A **162(4)** 491–502.

[18] Calvert, G., Spence, C., & Stein, B. E. (2004) *The Handbook of Multisensory Processes* (MIT Press, Cambridge, MA).

[19] Carr, C. E. & Christensen-Dalsgaard, J. (2008) Evolution of a sensory novelty: the tympanic ears and the associated neural processing, Brain Res Bull **75** 365–370.

[20] Carr, C. E. & Christensen-Dalsgaard, J. (2009) Evolution of the auditory system, in Goldstein B. (editor), *SAGE Encyclopedia of perception*, pp. 192–194 (SAGE Publications, Thousand Oaks).

[21] Carr, C. E. & Konishi, M. (1988) Axonal delay lines for time measurement in the owl's brainstem, Proc Natl Acad Sci USA **85** 8311–8315.

[22] Carr, C. E. & Konishi, M. (1990) A circuit for detection of interaural time differences in the brain stem of the barn owl, J Neurosci **10** 3227–3246.

[23] Carr, C. E., Maler, L., & Sas, E. (1982) Peripheral organization and central projections of the electrosensory nerves in gymnotiform fish, J Comp Neurol **211** 139–153.

[24] Carr, C. E., Soares, D., Smolders, J., & Simon, J. Z. (2009) Detection of interaural time differences in the alligator, J Neurosci **29** 7978–7990.

[25] Carriere, B. N., Royal, D. W., Perrault, T. J., Morrison, S. P., et al. (2007) Visual deprivation alters the development of cortical multisensory integration, J Neurophysiol **98** 2858–2867.

[26] 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.

[27] Christensen-Dalsgaard, J. & Manley, G. A. (2005) Directionality of the lizard ear, J Exp Biol **208** 1209–1217.

[28] Christensen-Dalsgaard, J. & Manley, G. A. (2008) Acoustical coupling of lizard eardrums, JARO **9(4)** 407–416.

[29] Chung, S., Pettigrew, A., & Anson, M. (1978) Dynamics of the amphibian middle ear, Nature **272** 142–147.

[30] 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.

Bibliography

[31] Clegg, J. C. (1968) *Calculus of variations* (Oliver and Boyd, Edinburgh).

[32] Cohen, H. A. (1993) *Course in Computational Algebraic Number Theory* (Springer, Heidelberg).

[33] Coles, R. B. & Guppy, A. (1988) Directional hearing in the barn owl (*tyto alba*), J Comp Physiol A **163** 117–133.

[34] 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.

[35] Copson, E. T. (1973) *An Introduction to the Theory of Functions of a Complex Variable* (Clarendon Press, Oxford).

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

[37] Dasika, V. K., White, J. A., Carney, L. H., & Colburn, H. S. (2005) Effects of inhibitory feedback in a network model of avian brain stem, J Neurophysiol **94** 400–414.

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

[39] 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.

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

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

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

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

[44] Dyson, M. L., Klump, G. M., & Gauger, B. (1998) Absolute hearing thresholds and critical masking ratios in the European barn owl: a comparison with other owls, J Comp Physiol A **182(5)** 695–702.

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

Bibliography

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

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

[48] Fay, R. R. (1988) *Hearing in Vertebrates: a Psychophysics Data Book* (Hill-Fay Associates, Winnetka).

[49] Feng, A. S. & Shofner, W. P. (1981) Peripheral basis of sound localization in anurans. Acoustic properties of the frog's ear, Hear Res **5(2-3)** 201–216.

[50] Fletcher, N. H. (1992) *Acoustic Systems in Biology* (Oxford University Press, Oxford).

[51] Fontaine, B. & Peremans, H. (2007) Tuning bat LSO neurons to interaural intensity differences through spike-timing dependent plasticity, Biol Cybern **97** 261–267.

[52] 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** 78 106.

[53] 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.

[54] 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.

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

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

[57] Friedel, P., Young, B. A., & van Hemmen, J. L. (2008) Auditory localization of ground-borne vibrations in snakes, Phys Rev Lett **100** 48 701.

[58] Funabiki, K., Koyano, K., & Ohmori, H. (1998) The role of GABAergic inputs for coincidence detection in the neurones of nucleus laminaris of the chick, J Physiol **508** 851–869.

[59] Fuzessery, Z. M., Wenstrup, J. J., & Pollak, G. D. (1985) A representation of horizontal sound location in the inferior colliculus of the mustache bat (Pteronotus p. parnellii), Hearing Res **20(1)** 85–89.

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

[61] 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.

Bibliography

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

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

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

[65] Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986) Neuronal population coding of movement direction, Science **233** 1416–1419.

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

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

[68] Gleich, O., Dooling, R. J., & Manley, G. A. (2005) Audiogram, body mass, and basilar papilla length: correlations in birds and predictions for extinct archosaurs, Naturwissenschaften **92** 595–598.

[69] Goldberg, J. M. & Brown, P. B. (1969) Response of binaural neurons of dog superior olivary complex: an anatomical and electrophysiological study, J Neurophysiol **32** 613–636.

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

[71] Gridi-Papp, M. & Narins, P. M. (2009) Environmental influences in the evolution of tetrapod hearing sensitivity and middle ear tuning, Integr Comp Biol **49(6)** 702–716.

[72] 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.

[73] Grothe, B. (2003) New roles for synaptic inhibition in sound localization, Nat Rev Neurosci **4** 540–550.

[74] Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008) Neural correlates of multisensory cue integration in macaque MSTd, Nat Neurosci **11** 1201–1210.

[75] Gutfreund, Y., Zheng, W., & Knudsen, E. I. (2002) Gated visual input to the central auditory system, Science **297** 1556–1559.

[76] Haberly, L. B. & Bower, J. M. (1989) Olfactory cortex: model circuit for study of associative memory, Trends Neurosci **12** 258–264.

[77] 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.

[78] 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.

[79] Hebb, D. O. (1949) *The organization of behavior* (Wiley, New York).

[80] Heffner, H. E. & Heffner, R. S. (1998) Hearing, In G. Greenberg and M. M. Haraway (Eds.), Comparative Psychology, A Handbook pp. 290–303.

[81] Heffner, H. E. & Heffner, R. S. (2003) Audition, In S. Davis (Ed.) Handbook of Research Methods in Experimental Psychology pp. 413–440.

[82] Heffner, H. E. & Heffner, R. S. (2007) Hearing ranges of laboratory animals, J Am Ass Lab An Sci **46** 11–13.

[83] Heffner, H. E. & Heffner, R. S. (2008) High-frequency hearing, In P. Dallos, D. Oertel, and R. Hoy (Eds.) Handbook of the Senses: Audition pp. 55–60.

[84] Helbig, H. B. & Ernst, M. O. (2007) Optimal integration of shape information from vision and touch, Exp Brain Res **179** 595–606.

[85] Helmholtz, H. (1860/1962) *"Handbuch der Physiologischen Optik"*, volume 3 (Southall, J. P. C. (Ed.), (English trans.), New York).

[86] Helmholtz, H. (1879) *Die Thatsachen in der Wahrnehmung* (August Hirschwald, Berlin).

[87] Higgs, D. M., Brittan-Powell, E. F., Soares, D., Souza, M. J., et al. (2002) Amphibious auditory responses of the american alligator (*alligator mississipiensis*), J Comp Physiol A **188** 217–223.

[88] Hill, K. G., Lewis, D. B., Hutchings, M. E., & Coles, R. B. (1980) Directional hearing in the japanese quail (*Coturnix Coturnix Japonica*): I. Acoustic properties of the auditory system, J Exp Biol **86** 135–151.

[89] Hodgkin, A. L. & Huxley, A. F. (1952) Quantitative description of membrane current and its application to conduction and excitation in nerve, J Physiol **117** 500–544.

[90] 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.

[91] Hubel, D. H. & Wiesel, T. N. (1977) Functional architecture of macaque monkey visual cortex, Proc Roy Soc Lond B **198** 1–59.

[92] 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, Heidelberg).

[93] Hürlimann, F., Kiper, D. C., & Carandini, M. (2002) Testing the Bayesian model of perceived speed, Vision Res **42** 2253–2257.

[94] 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.

[95] Hyson, R. L., Overholt, E. M., & Lippe, W. R. (1994) Cochlear microphonic measurements of interaural time differences in the chick, Hearing Res **81(1-2)** 109–118.

[96] Hyson, R. L., Reyes, A. D., & Rubel, E. W. (1995) A depolarizing inhibitory response to GABA in brainstem auditory neurons of the chick, Brain Res **677** 117–126.

[97] Izhikevich, E. M. (2007) *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting* (MIT Press, Cambridge, MA).

[98] Jazayeri, M. & Movshon, J. A. (2006) Optimal representation of sensory information by neural populations, Nat Neurosci **9** 690–696.

[99] Jeffress, L. A. (1948) A place theory of sound localization, J Comp Physiol Psychol **41** 35–39.

[100] 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.

[101] Johnson, D. H. & Dudgeon, D. E. (1993) *Array Signal Processing: Concepts and Techniques* (Prentice-Hall, Upper Saddle River, NJ).

[102] Johnson, K. O., Hsiao, S. S., & Yoshioka, T. (2002) Neural coding and the basic law of psychophysics, Neuroscientist **8** 111–121.

[103] Jørgensen, M. B., Schmitz, B., & Christensen-Dalsgaard, J. (1991) Biophysics of directional hearing in the frog *eleutherodactylus coqui*, J Comp Physiol **168** 223–232.

[104] Joris, P. X., Smith, P. H., & Yin, T. C. T. (1998) Coincidence detection in the auditory system: 50 years after Jeffress, Neuron **21** 1235–1238.

[105] Jost, J. & Li-Jost, X. (1998) *Calculus of variations* (Cambridge University, Cambridge).

[106] 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).

[107] Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000) *Principles of neural science* (McGraw-Hill, New York), 4th (international) edition.

[108] Kay, S. M. (1993) *Fundamentals of Statistical Signal Processing* (Prentice Hall, Upper Saddle River, NJ).

[109] Kempter, R., Gerstner, W., & van Hemmen, J. L. (1999) Hebbian learning and spiking neurons, Phys Rev E **59** 4498–4514.

[110] Kempter, R., Gerstner, W., van Hemmen, J. L., & Wagner, H. (1998) Extracting oscillations. Neuronal coincidence detection with noisy periodic spike input, Neural Comput **10** 1987–2017.

[111] 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 USA **98** 4166–4171.

[112] King, A. J. (1999) Sensory experience and the formation of a computational map of auditory space in the brain, Bioessays **21** 900–911.

[113] King, A. J. (2009) Visual influences on auditory spatial learning, Phil Trans R Soc B **364** 331–339.

[114] 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.

[115] 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.

[116] 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.

[117] King, A. J. & Parsons, C. H. (1999) Improved auditory spatial acuity in visually deprived ferrets, Eur J Neurosci **11(11)** 3873–3883.

[118] Klump, G. M., Windt, W., & Curio, E. (1986) The great tit's (*parus major*) auditory resolution in azimuth, J Comp Physiol A **158(3)** 383–390.

[119] Knudsen, E. I. (1982) Auditory and visual maps of space in the optic tectum of the owl, J Neurosci **2** 1177–1194.

[120] Knudsen, E. I. (2002) Instructed learning in the auditory localization pathway of the barn owl, Nature **417** 322–328.

[121] 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.

[122] 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.

Bibliography

[123] Knudsen, E. I. & Knudsen, P. F. (1985) Vision guides the adjustment of auditory localization in young barn owls, Science **230** 545–548.

[124] Knudsen, E. I., du Lac, S., & Esterly, S. D. (1987) Computational maps in the brain, Annu Rev Neurosci **10** 41–65.

[125] Koch, C. (1999) *Biophysics of Computation* (Oxford University Press, Oxford).

[126] Konishi, M. (1993) Die Schallortung der Schleiereule, Spekt Wissensch pp. 58–71.

[127] Köppl, C. & Manley, G. A. (1990) Periphal auditory processing in the bobtail lizard *Tiliqua rugosa* - II. Tonotopic organization and innervation pattern of the basilar papilla, J Comp Physiol **167** 101–112.

[128] Köppl, C. & Manley, G. A. (1990) Periphal auditory processing in the bobtail lizard *Tiliqua rugosa* - III. Patterns of spontaneous and tone-evoked nerve-fibre activity, J Comp Physiol **167** 113–127.

[129] Körding, K. & Wolpert, D. M. (2004) Bayesian integration in sensorimotor learning, Nature **427** 244–247.

[130] Krauzlis, R. J., Basso, M. A., & Wurtz, R. H. (1997) Shared motor error for multiple eye movements, Science **276** 1693–1695.

[131] Krauzlis, R. J., Liston, D., & Carello, C. D. (2004) Target selection and the superior colliculus: Goals, choices and hypotheses, Vision Res **44** 1445–1451.

[132] Krueger, L. E. (1989) Reconciling Fechner and Stevens: Toward a unified psychophysical law, Behav Brain Sci **12** 251–320.

[133] Kuhn, G. F. (1977) Model for the interaural time differences in the azimuthal plane, J Acoust Soc Am **82** 157–167.

[134] Laming, D. (1997) *The Measurement of Sensation* (Oxford University, Oxford).

[135] Lapicque, L. (1907) Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarization, J Physiol Pathol Gen (Paris) **9** 620–635.

[136] Leibold, C. & van Hemmen, J. L. (2005) Spiking neurons learning phase delays: How mammels may develop auditory time-difference sensitivity, Phys Rev Lett **94** 168 102.

[137] Lingenheil, M. (2004) *Theorie der Beuteortung beim Krallenfrosch*, Master's thesis, Technische Universität München.

[138] Luksch, H. (2008) Sensorimotor integration: Optic tectum, in M. D. Binder, N. Hirokawa, U. Windhorst, & M. C. Hirsch (editors), *Encyclopedia of Neuroscience* (Springer, Heidelberg).

[139] 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** RC 70–74.

[140] Luo, L. & Flanagan, J. G. (2007) Development of continuous and discrete neural maps, Neuron **56** 284–300.

[141] M., Aytekin, E., Grassi, M., Sahota, & F., Moss C. (2004) The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation, J Acoust Soc Am **116** 3594C3605.

[142] 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.

[143] Manley, G. A. (1972) Frequency response of the middle ear of geckos, J Comp Physiol A **81** 251–258.

[144] Manley, G. A. (1972) The middle ear of the *tokay gecko*, J Comp Physiol A **81** 239–250.

[145] Manley, G. A. (2002) Evolution of structure and function of the hearing organ of lizards, J Neurobiol **53** 202–211.

[146] Manley, G. A., Köppl, C., & Johnstone, B. M. (1990) Periphal auditory processing in the bobtail lizard *Tiliqua rugosa* - I. Frequency tuning of the auditory-nerve fibres, J Comp Physiol **167** 89–99.

[147] 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.

[148] Manley, G. A., Yates, G. K., Köppl, C., & Johnstone, B. M. (1990) Periphal auditory processing in the bobtail lizard *Tiliqua rugosa* - IV. Phase locking of auditory-nerve fibres, J Comp Physiol **167** 129–138.

[149] Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs, Science **275** 213–215.

[150] Michelsen, A. (1971) The physiology of the locust ear, J Comp Physiol **71** 63–101.

[151] Middlebrooks, J. C. & Knudsen, E. I. (1984) A neural code for auditory space in the cat's superior colliculus, J Neurosci **4** 2621–2634.

[152] Miller, K. (1970) Least squares methods for ill-posed problems with a prescribed bound, SIAM J Math Anal **1** 52–74.

[153] Mills, A. W. (1958) On the minimum audible angle, J Acoust Soc Am **30** 237.

[154] Modrow, D., Laloni, C., Doemens, G., & Rigoll, G. (2007) 3d face scanning systems based on invisible infrared coded light, ISVC (1), LNCS **4841** 521–530.

[155] Moiseff, A. & Konishi, M. (1983) Binaural characteristics of units in the owl's brainstem auditory pathway: precursors of restricted spatial receptive fields, J Neurosci **3** 2553–2562.

[156] Monsivais, P., Yang, L., & Rubel, E. W. (2000) GABAergic inhibition in nucleus magnocellularis: Implications for phase locking in the avian auditory brainstem, J Neurosci **20** 2954–2963.

[157] 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.

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

[159] Morse, P.M. (1948) *Vibration and sound* (McGraw-Hill, New York).

[160] Mosegaard, K. & Tarantola, A. (2002) Probabilistic approach to inverse problems, in W. H. K Lee, H. Kanamori, P. C. Jennings, & C. Kisslinger (editors), *International Handbook of Earthquake and Engineering Seismology*, chapter 16, pp. 237–265 (Academic, San Diego, CA).

[161] 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.

[162] Narins, P. M. & Capranica, R. R. (1978) Two-note call of the treefrog *eleutherodactylus coqui*, J Comp Physiol **127** 1–9.

[163] Nelson, M. E. & Bower, J. M. (1990) Brain maps and parallel computers, Trends Neurosci **13(10)** 403–408.

[164] 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.

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

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

[167] 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.

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

[169] Park, T. J. & Dooling, R. J. (1991) Sound localization in small birds: Absolute localization in azimuth, J Comp Psychol **105(2)** 125–133.

[170] Park, T. J., Grothe, B., Pollak, G. D., Schuller, G., et al. (1996) Neural delays shape selectivity to interaural intensity differences in the lateral superior olive, J Neurosci **16** 6554–6566.

[171] 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.

[172] Pickles, J. O. (1988) *An Introduction to the Physiology of hearing* (Academic, London), 2nd edition.

[173] Pollak, G. D., Burger, R. M., & Klug, A. (2003) Dissecting the circuitry of the auditory system, Trends Neurosci **26(1)** 33–39.

[174] Pouget, A., Dayan, P., & Zemel, R. S. (2003) Inference and computation with population codes, Annu Rev Neurosci **26** 381–410.

[175] Puetter, R. C., Gosnell, T. R., & Yahil, A. (2005) Digital image reconstruction: Deblurring and denoising, Annu Rev Astron Astrophys **43** 139–194.

[176] Putzar, L., Goerendt, I., Lange, K., Rösler, F., et al. (2007) Early visual deprivation impairs multisensory interactions in humans, Nat Neurosci **10** 1243–1245.

[177] Rauschecker, J. P. & Kniepert, U. (1994) Auditory localization behaviour in visually deprived cats, Eur J Neurosci **6(1)** 149–160.

[178] Rayleigh, L. J. (1907) On our perception of sound direction, Philos Mag **13(74)** 214–232.

[179] Reed, M. & Simon, B. (1972) *Methods of Modern Mathematical Physics, Vol. 1: Functional Analysis* (Academic Press, New York).

[180] Röder, B., Rösler, F., & Spence, C. (2004) Early vision impairs tactile perception in the blind, Curr Biol **14** 121–124.

[181] Röder, B., Teder-Salejarvi, W., Sterr, A., Rösler, F., et al. (1999) Improved auditory spatial tuning in blind humans, Nature **400** 162–166.

[182] Rosenfeld, D. (2002) New approach to gridding using regularization, Magn Reson Med **48** 193–202.

[183] Rosowski, J. J. & Saunders, J. C. (1980) Sound transmission through the avian interaural pathways, J Comp Physiol **136** 183–190.

[184] Rowland, B. A., Quessy, S., Stanford, T. R., & Stein, B. E. (2007) Multisensory integration shortens physiological response latencies, J Neurosci **27** 5879–5884.

[185] 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.

[186] 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.

Bibliography

[187] Rschevkin, C. (1963) *A Course of Lectures on the Theory of Sound* (Pergamon Press, Oxford).

[188] 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.

[189] 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.

[190] 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.

[191] Sato, T., Fukui, I., & Ohmori, H. (2010) Interaural phase difference modulates the neural activity in the nucleus angularis and improves the processing of level difference cue in the lateral lemniscal nucleus in the chicken, Neurosci Res **66** 198–212.

[192] Saunders, S. S. & Salvi, R. J. (1993) Psychoacoustics of normal adult chickens: thresholds and temporal integration, J Acoust Soc Am **94** 83–90.

[193] Schnupp, J. W. H. & Carr, C. E. (2009) On hearing with more than one ear: lessons from evolution, Nat Neurosci **12** 692–697.

[194] Schreiner, C. E. & Langner, G. (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization, J Neurophysiol **60** 1823–1840.

[195] Seung, H. S. & Sompolinsky, H. (1993) Simple models for reading neuronal population codes, Proc Natl Acad Sci USA **90** 10 749–10 753.

[196] Shambes, G. M., Gibson, J. M., & Welker, W. (1978) Fractured somatotopy in granule cell tactile areas of rat cerebellar hemispheres revealed by micromapping, Brain Behav Evol **15** 94–115.

[197] Shams, L., Kamitani, Y., & Shimojo, S. (2000) What you see is what you hear, Nature **408** 788.

[198] 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.

[199] Song, S., Miller, K. D., & Abbott, L. F. (2000) Competitive Hebbian learning through spike-timing-dependent plasticity, Nat Neurosci **3** 919–926.

[200] Stavenga, D. G. (2002) Reflections on colourful ommatidia of butterfly eyes, J Exp Biol **205** 1077–1085.

[201] 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.

[202] Stein, B. E. & Gaither, N. S. (1981) Sensory representations in reptilian optic tectum: Some comparisons with mammals, J Comp Neurol **202** 69–87.

[203] 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 Press, Cambridge, MA).

[204] Stein, B. E. & Meredith, M. A. (1993) *The Merging of the Senses* (MIT Press, Cambridge, MA).

[205] Stein, B. E. & Stanford, T. R. (2008) Multisensory integration: Current issues from the perspective of the single neuron, Nat Rev Neurosci **9** 255–266.

[206] Stürzl, W., Kempter, R., & van Hemmen, J. L. (2000) Theory of arachnid prey localization, Phys Rev Lett **84(24)** 5668–5671.

[207] 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.

[208] Szpir, M. R., Sento, S., & Ryugo, D. K. (1990) Central projections of cochlear nerve fibers in the alligator lizard, J Comp Neurol **295** 530–547.

[209] Takahashi, T. & Konishi, M. (1986) Selectivity for interaural time difference in the owl's midbrain, J Neurosci **6** 3413–3422.

[210] Takeda, M. & Goodman, J. W. (1986) Neural networks for computation: Number representations and programming complexity, Appl Optics **25** 3033–3046.

[211] Temkin, S. (1981) *Elements of Acoustics* (Wiley, New York).

[212] Thompson, S. P. (1882) On the function of the two ears in the perception of space, Philos Mag **13(83)** 406–416.

[213] Udin, S. B. & Fawcett, J. W. (1988) Formation of topographic maps, Annu Rev Neurosci **11** 289–327.

[214] 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.

[215] Van Brunt, B. (2000) *The calculus of variations* (Springer, Heidelberg).

[216] Van Hemmen, J. L. (2001) Theory of synaptic plasticity, in F. Moss & S. Gielen (editors), *Neuro-informatics, Neural Modelling*, volume 4 of *Handbook of Biological Physics*, pp. 771–823 (Elsevier, Amsterdam).

[217] Van Hemmen, J. L. (2002) The map in your head: How does the brain represent the outside world, ChemPhysChem **3** 291–298.

[218] Van Hemmen, J. L. & Schartz, A. B. (2008) Population vector code: a geometric universal as actuator, Biol Cybern **98** 509–518.

[219] 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 Press, Cambridge, MA).

[220] Von Waldeyer-Hartz, W. (1891) Ueber einige neuere Forschungen im Gebiete der Anatomie des Centralnervensystems (About some new researches in the field of anatomy of the central nervous system), Deutsche medicinische Wochenschrift **17** 1213–1218, 1244–1246, 1287–1289, 1331–1332, 1350–1356.

[221] Vossen, C., Christensen-Dalsgaard, J., & van Hemmen, J. L. (2010) Analytical model of internally coupled ears, JASA **128(2)** 909–918.

[222] Van der Waerden, B. L. (1957) *Mathematische Statistik* (Springer, Heidelberg).

[223] 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.

[224] 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.

[225] Wallace, M. T. & Stein, B. E. (1997) Development of multisensory neurons and multisensory integration in cat superior colliculus, J Neurosci **17** 2429–2444.

[226] Wallace, M. T. & Stein, B. E. (2007) Early experience determines how the senses will interact, J Neurophysiol **97** 921–926.

[227] Wandell, B. A. (1995) *Foundations of Vision* (Sinauer Associates, Sunderland, MA).

[228] Welker, C. (1971) Microelectrode delineation of fine grain somatotopic organization of SmI cerebral neocortex in albino rat, Brain Res **26** 259–275.

[229] Werner, Y. L. & Igic, P. G. (2002) The middle ear of gekkonoid lizards: interspecific variation of structure in relation to body size and to auditory sensitivity, Hear Res **167** 33–45.

[230] Werner, Y. L., Montgomery, L. G., Safford, S. D., Igic, P. G., et al. (1998) How body size affects middle-ear structure and function and auditory sensitivity in gekkonoid lizards, J Exp Biol **201** 487–502.

[231] Wever, E. G. (1968) The ear of the chameleon: *Chamaeleo senegalensis* and *Chamaeleo quilensis*, J Exp Zool **168** 423–436.

[232] 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.

[233] 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.

[234] 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.

[235] Wooldridge, J. M. (2002) *Introductory Econometrics. A Modern Approach* (South-Western College Pub, New York), 2nd edition.

[236] Y Cajal, R. (1905) *Estructura de los centros nerviosos de las aves* (Madrid).

[237] Y Cajal, R. (1954) *Neuron theory or reticular theory: Objective evidence of the anatomical unity of nerve cells (Translated by: Purkiss, U.M. and Fox, C.A.)* (Consejo superior de investigaciones cientficas instituto Ramon Y Cajal, Madrid).

[238] Yang, L., Monsivais, P., & Rubel, E. W. (1999) The superior olivary nucleus and its influence on nucleus laminaris: A source of inhibitory feedback for coincidence detection in the avian auditory brainstem, J Neurosci **19** 2313–2325.

[239] Zeil, J. & Hemmi, J. M. (2006) The visual ecology of fiddler crabs, J Comp Physiol A **192** 1–25.

[240] 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.

[241] Zhou, Y. T., Chellappa, R., Vaid, A., & Jenkins, B. K. (1988) Image restoration using a neural network, IEEE Trans Acoust Speech Sign Process **36** 1141–1151.