Physik Department
Technische Universität München

Auditory processing: From echo suppression to object formation

Moritz Bürck

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. M. Rief
Prüfer der Dissertation: 1. Univ.-Prof. Dr. J.L. van Hemmen
2. Univ.-Prof. Dr. Chr. Leibold,
Ludwig-Maximilians-Universität München

Die Dissertation wurde am 7.7.2010 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 23.9.2010 angenommen.

Preface

The auditory system has a remarkable characteristic that renders it especially appealing from a theoretician's point of view. Making use of only two one-dimensional quantities –the deflections of the two eardrums– it unfolds a complete three-dimensional world: our auditory scene. But how can such an auditory scene in its full dynamic be reconstructed based on what seems so little information? This question has plagued many scientists throughout time and space. Already in the early 1840s G. S. Ohm –the theoretician– and A. Seebeck –the experimentalist– had vivid discussions on the underlying strategy of the auditory system [139, 165]. One could easily believe that now, more than 150 years later, this debate is merely of historical interest and the topic is finally settled. It turns out, however, that even today's evidence cannot give a definite answer to the questions that kept Ohm and Seebeck involved back in the 19th century.

To cut a long story short: the auditory system is a truly challenging topic of research. This doctoral thesis approaches two important problems in auditory scene reconstruction. Namely, it addresses the question of how the auditory system can group different frequency components from one source together so as to identify a specific signal within a mixture of sounds, and how it can efficiently cope with acoustic echoes that degrade the signal. The solutions provided here are neuronally realizable and thus extend our understanding of auditory processing in animal and man.

The thesis consists of five chapters:

In the first chapter the concept of an auditory object is introduced. In a natural environment the auditory system picks up a mixture of sound which is separated into packages of frequency components that originate from the same source. Such a package is called an auditory object. There are many cues the auditory system uses for grouping the individual frequency components into an auditory object, the most important ones being onset times and temporal modulation. On a neuronal level, both are reflected in *coherent* activity. They are, however, degraded by reflections which permanently occur in a natural environment. These reflections –echoes– thus need to be coped with in auditory processing, or rather suppressed for the reliable extraction of information from auditory scenes.

In the second chapter an optimal model for echo suppression is presented based on the mathematical concept of error minimization. It suppresses echoes and extracts original signals in various echo scenarios, even in the absence of exact information on the specific echo form. The ensuing analysis allows to link echo suppression to auditory object formation. Moreover, the model can be implemented in a neuronal network that reproduces and extends the analytical results. The neuronal realization connects smoothly to the two common mechanisms of echo suppression, a fast monaural and a slower binaural one. Finally, the present model is the first to treat echo suppression as a sensory process that realizes a fundamental principle of neuronal information processing: stochastic optimality.

In the third chapter the concept used for echo suppression in the second chapter is extended to a framework for optimal stimulus reconstruction in space-time. Again, this framework can be implemented neuronally by means of a feedforward architecture, where different delays account for temporal aspects of stimulus reconstruction, and the network connectivity pattern covers the spatial aspects. Finally, the framework is condensed into a quick guide for non-physicists which explains how to apply the presented concept to arbitrary biological setups. An example in the spatial domain for such an application, that of optimal reconstruction of a blurred visual stimulus, completes the chapter.

In the forth chapter auditory object formation is addressed by a detailed mathematical analysis of two approaches to neuronal periodicity identification. One approach relies on excitatory–excitatory interaction and results in a band-pass characteristic via the neuronal analogon to autocorrelation. The approach can principally be realized in actual biological systems, i.e., it performs well when using neuronal parameters typical for the mammalian auditory system. Surprisingly, the limitation of the performance does not arise from the neuronal membrane time constants but mainly from the temporal precision of the connections between the neurons.

The alternative approach to neuronal periodicity identification is based on excitatory–inhibitory interaction. Here the band-pass characteristics vary systematically with the time constants of excitation and inhibition. Again the model relies on biologically plausible parameters only. It works best for excitatory and inhibitory neuronal couplings of equal strength, the so-called "balanced inhibition". Interestingly, the variation of a single parameter, the inhibitory time constant, can tune the system to different frequencies. In summary both approaches allow for the grouping of different frequency components with identical temporal modulation and hence are a basis for the neuronal formation of auditory objects.

In the fifth chapter a personal perspective of the discussed results is formulated. We hereby provide a "10,000 m-above-ground" perspective covering auditory processing from echo suppression to the formation of auditory objects and conclude this thesis with concrete suggestions for follow-up research. That is, we discuss the potential of adding feedback to optimal echo suppression –the ability to cope with a dynamic environment– as well as that of applying learning theory to periodicity identification –a possible explanation for the emergence of frequency-selectivity. Finally, to turn full circle: It seems Ohm and Seebeck both have been right.

ThankUall*

*In order of appearance: My family, J. Leo van Hemmen.
Paul Friedel, Andreas B. Sichert, Christine Voßen, Peter Neubäcker, and Dr. Frank N. Furter (a scientist).

Contents

Chapter 1

Fundamentals of auditory processing

Imagine all the people you know and a couple more together in one room. People are having drinks, they are chatting, moving, flirting, and amongst them, you. Catching a word here, dropping a sentence there, you glide through the masses and effortlessly recognize friend and foe . . . – in other words, a cocktail party; cf. Fig. 1.1.

The so-called "cocktail party scenario" [36] nicely illustrates several essential auditory phenomena. Even in a *mixture of sounds* we are able to pick up a specific sound, for instance the voice of a friend. Moreover, we are able to do so in a dynamically *changing environment*, in this case the cocktail party. At the same time we perceive this specific sound as *exactly the same* we know from a static environment without any distracting sounds. So far it remains unsolved how our auditory system achieves these tasks. It is known that the auditory system decomposes the sound arriving at the ear into a large number of frequency components in the cochlea. Which of these components have arisen from which source of sound, however, is not clear *a priori*, and of crucial interest. Namely, different frequency components originating from the same sound source need to be grouped into one perceptual entity so as to allow the auditory system to recognize the identity of a signal. This grouping of frequencies and the subsequent signal recognition directly leads us to a fundamental concept in cognitive auditory processing, that of an *auditory object*.

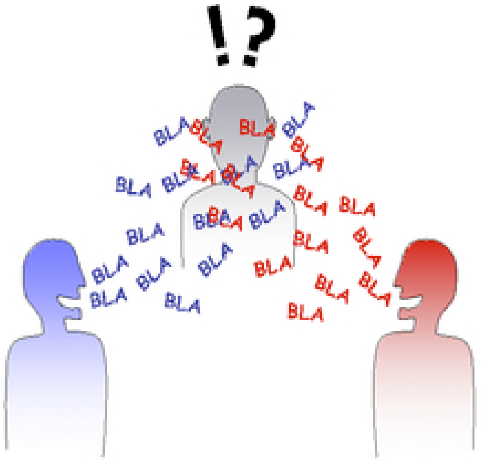

Figure 1.1: A “cocktail party”. Many different sounds are perceived simultaneously – nevertheless, we are able to effortlessly pick up a specific sound out of the mixture. We are, for instance, able to listen to the blue speaker solely, without getting distracted by the red one. This phenomenon is commonly referred to as the “cocktail party effect” and has been coined by Cherry in 1953 [36].

1.1 Definition of an auditory object

Objects as perceptual entity Before moving to auditory objects, the general concept of an object is worth pausing for a moment. What is an object and how can it be represented in the brain? Before modern science has approached these questions, the general notion of “an object” has been prone to intense philosophical considerations. Amongst them we pick George Berkeley’s because his thoughts are remarkably similar to what a modern neuroscientist may believe. Berkeley is of importance to philosophy because he denied the existence of any object [161]. According to him, objects do not exist independently of sensory experience. Rather, objects exist *because* they are perceived – they are a *mental event*. This mental event is a *perceptual entity* consisting of a bundle of characteristics, sensory perceptions. The different sensory perceptions dispose of a contiguity which leads us to mentally bind them into an “object” – a dispensable, as Berkeley says, conception since it does not add anything to the perceived characteristics. Furthermore, the conception is very subjective and ephemeral because perceptions are subject to characteristics of the individual state of mind of the observer – a very progressive point of view in the early 18th century.

Interestingly, Berkeley started the development of his philosophy with considerations on visual perception, namely “A New Theory of Vision”, first published in 1709. We see (*sic!*) that vision plays a dominant role in the development of thinking – not only Berkeley’s, but everybody’s: the Oxford Dictionary of English defines an object as “a material thing that can be seen and touched” [1]. Everybody has a concept of visual objects based on edges, color, movement, etc. A tremendous amount of research has been done on vision and the formation of visual objects, or else, the automatic segmentation of visual scenes into objects has been topic of research up to the most complex problems such as that of the units of attention [137, 151]. So far the other senses have not been investigated in comparable depth. We can thus consider the visual system as a “primus inter pares” in the sensory systems.

Consequently, when studying auditory objects, we might wonder if we can profit from insights gained in research of visual objects. The definition of a visual object may seem trivial, as pointed out above. It is a generally fruitful approach to access a problem via the opposite. Hence here: what is *not* an object? Everything is an object, of course – depending on what we are looking (or listening☺) for. Compare figure 1.2: We see either two faces or a vase. Figure and ground are determined by perceptual grouping mechanisms, subject to our individual state of mind. This refers to Berkeley’s statement of objects as a mental event. It is not by coincidence that in the groundbreaking book “Auditory Scene Analysis”[1] Bregman speaks of “streams”, not objects [22]. A stream is an event, whereas an object in the common notion includes both the source and the event perspective. We consider a written “a”, for instance. The object can either be oddly shaped ink on paper or else a meaningful vowel. This ambiguity is true for auditory objects as well. Imagine a spoken vowel, an “a” again. *A priori* the question of what the object is here, the voice as vibrating air or the vowel as meaningful symbol, cannot be answered, and we see that an (auditory) object is a perceptual entity categorized according to the task at hand.

Objects as coherent neuronal activity A “perceptual entity” is rather a vague notion we want to substantiate in the following. As initially mentioned, the problem is the grouping of different frequency components originating from the same sound source into the above perceptual entity. This is commonly referred to as the “binding problem”. An intuitive approach would be the assumption that whatever originates from the same location in space belongs to the same source. On an empirical level this hypothesis can easily be disproved by listening to a mono recording of a choral or concert, for instance. Even though arising from the same location, one

[1] This book is of visionary character since Bregman succeeded in unifying vast amounts of different, mostly experimental, research into the shared vision of “Auditory Scene Analysis”.

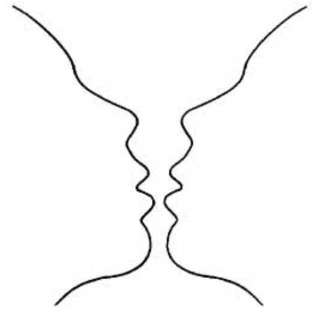

Figure 1.2: What is figure, what is ground? An object is a perceptual entity that can be categorized according to the task at hand. This categorization is strongly affected by the individual state of mind. Accordingly, we either identify a cup or two faces.

speaker, we have no difficulty in distinguishing different contributions to the whole. Furthermore, similar to visual processing, there are different pathways for handling the "what" and "where" of a signal in the auditory cortex [117,120,152,182], which hints at a parallel processing of object identity and object location. This, however, questions the notion of spatial location as a binding cue on a logical level. On an empirical level, there is evidence showing that the formation of an object must occur even *before* its localization [129]. Furthermore, experimental results suggest that enhancing object formation reduces misallocation of acoustic features across objects [40, 42, 108, 167]. Taken together, spatial location is not a good means of solving the binding problem.

A first step towards a solution to the binding problem would be, for example, to identify an auditory analogy to edge and contour. Typically, natural acoustic signals are a superposition of comodulated frequencies [135] that carry information about their source by frequency content and its fluctuations, the temporal structure, often called amplitude modulations [108]. These amplitude modulations of the signal persist in neuronal fluctuations of activity in the auditory nerve [92] and thus seem to underly a grouping mechanism that has been reported to rely on neuronal periodicity detection after the preprocessing by the cochlea [11, 25]. On the psychophysical side, there is ample evidence that common amplitude modulations serve to bind different frequencies together. In human auditory perception they are vital to speech recognition [166, 175], identification of acoustic events (the initially mentioned "cocktail party effect") [22, 36, 196], the perception of pitch [9, 92], and the "missing fundamental" effect where amplitude modulations evoke the percept of a non-existent frequency that matches the frequency of the amplitude modulations [9, 173]. In human communication, the amplitude modulations are superimposed on speech due to resonant frequencies in the human vocal tract [4]. Consequently, as every human vocal tract is different, so are the individual amplitude modulations that allow conclusions, for instance, on the speaker's size and sex, and are called "voicing frequency" [4, 171, 172]. Since this voicing frequency is imposed on any sound originating in the vocal tract, and in addition varies from speaker to speaker, it is a natural means for binding different frequency components that belong together. The importance of temporal information in speech is underlined by the nearly perfect recognition of speech under conditions of greatly reduced spectral information. With only three bands of noise modulated by the temporal envelopes of speech, the recognition rate of sentences is still above 80% [166].

Temporal information extraction is important not only for the processing of speech. In the animal kingdom, several species of echo-locating bat discriminate different insect species by their characteristic wing beat frequency that leads to a species-specific time-varying Doppler shift in the echo [164]. This time-varying Doppler shift is a realization of species-specific amplitude modulations in each frequency component of the echo; therefore the bats discriminate auditory objects by amplitude modulations.

Amplitude modulations, however, are not the only cues the auditory system takes advantage of to solve the binding problem, that is, to form auditory objects. On the psychophysical level, Bregman and Yost have subsumed seven cues for auditory object formation. They conclude that –ordered by importance– onset time, temporal modulation (i.e. amplitude modulations), offset time, duration, spectral content, level, and location in space determine what we perceive as an "auditory object" [22, 195]. What remains to investigate is the *neuronal mechanism* the auditory system employs to bind the different frequency components. In the visual system, there are basically two solutions to the "binding problem", binding by synchrony and binding by enhanced firing rate. The former establishes feature binding through neuronal synchrony in different areas of the brain processing different aspects of the same object; the latter provides feature binding through a convergent processing in higher areas of the brain (for a review see [157]). Focussing on the two most important cues for auditory object formation, common onset and common amplitude modulations, we see that they have in common a fixed phase relation. Neuronal activity in the auditory brainstem is indeed locked to the phase of the low-frequency components and the (amplitude-modulated) envelope of a sound [92, 99, 108, 181]. The phase locking, however, is not preserved in the ascending auditory pathway [160] but converted into a more stable code, increased local neuronal activity, before the inferior colliculus [69]. This conversion, the identification of neuronal periodicity and hence a strategy to form auditory objects in a neuronal network, is subject of chapter 4.

In summary, we state that an object is a perceptual entity composed of different sensory percepts. These percepts are grouped together by binding cues that mark the individual percepts as belonging to the same source. In the auditory system, common amplitude modulations are a sufficient binding cue for different frequency components stemming from one sound source. These amplitude modulations are reflected by coherent neuronal activity in the auditory brainstem. Within the auditory brainstem this coherent, phase-locked neuronal activity is converted into locally increased neuronal activity. In chapter 4 we provide two neuronal strategies that convey a *phase code* into a *rate code* and thus realize the formation of auditory objects by the neuronal identification of periodic neuronal activity.

1.2 Necessity of echo suppression

Signal degradation by echoes Following the above argument for the formation and separation of auditory objects, we could think we have now solved the cocktail party problem. If it were true you would not be reading this sentence. In the setting of the cocktail party we considered the different sources of acoustic signals, "streams", our auditory objects. What we have neglected so far is the influence of the (possibly changing) environment. In any natural environment sound is reflected. This phenomenon is usually referred to as reverberation and mainly known from large halls such as railway stations or the refectory, where the presence of many speakers plus reverberation leads to a setting very similar to the initially mentioned cocktail party. Reverberation is a consequence of sound propagating not only along the direct path from sound source to listener but also along any possible, indirect path that of course includes reflections. Obviously, the indirect paths are longer than the direct path so that a signal is followed by countless attenuated repetitions of itself; cf. Fig. 1.3. These repetitions are referred to as *echoes*.

The word "echo" itself stems from Greek mythology. It is the name of a nymph whom Zeus ordered to distract his wife Hera so that he had time for his affairs. Betrayed Hera cursed her so that she was forced to repeat whatever was said to her, hence our usage of her name. More commonly, when hearing the word "echo" we think of mountains, maybe a cathedral, or a tiled bath. There, multiple reflections are audible in contrast to most everyday situations due to large distances between walls or highly reflective surfaces, respectively. This criterion is also fulfilled in modern cities with skyscrapers, but there the echoes remain unheard because of the high amount of background noise. As anybody experiences from time to time, for instance when talking via voice-over-IP or a bad mobile phone, consciously perceived echoes are very annoying.

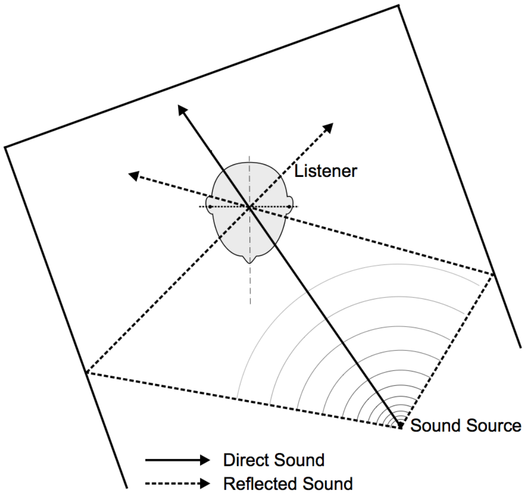

Figure 1.3: What is echo, what is sound? Sound propagates along the direct and any possible path. The resulting, misleading spatial information of the reflections –echoes– needs to be suppressed.

Echo characteristics The invention of the telephone in 1875 may indeed be the reason why algorithms for echo suppression have a long tradition in the field of engineering. The first telephones were wall-mounted with a fixed microphone into which one had to speak directly. This decoupled microphone from speaker and minimized the occurrence of dynamically changing echoes arising from an unpredictable position of the speaking person in relation to the microphone. The echoes arising within the wire can, in contrast, due to their static nature, be suppressed by relatively simple means (such as in the 1957 Bell System "Speakerphone" or by the 1960 "least mean square algorithm"; cf. Chap. 2) and hence were the first echoes to be dealt with in technical systems [76]. Nowadays, technology is much more advanced. The most sophisticated application of echo suppression algorithms in our every day life is probably echo cancellation during hands-free telephony. Here loudspeaker and microphone need to be decoupled to avoid back-coupling, and, in addition, room echoes in a dynamic environment need to be suppressed.

Concerning the suppression of echoes inside a room, the "room impulse response" (RIR) is one of the most important concepts. As the name indicates, the RIR describes the impulse response, i.e., the echoes within a room originating from a single click. A typical RIR is displayed in figure 1.4, where we clearly discern the initial click followed by many other clicks that decay exponentially in amplitude. A natural measure for the RIR is the reverberation time RT_{60}, which is defined as the time required for reflections of a direct sound to decay by 60 dB below the level of the

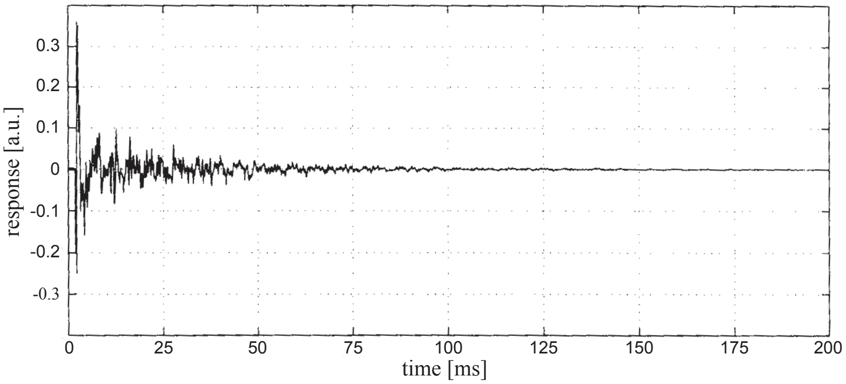

Figure 1.4: Room impulse response (RIR) measured in an office in arbitrary units [a.u.]. The initial click is followed by many other clicks, reflections of the initial click that decay exponentially in amplitude and degrade the signal.

direct sound. Just for a reminder, a decay of 60 dB is equivalent to a drop of sound pressure p to one thousandth of the initial value or, since sound energy $E \sim p^2$, to a drop of E to one millionth of the initial value. Thus, an extremely reverberating "live" environment such as a cathedral is characterized by a large reverberation time of several seconds. A large reverberation time is bad for speech understanding but well suited for music, especially organs. An exemplary value is $RT_{60} = 1.7\,\mathrm{s}$ for Carnegie Hall, New York [10], but it can reach more than 4 s as e.g. in Notre Dame de Kispest, Budapest [168], famous for its organ music[2]. The other extreme would be an anechoic chamber without any echoes, often referred to as acoustically "dead" environment, with a reverberation time of zero seconds. Reverberation times of our everyday environment lie in between these extremes; a desirable reverberation time for a typical living room, for instance, is about 0.4 s [159]. We have to keep in mind, though, that the RIR is merely a statistical description of the room acoustics that is greatly influenced by position and orientation of both sound source and listener as well as by room geometry and size.

There are several characteristics that determine the RIR. Most important, the laws for the reflection of sounds resemble the laws for the reflection of light – namely, the angle of incidence equals the angle of reflection; cf. Fig. 1.3. Similar to light, the refraction of sound is determined by the size of the structure of the reflecting surface as compared to the wavelength of the sound. This leads to frequency-dependent

[2] For connoisseurs: The organ was originally made in 1927 by Rieger Bros., opus 2256. It was reconstructed between 1995 and 2002 by László Varga according to the plans and direction of Bertalan Hock.

damping in natural rooms determined by the density of the reflecting material or medium. A tiled bath, for instance, reflects both low and high frequencies with high amplitude; when covered with a curtain only low frequency reflections will remain. Nevertheless, we have to keep in mind that the frequency components of an echo do not influence each other, i.e., any component present within the echo has already been present in the original sound. Solely the frequency-dependent damping of the room-at-hand modifies echoes, which are therefore referred to as *frequency-specific*. One exception is the Doppler shift which, although e.g. exploited by some types of bats [164], can be neglected in natural environments.

Biological echo suppression So, echoes, although sometimes hearable and sometimes not, are physically present in any natural environment all the time. In large rooms such as a church we can register them consciously as a separate auditory event. Our auditory system, however, is not optimized with respect to listening to chorals. It is optimized for survival, and survival in a possibly hostile environment (not only the cocktail party) depends on identifying and *localizing* friend and foe fast and reliably. Since in a wide space we perceive echoes with their confusing spatial information, this environment is obviously not the forte of our auditory system. In small rooms, in contrast, we do not consciously perceive echoes and their location but we can actually exploit them subconsciously –without our even noticing their existence– for an amplification of the signal; then without the additional, misleading spatial information. This makes sense for localization tasks in a small, unclear environment with many different physical objects and, consequently, many reflections. Here the auditory system is of unique importance since it is, in contrast to the visual system, omnidirectional and extremely fast (auditory neuronal time constants lie in the range of milliseconds, whereas the visual ones lie in the range of tenths of milliseconds).

For localization the auditory system mainly uses the direct sound, as proven by many behavioral and neurophysiological experiments [12, 48]. This phenomenon is known as the “law of the first wavefront” or the “precedence (formerly: Haas) effect” [63, 74]. In small and unclear environments the echoes arrive very soon after the first wavefront and are, as described in the last paragraph, not perceived consciously, hence suppressed neuronally. A small experiment gives us valuable insight into the strategy our auditory system is using for this suppression. Given we are listening to a speaker in a large lecture room. We do not perceive any echo. Now, we cover one ear. With a short delay, the speaker’s voice will suddenly sound more echoic. In a small seminar room, however, this will not work [191]. First, this reveals that there are two mechanisms for neuronal echo suppression, a monaural and a binaural one. Second, since the “switching off” of the binaural mechanism by covering one ear does not affect our perception in a small room where we have only fast echoes, the binaural mechanism is slower than the monaural one. The existence of a fast monaural and a slower binaural neuronal mechanism for echo suppression was already stated by von Békésy and Koenig more than 50 years ago [12]. The entire experiment gives a feeling for how the world would sound without echo suppression. We know echoes as a desired effect –for improving music perception in a concert hall or as an architectural feature like in the main hall of the Pinakothek der Moderne, Munich– but living in a world completely without echo suppression would be similar to living in a busy station hall. Obviously, we can consider ourselves lucky to not consciously perceive at least the fast echoes.

A mathematical estimate based on our ear-covering experiment tells us that the time span during which we do not perceive echoes covers the range of tenths of milliseconds. This means that in order to hear the reflection of a 0.2 s call as a completely separated echo we need a distance of about 35 m from the reflecting walls. Tenths of milliseconds are a lot of time in the auditory system since, as we remember, typical neuronal time constants lie in the range of milliseconds here. Therefore we can expect quite sophisticated neuronal operations to be involved. To structure our understanding of biological echo suppression, we introduce the term "echo threshold" for the maximal delay at which a perceptual fusion of signal and echo occurs. We know from psychophysical experiments that echo thresholds for impulsive stimuli typically lie in the range from $5 - 10$ ms [12, 116]. This makes them significantly shorter than the values of up to 30 ms reported for long-duration stimuli such as continuous speech or music [12, 74, 179]. We can state as a rule of thumb that echo thresholds for ongoing stimuli are about a factor five longer than echo thresholds for impulsive stimuli. This rule, however, does not hold exactly since there are ways and means to manipulate echo thresholds. Strong contextual cues in speech and music for example can extend echo thresholds [39, 75]. Alternatively, an abrupt change of source and echo location can reduce the echo threshold, a phenomenon known as the "Clifton effect" [38, 39]. In a specific setting even echo thresholds as low as $7 - 8$ ms, similar to clicks, have been measured using 500-ms complex tones [179]. These findings connect smoothly to the concept of a fast monaural and slow binaural echo suppression. The fast monaural part of echo suppression is hard-wired in the cochlear nucleus, the first nucleus in the auditory pathway and the only nucleus that receives purely monaural input [26, 191]. The slow binaural part of echo suppression, on the other hand, is very flexible and adaptive. There are models for binaural suppression that imitate the complete auditory brainstem [33, 197] and that are able to reproduce, for instance, the "Clifton effect" [197]. Such models help to pinpoint the functions of different centers within the auditory brainstem – it has been shown, for example, that a persistent inhibition in the dorsal nucleus of the lateral leminiscus, the last center in the auditory brainstem before the inferior colliculus that in turn projects to the auditory cortex, is sufficient for an ideal observer to identify echoes and accordingly exhibit echo suppression [146]. The hypothesis of a flexible, adaptive filtering mechanism for selective auditory processing in the auditory brainstem is supported by strong evidence for active suppression of irrelevant inputs [67]. The importance of echoes and of their suppression as irrelevant input is underlined by the proposition that the lagging inhibition in the mammal medial superior olive, one of the first centers in the auditory brainstem for binaural processing and commonly recognized as localization mechanism [21,71,125], originally evolved for the suppression of reverberation and echoes, and only later in evolution has been re-used for localization purposes [73].

Concluding, we state that echoes are, although mostly inaudible, omnipresent. Sometimes desired, such as in concert halls or alike, they usually are an undesired artifact degrading auditory perception. In contrast to the technical applications developed in telecommunication engineering our auditory system is capable of efficiently suppressing echoes even in dynamic environments. Hereunto it employs two distinct mechanisms for echo suppression, a fast monaural one and a slower binaural one. In the next chapter we will apply a general theoretical framework for optimal stimulus reconstruction (see Ch. 3) to auditory processing. The resulting model provides a unified access to established technical algorithms and features both a constant fast suppression and a slower part of the suppression that depends on the environment – a remarkable similarity to the biological combination of monaural and binaural mechanism for echo suppression. Thus the basic layout of biological and technical echo suppression can be founded on the mathematical principle of stochastic optimality.

Chapter 2

Optimal echo suppression

Any being is connected to the outside world through sensory systems such as vision or audition. As we could see in the last chapter, in a natural environment the information the auditory system provides is corrupted by reflections referred to as echoes. The sensory response therefore has to be processed in order to reflect the true characteristics of the outside world. In other words, the echoes need to be *suppressed*, and, if possible, *optimally*.

Optimality, vague as it is, asks for a precise definition. When extracting a signal from a specific sensory response we want the extracted signal to be as similar as possible –or even *identical*– to the original one. In a real-world scenario identity cannot be achieved because of noise and limitations of any sensory system. We can, however, minimize the difference between extracted and original signal and then define the model that features the least possible difference between original and extracted ("reconstructed") signal to be the *optimal* model. This approach of *error minimization* has already been applied successfully to sensory processing [90] such as found in the clawed frog and the pit viper [58, 170] as well as to neuronal information processing as found, e.g., in the multimodal context [45,46,118,148]. In the present work, we link the general principle of mathematical optimality to known biological mechanisms of echo suppression.

Namely, we will see that our approach allows to answer a set of fundamental questions concerning the neuronal substrate of echo suppression. Obviously, we expect delays and suppression, that is, inhibition, to play an essential role in our neuronal model setup. For a quantitative understanding, however, more specific questions need to be asked. What are the delays that play a role? What amount of inhibition is needed for which delay? Is not only inhibition but also excitation important for stimulus reconstruction? How does the form of the echo shape the model? How well does, e.g., monaural echo suppression alone work for stimulus reconstruction? All these questions can and will be answered by the approach of *optimal echo suppression.*

It is important to indicate that our approach is of purely theoretical character. We start with the principle of stochastic optimality, from where we develop a mathematical model for echo suppression which we then successfully transfer to a neuronal setup. This being a generic approach we do not aim at explaining specific anatomical details such as parts of the auditory brainstem involved in specific monaural or binaural mechanisms for echo suppression. Instead, we generate a conceptual insight into the optimal strategy for auditory signal enhancement and echo suppression as a whole. Finally, by comparing our results with what is known about biological echo suppression as well as with its correlate in the technical field, de-reverberation, we conclude with a comprehensive view on auditory signal reconstruction.

2.1 Derivation of the optimal model

The derivation of the analytical framework of the model is based on the general framework of signal reconstruction by an inverse transformation; cf. Ch. 3. Here we ultimately aim at a neuronal implementation of this framework for temporal processing as found in the auditory system. To this end, we discretize the problem in time and then concentrate on the basic conditions for signal reconstruction. By deriving the necessary neuronal connections we form the basis for a smooth integration of our model into the existing knowledge on auditory processing such as related physiological and psychophysical results.

The higher auditory system receives a sensory response $r(t)$ that can be described by means of a convolution of the original auditory signal $s(t)$ and an echo function $h(t)$ [105]

$$r(t) = \int_{-\infty}^{\infty} s(t-\tau)h(\tau)d\tau \tag{2.1}$$

with t denoting time. For the sake of simplicity we take the sensory response $r(t)$ as the amplitude of the complete signal, that is we leave out the existence of two ears as well as frequency decomposition taking place in the cochlea and study a single channel model. The echo function $h(t)$ corresponds to the “room impulse response” we introduced in chapter 1.2 and depends on the physical surrounding. Due to (2.1) the acoustic signals that cause the sensory response $r(t)$ are a linear superposition of the original auditory signal $s(t)$ and the respective echoes. Therefore, given a sufficient number of measurements, the response $r(t)$ encodes all original information. To decode this information, though, we must find a filter function $l(t)$ that suppresses echoes contained within the sensory response $r(t)$ so as to compute the reconstruction $\hat{s}(t)$ of the original signal $s(t)$. That is,

$$\hat{s}(t) = \int_{-\infty}^{\infty} r(t-\tau) l(\tau) d\tau \ . \tag{2.2}$$

We now discretize the problem in time. Every signal is then represented by a vector containing the values of the signal function sampled at discrete points in time. The convolution becomes a matrix multiplication and we denote the matrices corresponding to the echo kernel $h(t)$ and the filter kernel $l(t)$ by $\mathcal{H}$ and $\mathcal{L}$. We can thus rewrite (2.1) and (2.2) so as to read

$$\begin{aligned} r(t) = \int_{-\infty}^{\infty} s(t-\tau) h(\tau) d\tau \Rightarrow r_t = \mathcal{H}_t{}^\tau s_\tau \Rightarrow \mathbf{r} = \mathcal{H}\mathbf{s} \text{ and} \\ \hat{s}(t) = \int_{-\infty}^{\infty} r(t-\tau) l(\tau) d\tau \Rightarrow \hat{s}_t = \mathcal{L}_t{}^\tau r_\tau \Rightarrow \hat{\mathbf{s}} = \mathcal{L}\mathbf{r} \ . \end{aligned} \tag{2.3}$$

The coefficients $\mathcal{H}_t{}^\tau$ and $\mathcal{L}_t{}^\tau$ of the matrices $\mathcal{H}$ and $\mathcal{L}$ are describing how a value of the incoming signal at the time $t-\tau$ is mapped onto the output signal at time t.

Unfortunately real life is not that simple. Any sensory system has to cope with uncertainties no matter whether they come from measurement errors, variances within the assumed physical transmission process, or simply through the fact that space and time are continuous quantities that sensors and especially neurons cannot continuously represent. Of main interest to us here are the temporal dynamics which are restricted for technical sensors by the fact that they typically average over a specific amount of time, and for neurons by their refractoriness.

Altogether, the input-output relation within any system will be corrupted by errors or noise. We thus rewrite (2.3) by adding the noise term χ accounting for measurement and transmission failures to our sensory response. In addition, our signal of interest may, and in general will, be disturbed by complementary signals we pool and refer to as background noise. In our model we hence add a random variable ξ to the signal. All things considered we arrive at

$$r_t = \mathcal{H}_t{}^\tau (s_\tau + \xi_\tau) + \chi_t \ . \tag{2.4}$$

The filter function $\mathcal{L}$ in (2.3) thus needs to suppress not only the echo but also needs to cope with the noise. The task is now to find the best possible, that is, the *optimal* values for the coefficients ${\mathcal{L}_t}^\tau$ of the reconstruction matrix in (2.3). We define the expectation value of the quadratic error between the original auditory signal $\mathbf{s}$ and its reconstruction $\hat{\mathbf{s}}$ through

$$\Xi_q := \big\langle (s_t - \hat{s}_t)(s^t - \hat{s}^t) \big\rangle \ , \tag{2.5}$$

where we use the common tensor notation with upper and lower indices. We then insert (2.4) into the second equation of (2.3) and substitute the result for the reconstructed signal $\hat{s}_t$ into (2.5). Finally, minimizing Ξ with respect to ${\mathcal{L}_t}^\tau$ gives us a system of equations ($\partial \Xi_q / \partial {\mathcal{L}_\mu}^\nu = 0$) for coefficients of the reconstruction matrix ${\mathcal{L}_t}^\tau$. The optimal reconstruction matrix is thus the one with minimal expectation value for the quadratic error Ξ, that is, it is optimal in the least-square sense.

Before minimizing Ξ we need to discuss some of the terms included in (2.5). Of course, the input signal s_t is deterministic, and we would not expect any problem. But we do not know the exact values of $s(t)$ *ex ante.* We can overcome this problem by looking at "biologically relevant" signals. Such a signal belongs to a class of signals that we denote as "typical". Consequently a specific sensory signal is a concrete realization of a class of typical signals. That is, the input signal s_t is a stochastic quantity with a defined mean μ_s.

Furthermore, we take all appearing temporal cross-correlations to be zero. So the signal value at one specific point in time does not tell anything about the signal value in the next or any other time frame. Assuming a time-independent input makes the reconstruction more difficult since we assume knowing less than we actually do. We therefore can call our model a *minimal ansatz.* Consequently the results we will obtain later can be improved by including the correlation information we now disregard. Similarly we take both types of noise to be independent at different points in time, each with standard deviation σ_χ and σ_ξ, respectively. Please note that the allocative function for the noise need not be Gaussian for the model to work. Therefore only the autocorrelations remain and are given by

$$\begin{aligned} \langle s_\mu s_\nu \rangle &= \mu_s^2 \delta_{\mu\nu}, \\ \langle \chi_\mu \chi_\nu \rangle &= \sigma_\chi^2 \delta_{\mu\nu}, \quad \text{and} \\ \langle \xi_\mu \xi_\nu \rangle &= \sigma_\xi^2 \delta_{\mu\nu}, \end{aligned} \tag{2.6}$$

with $\delta_{\mu\nu}$ as Kronecker delta.

Of course one might argue that an echo is nothing but an intrinsic correlation of the auditory signal, and that we therefore should not assume (2.6). We, however, treat the echo by means of the special structure of the echo function $h(t)$ defined in (2.1).

We minimize (2.5) and obtain a linear system of equations for the coefficients of the filter function $\mathcal{L}$ depending on the echo function $\mathcal{H}$ only,

$$\mathcal{L}^{\mu}{}_{\gamma}\left[\mathcal{H}_{\nu}{}^{\delta}\mathcal{H}^{\gamma}{}_{\delta}\left(1+\eta^{2}\right)+\sigma^{2}\delta^{\gamma}_{\nu}\right]=\mathcal{H}_{\nu}{}^{\mu} \tag{2.7}$$

where $\mathcal{H}_{\nu}{}^{\delta}$ is the transpose of the matrix $\mathcal{H}^{\gamma}{}_{\delta}$. The dimensionless parameters $\sigma := \sigma_{\chi}/\mu_{s}$ and $\eta := \sigma_{\xi}/\mu_{s}$ correspond to inverse signal-to-noise ratios of σ –the mean signal strength to the variance of the detector measurement errors– and of η –the mean signal strength to the variance of the accustic noise. The filter function $\mathcal{L}$ now matches the echo function $\mathcal{H}$ and allows calculating the reconstruction $\hat{\mathbf{s}}$ of the original auditory signal $\mathbf{s}$ by means of the sensory response $\mathbf{r}$ only. Since we have neither specified the original signal nor used any information such as specific input correlations (2.6), our algorithm can reconstruct *any* arbitrary signal. To get an optimal reconstruction performance in the case of noise we simply need to adjust the two model parameters σ and η. If we have access to the noise levels σ_{χ} and σ_{ξ} as well as the typical value of the original input strength μ_{s}, the definition of σ and η gives us good estimates of these values.

Before applying the model to actual scenarios we need to consider that, as stated before, every physical environment has its own echo function. In our approach we simplify these widely varying echo functions to some typical types of echoes – in other words, we generalize. This process of generalization is necessary since, in the end, we want to provide a framework for a neuronal system that cannot afford many different sets of neuronal wiring but rather needs one single neuronal circuitry to cope with every possible situation. We therefore desire the underlying mathematical algorithm to be robust against variations of the echo so as to allow one single neuronal wiring to deal with a large set of different situations. Thus we analyze the behavior of the model performance both in absence and presence of noise, and in case of filter functions that do alternately match or not match the echo function.

Very much to our benefit, variations in the physical transmission –the echo function $\mathcal{H}$– can mathematically be included in the noise parameter χ. This leads to one of the main advantages of our reconstruction algorithm. In contrast to the common Wiener filter (for a review on linear filtering see [94]) that corresponds to our model with fixed $\eta = 0$ and $\sigma = 1$ [91, 149, 156, 158], we can adjust η and σ to the most probable situation. That is, we can adjust our model to different physical environments, noise level, and input strength. Hereby we gain an important advantage over conventional techniques as our model is robust against variations in the real noise and fits most of the possible natural situations. This characteristic will be of special importance in section 2.3 where we come to the neuronal implementation of the proposed architecture.

2.2 Model analysis for archetype echoes

We now analyze the mathematical model derived in the previous section so as to observe the intrinsic capabilities and characteristics of our approach. To this end we take a delta function in the time domain as input signal. This function corresponds to a click. In the sensory response it then appears slightly smeared out and is followed by an echo as defined by the echo function at hand. As our echo function $\mathcal{H}$ is the response of an environment, a room, to an acoustical impulse, it corresponds to the "room impulse response" (RIR) referred to in our introduction. For the sake of generality, we reduce the variety of echo functions $\mathcal{H}$ to three archetype forms in the following.

The first echo function corresponds to the simplest environment featuring an echo, which would be a single solid wall in a free space. Here, the echo would be a simple, weakened repetition of the original signal. Hence for the first echo function we assume a single, discrete reflection of a click and label the echo "d" for "discrete". In matrix notation [defined through $\mathbf{r} = \mathcal{H}\mathbf{s}$; cf. (2.3)], the echo function is hence given by

$$\mathcal{H}^{\mathrm{d}}_{\mu\nu} = \exp\left[-\frac{(\mu-\nu)^2}{2\kappa^2}\right] + \exp\left[-\frac{(\mu-\nu+\beta_{\mathrm{d}})^2}{2\kappa^2}\right] . \tag{2.8}$$

Here and in (2.9)–(2.10) all constants are real numbers ($\in \mathbb{R}$), μ and ν denote the rows and columns of $\mathcal{H}$, and $(\mu-\nu)$ indicates the relative discretized time difference. In (2.8) the parameter β_d marks the delay between signal and echo, and κ typifies how signals and echoes get broadened.

The second echo function mimics a typical room impulse response where we have not one but many walls. Hence the signal is followed by a short silence and many subsequent reflections [105]. We label this echo function "r" for "realistic" with

$$\mathcal{H}^{\mathrm{r}}_{\mu\nu} = \exp\left[-\frac{(\mu-\nu)^2}{2\kappa^2}\right] + (\mu-\nu+\beta_{\mathrm{r}})\exp\left[-\frac{(\mu-\nu+\beta_{\mathrm{r}})}{2\gamma^2}\right] . \tag{2.9}$$

The second term is a common alpha function where parameter β_r marks the delay between signal and echo. We choose it so that the reflections reach their maximum at about 20 ms after the click. The constants κ and γ denote how signals and echoes, respectively, get broadened.

For the third echo function we assume the closed space to be even more restricted and hence do not suppose any gap between signal and reflections. Here we take an exponential decay of the reflections and therefore mark it with “e”,

$$\mathcal{H}^{\mathrm{e}}_{\mu\nu} = \begin{cases} \exp\left[-\frac{(\mu-\nu)^2}{2\kappa^2}\right] & \text{if } \mu-\nu>0 \\ \exp\left[-\frac{(\mu-\nu)}{\kappa'}\right] & \text{if } \mu-\nu<0 \end{cases}, \tag{2.10}$$

the constant κ denoting how signals get broadened and κ' being a measure for the decay of the exponential tail.

By the above discretization into three exemplary echo functions we follow exemplary RIRs [76] and cover the full range of possible complex echo functions, i.e., explicitly assuming the reflecting boundaries at very far (case “d”), normal (case “r”), and close (case “e”) distance. The form of the original auditory signal $\mathbf{s}$, the detector response with echo $\mathbf{r}^{\mathrm{d,r,e}}$ as well as the echo matrix $\mathcal{H}^{\mathrm{d,r,e}}$ are depicted in figure 2.1.

To test our model, we have calculated input and reconstruction matrix for each of the three echo functions defined by (2.8) – (2.10). In the following we present an alternative, abridged version of the reconstruction matrix that is used henceforward. We evaluate its performance by comparing reconstructed and original signal for various conditions, including noise and a “mismatch” condition where echo function $\mathcal{H}$ and filter function $\mathcal{L}$ do not match; cf. Figs. 2.2 and 2.3.

The abridged version of the kernel does not, in contrast to the complete version, evaluate the sensory response $\mathbf{r}$ at times prior *and* subsequent to a specific moment t in time for the reconstruction of the original signal $\mathbf{s}$ at t. In other words, where the complete kernel takes advantage of *future* inputs the abridged version only uses the sensory response prior to t, which enables processing in *real time*. This is depicted in the top row of figure 2.2 where we see both abridged and unabridged reconstruction kernels for discrete, realistic, and exponential echo functions. The unabridged versions feature non-zero values for positive, i.e., future times whereas the abridged versions do not. The second row of figure 2.2 shows the reconstructed signals for both abridged and unabridged reconstruction kernels. The complete reconstruction kernels result in an almost perfect signal reconstruction with, if any, minimal artifacts. Using the abridged kernels leads to noticeable artifacts and a lower amplitude of the reconstructed signal; the overall quality of the reconstruction, however, is still very good. Additive Gaussian noise χ in the sensory response, i.e. assuming a noisy sensor, does not significantly change the quality of the reconstruction; see bottom row Fig. 2.2.

As stated in the derivation of our model in section 2.1, the result of an *error* or a *variation* of the echo function, say $\mathcal{H}_{\mu\nu} + \Delta\mathcal{H}_{\mu\nu}$, can mathematically be treated

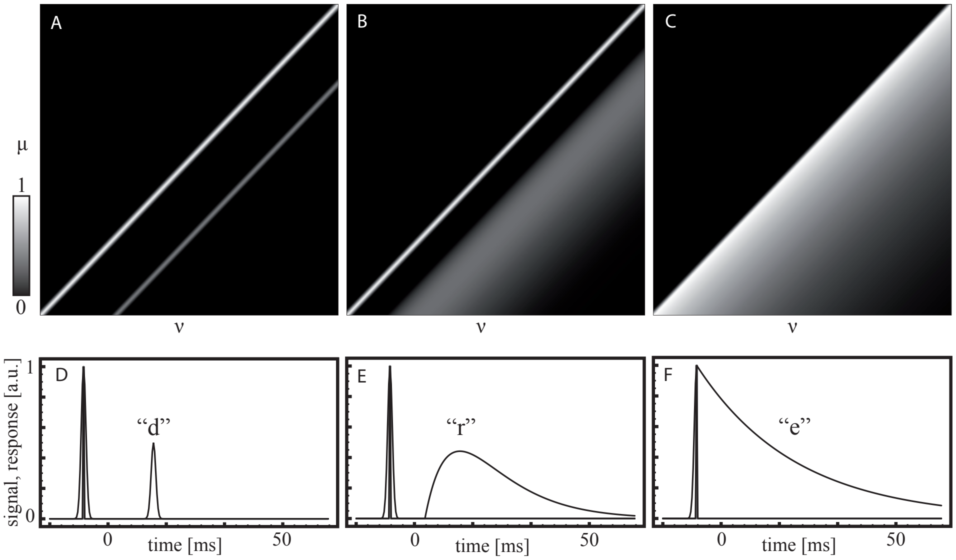

Figure 2.1: Three variations of echo functions $\mathcal{H}$ with the original signal $\mathbf{s}$ and the resulting sensory response $\mathbf{r}$. Here we have reduced all possible echo functions to three cases, viz., discrete “d” (left), realistic “r” (middle), and exponential “e” (right) echo, as explained in Sec. 2.2. The upper graphs (A–C) depict the discretized echo functions $\mathcal{H}^{\mathrm{d,r,e}}$ ($black = 0$, $white = 1$) in matrix notation as defined by (2.8) – (2.10). The lower graphs ($D - F$) show the original signal $\mathbf{s}$ (a delta function, filled line) and resulting sensory responses $\mathbf{r}^{\mathrm{d,r,e}}$ (solid line) corresponding to the auditory system given by $\mathbf{r} = \mathcal{H}\mathbf{s}$ in arbitrary units.

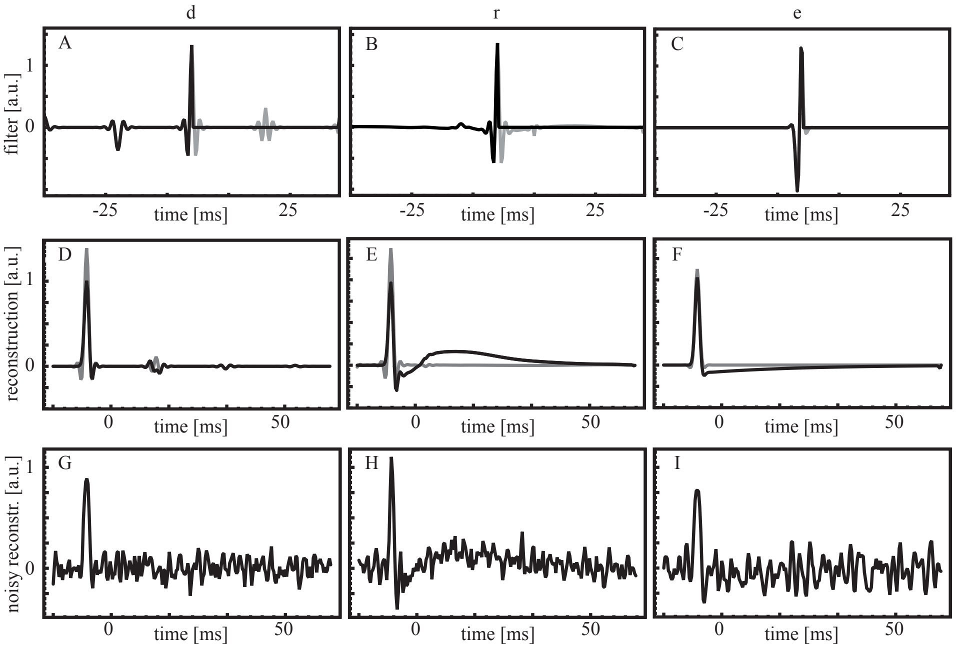

Figure 2.2: Reconstruction functions $\mathcal{L}$ and reconstructed signals $\hat{\mathbf{s}}$. Graphs $(A - C)$ represent the complete reconstruction functions $\mathcal{L}^{\mathrm{d,r,e}}$ as defined in (2.7) (grey) and the abridged version (black). The versions d, r, and e always appear as left, middle, and right. The reconstructed signals $\hat{\mathbf{s}}$ obtained through both abridged (black) and unabridged (grey) reconstruction functions are shown in graphs $(D - F)$. The complete reconstruction function $\mathcal{L}$ leads to higher peak amplitudes and less artifacts in the reconstructed signals. A noisy sensor still allows for a reliable reconstruction as depicted in $(G - I)$ with 10% noise as compared to input strength. All graphs have been expressed in arbitrary units [a.u.]. Original signal and sensory responses are as in Fig. 2.1.

as a contribution to the error χ defined in (2.4). Thus our model should react to variations of the echo function in a similar manner as to noise. Again such kind of error shows correlations, and we should change the definition of the expectation value $\langle\chi_\mu\chi_\nu\rangle$ defined in (2.6). Since there is no universal rule for the variation of the echo function in *any* arbitrary natural environment, we do not exploit this information and turn back to our minimal ansatz for showing the universality of the model presented here.

We have tested the ability of the model to suppress echoes even in case of a mismatch between echo function $\mathcal{H}$ and reconstruction filter $\mathcal{L}$. In doing so we have used the wrong reconstruction kernel for reconstructing the signal and have compared model performance with the "matching" condition. All nine possible cases are depicted in figure 2.3 and show that our reconstruction filters (mainly $\mathcal{L}^{\mathrm{r}}$) are able to cope with extreme variations –namely the wrong echo– and effectively enhance the signal. This flexibility is a prerequisite for any biological system that cannot afford a specific neuronal wiring for every possible echo (suppression-) scenario.

In our model we can tune the flexibility by choosing σ appropriately. For obtaining the results of figure 2.3 we have used an increased σ as compared to figure 2.2 ($\sigma = 5$ vs. $\sigma = 1$). Figure 2.4 visualizes the influence of σ on the reconstruction kernel for discrete, realistic, and exponential echo function. Compared to the initial kernel the modified kernel loses some of the peaks; it appears "softened" or smeared out. In particular, the complex form of the "realistic" kernel reduces to two distinct inhibitory regions; cf. Fig. 2.4. This loss of fine structure renders the three reconstruction kernels, especially for realistic and exponential echo, more similar to each other and explains why they perform better in the "mismatch" condition. The flexibility of our ansatz, the possibility of varying σ, is, as will be explained in detail in the discussion of section 2.4, a unique property that distinguishes the present model from a Wiener filter, that is, our model with fixed $\sigma = 1$ [91, 149, 156, 158].

Finally, our mathematical ansatz allows to systematically manipulate the filter function. That is, to set selected entries in the filter kernel to zero and observe the influence on signal extraction. Specifically, we have checked the impact of each of the two inhibitory regions of the "realistic" reconstruction kernel for $\sigma = 5$ as visible in figure 2.4. It turns out that the first, fast inhibitory region sharpens the contour of the signal peak and reduces the exponentially decaying tail of the echo. The second, slow inhibitory region reduces the steepness of the echo onset that is important for auditory object formation; for details we refer to the discussion in section 2.4.

In summary, the method of optimal echo suppression gives a good reconstruction of the original signal for different echo functions $\mathcal{H}$. The algorithm is able to reconstruct a signal in real time since it does not require an integration time. The reconstruction

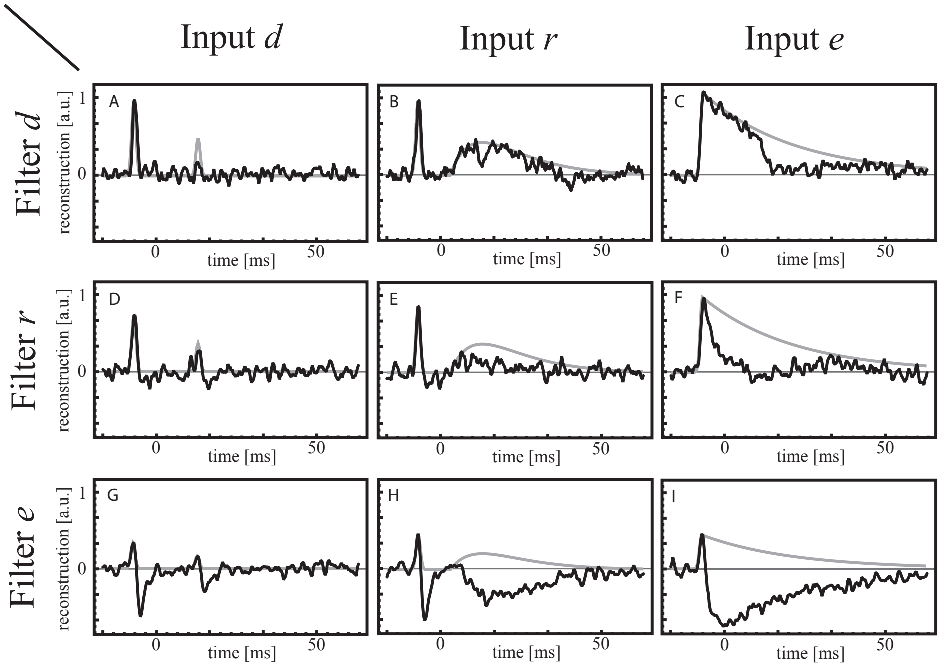

Figure 2.3: Flexibility of the model. All graphs show the reconstructed signal $\hat{\mathbf{s}}$ (black) and the actual echo function $\mathcal{H}$ (grey) in arbitrary units [a.u.]. In each column a different echo function $\mathcal{H}$ is used to calculate the sensory response whereas the reconstruction filter $\mathcal{L}$ is varied in each row. Echo function and reconstruction filter match on the diagonal (A, E, I). Even using a filter function different from the actual echo function (B, C, D, F, G, H) can lead to reasonable results if σ is chosen appropriately. Here $\sigma = 5$. In the case of function mismatch increased σ leads to results much better than the initial $\sigma = 1$, which corresponds to the Wiener filter [91, 149, 156, 158].

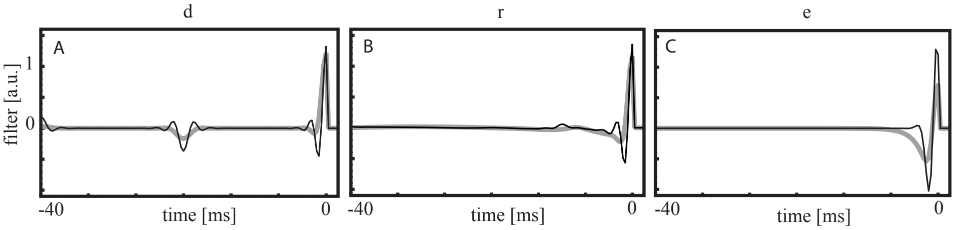

Figure 2.4: Influence of the noise's standard deviation σ on the reconstruction filter $\mathcal{L}$. The kernel of $\mathcal{L}^{\mathrm{d,r,e}}$ for $\sigma = 1$ (black) and $\sigma = 5$ (grey) in arbitrary units. As σ increases, the kernels get smeared out and lose fine structure. Simultaneously the reconstruction gains generality, which is advantageous for reconstructing a signal deformed by an unknown echo function; see Fig. 2.3. Original signal and sensory response are as in Fig. 2.1. We note that the reconstruction kernels correspond to the receptive fields of our neuronal model; cf. part 2.3.

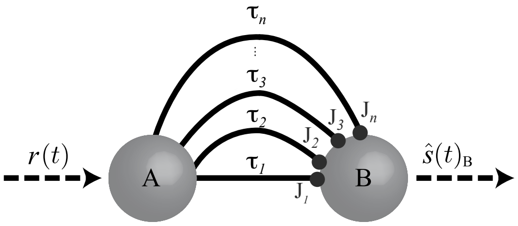

Figure 2.5: Neuronal setting. A noisy response $r(t)$ containing an echo is used to stimulate a set of neurons A, the so-called detector neurons. The resulting spikes propagate through delay lines with different delays $\tau_{1,..,n}$ and stimulate the neuronal output population B through the corresponding synaptic connection strengths $J_{1,...,n}$. The output spike train $\hat{s}(t)_B$ of the neurons B represents the original signal with suppressed echo and reduced noise.

is robust to noise, and the tuning of a single parameter σ lets the model even cope with a mismatch setting where assumed and actual echo function differ. This flexibility distinguishes our approach from common methods.

2.3 Neuronal implementation via receptive fields

A further uniqueness of our approach lies in the ease-of-transfer to a neuronal realization. The rows of the filter matrix $\mathcal{L}$, the reconstruction kernel, are essentially a temporal receptive field. The receptive field of a sensory neuron is defined as the region of space, or, in the case of a temporal receptive field, time in which the presence of a stimulus alters the activity of the neuron. Therefore we can use the value of the filter kernel at a specific time τ as the synaptic coupling strength for a neuronal feedforward connection with delay τ. Hereby our framework allows a one-to-one match of the optimal filter kernel to the neuronal connections including delay and synaptic strength. Sampling the complete kernel (see Fig. 2.4) and mapping it step-by-step to a feedforward network as depicted in figure 2.5 results in a neuronal model that suppresses the echo and extracts a signal just as in the mathematical model. We hereby do not state that echo suppression is done in biology by means of a single feedforward network. We rather give an existence proof of our model in neuronal hardware – the feedback found in the auditory brainstem could, for instance, be used to tune the parameters in a setting very similar to our feedforward approach. Moreover, we find connections to mono- and binaural echo suppression, which we are thus able to integrate conceptionally into the field of optimality and object formation found in many sensory systems in the discussion in the next section.

To account for the neuronal preprocessing and to obtain realistic spike trains, we have used noisy responses $r(t)$ from (2.4) as rate function for Poisson detector neurons [79] (neuron population A in Fig. 2.5). These neurons encode the input by producing concrete spike trains $r(t)_A$ according to an inhomogeneous Poisson process with the rate function $r(t)$. An inhomogeneous Poisson process with time-dependent rate function, $r(t)$ here, is defined by three properties. First, the probability of finding a spike between t and $t + \Delta t$ is $r(t)\,\Delta t$. Second, the probability of finding two or more spikes in this interval is $o\,(\Delta t)$, which means that we ignore their occurrence for small Δt. Third, events in disjoint intervals are independent, i.e., a Poisson process has independent increments.

As for the output neuron population B, we model them twice. In the first case, we take a set of Poisson neurons, driven by an inhomogeneous Poisson process where the density function $\hat{s}(t)$ of the spike train $\hat{s}(t)_B$ of the output neurons B is a linear function of the density function $r(t)$ of the input spike train $r(t)_A$. Thus no effect except neuronal noise masks the performance of our model, and echoes are suppressed just as in the mathematical model. But also in the second, more realistic and thus more interesting case where we take leaky integrate-and-fire neurons [66] (LIF) to model the output neurons, the signal is extracted reliably; see Fig. 2.6. When comparing top $[r(t)_A]$ and bottom $[\hat{s}(t)_B]$ row of figure 2.6 it becomes apparent that the choice of σ has a considerable impact on model performance. Again, just as in the mathematical model, an increased σ results in an improved echo suppression. Furthermore, we see that additional features of real neurons such as spontaneous rate, membrane capacity, neuronal time constants, or threshold behavior do not degrade the performance. Quite on the contrary, the threshold behavior is actually respon-

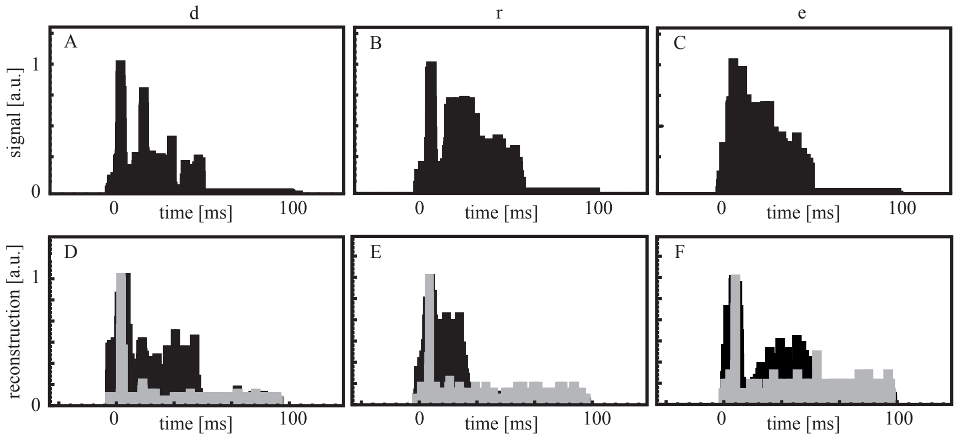

Figure 2.6: Normalized neuronal input and output activity in arbitrary units [a.u.]. The upper row depicts the activity $r(t)_A$ of the detector neurons A for the different echo functions $\mathcal{H}^{\mathrm{d,r,e}}$, the lower row the corresponding activity $\hat{s}(t)_B$ of the output neurons B for two sets of reconstruction parameters σ. We have used a population of 140 Poisson neurons to represent the input as calculated with the mathematical model (noise present) and 140 Leaky-Integrate-Fire neurons as output population. The kernel of our filter function samples the neuronal connection strength between input and output neurons; cf Fig. 2.5. For better comparability we have normalized the neuronal activity to 1. The lower row depicts neuronal activity of the output population with $\sigma = 1$ (black bars) and $\sigma = 10$ (D), 5 (E), and 10 (F) (grey bars). Choosing the right σ enhances the model performance as compared to $\sigma = 1$ that corresponds to the Wiener filter. The spontaneous firing rate of the output neurons is roughly 15 Hz, and the noise level is 5% of the input signal strength.

sible for an enhanced echo suppression of the nonlinear LIF model as compared to the linear Poisson setup.

In summary, our model for optimal echo-suppression is easy to implement neuronally, and this neuronal implementation works very well. The model thus fulfills the necessary prerequisites for a true realization in the auditory system as we discuss in the next section.

2.4 Conjunction with technics and biology

Based on the mathematical concept of error minimization, we have introduced a model for the real-time extraction of an acoustic signal from a corrupted sensory response. The model successfully extracts the original signal for three archetypes of echoes that cover the whole range of echo functions by mimicking a very large, a normal, and a very small room. The model can cope with acoustic as well as sensory noise and can even extract the original signal in the "mismatch" condition where actual and assumed echo function differ from each other. This proof of robustness justifies the consideration of our approach as a mathematical ansatz for a universal model of echo suppression.

We now focus on the links, differences, and advantages of our model in comparison to common techniques and established technical algorithms. From the technical point of view, our paper describes a single channel de-reverberation algorithm whereat we now provide a short overview.

In single channel de-reverberation one can distinguish three major groups of algorithms based on de-convolution, blind de-convolution, and suppression. The first group assumes known echo functions (i.e. RIR) and deals with approximate and fast methods for inverting it [76,131,134,142], as real echo functions are usually not exact invertible [134]. The second group tries to find a filter which accounts for a certain criterion, for instance the mean power spectral density of speech or noise reduction, and indirectly reduces the reverberation [52, 55, 68, 76, 185, 193, 194]. As in the previous case this filter converges to an approximation of the echo function inversion. The third group estimates the amount of echo power spectral density [19, 76, 112] and uses suppression methods to perform, for instance, minimum mean square estimation of the original signal. In the state-of-the-art algorithms for echo suppression, different methods are combined [61, 76, 78, 192].

Our model belongs to the first group and consequently is related to other techniques that solve the inverse problem [162] and apply linear filtering [94] such as e.g. the Wiener filter [91, 149, 156, 158]. Actually, our model generalizes the Wiener filter which is included in our model as a special set of parameters [149] ($\eta = 0$ and $\sigma = 1$). This corresponds to the situation where nothing is known about typical stimulus strength and noise level ($\mu_s \sim \sigma_\chi$).

Furthermore, the advantage of our approach to explicitly specify and link the model parameters to realistic quantities such as "typical" signal strength μ_s, input noise α and detector noise χ, or variations within the echo functions, distinguishes our model from common techniques. The technique of the pseudo-inverse and the Maximum Likelihood Estimation [91, 149, 183], which is realized *neuronally* as a fundamental computational principle in several areas of the nervous system [45, 46, 53, 83, 90, 118, 148], are namely to be mentioned here. As a consequence, our model implements the underlying concepts in a more universal way than the techniques described above by achieving signal extraction without the need for an external provision of model parameters.

Moreover, unlike realizations of the conventional techniques we mentioned, the conceptual design of our approach does not require any integration window and thus allows real-time information processing. The combined features of the proposed approach –high flexibility, well accessible parameters, and real-time processing– make technical application attractive and once more highlight the advantages of biomimetic concepts.

From the neuro-computational point of view the biological implementation of the model is straightforward. Initially, we have posed the question as to which neuronal delays play a role in echo suppression and what their synaptic coupling strengths are. Furthermore, we have been interested in the relevant time scale. By minimizing the quadratic error within the framework of the functional description of echoes we have gained a filter function or, in neuronal terms, a temporal receptive field that answers these questions. The figures 2.2 and 2.4 illustrate such receptive fields and give a quantitive answer to the above questions.

The resulting neuronal parameters, strength and delay of the synaptic connections, follow a very plausible pattern, especially when using the "realistic" echo function. The most obvious feature here is the emergence of the two distinct time scales of suppression already described in the last section. First, there is a direct, fast inhibition with a clear maximum at about 2 ms that lasts about 5 ms. Second, a delayed shallow inhibitory region follows at about 10-17 ms that does not have such a clear maximum as the fast one and is therefore called "slow inhibition". In the model, the fast inhibition sharpens the contour of the peak and reduces the exponentially decaying tail of the echo. That is, it reduces the echo amplitude but is not sufficient for stimulus reconstruction. Because of the fast time scale in a biological setting it has to be realized *monaurally.* The slow inhibition reduces the steep onset of the neuronal response to the echo. Since auditory object formation strongly relies on the onset of acoustic signals [22,195], it thus hinders the formation of the echo as aseparate object and a conscious echo perception. As stated before, the form of the slow inhibition varies with the shape of the echo, that is, with the environment at hand. That is why a flexible, binaural neuronal realization that can be modified by top-down processes makes sense here.

The above findings fit very well into today's picture of auditory processing. The detection of gap in auditory signals, auditory contrast enhancement, and echo suppression exploit properties of temporal receptive fields, more precise, delayed inhibition [28, 29, 72, 73, 188]. As for echo suppression, it is commonly understood that it is indeed in part monaural and in part binaural; cf. Sec. 1.2.

The monaural part of biological echo suppression corresponds to the fast inhibition in our model. It has a maximum at about 2 ms [77], is realized physiologically in

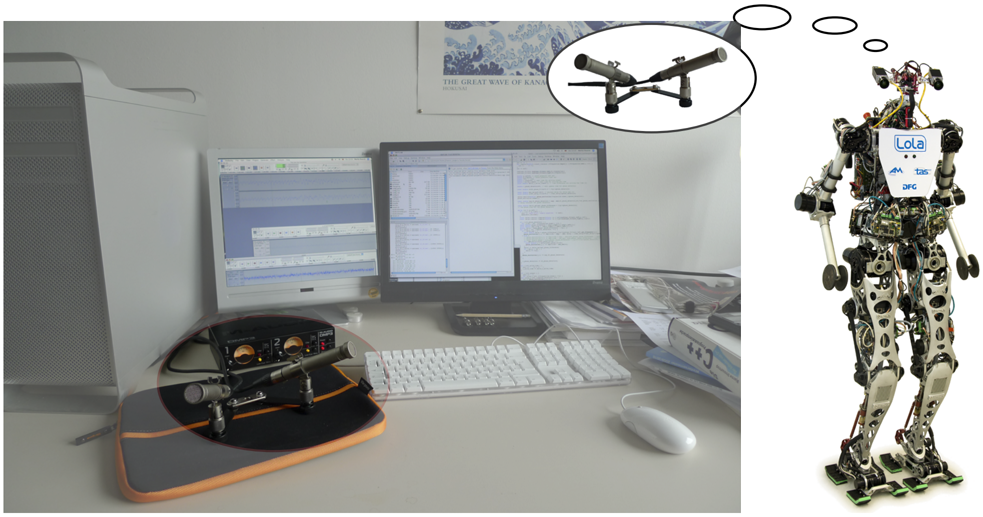

Figure 2.7: Experimental setup and future application. A pair of Oktava MK 012-01 microphones is used to record test sounds in a natural echoic environment, our office. The goal is to make our framework for echo suppression suitable for real-time processing in a multisensory robotic system, LOLA. Picture of LOLA by courtesy of the Institute of Applied Mechanics, TUM.

the cochlear nucleus [191], and has been theoretically analyzed before by Bürck and van Hemmen [28]. As for the binaural part of echo suppression, things are more complicated by nature. In general, binaural echo suppression is slower than the monaural one [12,191]. Furthermore, it has been shown that a long-lasting inhibition in the dorsal nucleus of the lateral leminiscus, called *persistent inhibition*, is involved in echo suppression, and stems from binaural mechanisms. Under *in vivo* conditions it persists up to 17 ms after stimulus offset, which suggests a correspondence of persistent inhibition and the slow inhibition in our model [146]. Interestingly, Pecka et al. [146] emphasize that their elaborate neuronal model can reproduce the "Clifton Effect" where, depending on the circumstances, echoes are alternately perceived as separate objects or not [38]. So the binaural model of Pecka et al. is able to suppress the conscious localization of the echo but at the same time allows the unconscious perception of its presence in some situations [146]. In other words, a binaural inhibitory mechanism does prevent the formation of a localizable auditory object – just as our "slow inhibition" hinders the perception of the echo as an object by reducing the steepness of the echo onset.

The analogy between the results of our framework Sec. 2.3 and biological echo suppression suggests a biomimetic application of our approach. Figure 2.7 shows the experimental setup we are currently using for evaluating our framework in a natural echoic environment, our office. The gained insights are aimed at an application in the humanoid walking robot LOLA of the Institute of Applied Mechanics, TUM. LOLA currently is only equipped with two cameras as sensory system. In cooperation with Dipl.-Ing. Thomas Buschmann the addition of two stereo microphones as shown in Fig. 2.7 will not only add the capability of omnidirectional sound source localization to the robot but also the potential for multimodal navigation and object identification. Such a multimodal approach is advantageous in an adverse environment, for instance when it comes to the identification of both spoken voice and corresponding speaker in a dynamic setting such as the cocktail party scenario.

In summary, our model can reliably extract an auditory signal that has been corrupted by different echo functions. Namely, the neuronal implementation gives reliable results and hence links smoothly to known psychophysical and physiological phenomena. Furthermore, our approach connects to established algorithms for echo suppression in technical systems. The approach thus underlines the power of a universal mathematical principle, that of stochastic optimality, applied to a biologically motivated problem leading to a shared understanding of both biological and technical solutions. The model can in principle be extended towards working in a dynamic environment by a flexible adaptation of the two model parameters. This is to be realized in cooperation with the Institute of Applied Mechanics for making LOLA a multimodal robot. The massive feedback projections that exist in the auditory brainstem could serve as a model for such a parameter adaptation. Having set up a general theoretical framework for echo suppression, a –as we could see earlier in section 1.2– necessary prerequisite for the extraction of auditory objects, we now extend the presented framework to processing in the spatial domain in the next chapter before dealing with the formation of auditory objects in chapter 4.

Chapter 3

Framework for optimal stimulus reconstruction in space-time

Our approach for optimal echo suppression in chapter 2 can be generalized into a more universal framework for optimal stimulus reconstruction in space-time [27]. We do this by first defining the generalized problem of stimulus reconstruction in space-time and then solving the ensuing ansatz by minimizing the expectation value of a squared error between estimated and real signal, similar to our proceeding in the last chapter. We show that the generalized framework can be implemented neuronally and provide a step-by-step guidance for the application of our framework to arbitrary biological sensory systems. We complete our extension of the optimal approach of chapter 2 to spatial processing with visual processing as an examplary application.

3.1 Definition of the generalized problem

Generally speaking, an object generates a stimulus $s^{\boldsymbol{x}}(t)$ varying in time t and position $\boldsymbol{x}$ 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 movement at a particular position within the visual field.

The signal induces a response $r_i(t)$ in a set of N sensory detectors. Depending on the problem at hand a single detector i with $0 \leq i \leq N$ can be a complete sensory organ, such as the hearing system as a whole we considered in the last chapter, 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

$$r_i(t) = \int_{\text{all space}} \mathrm{d}\boldsymbol{x} \int_{-\infty}^{t} \mathrm{d}\tau \; s^{\boldsymbol{x}}(\tau) h_i^{\boldsymbol{x}}(t-\tau) \tag{3.1}$$

where the *transfer function* $h_i^{\boldsymbol{x}}(t)$ incorporates the physics of signal transmission and detection, compare (2.1). The transfer function can be different for each detector i. Auditory transfer functions, for example, incorporate the position of sound source and ear with respect to the head midline, and therefore differ between right and left ear. In general, we can safely assume that $h_i^{\boldsymbol{x}}(t) = 0$ for large values of $|\boldsymbol{x}|$ and t. This reflects our intuition that events occuring far away or long ago will not influence the state of a sensor. We will need this property later on. Moreover, since any detector can only react to temporal-causal, i.e., past signals we set $h_i^{\boldsymbol{x}}(t) = 0$ for $t < 0$. We can then rewrite the response function (3.1) with adapted integration limits as a convolution with respect to time,

$$r_i(t) = \int \mathrm{d}\boldsymbol{x} \int_{-\infty}^{\infty} \mathrm{d}\tau \; s^{\boldsymbol{x}}(\tau) h_i^{\boldsymbol{x}}(t-\tau) =: \int \mathrm{d}\boldsymbol{x} \; (s^{\boldsymbol{x}} \star h_i^{\boldsymbol{x}})(t) \; . \tag{3.2}$$

We see that (2.1) is a special case where we focus on the temporal aspect and omit spatial information. Equation (3.2) describes the response of an ideal system. In biological systems the quality of the detector response is limited by at least three factors.

First, information may get lost during the transfer from the outside object to the inside sensory system. Second, noise influences all steps in the detection and reconstruction process [54]. Finally, limitations of the neuronal hardware, for instance, the limited dynamic range of receptors, constrain possible solutions; see Sec. 3.4 for details.

Within our mathematical model we incorporate these three restrictive factors by introducing additional noise terms. Accordingly, a term describing background noise $\xi^{\boldsymbol{x}}(t)$ must be added to the signal. Furthermore, we assume that transfer function and sensory response are hampered by additional noise terms $\eta_i^{\boldsymbol{x}}(t)$ and $\chi_i(t)$, respectively. Consequently (3.2) is modified so as to read

$$r_i(t) = \int \mathrm{d}\boldsymbol{x} \; [(s^{\boldsymbol{x}} + \xi^{\boldsymbol{x}}) \star (h_i^{\boldsymbol{x}} + \eta_i^{\boldsymbol{x}})](t) + \chi_i(t) \; . \tag{3.3}$$

To reconstruct the estimated signal from the detector responses $r_i(t)$, the above transformation must be "inverted" in some appropriate way. We therefore calculate

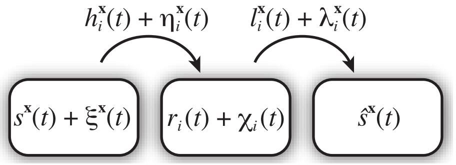

Figure 3.1: *Physical mapping*: signal $s^{\boldsymbol{x}}(t)$ with background noise $\xi^{\boldsymbol{x}}(t)$ is mapped onto a noisy receptor response $r_i(t) + \chi_i(t)$ through the noisy transfer function $h_i^{\boldsymbol{x}}(t) + \eta_i^{\boldsymbol{x}}(t)$. *Optimal stimulus reconstruction*: the (possibly noisy) inverse transfer function $l_i^{\boldsymbol{x}}(t) + \lambda_i^{\boldsymbol{x}}(t)$ gives an estimate $\hat{s}^{\boldsymbol{x}}(t)$ of the signal.

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

Table 3.1: Functions and error terms describing detection and processing of sensory information.

the time-dependent inverse transfer functions $l_i^{\boldsymbol{x}}(t)$ between detector i and the map at position $\boldsymbol{x}$. When applying $l_i^{\boldsymbol{x}}(t)$ to the receptor responses at i we obtain the estimate

$$\hat{s}^{\boldsymbol{x}}(t) = \sum_i [r_i \star (l_i^{\boldsymbol{x}} + \lambda_i^{\boldsymbol{x}})](t) \tag{3.4}$$

of the original signal $s^{\boldsymbol{x}}(t)$, analogously to (2.2) in chapter 2. Here the hat on $\hat{s}^{\boldsymbol{x}}(t)$ denotes a reconstruction and the term $\lambda_i^{\boldsymbol{x}}(t)$ represents the noise due to the concrete realization of the theoretical inverse transfer function. We note that in contrast to elsewhere [144, 150] the present model is non-iterative. This will result in a purely feedforward network structure when it comes to a neuronal realization in section 3.4.

Figure 3.1 illustrates the complete mathematical procedure of sensory information processing. All the relevant terms are summarized in table 3.1. In the next section we will indicate how to calculate inverse transfer functions $l_i^{\boldsymbol{x}}(t)$ that enable optimal signal reconstruction.

3.2 Framework for optimal reconstruction

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

This is possible because all quantities and functions (cf. Fig. 3.1) involved in both the process of physical mapping and the neuronal process of optimal signal reconstruction are self-averaging; cf. Sec. 3.4. The mathematical definition of self-averaging allows for a description in terms of expectation values [27].

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

$$\begin{aligned} E\{\boldsymbol{l}^{\boldsymbol{x}}(t), t\} &:= \left\langle \int_{t-T}^{t} \mathrm{d}t' \int \mathrm{d}\boldsymbol{x} \ \left[s^{\boldsymbol{x}}(t') - \hat{s}^{\boldsymbol{x}}(t')\right]^2 \right\rangle \\ &= \int_{t-T}^{t} \mathrm{d}t' \int \mathrm{d}\boldsymbol{x} \ \left\langle \left[s^{\boldsymbol{x}}(t') - \hat{s}^{\boldsymbol{x}}(t')\right]^2 \right\rangle \ . \end{aligned} \tag{3.5}$$

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

To be mathematically precise, an expectation value is an integral on a probability space with respect to a probability measure p. For arbitrary functions f and g, if $\langle |f-g|^2 \rangle = 0$ then $f = g$ *with respect to* p or, physically, looking at the world through p’s glasses: what p finds important pops up clearly whereas what p finds “irrelevant” has hardly any weight. The latter need not correspond to what we “think” ourselves; see van der Waerden [186].

A quadratic form of the error term has been proven to be a reasonable and practical choice in many physical optimizing problems; see, e.g., [130]. 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 [27, 91, 98].

Mathematically, the error (3.5) is a functional assigning to every set of inverse transfer functions one specific value. Minimization of functionals in the above integral form is a central and well-studied aspect of the calculus of variations [24, 37, 65, 93]. For the present situation the first variation with respect to every inverse transfer function $l_j(\boldsymbol{x}, t')$ is to vanish. That is,

$$\frac{\partial \left\langle [s^{\boldsymbol{x}}(t') - \hat{s}^{\boldsymbol{x}}(t')]^2 \right\rangle}{\partial\, l_j^{\boldsymbol{x}}(t')} = 0 \qquad \text{for every } j. \tag{3.6}$$

In order to solve (3.6), we have to substitute (3.4) for the estimate $\hat{s}^{\boldsymbol{x}}(t)$ and replace $r_i(t)$ by its description (3.3). Expanding the square, we encounter expectation values of products consisting of varying combinations of noise and signal terms. Here we assume that all noise terms as well as the signal itself are stochastically independent of each other so that the expectation of a product of independent term factorizes; for instance,

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

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

$$\left\langle f^{\boldsymbol{x}}(t)f^{\boldsymbol{x}'}(t')\right\rangle = \delta(\boldsymbol{x}-\boldsymbol{x}')\delta(t-t')(\mu_f^2+\sigma_f^2) \tag{3.7}$$

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

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

$$\left\langle \xi^{\boldsymbol{x}}(t)\xi^{\boldsymbol{x}'}(t')\right\rangle = \delta(\boldsymbol{x}-\boldsymbol{x}')\delta(t-t')\sigma_\xi^2 \ , \tag{3.8a}$$

$$\left\langle \chi_i(t)\chi_j(t')\right\rangle = \delta_{ij}\delta(t-t')\sigma_\chi^2 \ , \tag{3.8b}$$

$$\left\langle \eta_i^{\boldsymbol{x}}(t)\eta_j^{\boldsymbol{x}'}(t')\right\rangle = \delta_{ij}\delta(\boldsymbol{x}-\boldsymbol{x}')\delta(t-t')\sigma_\eta^2 \ , \tag{3.8c}$$

$$\text{with} \quad |\boldsymbol{x}| < x^{\max} \quad \text{and} \quad 0 < t < t^{\max} \ . \tag{3.8d}$$

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

The autocorrelation (3.7) of the signal $s^{\boldsymbol{x}}(t)$ 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, we choose the expectation value μ_s^2 of the signal as the appropriate quantity and therefore take σ_s^2 zero,

$$\left\langle s^{\boldsymbol{x}}(t)s^{\boldsymbol{x}'}(t')\right\rangle = \delta(\boldsymbol{x}-\boldsymbol{x}')\delta(t-t')\ \mu_s^2 \ . \tag{3.9}$$

While (3.8) incorporates reasonable assumptions for all noise terms, the correlation (3.9) for the signal is a strong hypothesis. Signals are characterized by spatio-temporal continuity. That is, objects and their corresponding signals usually do not disappear from one point in time or space to the next. A Gaussian correlation term

$$\left\langle s^{\boldsymbol{x}}(t)s^{\boldsymbol{x}'}(t')\right\rangle = A\exp\left(-\left|\boldsymbol{x}-\boldsymbol{x}'\right|^2/(2\sigma_x^2)\right)\exp\left(-\left|t-t'\right|^2/(2\sigma_t^2)\right)\ , \tag{3.10}$$

for instance, can take into account correlations between neighboring points in space and time. Here σ_x and σ_t are typical spatial and temporal correlation scales. The application of such a Gaussian correlation, however, does not greatly alter the further derivation but only smoothens the final estimated signal; for details see [27]. For reasons of clarity, we will therefore stick to the relation (3.9).

Returning to the (3.6) we have to solve it, and in so doing apply the correlations (3.8) and (3.9) so as to arrive at

$$l_j^{\boldsymbol{x}}(t)\left[\sigma_\chi^2+(\mu_s^2+\sigma_\xi^2)\int_{\substack{|\boldsymbol{y}|<y^{\max}\\ 0<\tau<t^{\max}}}\mathrm{d}\boldsymbol{y}\mathrm{d}\tau\ \sigma_\eta^2\right]+(\mu_s^2+\sigma_\xi^2)\sum_i\int\mathrm{d}\boldsymbol{y}\ \left[(h_i^{\boldsymbol{y}}\star l_i^{\boldsymbol{x}})\circ h_j^{\boldsymbol{y}}\right](-t) = \mu_s^2 h_j^{\boldsymbol{x}}(-t)\ ; \tag{3.11}$$

again we refer to [27] for an extensive calculation. The open circle $\circ$ denotes the autocorrelation integral

$$(a\circ b)(t) := \int_{-\infty}^{\infty}\mathrm{d}\tau\ a(\tau)b(t+\tau)\ . \tag{3.12}$$

In order to simplify (3.11), as in the last chapter, we define two new noise measures,

$$\tau^2 := \frac{\sigma_\xi^2}{\mu_s^2} \tag{3.13}$$

and

$$\sigma^2 := \frac{\sigma_\chi^2}{\mu_s^2}+\int_{\substack{|\boldsymbol{y}|<y^{\max}\\ 0<\tau<t^{\max}}}\mathrm{d}\boldsymbol{y}\mathrm{d}\tau\ \frac{\sigma_\eta^2(\mu_s^2+\sigma_\xi^2)}{\mu_s^2}\ . \tag{3.14}$$

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

In order to further simplify (3.11) we switch to Fourier space, where convolution (3.2) and correlation (3.12) become ordinary multiplications combined with complex conjugations. Denoting Fourier transforms by capital letters and the complex conjugation by an overline, (3.11) simplifies to

$$\sum_i L_i^{\boldsymbol{x}}\left[\sigma^2\delta_{ij}+(1+\tau^2)\int\mathrm{d}\boldsymbol{y}\ H_i^{\boldsymbol{y}}\overline{H_j^{\boldsymbol{y}}}\right]=\overline{H_j^{\boldsymbol{x}}} \tag{3.15}$$

where we have used (3.13) and (3.14).

Equation (3.15) is the main result of our derivation. In principle, it allows us to calculate the inverse transfer functions $L_i^{\boldsymbol{x}}$ for optimal signal reconstruction. A calculation of the second variation confirms that the inverse transformation we have found indeed minimizes the error [27]. For convenience we will introduce an alternative notation in the next section.

3.3 Alternative notation using matrices

To rewrite (3.15) in a more practical notation we introduce, very similar to (2.3), "matrices" $\mathcal{H}$ and $\mathcal{L}$ by putting

$$\mathcal{H}_{[ix]} = H_i^{\boldsymbol{x}} \ , \qquad \mathcal{L}_{[xi]} = L_i^{\boldsymbol{x}} \ . \tag{3.16}$$

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

In addition, we introduce the covariance matrix $\mathcal{C}(\boldsymbol{R})$ of the receptor response $\boldsymbol{R}$ as described, e.g., in [91, 98]. In our case we find

$$\begin{aligned} \mathcal{C}(\boldsymbol{R}) &:= \left\langle (\boldsymbol{R} - \langle \boldsymbol{R} \rangle)\overline{(\boldsymbol{R} - \langle \boldsymbol{R} \rangle)}^T \right\rangle & (3.17) \\ &= \mu_s^2 \left(\sigma^2 \mathbb{1} + \tau^2 \overline{\mathcal{H}} \cdot \mathcal{H}^T \right) & (3.18) \end{aligned}$$

where the superscript T denotes the matrix transpose and $\mathbb{1}$ the identity matrix. Equation (3.15) now simplifies to

$$\mathcal{M} \cdot \mathcal{L}^T = \overline{\mathcal{H}} \text{ with } \mathcal{M} := \mu_s^{-2} \mathcal{C} + \overline{\mathcal{H}} \cdot \mathcal{H}^T \ . \tag{3.19}$$

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

$$\mathcal{L} = \left(\mathcal{M}^{-1} \overline{\mathcal{H}} \right)^T = \overline{\mathcal{H}}^T \left(\mu_s^{-2} \mathcal{C} + \mathcal{H} \cdot \overline{\mathcal{H}}^T \right)^{-1} \ . \tag{3.20}$$

This equation gives a unique solution for the optimal reconstruction for any given set of transfer functions and noise constants (σ, τ). Using (3.4) in matrix form we find

$$\hat{\boldsymbol{S}} = \mathcal{L} \cdot \boldsymbol{R} \tag{3.21}$$

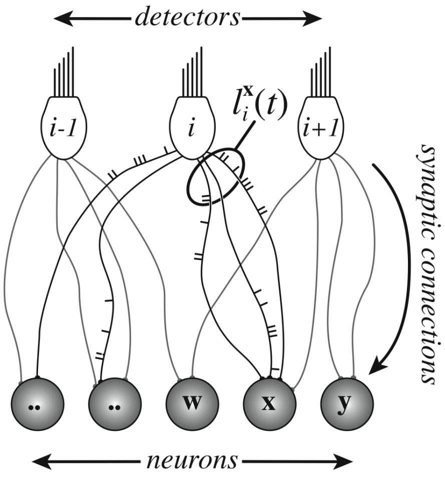

Figure 3.2: Neuronal realization of optimal stimulus reconstruction. Each sensor (here hair cells labeled i) connects to several neurons that represent sensory space. These neurons (encoding the location $\boldsymbol{x}$) may receive (multiple) connections from each sensor. Each connection has a well-defined strength and temporal delay t. In this way, the transformation $l_i^{\boldsymbol{x}}(t)$ can be reliably represented in a neuronal network [58].

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

3.4 Neuronal realization of the framework

In section 2.3 we transferred the mathematical model for optimal echo suppression to a neuronal realization. We now translate the general mathematical algorithm of optimal stimulus reconstruction into a concrete neuronal context and justify the validity of this translation. We therefore verify whether the assumptions we have employed in the above derivation are fulfilled in neuronal processing. That is, we check whether the neuronal quantities and functions of optimal stimulus reconstruction are *self*-averaging. To this end we note on the one hand that firing of neurons is correlated with neuronal input and that neuronal noise can be described by a stochastic process, e.g., a Gaussian one; we will see below why. Our framework can cope with any distribution of neuronal noise as long as the mean is zero. On the other hand the optimal inverse transfer functions $l_i^{\boldsymbol{x}}(t)$ are *learned* synaptic connections between the internal representation of a sensory modality and the corresponding sensory input, hence reflect properties of the underlying learning process. Effective learning is slow because it needs many independent repetitions. Accordingly time scales for learning and individual realizations of an external signal can be separated. In other words, learning is a self-averaging process where only *averaged* quantities enter by the very nature of the process; see [101]. Quantities and functions within the physical mapping process are self-averaging as well [27]. In conclusion, the conditions needed to exploit the mathematical framework as derived above are fulfilled.

As a consequence, our neuronal realization of optimal echo suppression is valid, and we can even translate the general inverse transfer functions $l_i^{\boldsymbol{x}}(t)$ into neuronal hardware. In such an architecture, the actual spatio-temporal processing is performed by the synaptic connections between neurons and detectors. Spatial processing is governed by the topographic structure of the network that defines which detector is connected to which neuron. Temporal processing on the other hand is determined by the distribution of delays within the set of connections. Figure 3.2 shows an example of such a neuronal setup.

In the present derivation we have already taken into account the discrete character of detectors and the ensuing representation 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 $\lambda_i^{\boldsymbol{x}}$. That is, we are left with the temporal discretization of the inverse transfer functions $l_i^{\boldsymbol{x}}(t)$. This discretization is realized by a sampling procedure where a number of dendrites with appropriate delays is chosen to represent the complete $l_i^{\boldsymbol{x}}(t)$. It has indeed been shown that a limited number of synaptic connections suffices to sample the time course of $l_i^{\boldsymbol{x}}(t)$ [58]. Even more so, the response of the neurons representing sensory space is robust with respect to the sampling method of the temporal delays [115] as well.

Consequently, as illustrated by figure 3.2, our unified framework can be implemented by means of a simple feedforward network of excitatory and inhibitory connections in order to form a neuronal representation of an arbitrary input [58, 115, 170]. It does not, however, explain how such a connectivity pattern is established in a real biological system. Here the correct synaptic connections have to be *learned*. It has been shown [56, 60] that a teacher such as the visual system can generate correct synaptic strengths so that a representation can indeed develop in other modalities by means of (supervised) STDP [27]. Thanks to the present method we can compare the learned connectivity pattern with the optimal one as given by (3.15) and (3.20).

A meaningful comparison of the mathematically optimal network architecture with an actual biological setup, though, may not be straightforward. In real biological systems, the reduction of the error to its minimum as in (3.5) –that is, realizing the optimal connectivity– may not be possible because of neuronal limitations. The limited neuronal accuracy that results can be included into our framework by reducing the error only below a certain error threshold, which may even vary in space. For instance, the sampling arrays of animal eyes are non-uniform, with different parts of the visual field being sampled with different spatial and spectral resolution [85, 178, 198]. Such a focus on specific spatio-temporal domains can mathematically be realized by introducing a positive weighting function into the integral in (3.5). Accordingly, when reducing the global error below a certain threshold, the areas within the focus of the weight function have to reach a higher level of optimization, i.e., of resolution, than the rest.

Taken together, the formalism of optimal stimulus reconstruction provides us with an optimal neuronal connectivity pattern for stimulus reconstruction in space-time. The optimal echo suppression we have investigated in chapter 2 is one potent example for stimulus reconstruction in time, but the framework can be extended towards many possible applications in understanding neuronal processing of sensory signals. We now want to fathom the capabilities of the framework by providing an easy how-to guide for the application of our framework to other sensory systems and illustrate this guide with an exemplary application of our framework to visual processing.

3.5 Exploring space-time as non-physicist

Up to now we have shown that an optimal connectivity pattern between sensory system and signal representation can be calculated [Fig. 3.1 and Eq. (3.15)] and that it can be realized neuronally (Fig. 3.2). We now focus on concrete applications of our framework. To this end, we provide a *simple* how-to that summarizes the mathematical concepts discussed above. Following this recipe step by step we then demonstrate how to arrive at optimal stimulus reconstruction not only in the temporal but also in the spatial domain.

We bring to life the generalized mathematical framework by presenting a quick guide that allows also the mathematically untrained to find the optimal network connectivity in a realistic biological setup:

- First, we derive the transfer function $h_i^{\boldsymbol{x}}(t)$ that determines the response of the detector i to a stimulus pulse that occurred t time units ago at position $\boldsymbol{x}$.
- Next, we calculate the Fourier transform $H_i^{\boldsymbol{x}}$ of the transfer function $h_i^{\boldsymbol{x}}(t)$.
- We choose suitable values of τ and σ. In general the noise-to-signal ratio τ can be assumed to be much smaller than 1 for any measurable signal. In contrast, σ needs to be estimated in dependence upon the situation at hand [57,58,170].

- We then calculate the matrix entries M_{ij} as given by Eq. (3.19) and invert the model matrix $\mathcal{M}$.
- We multiply the inverted matrix $\mathcal{M}^{-1}$ by the vector $\overline{H_i^{\boldsymbol{x}}}$ so as to find the input connection strengths $L_i^{\boldsymbol{x}}$.
- Finally, we calculate the inverse Fourier transform of $L_i^{\boldsymbol{x}}$ so as to find the connection strengths $l_i^{\boldsymbol{x}}(t)$.

In the following we will demonstrate the power of the above how-to through an example, the derivation of optimal reconstruction in the spatial domain in the visual system.Within the visual system each sensory neuron is basically tuned to a particular spatial position. In mathematical terms, every retinal neuron i receives input from a spatial position $\boldsymbol{x}_i$, its preferred position, and neighboring positions within a region determined by resolution ρ. The transfer function corresponding to such a sensory system is

$$h_i^{\boldsymbol{x}}(t) = \exp\left(-\frac{|\boldsymbol{x}-\boldsymbol{x}_i|^2}{2\rho^2}\right)\delta(t)\ , \tag{3.22}$$

and its Fourier transform reads

$$H_i^{\boldsymbol{x}} = \exp\left(-\frac{|\boldsymbol{x}-\boldsymbol{x}_i|^2}{2\rho^2}\right)\ . \tag{3.23}$$

Within our exemplary setup we assume that the signal position $\boldsymbol{x} = (u, v)$ encodes positions $u, v \in [-1/2, 1/2]$. As a reminder, we have rescaled positions so as to make them dimensionless and fit in the square $[-1/2, 1/2]^2$. From the above ansatz (3.22) and (3.15) we calculate the matrix components

$$\begin{aligned} M_{ij} &= \sigma^2\delta_{ij} + (1+\tau^2)\exp\left(-\frac{|\boldsymbol{x}_i-\boldsymbol{x}_j|^2}{4\rho^2}\right) \\ &\times \left[\operatorname{erf}\left(\frac{u_i+u_j-1}{2\rho}\right) - \operatorname{erf}\left(\frac{u_i+u_j+1}{2\rho}\right)\right] \\ &\times \left[\operatorname{erf}\left(\frac{v_i+v_j-1}{2\rho}\right) - \operatorname{erf}\left(\frac{v_i+v_j+1}{2\rho}\right)\right] \end{aligned} \tag{3.24}$$

where $\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}}\int_0^x \exp\left(-y^2\right)dy$ is the error function. To find the connection strengths $l_i^{\boldsymbol{x}}$, we numerically calculate the model matrix $\mathcal{M}$ for a discretized space and parameters $\sigma = 1$ and $\tau = 0$. With the matrix $\mathcal{M}$ we then determine the

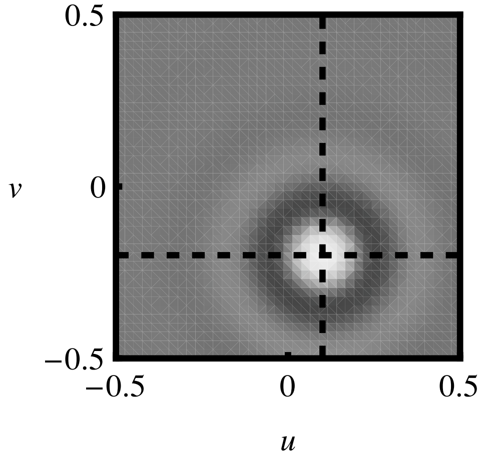

Figure 3.3: Spatial receptive field. Connection strengths to a neuron encoding the position $(u, v) = (0.1, -0.2)$. The sensory neurons are distributed on a 40×40 grid with preferred positions $u, v \in [-1/2, 1/2]$ and a tuning curve width $\rho = 0.9$. We chose $\sigma = 1$ and $\tau = 0$. A clear center-surround receptive field emerges. Receptor neurons that have a preferred position matching that of the encoding neuron have excitatory connections (white spot). Receptor neurons having a slightly off-set position *inhibit* the encoding neuron (dark circle). Neurons with preferred positions far away from the encoding neuron have connection strength zero (gray).

connection strengths $\boldsymbol{L}$. By an inverse Fourier transformation we can numerically obtain $l_i^{\boldsymbol{x}}$ for each position $\boldsymbol{x}$ as shown in figure 3.3. Here the connections from all receptors to the encoding neuron i, i.e., its receptive field, are plotted for an arbitrary preferred position $\boldsymbol{x}_i = (0.1, \ -0.2)$. Clearly, the receptors encoding the preferred position have strong projections to the encoding neuron (bright spot in Fig. 3.3) but, interestingly, the receptors that encode slightly differing locations contribute negatively (dark circle in Fig. 3.3).

Such a center-surround profile is called “Mexican hat” and is, e.g., realized by lateral inhibition, a well-known phenomenon first described by Mach [119] in the visual system in 1866. Up to now this mechanism, studied in the mammalian visual system [96, 187], has been discovered as well in, for instance, insect vision [89], snake infrared vision [170, 177], electric field detection in electric fish [169], and surface wave detection in the back swimmer [133].

In summary, the above examples show the importance of a center-surround receptive field, the natural consequence of our model. Our approach thus explains lateral inhibition as an *optimal* strategy for the reconstruction of a spatially blurred stimulus. Therefore we can record that the mathematical concept of optimality is a powerful tool for linking characteristics of the physical environment and the biological sensory apparatus to the characteristics of the neuronal processing of the corresponding sensory information. This linking can be consolidated by the above quick guide so that even the mathematically untrained can apply our framework to the specific need at hand.

In the context of this thesis, stimulus reconstruction in space-time leads to the issue of object formation. Any object exists in space-time and hence generates spatio-temporal stimuli. The reconstruction of these stimuli alone, however, is not sufficient; for further processing objects need to be *identified.* In the next chapter we mathematically analyze different fundamental strategies for the identification of signal periodicity in neuronal systems so as to allow for auditory object formation.

Chapter 4

Signal periodicity and auditory object formation

In section 1.1 we emphasized the important role of common amplitude modulations of different frequency components for auditory object formation. In the following, we will discuss how such amplitude modulations (AM) can be identified neuronally. We start with a conceptual overview on periodicity in the neuronal and acoustic context.

4.1 Periodicity in neuronal and acoustic activity

The setup we study is of universal interest: it aims at explaining not only the processing of amplitude modulations in mammals but also the more generic identification of periodicity in neuronal signals. Mammals feature a cochlea as vibration-sensitive organ that decomposes acoustic signals into their constituting frequencies. Each frequency is further processed in a distinct, frequency-specific neuronal channel. Hence periodic modulations in the neuronal activity of such a channel reflect amplitude modulations of the signal. Other animals such as spiders, frogs, or surface feeding fish detect vibratory signals as well. Even though their vibration-sensitive organs, in contrast to the cochlea, do not display frequency specificity [6,7,41,95], these animals can distinguish the frequency of vibratory signals [15,18,50,51,97,128] on the surface of water [5,13,16,107], in spider webs [123], or on plant leaves [5,81]. Hence, in animals without a cochlea or other frequency-specific organs the periodicity of vibratory signals needs to be identified *neuronally*, similarly to the identification of amplitude modulations in mammals with a cochlea.

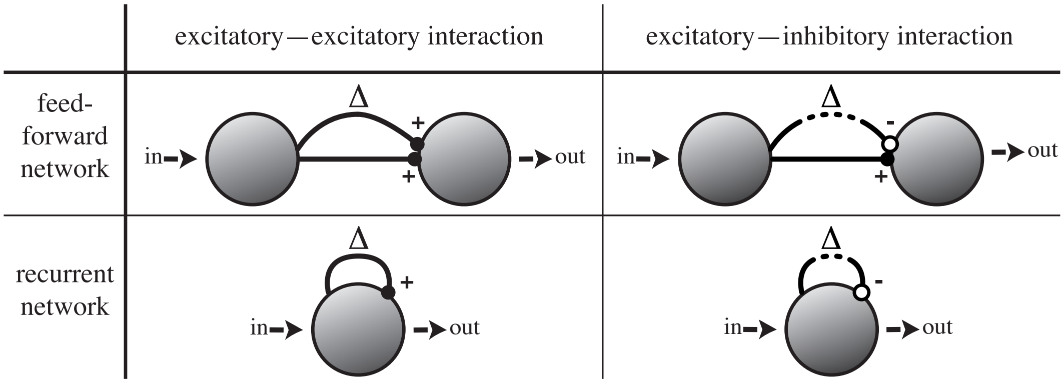

Figure 4.1: Overview of neuronal circuits for periodicity identification. There are four groups of neuronal networks that can identify amplitude modulations (AM). Identification of AM either arises from excitatory–excitatory (left) or excitatory–inhibitory (right) synaptic interaction, and both types of interaction can be realized by means of a feedforward (top) or a recurrent (bottom) network. The delay lines Δ, crucial for periodicity identification, are indicated as excitatory (+) or inhibitory (−) connections. The resulting four archetypes of neuronal networks are topic of the present chapter.

How does one identify periodicity in neuronal activity? As mentioned in section 1.1 it is known that there are neurons selectively responding to specific modulation frequencies [92, 163, 176]. The question is how such a selectivity, a neuronal band-pass characteristic, can be explained. On the level of single neurons a band-pass response can emerge from the cell membrane dynamics. Here the spike-generating mechanism can induce oscillations of the membrane potential that follow a spike and thus enhance the firing at certain instants of time after the first spike [88]. Alternatively, inhibitory input can cause such an oscillation of the membrane potential, the so-called "post-inhibitory rebound" [110]. On the level of neuronal circuitry a band-pass characteristic can be realized by either excitatory–excitatory synaptic interaction or excitatory–inhibitory synaptic interaction, where each type of interaction can be realized in a recurrent or a feedforward network; cf. Fig. 4.1. As we will see in the first part of this chapter, the excitatory–excitatory interaction basically works like a coincidence detector where two spikes can only evoke neuronal activity if they arrive at a neuron simultaneously, that is, if they arrive in phase. The timing of the spikes can either arise from delays [59, 114] –the neuronal analogon to autocorrelation– or from "chopper neurons" [127], neurons that produce a series of well-timed spikes. Similarly, a band-pass characteristic arises if a single excitatory spike strong enough to evoke neuronal activity is combined with a delayed inhibitory spike that arrives in anti-phase to the excitatory input [70]. Furthermore, band-pass

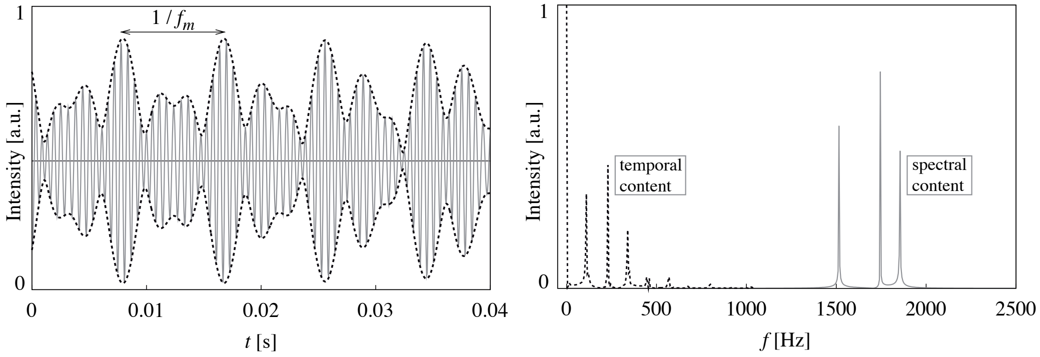

Figure 4.2: The left panel shows a complex signal wave form, composed of three frequency components (solid grey) and its envelope, the instantaneous amplitude (dotted black) in arbitrary units [a.u.]. The interference of the frequency components causes amplitude modulation on a slow scale. One of the modulation frequencies f_m ($\sim$ 100 Hz) has been indicated in the plot.
The right panel shows the Fourier transform of the signal (solid grey) and the envelope (dotted black) in arbitrary units [a.u.]. Although the signal consists of three frequencies in the $1500 - 2000$ Hz range, the envelope shows only slow variations, mainly below 500 Hz. As the envelope has a non-zero mean value the Fourier spectrum shows an additional peak at 0 Hz.

characteristics within such an excitatory–inhibitory setup can also arise from different time constants for excitation and inhibition [29, 136]. This will be elaborated in the second part of this chapter.

Before presenting the various models for the neuronal identification of signal periodicity, it is important to review some general characteristics of vibratory signals. We have to bear in mind that vibratory signals are not limited to air-borne sound but may propagate in a variety of substrates such as sand [2, 23], the water surface [14, 17], spider webs [106, 124], or leaves [121]. All vibratory signals consist of a time-dependent change in pressure or medium deflection. Even with no explicit modulation present in a signal, interference effects between different frequency components of a natural signal usually lead to complex signal wave forms; cf. Fig. 4.2. In such a complex signal wave form, fast periodical variations in signal strength are normally designated as *spectral* content, or frequencies; slow variations are denoted as *temporal* content. The slowly varying amplitude of the signal is called *envelope*. In general the distinction between temporal and spectral content is a matter of convention. In our setting, we consider all periodic signal fluctuations which can be resolved *neuronally* as temporal content, that is, signal variations with frequencies lower than approximately 500 Hz are temporal.

In the following, we will mathematically describe two fundamental neuronal architectures for detecting signal periodicity, one based on excitatory–excitatory and one based on excitatory–inhibitory interaction. We have restricted ourselves to a minimalistic implementation of the models and do not take into account specific physiological details. There are two reasons for doing so. First, discussing simple models allows a detailed mathematical treatment leading to a thorough comprehension of the capacities and limitations of the circuitry. Second, since periodicity identification is a capability of many animals, it is important to understand general mechanisms rather than any specific realization.

4.2 Identification of signal periodicity in an excitatory–excitatory setup

We start with the analysis of periodicity identification in the purely excitatory setup, without any inhibitory connections.

4.2.1 Model essence: excitatory delay lines

The goal of our models will be to identify slow fluctuations present in a specific input signal. Mathematically, periodic features of a signal $s(t)$ can be detected by calculating its autocorrelation χ (see e.g. [126]), defined by

$$\chi(\Delta) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{d}\tau \; s(\tau)s(\Delta+\tau) \; . \tag{4.1}$$

The autocorrelation has maxima for correlation times Δ corresponding to the frequencies present in the signal, but also for the periods of the envelope fluctuations, cf. Fig. 4.2. The above calculation immediately suggests two neuronal mechanisms for detecting periodicity as illustrated in figure 4.3.

The first model consists of a neuron that receives an input signal $s_{\text{in}}(t)$. As the neuron spikes, the output spike is fed into a pathway that ultimately projects onto the neuron itself with a particular delay Δ, corresponding to the correlation time above. This pathway need not be a direct connection from the axon onto the neuron's own dendritic tree. One or more processing steps may occur before the output from the neuron returns but a well-defined delay needs to be associated with the pathway. This matter will be further discussed in section 4.2.4. Because of the delay loop, the neuron detects correlations on a time scale Δ. An array of such neurons, all

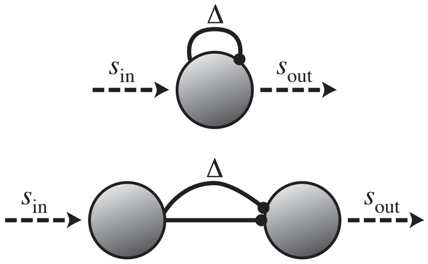

Figure 4.3: There are basically two ways to extract frequency or timing information from a signal relying on excitatory–excitatory neuronal interaction with spiking neurons. The first method (upper panel) uses a recurrent loop with time delay Δ. This we call the *recurrent model.* The neuron is driven by a continuous input function s_{in}. If the neuron emits a spike at time $t = t_0$, the firing probability is enhanced at time $t = t_0 + \Delta$. Signal periodicity with characteristic time Δ then leads to a higher number of spikes in the output signal s_{out}.
The second method (lower panel) is based on the same idea, but uses a feedforward network, and is called the *feedforward model.* The first neuron, again driven by s_{in}, sends two spikes to the output neuron with a delay differing by an amount Δ, e.g., using interneurons. Again, correlations in the input signal with period Δ lead to an augmented firing probability for the output neuron.

with different Δ, can then function as a periodicity analyzer. We call this model the *recurrent model.*

The second model consists of a two-neuron network. If the input neuron fires, its spikes are fed into two pathways to the output neuron. The temporal durations of these pathways differ by an amount Δ. The output neuron will have a high firing probability if spikes arrive from the two different pathways at the same time. Again, the network reveals correlations on time scale Δ. We call this model the *feedforward model.*

Both types of networks have been discussed before in the literature. To our knowledge, the first author to propose a network of delay lines to detect signal periodicity was Licklider [114]. More recent work on feedforward-like models has been done by Borst et al. [20] as well as Meddis and O'Mard [127]. Both articles presented a very detailed model, based on specific properties of neuronal circuitry found in the mammalian auditory system. Cariani [34, 35] discussed a recurrent-like model. He, however, used an overly simple model in which formal neurons manipulating strings of 0s and 1s are used. None of these authors have provided a detailed mathematical analysis of their models. In this section, we provide this missing analysis and use fairly realistic neuron models for our simulations without settling on a specific neuronal architecture. We start with analyzing the characteristics of our models in more detail.

Detailed description of the recurrent model

The recurrent model consists of N_{out} output neurons that all receive the same external continuous input $s_{\text{in}}(t)$. All input neurons have a recurrent connection that feeds output spikes back into the neuron itself. The recurrent spikes are characterized by a delay Δ that is different for each neuron and has a synaptic coupling strength J. The feedback current is described by a general function g (see also section 4.2.2) for which we will take an α-function in our simulations [66]

$$g(t) = \frac{t - t_0 - \Delta}{\tau^2} e^{-(t-t_0-\Delta)/\tau} \theta(t - t_0 - \Delta) \ . \tag{4.2}$$

The width of the α-function is given by τ, t_0 is the spiking time of the neuron, and θ denotes the Heaviside step function, i.e., $\theta(t) = 0$ for $t < 0$ and $\theta(t) = 1$ for $t \geq 0$.

The neurons are simulated as leaky integrate-and-fire (LIF) neurons [66]. Their firing dynamics are governed by a differential equation for the membrane potential V,

$$\frac{dV}{dt} = -(V - V_0)/\tau_{\text{mem}} + \frac{1}{C_{\text{mem}}} \left(I_{\text{ext}} + I_{\text{noise}} \right) \ . \tag{4.3}$$

The potential changes under influence of an external input current I_{ext} that drives the neuron. If there is no input current the potential relaxes to a resting value V_0 with characteristic membrane time constant τ_{mem}. The last term, I_{noise}, accounts for internal noise of the neuron that will be needed in the case of the recurrent model. The constant C_{mem} is the membrane conductance of the neuron determining how effectively the current can change the membrane potential.

If the potential in (4.3) reaches a certain threshold value V_θ a spike occurs and the potential is reset to a value V_R. Refractoriness of the neuron can be taken into account by disallowing the neuron to fire for a certain period after spiking, by changing the threshold voltage temporarily to a higher value, or by temporarily ignoring the input current (see also [66]).

To get the model to work the output neurons must fire a first spike to start with, since the feedback loop needs input, which can only come from the neurons themselves. It is not possible to use supra-threshold input since this would imply that *all* output neurons would fire in response to the input, regardless the length of their delay loop. The solution is to use subthreshold input with added internal neuronal noise. Every now and then the neuron will fire. But only if the delay loop length has the right value the neuron will be able to resonate in response to the input. The mechanism described here is called *stochastic resonance*, reviewed in detail elsewhere [62].

Detailed description of the feedforward model

The feedforward model consists of N_{in} input neurons, which we simulate as Poisson neurons [79]. That is, we assume the firing of the input neurons to be a statistical process, an *inhomogeneous* Poisson process. Such a Poisson process is defined by three properties as we have mentioned in Sec. 2.3. First, the probability of finding a spike between t and $t+\Delta t$ is $\lambda(t)\,\Delta t$, so $\lambda(t)$ is the time-dependent firing probability density or rate function. Second, the probability of finding two or more spikes in $[t; t+\Delta t[$ is $o\,(\Delta t)$, which means that we ignore their occurance for small Δt. Third, events in disjoint intervals are independent, i.e., a Poisson process has independent increments.

The Poisson input neurons are driven by an external input $s_{\text{in}}(t)$. If one of the input neurons fires, its spike is fed into an axon branching off to N_{out} different output neurons. One spike reaches the output neurons directly, and another spike resulting from the same event reaches the output neurons with a delay Δ. A specific delay Δ is associated with every output neuron; in this way, every output neuron will turn out to encode a particular frequency $f = 1/\Delta$.

The output neurons are simulated as leaky integrate-and-fire (LIF) neurons *without noise*, contrary to those in the recurrent model. If a spike is emitted at time $t = t_0$ by any of the input neurons it leads to two postsynaptic current injections arriving at the output neurons, again in the form of α-functions,

$$\varepsilon_{\text{direct}} = J\frac{t-t_0}{\tau^2}e^{-(t-t_0)/\tau}\theta(t-t_0) \tag{4.4}$$

and

$$\varepsilon_{\text{delayed}} = J\frac{t-t_0-\Delta}{\tau^2}e^{-(t-t_0-\Delta)/\tau}\theta(t-t_0-\Delta)\ . \tag{4.5}$$

The former spike travels to the output neuron without delay, and the latter arrives with a delay Δ. The synaptic coupling strength is again given by the parameter J.

4.2.2 Analysis: delay and frequency selectivity

In this section we will mathematically discuss the behavior of the two types of excitatory–excitatory periodicity detector. Explicit analysis of LIF neurons is in general already quite difficult (for an extensive review, see [30, 31]). We will see, however, that no explicit analysis of LIF neuron dynamics is needed to gain valuable insight into the dynamics of our models. In fact, the key properties of the models are independent of the specific type of neurons that are used.

Recurrent model

The problem of analytic calculations using integrate-and-fire neurons lies in the nonlinearity of the spike generation. In the case of the recurrent network we are discussing here, the problem is even more difficult than usual since the feedback introduces an extra complication into the system. We therefore simplify our discussion by considering Poisson neurons again. We will later compare the findings obtained here with the simulations in section 4.2.3 to see whether the calculations using Poisson neurons can serve to understand the dynamics of the LIF neurons used in section 4.2.3.

We can describe the rate function λ of a single Poisson neuron projecting back to itself with a particular delay time Δ by the integral equation

$$\begin{aligned}\lambda(t) &= s_{\text{in}}(t) + J \int_{-\infty}^{\infty} \mathrm{d}s \; g(s;\Delta)\lambda(t-s) \\ &= s_{\text{in}}(t) + J(g \star \lambda)(t) \; .\end{aligned} \tag{4.6}$$

The rate function consists of the sum of the external input s_{in} and the delayed input from the recurrent loop, "smeared out" due to the finite width of kernel g. The feedback strength is given by J, and we choose g to ensure causality [$g(t) = 0$ if $t < 0$] and to have unit weight

$$\int_{-\infty}^{\infty} \mathrm{d}t \; g(t) = 1 \; . \tag{4.7}$$

In (4.6) the convolution integral of λ with the kernel g assumes that we may use the expectation value of the firing rate λ to describe the neuron output instead of a specific realization of the output. We thereby ignore the "spiky" character of the neuron output. This approach is only correct for very high firing rates or, mathematically equivalent, a large number of Poisson neurons with a low firing rate. The total amount of output spikes must be high enough so that the output signal is reliably sampled by the output spikes. An example of such a smooth convolution of λ and g is shown in figure 4.4.

To solve (4.6) for the output firing rate λ we take the Fourier transform of the equation. The Fourier transform of a function h is defined by

$$H(\omega) = \mathcal{F}\left[h(t)\right](\omega) := \int_{-\infty}^{\infty} \mathrm{d}t \; e^{-i\omega t} h(t) \tag{4.8}$$

and has the useful property that, when transformed, a convolution becomes an ordinary product. Denoting the Fourier transform of each input term by a capital

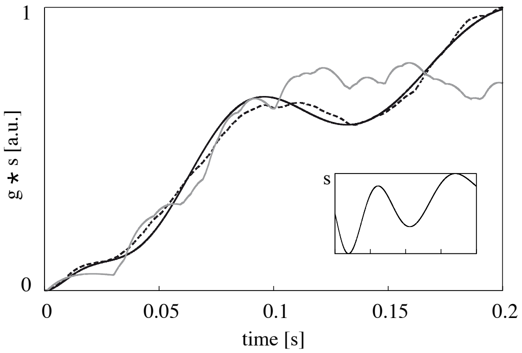

Figure 4.4: Stochastic fluctuations in the response of a Poisson neuron are smaller if the firing rate is higher. The convolution of a signal s (inset) with the response kernel g in black is compared to two explicit realizations of the firing process (normalized in arbitrary units [a.u.] for comparison). The grey curve was obtained using about 40 spikes, the black dotted curve results from about 300 spikes. Clearly, a high firing rate (or, mathematically equivalent, a large population of Poisson neurons) is needed for (4.6) to apply.

letter we obtain

$$\Lambda(\omega) = S_{\text{in}}(\omega) + JG(\omega)\Lambda(\omega) \ . \tag{4.9}$$

The solution is then given by

$$\Lambda = \frac{S_{\text{in}}}{1 - JG} \ . \tag{4.10}$$

The solution as a function of time can then be found by taking the inverse Fourier transform

$$\lambda(t) = \mathcal{F}^{-1}\left[\Lambda(\omega)\right](t) := \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega \ e^{i\omega t}\Lambda(\omega) \ . \tag{4.11}$$

Given any input function s_{in} and response function $g(t)$ we can now explicitly calculate the firing probability of the neuron. In our simulations we will use an α-function for g to model the response function [see (4.2)]. The Fourier transform of this response function is given by

$$G(\omega) = \frac{e^{-i\omega\Delta}}{(1 + i\omega\tau)^2} \ . \tag{4.12}$$

Since we are interested in identifying periodicity, we must know which frequency f corresponds to a certain delay time Δ. A first guess would be to set $f = 1/\Delta$; but since the response function transforms the recurrent signal this relation cannot be expected to hold exactly. We therefore consider the response of the system to an incoming pure sine wave of frequency f and find the corresponding Δ that maximizes the amplitude of the response. We then have an explicit connection between the delay Δ and the signal frequency that is decoded optimally through this delay.

For harmonic input given by

$$s_{\text{in}}(t) = A\cos(\omega t) = A\cos(2\pi f t) \tag{4.13}$$

we calculate the response to be

$$\lambda(t) = L\cos(\omega t + \phi) \ , \tag{4.14}$$

where ϕ is a phase that is not relevant for our further calculations and L is an amplitude given by

$$L = \frac{2(1+\xi^2)^2}{\sqrt{J^2 + (1+\xi^2)^2 - J\left[2(1-\xi^2)\cos(\omega\Delta) - 4\xi\sin(\omega\Delta)\right]}} \ , \tag{4.15}$$

with the definition $\xi := \omega\tau$. The amplitude L of the response is maximal if the relation

$$2(1-\xi^2)\cos(\omega\Delta) - 4\xi\sin(\omega\Delta) = 0 \tag{4.16}$$

holds. The delay must therefore satisfy

$$\Delta = \omega^{-1}\left[\arctan\left(\frac{2\xi}{\xi^2-1}\right) + n\pi\right] \ , \tag{4.17}$$

with $n = 1$ if $\xi > 1$ and $n = 2$ for $\xi < 1$. If the width of the kernel g approaches zero ($\xi \to 0$) this relation indeed reduces to

$$\Delta = \frac{2\pi}{\omega} = \frac{1}{f} \ . \tag{4.18}$$

As a more complicated and realistic example let us consider an input of the form

$$s_{\text{in}}(t) = \int_0^\infty \mathrm{d}\sigma \ B(\sigma)\cos\left[\sigma t + \phi(\sigma)\right] \ . \tag{4.19}$$

Instead of a single harmonic component we now describe the input by a distribution of input frequencies with arbitrary amplitude and phase. The Fourier transform of such an input is given by

$$S_{\text{in}}(\omega) = \int_0^\infty \mathrm{d}\sigma \ B(\sigma)\pi e^{i\phi(\sigma)\omega/\sigma} \times \left[\delta(\sigma-\omega) + \delta(\sigma+\omega)\right] \ , \tag{4.20}$$

with $\delta(.)$ the Dirac delta function. Plugging this result into (4.10) and (4.11) gives the solution for the firing rate of the output neuron

$$\lambda(t) = \int_0^\infty \mathrm{d}\sigma \; B(\sigma)\mathcal{R}\left[\frac{e^{i(\phi(\sigma)+\sigma t)}}{1 - Je^{-i\sigma\Delta}/(1+i\sigma\tau)^2}\right] \;, \tag{4.21}$$

where $\mathcal{R}[x]$ denotes the real part of x. We note that the signal function (4.19) need not be positive, although a negative firing rate certainly does not make sense for a Poisson neuron. We therefore always use half-wave rectified signals in the simulations. Unfortunately, exact calculations are not feasible in this case. In spite of this drawback (4.21) captures the essence of the network response. If the solution (4.21) is plotted for various input spectra $B(\sigma)$ the amplitude of λ is largest if the length of the delay loop Δ corresponds to a frequency that is present in the input signal. Due to its complicated form, (4.21) is of limited practical use. A better way to gain insight into the model dynamics are the numerical simulations we provide in section 4.2.3.

Feedforward model

For the analytic description of the feedforward model we will use an input population of Poisson neurons which are, just as before, driven by an input $s_{\text{in}}(t)$ identical for each neuron. We consider LIF neurons as output neurons. Every output neuron receives input from the Poisson neurons via two distinct pathways: a direct connection and a connection with a delay Δ which is different for each output neuron.

If we calculate the expectation value of the current that arrives at the output neurons a sinusoidal function results. The response of LIF neurons to harmonic input is difficult to calculate but several exact results have been presented by Burkitt [32] as will be discussed below.

We start our calculations by considering input given by

$$s_{\text{in}}(t) = \frac{A}{2}\left[1 + \cos(\omega t)\right] \;. \tag{4.22}$$

The input current that one output neuron with a particular delay time Δ receives from the set of N_{in} input neurons is given by [referring to (4.4) and (4.5)]

$$\varepsilon_{\text{total}} = \varepsilon_{\text{direct}} + \varepsilon_{\text{delayed}} \;. \tag{4.23}$$

As a consequence of the Poisson nature of the input neurons, the expectation value of the current to the output neurons is given by [79]

$$\langle I \rangle = \int_{-\infty}^{\infty} \mathrm{d}s \; s_{\text{in}}(s) \; \varepsilon_{\text{total}}(t-s) \tag{4.24}$$

and the variance of the current is given by

$$\mathrm{var}_I = \int_{-\infty}^{\infty} \mathrm{d}s \; s_{\mathrm{in}}(s) \; \varepsilon^2_{\mathrm{total}}(t-s) \; . \tag{4.25}$$

Equations (4.24) and (4.25) can be evaluated exactly for the given input function (4.22). The results are

$$\langle I \rangle = N_{\mathrm{in}} A J \left\{ 1 + \frac{\cos(\omega\Delta/2)}{(1+\xi^2)^2} \Big[(1-\xi^2) \cos\big(\omega(t-\Delta/2)\big) + 2\xi \sin\big(\omega(t-\Delta/2)\big) \Big] \right\} \tag{4.26}$$

with $\xi = \omega\tau$. The amplitude (current arriving at the output neuron) is thus maximal for integer

$$\Delta \cdot \frac{\omega}{2\pi} \in \mathbb{N} \; . \tag{4.27}$$

That is, a maximal response of the output neurons is to be expected if the input frequency matches the delay of the system. If the input signal contains a periodicity with frequency f^* the neuron with a delay time Δ^* corresponding to this frequency will respond optimally. All neurons sensitive to a *subharmonic* frequency (f^*/n, with $n \in \mathbb{N}$) will also respond, as can be seen from (4.27). This is because an input signal with a periodicity f^* is automatically also periodic with frequency f^*/n.

The variance of the current is given by

$$\begin{aligned} \mathrm{var}_I = & \frac{N_{\mathrm{in}} A J^2}{\tau} \left\{ \frac{1}{4} + \frac{2\cos(\omega\Delta/2)}{(4+\xi^2)^3} \right. \\ & \left. \times \Big[(8-6\xi^2) \cos\big(\omega(t-\Delta/2)\big) + \xi(12-\xi^2) \sin\big(\omega(t-\Delta/2)\big) \Big] + M \right\} \end{aligned} \tag{4.28}$$

where M is given by

$$\begin{aligned} M = & e^{-\Delta/\tau} \left\{ \frac{1-\Delta/\tau}{4} + \frac{1}{(4+\xi^2)^3} \right. \\ & \times \left[\left(16 - 8\xi^2 + \frac{\Delta}{\tau(16-\xi^4)} \right) \cos\big(\omega(t-\Delta)\big) \right. \\ & \left. \left. + 2\xi \left(12 - \xi^2 + \frac{2\Delta}{\tau(4+\xi^2)} \right) \sin\big(\omega(t-\Delta)\big) \right] \right\} \; . \end{aligned} \tag{4.29}$$

In order to allow correct periodicity detection, the time scale of the periodicity must clearly exceed the time scale τ of the individual current response functions ϵ.

We thus expect the system to work best if the relation $\Delta \gg \tau$ holds, meaning that the time scale of the periodicity is much larger than that of the post-synaptic response. In the auditory system we can expect this condition to hold. M can then be neglected because of the exponential prefactor $e^{-\Delta/\tau}$ in (4.29). If low-frequency input is presented, we have $\omega \ll 1/\tau$ and thus $\xi = \omega\tau \to 0$. The current and its variance are then given by

$$\langle I \rangle = N_{\text{in}} A J \left[1 + \cos(\omega\Delta/2) \cos\big(\omega(t - \Delta/2)\big) \right] \tag{4.30}$$

and

$$\text{var}_I = \frac{4 N_{\text{in}} A J^2}{\tau} \left[1/16 + \cos(\omega\Delta/2) \cos\big(\omega(t - \Delta/2)\big) \right] \ . \tag{4.31}$$

The relative variation of the current is proportional to

$$\frac{\delta I}{I} = \frac{\sqrt{\text{var}_I}}{I} \propto (N_{\text{in}} A \tau)^{-1/2} \ , \tag{4.32}$$

which also holds if we do *not* assume $\Delta \gg \tau$ and $\xi \to 0$. As expected, the current is less sensitive to random fluctuations if the number of input neurons or the input amplitude increases. The fact that the current fluctuates more if τ gets smaller can be attributed to a very short synaptic time scale enhancing the "spiky" character of the current. The system, however, does not become less reliable since a short post-synaptic current enables better coincidence detection by the output neurons [100, 102].

The expression (4.26) for the mean current, which is a good approximation if there are enough input neurons, shows that all output neurons receive a harmonic current. The amplitude of the current is largest if the delay matches the periodicity of the input signal. The response of integrate-and-fire neurons to harmonic input is difficult to calculate but it has been done for a slightly different system [32]. The results show that the periodicity of the input current is retained in the firing of the output neuron. This means that the output signal is phase-locked to the current. The vector strength VS, which can be defined as the absolute value of the first Fourier coefficient of the signal divided by the zeroth Fourier coefficient, measures the amount of synchronization or phase locking. For perfect phase locking VS= 1. For a random distribution of phases (complete absence of phase locking) we find VS= 0. In the setups we have discussed here, VS tends to be larger in the output neuron than in the current itself.

Summary of analytical results

The most important mathematical property of the recurrent model is the relation between loop delay Δ and optimal coding frequency f. Naïvely, one would expect the relation $\Delta = 1/f$ to hold. The recurrent loop does not, however, simply project the output back to the neuron. The response kernel h rather smears out the feedback, changing the exact timing of the recurrent input. The exact relationship between delay and coding frequency is given by (4.17) for synaptic responses in the form of an α-function.

In the feedforward model, the problem of finding the relation between Δ and f does not arise. Both the direct and the delayed pathway smear out the input spikes in the same manner, and therefore the relative timing of the two signals arriving at the output neuron is fixed. The most important result for the feedforward model is that the amount of phase locking (a measure for the accuracy of spike timing) actually increases. The output neurons thus fire more accurately than the input population does. This is in accordance with physiological findings in the mammalian auditory pathway [92].

For a true understanding of the models the mathematical description presented above does not suffice. Since the nonlinear process of spike generation cannot be taken into account, numerical simulations are needed to characterize the response of the models to realistic input. The next section discusses such simulations.

4.2.3 Implementation: neuronal effects and temporal jitter

In this section we discuss results obtained by numerical simulations. The neuronal networks as described in section 4.2.1 have been implemented through the C++ programming language. To test the performance of the models we have provided the networks with three different kinds of input: amplitude-modulated (AM) input, a Gaussian distribution of frequency components, and input mimicking the "missing fundamental" effect, as explained below. The response of the system was characterized by counting the number of output spikes that occurred during one second of input presentation as a function of the coding frequency of the output neuron. The coding frequency of the output neurons was calculated using (4.17) for the recurrent network and (4.27) for the feedforward network. We will see that in both networks the neurons encoding the periodicity present in the input signal respond maximally. The networks are thus able to convert a *phase code* into a *rate code* as we required in the introduction.

Half-wave rectification of the signals has always been performed before presenting them to the network. Hair cells, the basic receptor units of the ear and the lateral line system, depolarize following one direction of displacement and hyperpolarize if displacement is in the other direction [84]. Half-wave rectification is therefore automatically performed upon detection in many biological sensory systems.

parameter	**value**
number of output neurons	$N_{\text{out}} = 491$
output frequency range	10–500 Hz
synaptic time constant	$\tau_s = 1\,\text{ms}$
synaptic strength	$J = 2.5 \times 10^{-5}$
input normalization	$1/T \int_0^T \mathrm{d}t\ s_{\text{in}} = 300$
output neuron	
membrane time	$\tau_m = 1.25\,\text{ms}$
absolute refraction time	$\tau_{\text{refr}} = 1.0\,\text{ms}$
resting potential	$V_r = 0$
reset potential	$V_{\text{reset}} = V_r = 0$
threshold	$V_\theta = 1$
capacitance	$C_m = 1$

Table 4.1: Simulation parameters for the recurrent model.

The input signal to the network was normalized to deliver the same time-integrated input power in each case. Obviously, it is not realistic to expect external input to a vibration detection system to be normalized but several mechanisms of neuronal *gain adaptation* have been shown to exist; e.g., in the auditory pathway [43,87,174,189]. Such mechanisms are thought to keep neuronal firing rates within an optimal range. In our case power normalization is needed to keep the output firing under control. If the input power is too low, the output neurons cannot fire at all. If, on the other hand, the input power is too high all neurons will fire at a high rate and the discriminative capacity of the system is lost.

The numerical values of the parameters used in the computations are given in table 4.1 for the recurrent model. According to (4.3) internal noise of the neurons has been implemented by adding a noise term I_{noise} to the input of each neuron. In our simulations this noise term is given by

$$I_{\text{noise}} = \sum_{n=1}^{50} A_{\text{noise}} \cos(2\pi f^n_{\text{noise}} t + \phi^n_{\text{noise}}) \tag{4.33}$$

where the frequencies are chosen from a uniform distribution $f^n_{\text{noise}} \in [0 - 1000\,\text{Hz}]$. Phases are uniformly distributed in $\phi^n_{\text{noise}} \in [0 - 2\pi]$, and the amplitude of every component is given by $A_{\text{noise}} = 0.01/50$. For each neuron, independent noise is assumed and the noise is then added linearly to the input for each neuron.

As for the simulations using the feedforward model, the parameters used are summarised in table 4.2.

parameter	**value**
number of input neurons	$N_{\text{in}} = 25$
number of output neurons	$N_{\text{out}} = 491$
output frequency range	10–500 Hz
input neuron mean rate	20 Hz
synaptic time constant	$\tau_s = 1\,\text{ms}$
synaptic strength	$J = 3.5 \times 10^{-4}$
output neuron	
membrane time	$\tau_m = 1\,\text{ms}$
absolute refraction time	$\tau_{\text{refr}} = 0.25\,\text{ms}$
resting potential	$V_r = 0$
reset potential	$V_{\text{reset}} = V_r = 0$
threshold	$V_\theta = 1$
capacitance	$C_m = 1$

Table 4.2: Simulation parameters for the feedforward model.

Amplitude-modulated input

We consider two types of AM input signals. First, we present a modulated pure tone

$$s_{\text{in}}(t) = \frac{A}{2}[1 + \cos(2\pi f_m t + \phi)] \cos 2\pi f_c t \tag{4.34}$$

with modulation frequency $f_m = 50\,\text{Hz}$ or $f_m = 200\,\text{Hz}$ and random modulation phase ϕ. The carrier frequency is $f_c = 2000\,\text{Hz}$.

In the second case we consider noise by composing a signal from 50 sinusoidal components with frequencies f^n_{rand} chosen from a uniform distribution on $[0, 1000\,\text{Hz}]$ and random phases ϕ^n_{rand} with uniform distribution on $[0, 2\pi]$ so as to obtain

$$s_{\text{noise}}(t) = \sum_{n=1}^{50} \cos(2\pi f^n_{\text{rand}} t + \phi^n_{\text{rand}}) \ . \tag{4.35}$$

We then modulate this signal with modulation frequency $f_m = 50\,\text{Hz}$

$$s_{\text{in}}(t) = \frac{A}{2}[1 + \cos(2\pi f_m t + \phi)] \times s_{\text{noise}} \ . \tag{4.36}$$

In both cases the amplitude has been chosen in such a way that the rectified input signal is normalized appropriately.

The results of these simulations are displayed in figure 4.5. Obviously both network

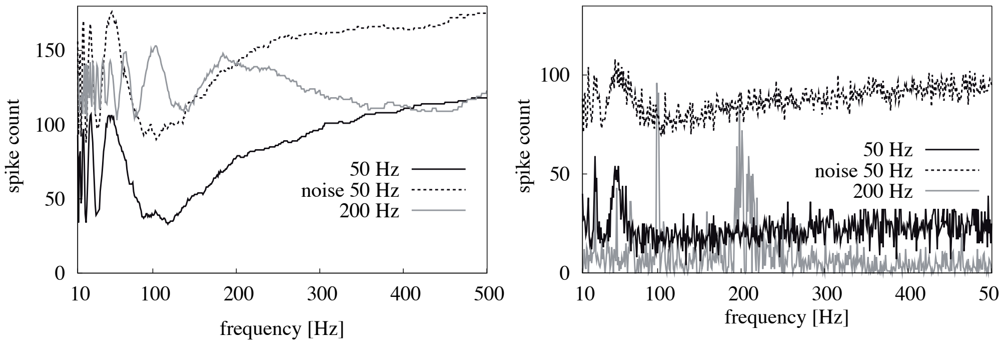

Figure 4.5: Response to AM input of an array of neurons, each with a different delay and corresponding frequency (horizontal axis). The total number of spikes in one second is shown vertically. Left panel: feedforward model; right panel: recurrent model. The peaks corresponding to the input periodicity clearly appear in the graphs. Evidently, both networks correctly identify the signals.

types succeed very well in detecting the periodicity of the input signal. Clear peaks in the response occur for the correct frequencies. The response peaks for the subharmonic frequencies are also distinctly recognizable. Although the response of the recurrent model is quite noisy, this drawback may be overcome easily by combining input from several close-by channels.

Gaussian frequency distribution

The second test for the feedforward and recurrent models consists of taking a *distribution* of frequencies as input. To mimic the real biological situation we have built an input signal from 30 different frequency components chosen randomly from a Gaussian probability distribution with a center frequency μ and a width σ. The components were added together with random phases, and the resulting signal was half-wave rectified and presented to the network. This signal can be considered as a rough model for a struggling insect on the water surface or in a spider web. The results of these simulations are displayed in figure 4.6. It is pretty evident that both frequency profiles with $(\mu, \sigma) = (30\,\mathrm{Hz}, 5\,\mathrm{Hz})$ and $(\mu, \sigma) = (100\,\mathrm{Hz}, 20\,\mathrm{Hz})$ are correctly identified by the two networks.

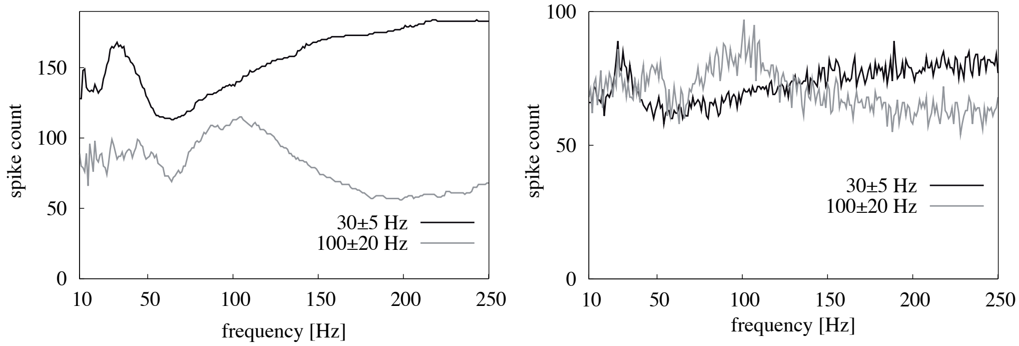

Figure 4.6: Response of an array of neurons to a distribution of frequencies. Input was presented to an array with neurons, each with a different delay and specific frequency (horizontal axis). The total number of spikes in one second is shown vertically. Left panel: feedforward model; right panel: recurrent model. Similar to Fig. 4.5 the signals are reliably identified.

Missing fundamental

If several pure tones with a *common* fundamental frequency are presented to a listener, the subject often perceives a tone with a pitch corresponding to this fundamental frequency, even though the fundamental frequency itself is not present in the input signal. Nonetheless a clear neuronal representation of this frequency is formed by the subject. To mimic such an experiment we give both models input consisting of three harmonics

$$s_{\text{in}}(t) = \sum_{n=1}^{3} \cos(2\pi f_n t + \phi_n) \ , \qquad (4.37)$$

with $f_1 = 200\,\text{Hz}$, $f_2 = 300\,\text{Hz}$, $f_3 = 400\,\text{Hz}$ and the phases random. The response of the feedforward model is shown in figure 4.7, together with the response to a pure tone of 100 Hz. Although the peak is not as clear as with pure tone stimulation, a pitch of 100 Hz is still easily recognizable.

Because of the noisy response, the “missing fundamental effect” is not reproduced very well by the recurrent model. A very good response can sometimes be obtained but this crucially depends on the precise values of the phases ϕ_n, which is not realistic biologically. Results for the recurrent network are therefore not shown here.

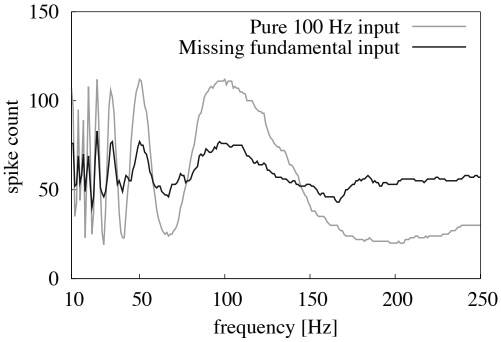

Figure 4.7: Response of the feedforward network to "missing fundamental" input as in (4.37) with three frequencies 200, 300 and 400 Hz compared to the response to a pure 100 Hz tone. The peak at 100 Hz is clearly recognizable.

Phase locking

A very important concept in auditory or vibratory processing is phase locking. Phase locking describes the capability of neurons to spike preferentially at a specific phase of the input signal. Phase locking is especially important to extract precise temporal clues from a signal; for instance, in sound localization [73,138]. The amount of phase locking is characterized by the vector strength VS as discussed above.

For AM noise input, as in (4.36), the vector strength has been displayed in figure 4.8. Interestingly, phase locking is quite good in the recurrent model although the output firing rate fluctuates a lot. This behavior results from the subthreshold input dynamics of the recurrent model. Only the presence of noise in the input assures that every now and then a spike occurs. The occurrence of a spike is of course much more likely if the input amplitude is large, and consequently the output firing tends to be phase-locked to the input periodicity. For the feedforward model phase locking is good if the decoding frequency of the output neurons matches the periodicity of the input. Again, spike generation is most likely when the input amplitude is large *and* the delay time matches the frequency of the input signal – phase locking results. Remarkably, for both models the phase locking is significantly stronger in the output signal than in the input signal.

Temporal jitter of delays

The ability to identify signal periodicity crucially depends on the timing of the delays Δ. We therefore investigate the effect of temporal jitter in the delays on

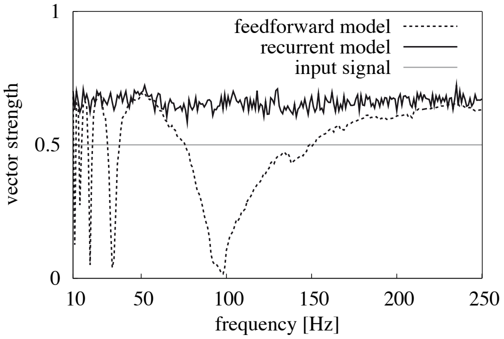

Figure 4.8: Phase locking strength as a function of best frequency for the feedforward and the recurrent model. 50 Hz modulated input as in (4.36). Vector strength of the input signal is 0.5, as indicated by the horizontal line. The output phase locking is stronger than the input phase locking in the relevant frequency range.

identification performance. We present four different pure tones with frequency f_{in} to both networks and add stochastic jitter to the delay time for every emitted spike. The jitter is Gaussian-distributed with mean 0 and a standard deviation from 0.2 ms to 20 ms. For each trial (a specific combination of input frequency and jitter strength) we calculate the *selectivity* Q defined by

$$Q = \frac{\left|\sum_j r_j e^{2\pi i \Delta_j f_{\text{in}}}\right|}{\sum_j r_j} \ . \tag{4.38}$$

Here Δ_j is the temporal delay corresponding to output neuron j, and r_j is its firing rate. This definition again has the form of a vector strength. If the output firing rate peaks for neurons with the correct delay ($\Delta_j f_{\text{in}} \in \mathbb{Z}$) the value of the numerator in (4.38) will be large. If much temporal jitter is present all output neurons will respond, even if their delay does *not* match the input signal frequency. In this case the phases in the numerator of (4.38) will cancel out, and Q will have a low value.

In figure 4.9 the selectivity for different input frequencies and jitter magnitudes is plotted, normalized to the selectivity without jitter. As could be expected, the selectivity deteriorates if jitter is present in the delays. For high input frequencies the sensitivity to temporal jitter is largest. For low frequencies, say $\lesssim 25$ Hz, both models are quite robust and can cope with temporal jitter up to ~ 10 ms. A jitter of about 20% of the input periodicity leads to a 50% decrease in selectivity. The amount of jitter thus determines the fastest input periodicity that can still be identified. For a temporal jitter of 1 ms this upper limit is approximately 200 Hz.

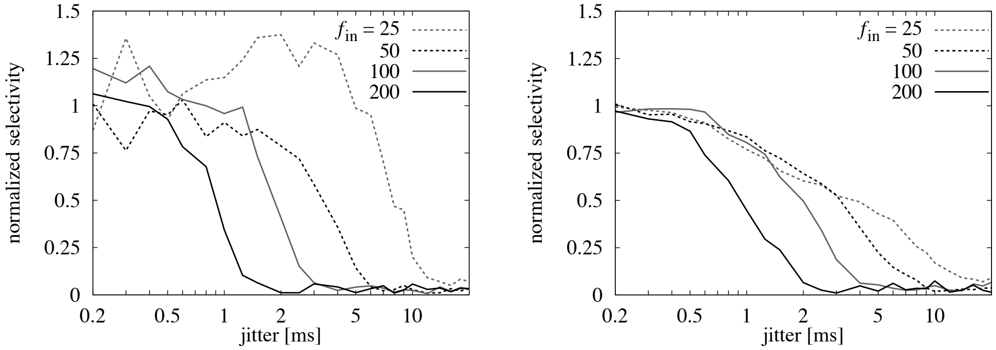

Figure 4.9: The selectivity as defined in (4.38) for the feedforward (left) and recurrent (right) model for several input signal frequencies f_{in} as a function of jitter. The selectivity Q is normalized with respect to the value in the absence of jitter. As expected, increasing jitter leads to a decrease in selectivity. For both the feedforward and the recurrent model a jitter of 20% of the input period leads to a 50% decrease in selectivity.

4.2.4 Discussion: limits of the excitatory setup

In the present section we have quantitatively analyzed two different models for periodicity detection based on excitatory–excitatory interaction. We have shown that both a feedforward architecture and a recurrent loop architecture can be used to extract periodic modulation from input signals. Furthermore, we have provided an extensive mathematical characterization. It has been shown that for both approaches the basic constraints are the same.

As expected, neuronal time constants are a limiting factor for recognizing the periodicity of the input modulation. The width of the post-synaptic current response presents a fundamental limit to the delay time that can be detected. It limits modulation recognition to about $\lesssim 1000\,\text{Hz}$. Indeed the experimental literature tells us that AM sensitivity reaches frequencies as high as $1000\,\text{Hz}$. The vast majority of neurons, however, is sensitive to modulation frequencies in the range of $10-300\,\text{Hz}$, most of them lying in the even more restricted range of $30 - 100$ Hz. This finding is valid for various animals [104, 109, 153–155]. Relevant biological stimuli on water surface and in spider webs also tend to contain most of their information in the low frequency range $\lesssim 250\,\text{Hz}$ [17, 106]. Thus the limitations imposed by neuronal time constants are not the essential ones.

A better explanation for the reported frequency range is the restriction arising from the limited accuracy of the delay lines. For both approaches the capability of distinguishing different frequencies crucially depends on well-known and constant delay times Δ. Only then is it possible to reliably assign a particular frequency to the output neurons. In reality the time it takes for the signal to propagate along the delay line may, however, vary. Using a deviation of $\delta\Delta \approx 0.5\,\mathrm{ms}$ cuts down the accessible detection range to about 200 Hz, a value reasonably close to the above 300 Hz.

According to section 4.2.1, the delay times Δ need not necessarily arise from a direct connection between two neurons but can be the result of a number of interneurons. Consequently, these interneurons then have to be driven by a very reliable synapse. Every input spike should trigger an output spike, and the delay between input and output spike should be fixed, as it usually is. A very prominent example of such a reliable “one-to-one” synapse in the auditory pathway is the so-called *Calyx of Held* at the end of the auditory nerve. Although this specific type of reliable synapse is only found in the lower auditory pathway its existence demonstrates that fast and reliable synapses are present in the auditory system, a neuronal system of exceptional acuity. For example, in the mammalian auditory brain-stem nuclei neurons can preserve the relative timing of action potentials passed through sequential synaptic levels [184]. In the avian auditory system, too, single presynaptic stimuli can produce short (and thus precise) suprathreshold spikes with a time constant of about 0.5 ms resulting in reliable information transmission [200]. Another possibility to reliably transfer precisely-timed signals is the use of synfire chains [49]. Depending on the input strength, synfire chains can relay information with a temporal precision around 1 ms, accurate enough for use in long-delay feedback and feedforward loops.

Another limitation common to both feedforward and recurrent circuitries is that they detect only the highest modulation frequency components in any signal. Since activity in high-frequency channels also excites low-frequency channels it is not possible to distinguish subharmonics of a high-frequency signal from a direct low-frequency input. The known phenomenon of the missing fundamental fits well into the behavior of such a simple network for periodicity extraction. Equivalent to the above is the fact that every neuron responds not only to its own specific frequency but also to all of its harmonics. Consequently the perceived similarity between tones one octave apart from each other [44, 47, 86] and the interference of harmonic target-distractor combinations at low frequencies [25] are a natural side-effect of the proposed architecture.

We conclude that in the excitatory–excitatory approach neuronal time constants do not limit model performance. Instead, in real biological systems the limiting factor will be the accuracy of the delay Δ. Since relatively long and well-defined delay times Δ can be realized by means of interneurons, this presents no fundamental problem to our model. The fact that every neuron responds not only to its own specific frequency but also to all of its harmonics is to be considered a feature rather than a limitation.

Before proceeding to the excitatory–inhibitory setup, we have to state that the two models, the recurrent and the feedforward one, differ in their behavior as far as their robustness is concerned. By design, the recurrent network is much more susceptible to noise and, as a consequence, can be disturbed by noise more easily than the feedforward model. This is a common problem of excitatory recurrent networks in general since in such networks perturbations tend to amplify themselves. Consequently we will mainly focus on a feedforward network and only briefly discuss the recurrent network in the excitatory–inhibitory setup we treat in the next section.

4.3 Identification of signal periodicity in an excitatory–inhibitory setup

Similar to the SFIE (same frequency inhibition and excitation) model proposed by Nelson et al. [136] the approach we will develop in the following is based on a bandpass characteristic arising from different time constants for excitatory and inhibitory postsynaptic potential (PSP). This is possible because every synapse is a low-pass filter with the "cut-off" frequency determined by the neuronal time constant τ of the PSP. A larger τ will lead to a lower "cut-off" frequency of the synapse. According to this consideration the combination of an excitatory synapse with small τ_{exc} and an inhibitory synapse with large τ_{inh} projecting to the same population of neurons will lead to a bandpass characteristic which is then governed by absolute value and difference of the excitatory and inhibitory time constants.

4.3.1 Model essence: inhibitory time constants

Below we provide a detailed analysis of two minimal models for periodicity identification on the basis of excitatory–inhibitory interplay. The models are "minimal" in that they feature two neurons or neuron populations at most, and only two synapses, one inhibitory and one excitatory; cf. Fig. 4.10.

Analogous to the considerations in the last section, the first model consists of two neurons or neuron populations [59]. If the input neuron (population) fires a spike, it is fed into two distinct pathways leading to the output neuron (population). One pathway will project onto the output neuron via an excitatory synapse, the other, delayed pathway via an inhibitory synapse. In a biological realization the delayed

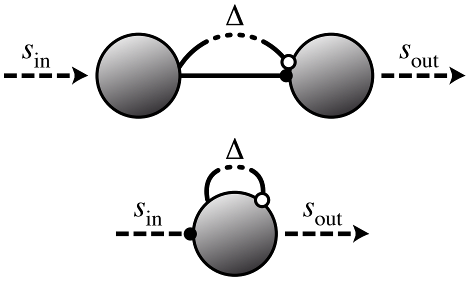

Figure 4.10: Similar to the setup in the last section, there are two ways of extracting frequency or timing information neuronally from a signal using excitatory–inhibitory interaction. As before, the first method (upper panel) uses a feedforward network, and is called *feedforward model*. The input neuron, driven by a continuous input function s_{in}, sends two spikes to the output neuron, one via an excitatory (closed circle), the other, delayed one, via an inhibitory synapse (open circle). Depending on excitatory and inhibitory time constants, certain temporal correlations in the input signal lead to an augmented firing probability for the output neuron. The second method (lower panel) is based on the same idea, but uses an inhibitory recurrent loop with time delay Δ. This we call the *recurrent model*. The neuron is driven again by s_{in}, this time via an excitatory synapse. If the neuron emits a spike at time $t = t_0$, its firing probability is reduced at time $t = t_0 + \Delta$ because of inhibitory feedback. Depending on excitatory and inhibitory time constants, a certain signal periodicity leads to a higher number of spikes in the output signal s_{out}. In a biological realization both systems will feature at least one additional neuron that forwards the inhibitory signal.

pathway will consist of at least one reliable interneuron. Certain combinations of delay, inhibitory and excitatory time constants, as well as the strength of the synapses will lead to maximal firing rates for different frequencies. We call this model the *feedforward model*.

The second model consists of a single neuron (or, again, a neuron population) that receives an input signal via an excitatory synapse. If the neuron spikes, the output spike will be fed into a pathway (again biologically realized by an interneuron) that projects back to the neuron itself with a particular delay. The spike will result in an *inhibitory* PSP characterized by its strength and a time constant different from the excitatory one. Such a setting leads, as we will see in the following sections, to a maximal firing rate for one specific frequency. A set of neurons, each with different time constants and coupling strengths, should then act as a frequency analyzer. We call this model the *recurrent model*.

Detailed description of the feedforward model

Just as in the excitatory–excitatory setting, the feedforward model features an input neuron population of Poisson neurons [79]. The Poisson input neurons are driven externally by a function $s_{\text{in}}(t)$ and form a simple input stage for the model, similar to, e.g., the auditory nerve. If any of the input neurons fires, its spike is fed into two pathways, one excitatory and one inhibitory, to an output neuron population – Poisson neurons again. The excitatory spike reaches the output neurons directly, the inhibitory spike is delayed by Δ due to interneurons. We note that in principle the delay Δ could be negative, that is, the excitatory spike could be delayed more than the inhibitory one by excitatory interneurons. Since we want to keep the setup simple, and in biological systems excitatory signals usually are converted into inhibitory signals by means of inhibitory interneurons, delayed inhibition is a reasonable assumption. The connection to every output neuron population is therefore described by a specific combination of inhibitory time constant τ_{inh}, delay Δ and inhibitory coupling strength J_{inh} on the one hand, and excitatory time constant τ_{exc} and excitatory coupling strength J_{exc} on the other hand.

If a spike is emitted at time $t = t_0$ by any input neuron it leads to two postsynaptic responses ε in the output neuron. Again we will model the postsynaptic responses with weighted α-functions [66]. The excitatory connection in the excitatory–inhibitory setup corresponds to the direct connection in the excitatory–excitatory setup, and its postsynaptice response is hence, similar to (4.4), described by

$$\varepsilon_{\text{exc}} = J_{\text{exc}} \frac{t - t_0}{\tau_{\text{exc}}^2} e^{-(t-t_0)/\tau_{\text{exc}}} \theta(t - t_0) \ . \tag{4.39}$$

Analogously to (4.5) we get

$$\varepsilon_{\text{inh}} = J_{\text{inh}} \frac{t - t_0 - \Delta}{\tau_{\text{inh}}^2} e^{-(t-t_0-\Delta)/\tau_{\text{inh}}} \theta(t - t_0 - \Delta) \ . \tag{4.40}$$

for the inhibitory postsynaptic response. Here J is the synaptic weight, positive for excitatory and negative for inhibitory synapses, t_0 the spiking time of the presynaptic neuron, τ determines the width of the α-function, and Δ is the delay of the inhibition. θ denotes the Heaviside step function [$\theta(t) = 0$ if $t < 0, \theta(t) = 1$ if $t \geq 0$].

Detailed description of the recurrent model

The recurrent model consists of Poisson output neurons that are driven by the continuous input function $s_{\text{in}}(t)$ convoluted with the excitatory postsynaptic response.

All neurons feature a recurrent connection that feeds output spikes back into the neuron. The recurrent connection is characterized by a specific combination of inhibitory time constant τ_{inh}, delay Δ, and inhibitory coupling strength J_{inh}. Again, inhibitory and excitatory currents are described by α-functions of the form (4.39) and (4.40).

4.3.2 Analysis: tuning of time constants

We are now going to mathematically discuss the behavior of the two types of periodicity detectors in more detail.

Feedforward model

We mimic a realistic, usually half-wave rectified periodic signal by a shifted cosine similar to the (positive) envelope of an AM signal just as in (4.22). As a consequence of the properties of a Poisson neuron, this input function then describes the inhomogeneous firing probability density λ_{in} of the input neuron,

$$s_{\mathrm{in}}(t) = \frac{A}{2}\left[1 - \cos(2f\pi t)\right] = \lambda_{\mathrm{in}}(t) \ . \tag{4.41}$$

As in the last section, the total response $\varepsilon_{\mathrm{total}}$ of one specific output neuron to the input neuron activity is given by [referring to (4.39) and (4.40), and (4.23), respectively]

$$\varepsilon_{\mathrm{total}} = \varepsilon_{\mathrm{exc}} + \varepsilon_{\mathrm{inh}} \ . \tag{4.42}$$

Contrary to our analysis of the excitatory–excitatory setup we stick with Poisson neurons as output neurons. Hence, similar to (4.24), the firing probability density λ_{out} of the output neurons is then given by

$$\lambda_{\mathrm{out}}(t) = \int_{-\infty}^{\infty} \mathrm{d}s \ s_{\mathrm{in}}(s) \ \varepsilon_{\mathrm{total}}(t-s) \ . \tag{4.43}$$

Equation (4.43) can be evaluated exactly for the given input function (4.41) with result

$$\begin{aligned} \lambda_{\mathrm{out}}(t) =& \frac{1}{2} J_{\mathrm{exc}} \left[\frac{\left(1+4\zeta_{\mathrm{exc}}^2\right)^2 + \left(-1+4\zeta_{\mathrm{exc}}^2\right)\cos(2f\pi t) - 4\zeta_{\mathrm{exc}}\sin(2f\pi t)}{(1+4\zeta_{\mathrm{exc}}^2)^2} \right] \\ &+ \frac{1}{2} J_{\mathrm{inh}} \left[\frac{\left(1+4\zeta_{\mathrm{inh}}^2\right)^2 + \left(-1+4\zeta_{\mathrm{inh}}^2\right)\cos[2f\pi(t-\Delta)] - 4\zeta_{\mathrm{inh}}\sin[2f\pi(t-\Delta)]}{\left(1+4\zeta_{\mathrm{inh}}^2\right)^2} \right] \end{aligned} \tag{4.44}$$

where we assumed $A = 1$ and $\zeta_j = f\pi\tau_j$ for $\tau_j = \tau_{\text{exc}}$ and τ_{inh}, respectively. The symmetry between excitation and delayed inhibition is obvious.

In order to analyze (4.44), it is desirable to reduce the number of free parameters. We therefore set $J_{\text{exc}} = 1$ in the following. Furthermore, it is easy to see that (4.44) is of the form $\lambda_{\max}(J_{\text{inh}};\, \Delta; \tau_{\text{exc}}; \tau_{\text{inh}}) * \cos(2f\pi t + \phi)$, ϕ being a phase shift of no further interest. It is thus sufficient to consider the amplitude $\lambda_{\max}$ to obtain an understanding of the system:

$$\begin{aligned}\lambda_{\max} = \frac{1}{2} + \frac{J_{\text{inh}}}{2} + \Big\{ &\frac{1}{2\left(1+4\zeta_{\text{exc}}^2\right)^2\left(1+4\zeta_{\text{inh}}^2\right)^2} \times \\ &\Big[\left(J_{\text{inh}} + 4J_{\text{inh}}\zeta_{\text{exc}}^2\right)^2 + \left(1+4\zeta_{\text{inh}}^2\right)^2 + 2J_{\text{inh}}\big(1 + 16\zeta_{\text{exc}}^2\zeta_{\text{inh}}^2 \\ &\quad - 4\zeta_{\text{exc}}^2 + 16\zeta_{\text{exc}}\zeta_{\text{inh}} - 4\zeta_{\text{inh}}^2\big)\cos(2f\pi\Delta) \\ &\quad + 8J_{\text{inh}}(\zeta_{\text{exc}} - \zeta_{\text{inh}})\left(1 + 4\zeta_{\text{exc}}\zeta_{\text{inh}}\right)\sin(2f\pi\Delta)\Big]\Big\}^{\frac{1}{2}}. \qquad (4.45)\end{aligned}$$

Next, we want to get rid of J_{inh} as a free parameter. For an optimal performance of our model the maximum of the response should be a clear peak. We can minimize $\lambda_{\max}$ at the boundary of the range of frequencies we are interested in – that is, positive frequencies. If $\lambda_{\max}$ is minimal at the left and at the right border of the frequency range under consideration, the peak, somewhere in between these two limits, should be easy to distinguish. At f =1 Hz, $\lambda_{\max}$ is minimal for an inhibitory coupling strength J_{inh} of -1 to -0.99, depending on the parameters chosen. This is true for the complete range of accessed parameters; that is Δ, τ_{exc}, and τ_{inh} taking any value from 1 ms to 10 ms each. At the same time, the limiting value of (4.45) for $f \to \infty$ is $(1 + J_{\text{inh}})/2$. The optimal inhibitory coupling J_{inh} is therefore -1, the same absolute value as the excitatory coupling. This is called *balanced inhibition*; see Sec. 4.3.3.

In the following, we will take the delay Δ to be 2 ms. This assumption is equivalent to our concept of constructing a “minimal” model since, in order to turn an excitatory signal inhibitory, we need at least one interneuron. Two milliseconds are a reasonable time for a signal passing one neuron. At the end of this section we will discuss the influence of the delay and its variation on the behavior of the model.

Figure 4.11, left, illustrates the behavior of the solution (4.45) for four different parameter sets. We see that the solutions have a clear maximum for one *specific* frequency ranging from about 14 Hz (solution A) to approximately 140 Hz (solution D), depending on the combination of time constants τ_{exc} and τ_{inh}. Before analyzing (4.44) further, it is interesting to compare its behavior with numerical simulations

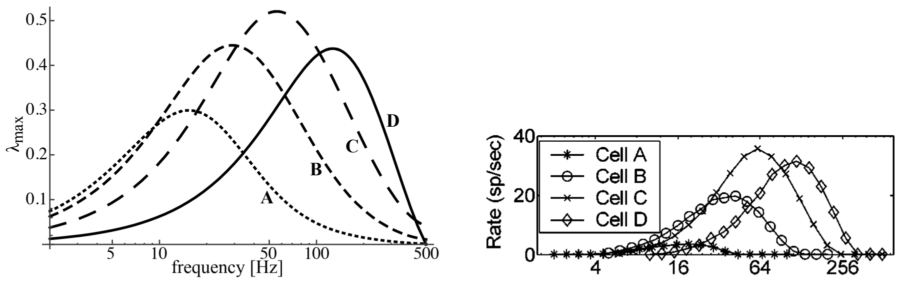

Figure 4.11: Frequency detection by excitatory–inhibitory networks. *Left*: time-invariant amplitude $\lambda_{\max}$ of the firing probability density against frequency of the input signal s_{in}. Four sets of parameters are shown, each resulting in a maximum of the amplitude at different frequencies. Characteristics of the solutions match numerical results from [136]; cf. right panel. The parameters except J_{inh} were taken from [136]: A(τ_{exc}; τ_{inh}) = (5 ms; 10 ms), B(2 ms; 6 ms), C(1 ms; 3 ms), D(1 ms; 1 ms); $\Delta = 2\,\text{ms}$; $J_{\text{inh}} = -1$. *Right*: absolute rate modulation transfer function of the SFIE model [136], rate versus frequency [Hz]. Four different model cells in the inferior colliculus have been simulated, every cell responding maximally to a certain modulation frequency of the signal. The match of analytical and numerical results for identical parameters is surprising since the SFIE model [136] is much more complicated than our setup.

published before [136]. In the latter, time constants as well as delay between excitation and inhibition used by us have led to almost identical results; see Fig. 4.11, right. It is noteworthy that, motivated by physiological findings, the setup of the model of Nelson and Carney [136] is much more complicated than ours: two subsequent stages of delayed inhibition and excitation with different coupling strengths featuring three cell populations (auditory nerve, cochlear nucleus, and inferior colliculus) and four synapses lead to quantitatively the same results regarding frequency selectivity.

Ideally, a maximum that is to be discerned clearly should have a big amplitude (in this respect, cell A in the right panel of Fig. 4.11 would be a bad example). As a consequence we are interested in those regions of our solution where the amplitude $\lambda_{\max}$ is maximal. Since an analytical solution is not feasible we will revert to a graphical approach.

Figure 4.12 shows the amplitude of the solution (4.45) for different time constants and frequencies. For low frequencies we can discern two distinct regions of maximal amplitude: amplitude is maximal when inhibitory and excitatory time constants have a maximal difference (dark areas). In the figure, the amplitude is minimal for $\tau_{\text{inh}} = \tau_{\text{exc}} + 0.5\Delta$ (bright area), but this relation only holds if Δ is small compared to the time scale of the frequency under consideration. As the frequency increases

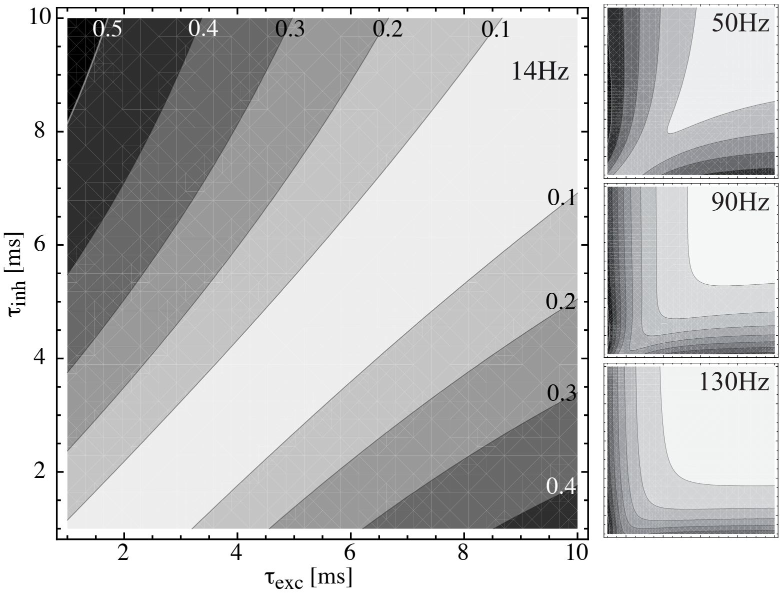

Figure 4.12: Amplitude of the response of the feedforward model for low and high-frequency signals with a fixed delay as a function of excitatatory and inhibitory time constant. Black stands for a large, white for a small amplitude. Big: low-frequency stimuli (here: 14 Hz) lead to two clearly separated areas of maximal response. The response is maximal for a large difference of τ_{exc} and τ_{inh}, and a larger inhibitory time constant results in a higher amplitude (~ 0.5 vs. ~ 0.4 in case of larger excitatory time constant). Right, top to bottom (units same as on the left): increasing frequency of the stimulus (here: 50, 90, and 130 Hz) leads to a merge of the two areas of maximal response and a decreasing amplitude. The amplitude is maximal when either excitatory or inhibitory time constant is very small. Here $\Delta = 2\,\text{ms}$, $J_{\text{inh}} = -1$.

(right side of Fig. 4.12, top to bottom), the two regions of maximal amplitude move towards the origin and merge. The overall amplitude shrinks but is still largest for one of the time constants being very small. At 130 Hz, finally, the amplitude maximum is reached at combinations of very small inhibitory with even smaller excitatory time constants.

The response magnitude dependence upon excitatory and inhibitory time constant as shown in figure 4.12 does not, however, elucidate how the frequency with maximal response amplitude depends on the combination of excitatory and inhibitory time constant and their respective delay. Since the derative of $\lambda_{\max}$ (4.45) with respect to inhibitory and excitatory time constants is not tractable analytically, we have to stick to a graphical solution once more. Figure 4.13 depicts the dependence of the maximum of (4.44) upon excitatory and inhibitory time constants. Generally, lower time constants lead to a maximum for higher frequencies. Lower frequencies can be accessed by larger time constants, leading to no strict cutt-off in the low-frequency range. The delay breaks the symmetry of the solution and results in an "anomaly" along the line $\tau_{\text{inh}} = \tau_{\text{exc}} + 0.5\Delta$ if $\Delta \ll 1/f$. Since the amplitude of the solution is minimal along this axis, useful maxima lie at small values of either the excitatory or inhibitory time constant. In principle every combination of a small excitatory with a larger inhibitory time constant has an equivalent combination of small inhibitory with larger excitatory time constant, but the discrimination ability for high frequencies is poorer (see maximum for 90 Hz and 130 Hz in Fig. 4.13). In addition, combinations of small excitatory with larger inhibitory time constants lead to higher amplitudes, so that our original idea of filtering and subtracting different frequencies with help of different time constants seems suggestive.

The considerations above are, however, only valid if the assumption of Δ being much smaller than $T = 1/f$ holds. If Δ is varied independently of f the landscape of the solution changes, as figure 4.14 illustrates, drastically.

Figure 4.14 shows the amplitude $\lambda_{\max}$ as a function of dimensionless time constants τ' and delay Δ'. We define dimensionless units x' as $x' = x/T$. For integer multiples of the cycle periods T of the signal the amplitude behaves very similarly to figure 4.12, big panel, viz., two distinct areas of maximal amplitude are separated by a diagonal of minimal response. The reason is that a delay of 2 ms is small compared to the cycle period of 14 Hz, ∼70 ms. Increasing the delay Δ (Fig. 4.14: to $0.25\,T$) shifts the axis of minimum response to the right; that is, to larger excitatory time constants. At the same time the maximum moves towards smaller inhibitory time constants. The very same behavior occurs when signal frequency is increased but the delay is kept constant. The increase of frequency from 14 Hz to 50, 90, and 130 Hz at a constant delay of 2 ms in figure 4.12 corresponds to an increase of the

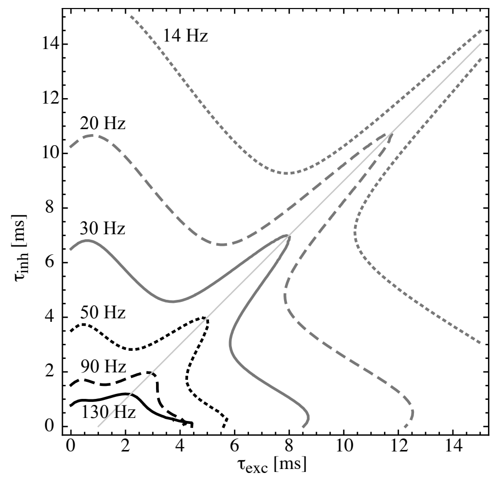

Figure 4.13: Contours of the maximal response amplitude in the τ_{inh}-τ_{exc}-plane for different signal frequencies with fixed delay. Black solid, dashed, and dotted line, grey solid, dashed, and dotted line: amplitude maxima for 130, 90, 50, 30, 20, and 14 Hz; thin grey line: $\tau_{\text{inh}} = \tau_{\text{exc}} + 0.5\Delta$. As the frequency increases, the maximal amplitude appears at smaller time constants. We note that the performance of the model can only be estimated in combination with the absolute amplitude; cf. Fig. 4.12. Here $\Delta = 2\,\text{ms}$, $J_{\text{inh}} = -1$.

delay from 0.028 T to 0.1, 0.18, and 0.26 T in the current setting. At a delay corresponding to half the cycle period of the signal, symmetry is restored, and a single maximum exists at $(\tau_{\text{exc}};\ \tau_{\text{inh}}) = (0;\ 0)$; that is, the PSPs behave like δ- instead of α-functions. Since at this particular delay the inhibitory signal operates in the valley of the excitatory signal, a minimal excitatory–inhibitory interference leads to a maximal response. The minimal interference is provided by δ-functions as PSPs. At a further increase of the delay the maximum wanders towards larger excitatory time constants, and a second maximum appears for small excitatory and large inhibitory time constants. For $\Delta = T$, the contour of the amplitude is finally symmetric again, featuring two clearly separated areas of maximal response.

Two considerations restrict our interest to the regime shown in the upper half of figure 4.14. First, in various animals most neurons that are sensitive to amplitude modulation are responding maximally to frequencies between 30 and 100 Hz. Second, the initial motivation for a model of neuronal frequency identification by means of inhibition has been the lack of evidence for delay lines with $\Delta > 10\,\text{ms}$ in biological systems, so only "short" delays are of interest to us. A delay of 4 ms, which is a value well within the range of physiological constraints, corresponds to 0.5 T at 125 Hz. In order to obtain a maximal response to amplitude-modulated stimuli

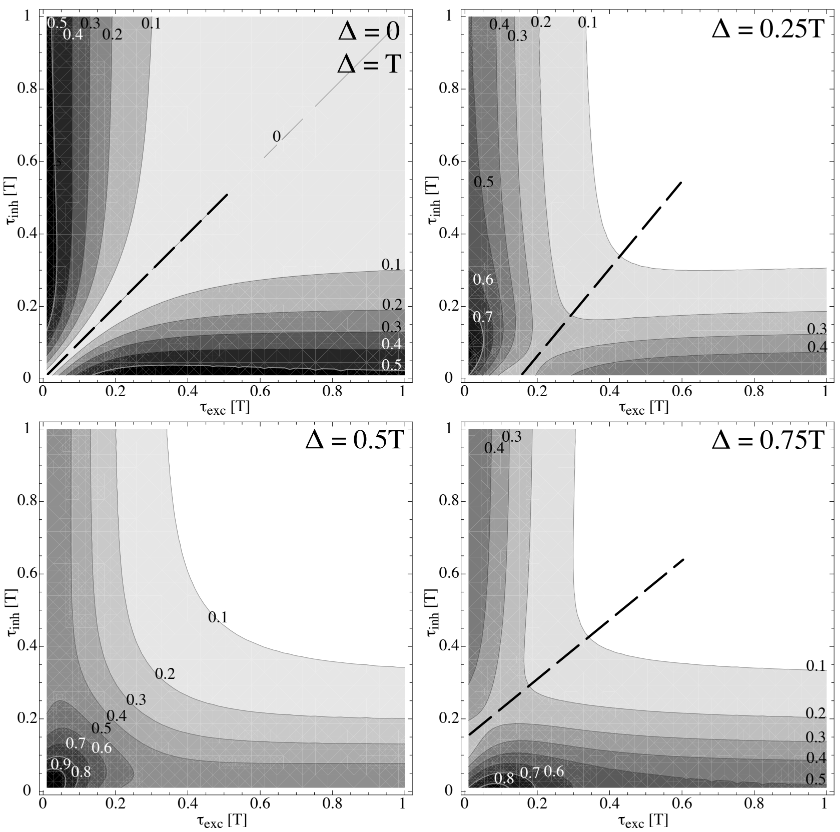

Figure 4.14: Influence of the delay on the amplitude in dimensionless units. Amplitude of the solution to (4.45) as a function of dimensionless excitatory and inhibitory time constant in cycle periods T of the signal. Upper left: in case of no delay or the delay matching exactly one period of the signal frequency, the solution is completely symmetric relative to excitatory and inhibitory time constants. Upper right: increasing delay shifts the axis of the minimum to larger excitatory time constants and the maximum to the origin. Lower left: a delay of $T/2$ leads to a maximal response for minimal excitatory and inhibitory time constants; that is, δ-functions as PSPs. Lower right: the axis of the minimum reappears at further increase of the delay, this time at larger inhibitory time constants. Since we are interested in low frequencies and delays of limited length, only the regime displayed in the upper panel is relevant. $J_{\mathrm{inh}} = -1$.

in this frequency range, it therefore makes sense to combine small excitatory with larger inhibitory time constants.

The delay can also be varied so as to allow a broader range of frequencies. A very short delay of $\Delta = 0.3\,\mathrm{ms}$ pushes the upper limit of about 140 Hz for a Δ of 2 ms to about 500 Hz. Longer delays extend the accessible frequency range to lower frequencies. Changing the delay from 2 ms to a Δ of 15 ms, for example, lowers the preferred frequency for $(\tau_{\mathrm{exc}}; \tau_{\mathrm{inh}}) = (1; 15.5)$ from 14 Hz to 10 Hz.

With a given delay we can take the excitatory time constant to be a very small value (e.g. 1 ms) and vary the inhibitory time constant in order to control the preferred frequency of our model; cf. Fig. 4.13. We thus arrive at a neuronal band-pass filter characterized by the biologically plausible variation of a single parameter, the inhibitory time constant.

The analytical calculations above have been verified by numerical simulations. As in the last section we have used a population of Poisson input neurons and LIF output neurons. The outcome matched our analytical results very closely. This was to be expected, since (4.43) does not only describe the firing probability density for Poisson neurons but also holds for the expectation value of an input current to LIF neurons; cf. (4.24). Interestingly, the phase locking of the output spikes has been increased by the model even further than in the excitatory–excitatory setup.

Recurrent model

The idea of a neuronal band-pass filter we developed in the last section can be compressed into an even simpler setup. One single population of neurons suffices if we use a recurrent inhibitory connection; see the bottom panel of Fig. 4.10. Again, we will consider Poisson neurons for our analytic calculations.

For sufficient neuronal activity [59] we can describe the rate function λ of a single Poisson neuron or neuron population projecting back to itself with a particular delay time Δ by an integral equation similar to (4.6), namely

$$\begin{aligned}\lambda(t) =& J_{\mathrm{exc}} \int_{-\infty}^{\infty} \mathrm{d}s\; g_{\mathrm{exc}}(s) s_{\mathrm{in}}(t-s) + J_{\mathrm{inh}} \int_{-\infty}^{\infty} \mathrm{d}s\; g_{\mathrm{inh}}(s;\Delta)\lambda(t-s) \\ =& J_{\mathrm{exc}}(g_{\mathrm{exc}} \star s_{\mathrm{in}})(t) + J_{\mathrm{inh}}(g_{\mathrm{inh}} \star \lambda)(t)\ .\end{aligned} \tag{4.46}$$

The rate function consists of the sum of the external input s_{in} and the delayed inhibitory input from the recurrent loop, both “smeared out” by the kernel g_{exc} and g_{inh}, respectively. The feedback strength is given by J_{inh}, and we choose g to be α-functions as in (4.39) and (4.40) so as to ensure causality and obtain unit weights.

To solve (4.46) we change to Fourier space where convolutions are ordinary products. The Fourier-transformed version of (4.46) reads

$$\Lambda(\omega) = J_{\text{exc}} G_{\text{exc}}(\omega) S_{\text{in}}(\omega) + J_{\text{inh}} G_{\text{inh}}(\omega; \Delta) \Lambda(\omega) \ , \tag{4.47}$$

where the Fourier transform of each input term is denoted by a capital letter. The solution is thus given by

$$\Lambda = \frac{J_{\text{exc}} G_{\text{exc}}}{1 - J_{\text{inh}} G_{\text{inh}}} S_{\text{in}} \tag{4.48}$$

and can be transformed back into a function of time by taking its inverse Fourier transform. This equation corresponds to (4.10) in the excitatory–excitatory setup.

In a way similar to (4.19) the last section, we mimick a half-wave rectified signal by a shifted cosine function

$$s_{\text{in}}(t) = \frac{1}{2} \left[B - \cos(2 f \pi t) \right] \tag{4.49}$$

where B denotes the shift of the cosine along the y-axis. This is a necessary precaution in order to avoid a negative rate function. We obtain a solution that is, just as described in the feedforward model by (4.14), of the form

$$\lambda(t) = \lambda_{\max}(B; J_{\text{exc}}; J_{\text{inh}}; \Delta; \tau_{\text{exc}}; \tau_{\text{inh}}) \cos(2 f \pi t + \phi) \ . \tag{4.50}$$

As before, ϕ is a phase shift of no further interest. For any finite solution we can find a B that can shift the solution to positive values and prevent a negative rate function. Since this shift does not affect the solution otherwise, we can as well forego the shift; that is, in the following we set $B = 0$ for the sake of convenience. In analogy to (4.15) we now turn to the time-invariant amplitude $\lambda_{\max}$ that is of interest for a characterization of the system,

$$\lambda_{\max} = \frac{J_{\text{exc}}}{\sqrt{2}} \frac{(4 f^2 \pi^2 \tau_{\text{inh}}^2 + 1)}{\sqrt{\Upsilon^2 + \Omega^2}} \tag{4.51}$$

where

$$\begin{aligned} \Upsilon = \sqrt{2} J_{\text{inh}} \left(-1 + 4 \zeta_{\text{exc}}^2 \right) + 2 \sqrt{\pi} \times & \\ \Big\{ \left[1 + 16 \zeta_{\text{exc}}^2 \zeta_{\text{inh}}^2 - 4 \left(\zeta_{\text{exc}}^2 + 4 \zeta_{\text{exc}} \zeta_{\text{inh}} + \zeta_{\text{inh}}^2 \right) \right] \cos(2 f \pi \Delta) & \\ + 4 (\zeta_{\text{exc}} + \zeta_{\text{inh}}) \left(-1 + 4 \zeta_{\text{exc}} \zeta_{\text{inh}} \right) \sin(2 f \pi \Delta) \Big\} & \end{aligned} \tag{4.52}$$

and

$$\begin{aligned} \Omega = -4 J_{\text{inh}} \sqrt{2} \zeta_{\text{exc}} - 2 \sqrt{\pi} \times & \\ \Big(\left(1 + 16 \zeta_{\text{exc}}^2 \zeta_{\text{inh}}^2 - 4 \left(\zeta_{\text{exc}}^2 + 4 \zeta_{\text{exc}} \zeta_{\text{inh}} + \zeta_{\text{exc}}^2 \right) \right) \sin(2 f \pi \Delta) & \\ + 4 (\zeta_{\text{exc}} + \zeta_{\text{inh}}) \left(-1 + 4 \zeta_{\text{exc}} \zeta_{\text{inh}} \right) \cos(2 f \pi \Delta) \Big) & \end{aligned} \tag{4.53}$$

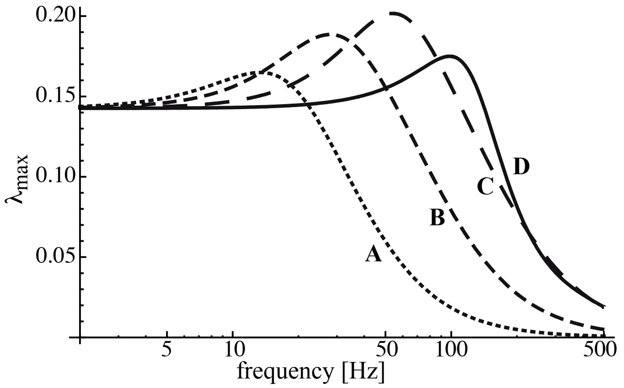

Figure 4.15: Frequency detection of the recurrent excitatory–inhibitory network for balanced inhibition in the form of the time-invariant amplitude $\lambda_{\max}$ of the rate function against the frequency of the input signal s_{in}. The parameter sets are identical to those of Fig. 4.11 and lead to a maximal response for virtually identical frequencies. The quality of the peaks is low as compared to the feedforward model. A smaller overall amplitude and a relatively high amplitude for low frequencies deteriorates the recurrent network performance. Parameter values are A(τ_{exc}; τ_{inh}) = (5 ms; 10 ms), B(2 ms; 6 ms), C(1 ms; 3 ms), D(1 ms; 1 ms); $\Delta = 2\,\text{ms}$; $J_{\text{inh}} = -1$.

with $\zeta_j = f\pi\tau_j$ for $\tau_j = \tau_{\text{exc}}$ and τ_{inh}. In order to reduce the number of parameters we set $J_{\text{exc}} = 1$ again.

Figure 4.15 illustrates the performance of the recurrent model. Parameter sets that are identical to the ones we have used in the example for the feedforward model (Fig. 4.11) lead to a very similar behavior, viz., maximal response at virtually identical frequencies. The peaks are, however, less clear since for low frequencies the amplitude does not drop as in the feedforward model. In addition, the overall amplitudes are lower.

For a quantitative understanding of the recurrent model, we proceed as in the last section and change to dimensionless units. We derive the inhibitory coupling strength $J_{\max}$ for which the dimensionless version of (4.51) is maximal,

$$J_{\max} = \sqrt{2\pi}\left[(1-\zeta_{\text{inh}}^2)\cos(2f\pi\Delta) - \zeta_{\text{inh}}\sin(2f\pi\Delta)\right] \tag{4.54}$$

By combining this inhibitory coupling strength with the dimensionless version of (4.51) we arrive at a $\lambda_{\max}$ in dimensionless units that is dependent only on the excitatory and inhibitory time constant as well as the delay,

$$\lambda_{\text{max}} = \frac{1}{\sqrt{2\pi}} \times \frac{1 + \zeta_{\text{inh}}^2}{(1 + \zeta_{\text{exc}}^2)[2\zeta_{\text{inh}} \cos(\zeta_\Delta) + (1 - \zeta_{\text{inh}}^2) \sin(\zeta_\Delta)]} \tag{4.55}$$

where $\zeta_\Delta = 2f\pi\Delta$. Obviously the excitatory time constant does not characterize the band-pass response of the model but simply scales the amplitude; we will not discuss this parameter in the following.

We can now easily derive a constraint for the relation between inhibitory time constant $\zeta_{\text{inh}} = f\pi\tau_{\text{inh}}$ and delay $\zeta_\Delta = f\pi\Delta$: Equation (4.55) is maximal if

$$\zeta_\Delta = \arctan\left(\frac{2\zeta_{\text{inh}}}{\zeta_{\text{inh}}^2 - 1}\right) + n\pi \tag{4.56}$$

with $n = 0$ if $\zeta_{\text{inh}} > 1$ and $n = 1$ if $\zeta_{\text{inh}} < 1$. For the inhibitory time constant approaching zero, that is, δ- instead of α-functions as PSPs, (4.56) reduces to $\zeta_\Delta = \pi$ or, in dimensional units,

$$\Delta = \frac{1}{2f} = 0.5T \ , \tag{4.57}$$

just as in the feedforward model where the amplitude is maximal for $(\tau_{\text{exc}}; \tau_{\text{inh}}) = (0; 0)$ if the delay is 0.5 T.

Equation (4.56) could be interpreted as if an arbitrary short delay could be compensated by an appropriate inhibitory time constant. This is not the case since, as a consequence of (4.54), such an arbitrary short delay would require a *very* large inhibitory coupling. For instance, a delay of $\Delta = 0.05T$ would result in $\tau_{\text{inh}} = T$ and $J_{\text{inh}} = -101$. But how far can we get with a realistic inhibitory coupling?

From figure 4.15 we see that restricting the inhibitory strenght to a balanced inhibition ($J_{\text{inh}} = -1$) as in the feedforward model still gives reasonable results. What is, however, the relation between parameter set and preferred frequency, the frequency for which the response of the model is maximal? Analytic insight is easy in dimensionless units but hard to transfer into dimensional units, so we will stick to a graphical approach as before.

The relation between excitatory time constant τ_{exc}, inhibitory time constant τ_{inh}, and preferred frequency is shown in figure 4.16. As in the feedforward model, lower inhibitory time constants lead to a maximum for higher frequencies. However, contrary to the feedforward model, there is no symmetry between combinations of large excitatory with small inhibitory and combinations of large inhibitory with small excitatory time constants. All maxima that are characteristic to a given frequency feature inhibitory time constants that are larger than the excitatory ones. Since,

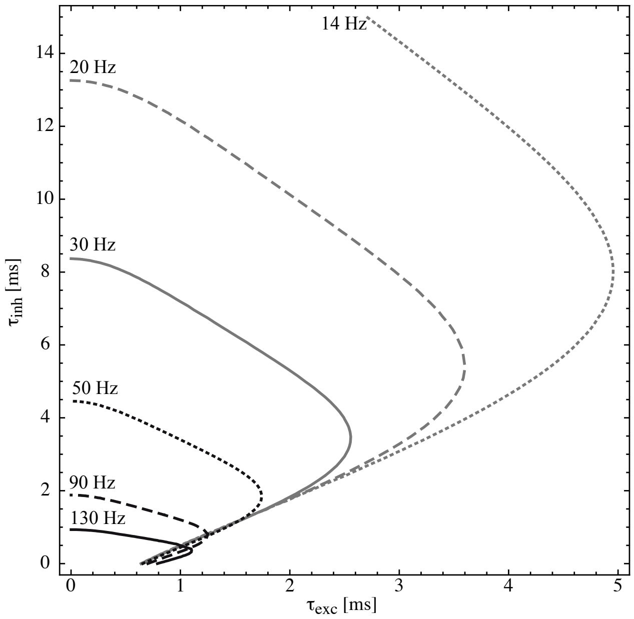

Figure 4.16: Contours of the maximal response amplitude in the $\tau_{\rm inh}$-$\tau_{\rm exc}$-plane for different signal frequencies with fixed delay for the recurrent model. Black solid, dashed, and dotted line, grey solid, dashed, and dotted line: amplitude maxima for 130, 90, 50, 30, 20, and 14 Hz. As the frequency increases, the maximal amplitude appears at smaller time constants. We note that the performance of the model can only be estimated in combination with the absolute amplitude; as in the feedforward model largest amplitudes are obtained for small excitatory time constants. In contrast to the feedforward model, all maxima that are characteristic for a given frequency feature inhibitory time constants that are larger than the excitatory ones. Here, $\Delta = 2\,\mathrm{ms}$, $J_{\rm inh} = -1$.

just as in the feedforward model, the amplitude of the solution is maximal for small excitatory time constants, it makes sense to choose the excitatory time constant as small as possible. The result is a system where –given delay and inhibitory strength fixed– the frequency response can again be tuned over one order of magnitude by a variation of the inhibitory time constant.

4.3.3 Discussion: potency of the mixed setup

As we have seen, a simple excitatory–inhibitory feedforward model can identify frequencies in the range of approximately ten to several hundred Hertz relying on biologically plausible parameters only, viz., short delays and balanced inhibition. The model works best for a very short, fixed excitatory time constant. Given a specific delay, the frequency where the response of the model is maximal –the preferred frequency of the model– can be varied by tuning the inhibitory time constant. Alternatively, the inhibitory time constant can be taken to be short and the model can be tuned by the excitatory time constant. A recurrent setup shows a behavior very similar to the feedforward model and can identify frequencies in the same range. The amplitude peaks, however, are shallow as compared to the response maxima in the feedforward model. Furthermore, in contrast to the feedforward model, a short excitatory time constant is necessary for the model to work. Again, the model can be tuned by choosing the appropriate inhibitory time constant.

Interestingly, the characteristics of the neuronal band-pass filter at hand are quite different from the initial conception we formulated in the introduction. The naïve picture of simply subtracting the envelopes of two low-pass-filtered signals does not explain the characteristics of the system. If the neuronal band-pass filter followed such a simple relation and we had defined the cut-off frequency as the frequency where the response of the system is half of the maximal response, the preferred frequency would be given by $f_{\mathrm{pref}} = 1/(4\pi^2 \tau_{\mathrm{exc}} \tau_{\mathrm{inh}})^{1/2}$. This would lead to plain hyperbola-like curves instead of the odd-shaped curves depicted in figure 4.13. Obviously an in-depth analytical description is crucial in order to arrive at a thorough understanding of the system.

Although motivated by our intention to create a “minimal model”, the delay of 2 ms chosen in the current calculations may appear arbitrary. Appearances, however, are deceiving as can be seen through experimental results of inhibition being delayed by 2.4 [188] and 2 ms [111] in the auditory cortex. In the auditory brainstem one can expect even shorter delays like 0.6 ms for the inhibition [190]. Thinking of the influence a short delay has on the preferred frequency of our model, these short delays fit the concept of the auditory brainstem dealing with higher frequency signal periodicity than the cortex. In fact, sensitivity for amplitude modulations up to 1000 Hz has been reported in the experimental literature [92]. However, neurons sensitive to modulation frequencies > 300 Hz are few and far between, while the majority of the neurons is confined to the range of 30-100 Hz. This finding is valid for various animals [104,109,153–155] so that, from a conceptual point of view, most of the AM sensitivity of neurons can be explained by our model.

Quite surprisingly, "balanced inhibition" (BI) turns out to be the optimal choice for the inhibitory coupling strength. Balanced inhibition denotes inhibitory input of approximately the same strength as the excitatory input. It has been observed at several locations and under various conditions, ranging from cat visual cortex [3, 132] to the cochlear nucleus in rats [145], and in ongoing as well as sensory-evoked neuronal activity [140]. A number of possible functions has been proposed for BI, but its actual purpose is still a matter of ongoing debate.

Our findings suggest an additional role for BI, namely in the processing of signal periodicity such as amplitude modulations and/or the processing of vibratory signals. The findings of single whisker deflections causing a sequence of excitation and BI in the rat barrel cortex [82], and BI changing the chopping frequency in chopper neurons in the very same animal [145] fit here nicely. Balanced inhibition has been proposed to account for enhancing temporal precision and regulating random background activity [188]. Furthermore, experimental evidence supports the importance of BI in the processing of frequency modulated tones [199].

The idea of BI acting as a kind of gate or filter between cortical areas [82] agrees with our present results in that the frequency range of our model covers the β (13-30 Hz) and γ (30-100 Hz) oscillations that are believed to contribute to the communication between different parts of the brain, and to attention, a related topic. Furthermore, BI is locked to noise envelopes in the cat auditory cortex, and locking is suppressed by low-level tones [111]. This can be taken as a hint towards BI playing a role in the attentional framework.

4.4 Neuronal binding and signal recognition

We have shown that without any specialized architectural features –a generic neuron model, short delays and a variation of time constants– the modulation frequency components of a signal can be resolved neuronally. In their remarkable simplicity, the models already show characteristics that are surprisingly consistent with experimental data. A small number of input neurons is enough to sample the input signal with sufficient accuracy. Importantly, the neuronal parameters necessary for periodicity identification lie in the range of milliseconds, comparable to the typical auditory time scale. Concerning the biological relevance, however, there is a problem with the purely excitatory approach. For the identifcation of low AM frequencies very long neuronal delays are required. Such neuronal delays have not been found at least in the mammalian auditory system in spite of extensive research. The role the delays play in the excitatory–excitatory setup can be taken over by chopper neurons, though, if spiking regularity meets the required precision; cf. Sec. 4.2.3. The excitatory–inhibitory approach by contrast *ab initio* only relies on neuronal building blocks well-documented in the auditory system, namely long inhibitory time constants and balanced inhibition. Thus both principles are biologically realizable for the initially discussed conversion of a phase code into a rate code. Regardless which strategy is the preferred one, the output of the networks can be further enhanced by post-processing mechanisms. Lateral inhibition, for example, can be used to sharpen the peaks in the frequency response.

At this point we want to emphasize that *extracting* the slowly-varying envelope from an input signal is easily accomplished in a biological system. Half-wave rectification of the signal and subsequent low-pass filtering suffice. This can be accomplished by simple means as a slow synapse that filters out all high frequency components. Even if the slow envelope of the signal has been extracted, however, the frequency of the envelope oscillations is still unknown. The advantage of the proposed models is their ability to *identify* the frequency of the envelope oscillations. Such an identification of envelope frequency is of great importance since the recognition of sounds depends on the capability of determining the periodicity of the input signal, which brings us back to the concept of auditory objects.

Namely, it is exactly this *identification* of periodicity that makes our models suitable for the neuronal formation of auditory objects. In concreto, our models connect the two main advantages of the concept of an object to neuronal mechanisms in the following way.

The first advantage, as we remember, is the idea of binding together different aspects, here frequencies, of one signal that belong together for efficient processing. As already discussed in the introduction, common amplitude modulations are a sufficient cue for achieving such binding; cf. Sec. 1.1. In our auditory system, the cochlea acts as an initial frequency decomposer that splits the neuronal processing of auditory signals into distinct channels that enable frequency-specific processing. Within such a frequency channel, we now can deploy a set of AM-identifying circuits, each with a different preferred AM frequency. Each frequency channel is thus further decomposed into AM channels. As a consequence different frequency components of one auditory signal with identical AM frequencies will lead to activity in the same

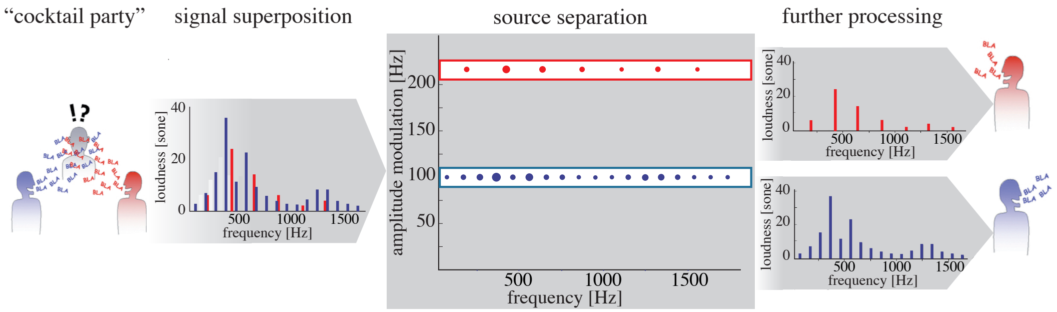
"cocktail party" signal superposition source separation further processing

Figure 4.17: The principle of auditory object formation based on common amplitude modulations. A superposition of two signals is illustrated by our initial cocktail party comic (left; cf. Fig. 1.1). The cochlea resolves frequency, but cannot identify the frequency components belonging to the individual sources. An array of neuronal periodicity detectors resolves the individual frequency components by decomposing each cochlear frequency channel into AM channels (grey box). Subsequently, the different auditory objects, here the red and the blue speaker, can be processed further (right).

AM channels. This is how we can realize grouping for auditory object formation by means of our models; cf. Fig. 4.17. In a subsequent stage our auditory system can then selectively process the frequency components of that very object, for instance by means of an attentional framework.

The second advantage of the concept of an object is that of signal recognition. As just described, binding allows to select frequency components belonging to the same source. Since each auditory object is consequently defined by its specific combination of contributing frequencies (bound together by identical AM), the identity of the auditory object can be recognized independently of the surrounding by memorizing this specific combination. This is indeed what is happening in our auditory system as we will expound in the following.

Both advantages are nicely illustrated by human speech. For this purpose we recall the "cocktail party" from the very beginning of the thesis. For a reminder, within a complex mixture of sounds, we want to extract what our vis-à-vis is trying to tell us. To keep things simple, we stick to the vowel "a" as an example. Analogously to what we have described above we need to first bind the right frequency components together and second recognize the vowel, that is, identify both the vowel and the speaker so as to understand him.

First, we address the idea of binding. We know that human speech consists of several frequency bands that are comodulated by a guttural frequency called the *voicing* *frequency* or *fundamental frequency*; cf. Sec. 1.1. In figure 4.18 we see the Fourrier transform of the vowel “a”. The voicing frequency is visualized by the distance between the bars that represent the isolated frequency components. In the top panel, each component is modulated by 100 Hz, and in the bottom panel by 220 Hz. As the voicing frequency differs from speaker to speaker, any frequency component belonging to an individual speaker can be grouped by means of our models[1]. We hence in our cocktail party scenario have arrived at the level where we do not have to deal with a mixture of sounds any more but where we dispose of frequency packages that originate from different sources. Thus, by focussing attention on specific AM frequencies, we can switch between these packages and separate contributions from different individuals. These isolated sound packages can be localized by standard mechanisms based on interaural time differences since the necessary information hereunto, the phase locking of the input, is preserved throughout the processing in our models. Auditory object formation would then occur *before* object localization. This agrees not only with our reasoning in the introduction of this thesis but also with previous experimental work showing that spatial separation of sounds is indeed linked to the comodulation of signal amplitude across several frequency channels. So we know which frequency components belong together and where they come from, but we do not know what vowel is pronounced and who the speaker is.

Second, we address the idea of signal recognition. Here our goal is to recognize the identity of the vowel, here “a”, as well as the identity of the individual pronouncing it. Figure 4.18 visualizes the situation: A signal, a spoken “a” as in “father”, consists of many modulated frequency components of different averaged intensity, or loudness, depicted as bars. The peaks ($*$) in the spectrum determine the identity of the vowel. The three peaks at about 750, 1100, and 2600 Hz are characteristic for an “a”. An “e” as in “heed”, for instance, would be determined by peaks at about 250, 2250, and 3250 Hz. So the identity of each vowel is defined by peaks at characteristic frequencies in a given frequency distribution. A distribution is not only determined by its peaks, though – the ratio of the different peaks in respect to each other is another important characteristic. It determines the identity of the individual who pronounces the vowel. We explain this by figure 4.18. The grey shaded areas in the top and bottom panel, the envelopes of the frequency distributions, indicate that different individuals pronounce the same vowel since the peaks are identical whereas their ratio is not. Contrariwise the black bars in the

[1] There is an alternative concept for frequency grouping: As we see in figure 4.18, our auditory system could exploit the relation between AM frequency and the distance of the relevant frequency components with respect to each other along the spectral axis. This relation, however, then needs to be learned for each and every possible combination of frequency components and AM frequency – a, given the ease of our approach, incredibly wasteful strategy that will hardly be realized in a biological system.

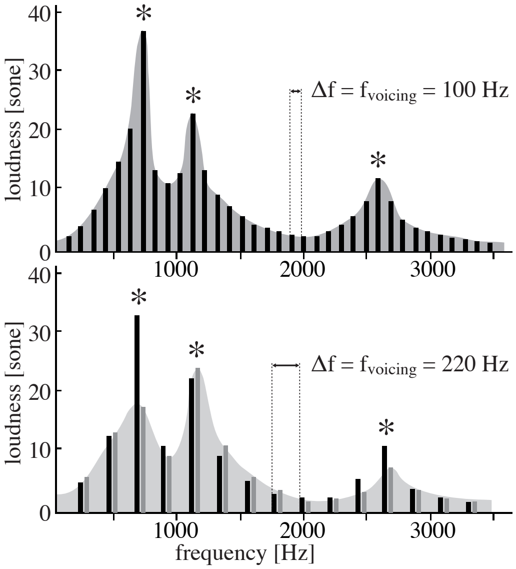

Figure 4.18: Exemplary spectra of the vowel "a" as in father in loudness versus frequency. The peaks of the spectra, the *formants*, are clearly discernible, and their position ($*$) marks the vowels as "a". The voicing frequency is reflected in the distance of the frequency components (bars). The voicing frequency is 100 Hz (top), and 220 Hz (bottom), respectively. The shaded areas mark the envelope of the spectrum, a characteristic identifying the individual who pronounces the vowel. The black bars in the bottom originate from the same speaker as the black bars in the top with a higher modulation frequency, whereas the grey bars in the bottom can be attributed to another individual. Adapted from [8].

bottom level can easily be attributed to the speaker of the top panel, only that he speaks the vowel with a higher modulation frequency.

In summary, everyday experience proves that our auditory system does not only exploit and memorize isolated features of auditory signals such as the identity of peaks that determine a vowel. Rather, complex contextual relations, for instance the relative ratio of the peaks, vitally contribute to auditory object recognition such as the attribution of speech to a specific, known individual. A necessary prerequisite for all this, however, is the correct grouping of frequencies stemming from one source, that is, a possibility to reliably access them. Our models give this possibility in a very simple and natural way, and hence are to be considered as archetypes for the formation of auditory objects in neuronal networks.

Chapter 5

Synopsis and research perspective

After zooming in from the general concepts of auditory objects and echoes of chapter 1 onto specific theoretical concepts and their neuronal realization in chapter 2 and 4, we now want to zoom out so as to obtain a "10.000 m-above-ground" perspective. From up there what we have gained in almost hundred pages can be condensed into two fundamental statements. First, the mathematical concept of stochastic optimality gives us a framework for evaluating biological strategies for echo suppression. Namely, we are now able to compare optimal environment-dependent temporal receptive fields to their neuronal counterparts in the auditory brainstem. Second, the neuronal limitations for the identification of signal periodicity allow for AM-based object formation. Again, we can now compare characteristics of neurons in the auditory brainstem such as the temporal jitter of spikes or neuronal time constants with the theoretical requirements and the performance of existing auditory circuitry.

We conclude this thesis with three personal remarks on promising sites for future research. The first one concerns the extension of our view on biological echo suppression towards a multimodal perspective. The second one is on the conceptual advantage of using many different frequencies for information transmission in a natural environment. The third one sheds light on a possible origin –learning via spike-timing-dependent plasticity– of neuronal circuitry for periodicity identification. In the final paragraph of this thesis we come back to Seebeck and Ohm.

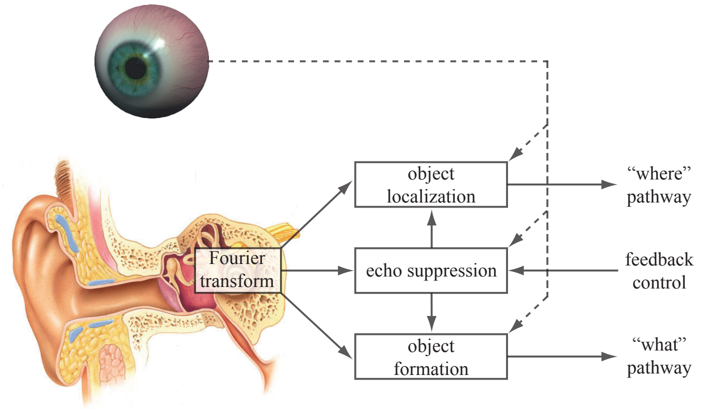

Figure 5.1: Auditory processing overview. Object localization, object formation, and echo suppression are the three core tasks the auditory system has to perform after the Fourier transform in the cochlea. Object localization and object formation both strongly rely on phase relations within the signal, and efficient echo suppression is necessary for the phase relations to remain intact. All three tasks can be supported in a multimodal approach, for instance, by vision providing top-down information about auditory source localization, identity, and echo structure as determined by the environment.

5.1 The multimodality of echo suppression

We remember that amplitude modulations are one of several cues for auditory object formation. Apart from AM an important cue, especially when it comes to short signals, is the "common onset". Common onset and common amplitude modulations can be subsumed to the unifying cue of a fixed phase relation, as we did in chapter 1. Interestingly, this concept –a fixed phase relation– immediately entangles spatial location. Namely, for sound source localization we compare the relative phases at the two ears that depend on the of the source to each ear. These two effects, phase relation for accessing both object identity and object location, are already complex in their combination and are further complicated by the existence of echoes. Echoes completely mess up the phase relation in an unpredictable way. Hence, the auditory brainstem has to deal with the phase relation of any signal concerning three different aspects at the same time: object formation, object localization, and echo suppression; cf. Fig 5.1.

While tremendous amount of research has been done on auditory signal localization, the two remaining aspects have been widely ignored. The present thesis, however, lays a cornerstone for understanding the formation of auditory objects in biological systems by mathematically analyzing intrinsic limitations and potency of neuronal periodicity identification. It turns out that in auditory localization and especially in auditory object formation, the handling of dynamic, environment-dependent echoes is a key ingredient. Our optimal model for echo suppression can, as we saw in chapter 2, be tuned to a high degree of generality and consequently deal with a variety of environments. We could extend our model, though, by adding the possibility to adjust the generality of the reconstruction filter to the actual need. Since our setup for echo suppression is of feedforward quality, the model in its present form cannot adapt itself to the environment-at-hand. Here, a feedback structure could enhance the capabilities of the presented framework by realizing a dynamic adaptation via real-time adjustment of the model parameters.

Such a feedback control is an evident step since, interestingly, the auditory system is known for massive feedback projections, some of which even reach the cochlea. Bearing our model in mind we easily can imagine feedback projections to several stages within, e.g., the auditory brainstem modifying the flexible components of the reconstruction function. These modifications of the processing pathway would then correspond to the above real-time adjustment of our model. Since our model is based on the mathematical concept of error minimization, the feedback loops within the auditory pathway would then basically convey an error code to the individual nuclei, a consideration that we find for instance in a recent approach to feedback in the auditory system by Gonzalo Otazu et al. [143].

An interesting extension of our model would be the inclusion of input from other modalities into auditory feedback signals. Since the environment shapes the echoes that we need to suppress, it makes sense to exploit this information as it is available from other senses. Vision, for instance, can provide a whole clutch of cues that reduce the possible echo characteristics. Such cues can be, e.g., the size of the room-at-hand, the listener's distance from the next wall, the speaker's distance from the next wall, or the speaker's size and sex. All this information then would influence our expectation on the degradation of the signal and, consequently, alter and improve signal reconstruction, i.e., echo suppression, via feedback control. The walking robot LOLA equipped with both microphones and cameras will be a nice example demonstrating the power of a biomimetic multimodal approach.

5.2 The spreading of information across frequencies

Another intriguing thought is the following: bearing chapter 3 in mind, namely the importance of amplitude modulations, we might wonder why there is so much ado about different frequency components belonging to one auditory object. We know that already three bands of amplitude-modulated noise give a very good sentence intelligibility. If most of the information is transmitted in the range of several hundred Hertz, the arising question is why most of the energy in our speech is contained in the frequency range of *thousands* of Hertz.

An appealing answer results from a simple consideration. Within the natural environment, obviously there are many sounds that may disturb a signal we want to communicate. So as to obtain a reliable signal communication it is a good strategy to *spread* the signal we want to transmit onto several transmission channels, in our case the carrier frequencies in the range of several thousand Hertz. Such a spreading of information is a common technique in modern communication systems and comes in various colors and forms depending on the space in which you spread the information. Since we are talking about the frequency domain, we give a short explanation of frequency-division multiplexing.

In contrast to single carrier modulation, where information is transmitted only by variations of phase and amplitude of a single carrier frequency, frequency-division multiplexing (FDM) extends this concept by using multiple subcarrier frequencies within one transmission channel. The total information to be sent in such a channel is usually divided between the various subcarriers. A good example of FDM is the current NTSC television multiplex. Frequency-division multiplexing offers an advantage over single-carrier modulation in terms of narrowband frequency interference since narrowband interference will only affect one of the subcarriers. The other subcarriers will not be affected by the interference, hence a more reliable information transmission is achieved.

So we can consider the fact that we spread information across several frequencies when communicating as a biological realization of FDM. Or rather, since speech has been there first, FDM is a technical realization of our natural communication strategy. Although the idea of FDM and similar technology has been around for a while in the field of technical application, its realization has been rendered possible rather lately by the invention of the microprocessor. Complementary, the emergence of the AM processing scheme in neuronal networks is an interesting question yet to be answered.

5.3 The learning of periodicity identification

As we discussed in chapter 4, it is straightforward to arrange the neuronal building blocks we have derived for periodicity identification into an array that allows to

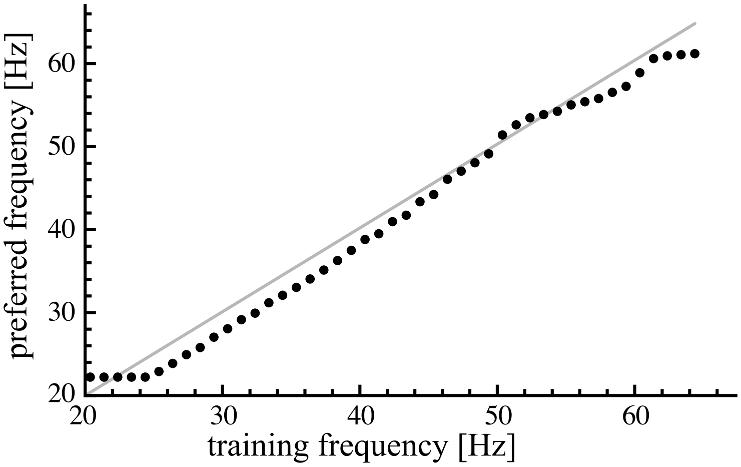

Figure 5.2: Learning of AM identifcation in an excitatory–inhibitory setup. Preferred frequency is plotted against training frequency in Hertz. The grey line marks the identity of training frequency and preferred frequency, the black dots our preliminary analytical results. We see a linear relation close to identity between the preferred frequency of the setup and the frequency used for training. For obtaining the results we have employed an adapted STDP learning rule that takes into account the location of a synapse along the dendritic tree.

access frequency components with identical AM. Such an array is then a map for amplitude modulations. For a reminder, a map is a neuronal representation of the external world, or, more precise, a neuronal representation of a specific feature of the external world – in our case of frequency components with similar AM. Maps in general are an important theoretical concept and play a major role in sensory processing [80, 103].

In the auditory system there are different kinds of maps, amongst them maps for frequency, interaural time difference, interaural amplitude difference, and even, not surprising to us after having learned so much about their importance, amplitude modulations [64, 122, 141, 147, 163, 180]. Their mere existence, however, does not tell us why they exist or how they can emerge. While we cannot provide a definite answer to the first question, at least the latter has been answered for some maps. Interaural time difference maps, for instance, can be achieved by applying a specific learning theory, spike-time-dependent plasticity (STDP) [113]. For details on the STDP learning theory we refer to elsewhere [79].

This learning theory can also be deployed to the formation of AM maps. A preliminary result for a subset of neurons in an excitatory–inhibitory setting is depicted in figure 5.2. To obtain the result displayed, we have combined one fixed excitatory with several learning inhibitory synapses. The learning is realized by an adapted STDP learning rule that, by taking into account the location of the synapses along the dendritic tree, allows the learning of different time constants, hence, according to section 4.3, the learning of different preferred frequencies. Such a setting can lead to a linear relation between, or even identity of the frequency that is used to train the network and the frequency the network responds to maximally, its preferred frequency; cf. Fig. 5.2. So our analytical calculations underlying figure 5.2 constitute a solid starting point for a mathematical explanation of the emergence of AM maps, a matter to be explored in detail in future work.

Last but not least, let's come back to poor old Seebeck and Ohm, both of whom had so furious battles about who is right and who is mistaken. In Seebeck's opinion, very similar to the ideas underlying chapter 3, our auditory system relies on periodicity cues for analyzing the auditory scene. He was an experimentalist and based his concepts on experiments with sirens, mainly. Ohm being a theoretician proclaimed that, since a Fourier transform of any signal will give an unambiguous mean of identifying this very signal, the auditory system will employ a Fourier transformation for auditory scene analysis. Now, 150 years later, we finally can settle the affair: They both have been right. As discussed in section 4.4 signal periodicity is used for solving the binding problem, the formation of auditory objects, and a Fourier transform is needed for the interpretation of the object, for instance, for recognizing a vowel.

Bibliography

[1] Oxford Dictionary of English, 2nd revised edn. Oxford University Press, Oxford (2005)

[2] Aicher, B., Tautz, J.: Signal transmission through the substrate. J Comp Physiol A **166**, 345–353 (1990)

[3] Anderson, J.S., Carandini, M., Ferster, D.: Orientation Tuning of Input Conductance, Excitation, and Inhibition in Cat Primary Visual Cortex. J Neurophysiol **84**, 909–926 (2000)

[4] Avendaño, C., Deng, L., Hermansky, H., Gold, B.: The analysis and representation of speech. In: S. Greenberg, W. Ainsworth, A. Popper, R. Fay (eds.) Speech Processing in the Auditory System, chap. 2, p. 63 ff. Springer, New York (2004)

[5] Barth, F.: Neuroethology of the spider vibration sense. In: F. Barth (ed.) Neurobiology of Arachnids, chap. 11, p. 203 ff. Springer, New York (1985)

[6] Barth, F.: The vibrational sense of spiders. In: R.R. Hoy, A.N. Popper, R.R. Fay (eds.) Comparative Hearing: Insects, chap. 7, p. 228 ff. Springer, New York (1998)

[7] Barth, F., Geethabali: Spider vibration receptors: Threshold curves of individual slits in the metatarsal lyriform organ. J Comp Physiol A **148**, 175–185 (1982)

[8] Benade, A.H.: Fundamentals of musical acoustics. Oxford University Press, New York (1976)

[9] Bendor, D., Wang, X.: The neuronal representation of pitch in primate auditory cortex. Nature **436**, 1161–1165 (2005)

[10] Beranek, L.L.: Music, Acoustics & Architecture. Wiley, New York (1962)

[11] Best, V., Gallun, F.J., Carlile, S., Shinn-Cunningham, B.G.: Binaural interference and auditory grouping. J. Acoust. Soc. Am. **121**(2), 1070–1076 (2007)

[12] Blauert, J.: Spatial Hearing. MIT Press, Cambridge, MA (1999)

[13] Bleckmann, H.: Prey identification and prey localization in surface-feeding fish and fishing spiders. In: J. Atema, R. Fay, A. Popper, W. Tavolga (eds.) Sensory Biology of Aquatic Animals, chap. 24, p. 619 ff. Springer, New York (1987)

[14] Bleckmann, H.: Reception of Hydrodynamic Stimuli in Aquatic and Semiaquatic Animals. Gustav Fischer, Stuttgart (1994)

[15] Bleckmann, H., Barth, F.: Sensory ecology of a semi-equatic spider *Dolomedes triton* II. The release of predatory behavior by water surface waves. Behavioral Ecology and Sociobiology **14**, 303–312 (1984)

[16] Bleckmann, H., Borchard, M., Horn, P., Görner, P.: Stimulus discrimination and wave source localization in fishing spiders (*Dolomedes triton* and *D. okefinokensis*). J Comp Physiol A **174**, 305–316 (1994)

[17] Bleckmann, H., Breithaubt, T., Blickhan, R., J., T.: The time course and frequency content of hydrodynamic events caused by moving fish, frogs and crustaceans. J Comp Physiol A **168**, 749–757 (1991)

[18] Bleckmann, H., Waldner, I., Schwartz, E.: Frequency discrimination in the surface-feeding fish *Aplocheilus lineatus* – A prerequisite for prey localization? J Comp Physiol A **143**, 485–490 (1981)

[19] Boll, S.F.: Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust., Speech, Signal Process. **27**, 113–120 (1979)

[20] Borst, M., Langner, G., Palm, G.: A biologically motivated neural network for phase extraction from complex sounds. Biol Cybern **90**, 98–104 (2004)

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

[22] Bregman, A.: Auditory Scene Analysis. MIT Press, Cambridge, MA (1990)

[23] Brownell, P.: Compressional and surface waves in sand: Used by desert scorpions to locate prey. Science **197**, 479–482 (1977)

[24] van Brunt, B.: The calculus of variations. Springer, Heidelberg (2000)

[25] Buell, T.N., Hafter, E.R.: Combination of binaural information across frequency bands. J. Acoust. Soc. Am. **90**(4), 1894–1900 (1991)

[26] Bürck, M.: Neurophysik der Echounterdrückung. Master's thesis, Technische Universität München (2005)

[27] Bürck, M., Friedel, P., Sichert, A.B., Vossen, C., van Hemmen, J.L.: Optimality in mono- and multisensory map formation. Biol Cybern **103**(1), 1–20 (2010)

[28] Bürck, M., van Hemmen, J.L.: Modeling the cochlear nucleus: A site for monaural echo suppression? J. Acoust. Soc. Am. **122**(4), 2226–2235 (2007)

[29] Bürck, M., van Hemmen, J.L.: Neuronal identification of signal periodicity by balanced inhibition. Biol Cybern **100**, 261–270 (2009)

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

[31] Burkitt, A.: A review of the integrate-and-fire neuron model: II. Inhomogeneous synaptic input and network properties. Biol Cybern **95**, 97–112 (2006)

[32] Burkitt, A., Clark, G.: Synchronization of the neural response to noise periodic synaptic input. Neural Comp **13**, 2639–2672 (2001)

[33] Cai, H., Carney, L.H., Colburn, S.H.: A model for binaural response properties of inferior colliculus neurons. I. A model with interaural time difference-sensitive excitatory and inhibitory inputs. J. Acoust. Soc. Am. **103**(1), 475–493 (1998)

[34] Cariani, P.: Neural timing nets. Neur Netw **14**, 737–753 (2001)

[35] Cariani, P.: Recurrent timing nets for auditory scene analysis. Proc Int Joint Conf Neural Netw pp. 1575–1580 (2003)

[36] Cherry, E.C.: Some Experiments on the Recognition of Speech, with One and with Two Ears. J. Acoust. Soc. Am. **25**(5), 975–979 (1953)

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

[38] Clifton, R.K.: Breakdown of echo suppression in the precedence effect. J. Acoust. Soc. Am. **82**(5), 1834–1835 (1987)

[39] Clifton, R.K., Freyman, R.L., Litovsky, R.Y., McCall, D.: Listeners' expectations about echoes can raise or lower echo threshold. J. Acoust. Soc. Am. **95**(3), 1525–1533 (1994)

[40] Cooke, M., Ellis, D.: The auditory organization of speech and other sources in listeners and computational models. Speech Comm **35**, 141–177 (2001)

[41] Coombs, S., Görner, P., Münz, H. (eds.): The Mechanosensory Lateral Line: Neurobiology and Evolution. Springer, New York, NY (1989)

[42] Culling, J.F., Summerfield, Q.: Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. J. Acoust. Soc. Am. **98**(2), 785–797 (1995)

[43] Dean, I., Harper, N., McAlpine, D.: Neural population coding of sound level adapts to stimulus statistics. Nat. Neurosci. **8**, 1684–1689 (2005)

[44] Demany, L., Semal, C.: Dichotic fusion of 2 tones one octave apart: Evidence for internal octave templates. J. Acoust. Soc. Am. **83**, 687–695 (1988)

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

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

[47] Deutsch, D.: Octave generalization of specific interference effects in memory for tonal pitch. Percept Psychophys **13**, 271–275 (1973)

[48] Devore, S., Ihlefeld, A., Hancock, K., Shinn-Cunningham, B., Delgutte, B.: Accurate sound localization in reverberant environments is mediated by robust encoding of spatial cues in the auditory midbrain. Neuron **62**, 123–134 (2009)

[49] Diesmann, M., Gewaltig, M.O., Aertsen, A.: Stable propagation of synchronous spiking in cortical neural networks. Nature **402**, 529–533 (1999)

[50] Elepfandt, A.: Localization of water surface waves with the lateral line system in the clawed toad (*Xenopus laevis* daudin). In: D. Varjú, H. Schnitzler (eds.) Localizatoin and Orientation in Biology and Engineering, p. 63 ff. Springer, New York, NY (1984)

[51] Elepfandt, A.: Wave frequency recognition and absolute pitch for water waves in the clawed frog *Xenopus laevis*. J Comp Physiol A **158**, 235–238 (1986)

[52] Ephraim, Y., Malah, D.: Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. IEEE Trans. Acoust., Speech, Signal Process. **32**, 1109–1121 (2004)

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

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

[55] Flanagan, J.L., Lummis, R.C.: Signal processing to reduce multipath distortion in small rooms. J. Acoust. Soc. Am. **47**(6), 1475–1481 (1970)

[56] Franosch, J.M.P., Lingenheil, M., van Hemmen, J.L.: How a frog can learn what is where in the dark. Phys. Rev. Lett. **95**, 78106 (2005)

[57] Franosch, J.M.P., Sichert, A.B., Suttner, M.D., van Hemmen, J.L.: 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 (2005)

[58] Franosch, J.M.P., Sobotka, M.C., Elepfandt, A., van Hemmen, J.L.: Minimal model of prey localization through the lateral-line system. Phys. Rev. Lett. **91**, 158101 (2003)

[59] Friedel, P., Bürck, M., van Hemmen, J.L.: Neuronal identification of acoustic signal periodicity. Biol. Cybern **97**, 247–260 (2007)

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

[61] Furuya, K., Kataoka, A.: Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction. IEEE Trans. Audio, Speech, Language Process. **15**, 1579–1591 (2007)

[62] Gammaitoni, L., Hänggi, P., Jung, P., Marchesoni, F.: Stochastic resonance. Rev. Mod. Phys. **70**, 223–287 (1998)

[63] Gardner, M.B.: Historical background of the haas and/or precedence effect. J. Acoust. Soc. Am. **43**, 1243–1248 (1968)

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

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

[66] Gerstner, W., Kistler, W.: Spiking Neuron Models. Cambridge University Press, Cambridge (2002)

[67] Giard, M.H., Fort A.and Mouchetant-Rostaing, Y., Pernier, J.: Neurophysiological Mechanisms of Auditory Selective Attention in Humans. Frontiers in Bioscience **5**, 84–94 (2000)

[68] Gillespie, B.W., Malvar, H.S., Florencio, D.A.F.: Speech dereverberation via maximum-kurtosis subband adaptive filltering. Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. **6**, 3701–3704 (2001)

[69] Griffiths, T.D., Uppenkamp, S., Johnsrunde, I., Josephs, O., Patterson, R.D.: Encoding of the temporal regularity of sound in the human brainstem. Nat. Neurosci. **4**(6), 633–637 (2001)

[70] Grothe, B.: Interaction of excitation and inhibition in processing of pure tone and amplitude-modulated stimuli in the medial superior olive of the mustached bat. J. Neurophysiol. **71**, 706–721 (1994)

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

[72] Grothe, B., Covey, E., Casseday, J.H.: Medial superior olive of the big brown bat: Neuronal responses to pure tones, amplitude modulations, and pulse trains. J. Neurohysiol. **86**, 2219–2230 (2001)

[73] Grothe, B., Neuweiler, G.: The function of the medial superior olive in small mammals: temporal receptive fields in auditory analysis. J Comp Physiol A **186**, 413–423 (2000)

[74] Haas, H.: Über den Einfluss des Einfachechos auf die Hörsamkeit von Sprache. Acustica **1**, 49–58 (1951)

[75] Hafter, E.R., Buell, T.N., Richards, V.M.: Onset coding in lateralization: Its form, site and function. In: G. Edelman (ed.) Auditory function, pp. 647–674. Wiley (1988)

[76] Hänsel, E., Schmidt, G. (eds.): Acoustic Echo and Noise Control – A Practical Approach. Wiley, Hoboken (2004)

[77] Harris, G.G., Flanagan, J.L., Watson, B.J.: Binaural interaction of a click with a click pair. J. Acoust. Soc. Am. **35**, 672–678 (1963)

[78] Hasan, M.K., Salahuddin, S., Khan, M.R.: Reducing signal-bias from MAD estimated noise level for DCT speech enhancement. Signal Process. **84**, 151–162 (2004)

[79] van Hemmen, J.L.: Theory of synaptic plasticity. In: F. Moss, S. Gielen (eds.) Handbook of Biological Physics (Vol.4), Neuro-informatics, Neural Modelling, pp. 771–823. Elsevier, Amsterdam (2001)

[80] van Hemmen, J.L.: The map in your head: How does the brain represent the outside world? Chem. Phys. Chem. **3**, 291–298 (2002)

[81] Hergenröder, R., Barth, F.: The release of attack and escape behavior by vibratory stimuli in a wandering spider (*Cupiennius salei* Keys.). J Comp Physiol A **152**, 347–358 (1983)

[82] Highley, M.J., Contreras, D.: Balanced Excitation and Inhibition Determine Spike Timing during Frequency Adaptation. J. Neurosci. **26**(2), 448–457 (2006)

[83] Hilis, J.M., Ernst, M.O., Banks, M.S., Landy, M.S.: Combining sensory information: Mandatory fusion within, but not between, senses. Science **298**, 1627–1630 (2002)

[84] Hudspeth, A., Corey, D.: Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli. Proc. Natl. Acad. Sci. USA **74**, 2407–2411 (1977)

[85] Hughes, A.: The topography of vision in mammals of contrasting life style: Comparative optics and retinal organization. In: F. Crescitelli (ed.) Handbook of sensory physiology, vol. VII/5, chap. 8, p. 613 ff. Springer, Berlin Heidelberg New York (1977)

[86] Humphreys, L.: Generalization as a function of method of reinforcement. J Exp Psych **25**, 361–372 (1939)

[87] Ingham, N., McAlpine, D.: Spike-frequency adaptation in the inferior colliculus. J Neurophysiol **91**, 632–645 (2004)

[88] Izhikevich, E.M.: Resonate-and-fire neurons. Neural Networks **14**, 883–894 (2001)

[89] Järvilehto, M.: The eye: Vision and perception. In: G.A. Kerkut, L.I. Gilbert (eds.) Nervous System: Sensory, *Comprehensive Insect Physiology, Biochemistry, and Pharmacology*, vol. 6, pp. 355–429. Pergamon, Oxford (1985)

[90] Jazayeri, M., Movshon, J.A.: Optimal representation of sensory information by neural populations. Nat. Neurosci. **9**(5), 690–696 (2006)

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

[92] Joris, P.X., Schreiner, C.E., Rees, A.: Neural Processing of Amplitude-Modulated Sounds. Physiological Review **84**, 541–577 (2004)

[93] Jost, J., Li-Jost, X.: Calculus of variations. Cambridge University, Cambridge (1998)

[94] Kailath, T.: A view of three decades of linear filtering theory. IEEE Transactions on Information Theory **20**, 146–181 (1974)

[95] Kalmijn, A.: Hydrodynamic and acoustic field detection. In: J. Atema, R. Fay, A. Popper, W. Tavolga (eds.) Sensory Biology of Aquatic Animals, chap. 4, pp. 83–130. Springer, New York, NY (1988)

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

[97] Käse, R., Bleckmann, H.: Prey localization by surface wave ray-tracing: Fish track bugs like oceanographers track storms. Experientia **43**, 290–293 (1987)

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

[99] Keller, C.H., Takahashi, T.T.: Localization and Identification of Concurrent Sounds in the Owl's Auditory Space Map. J. Neurosci. **25**(45), 10446–10461 (2005)

[100] Kempter, R., Gerstner, W., van Hemmen, J.L.: How the threshold of a neuron determines its capacity for coincidence detection. BioSys **48**, 105–112 (1998)

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

[102] Kempter, R., Gerstner, W., van Hemmen, J.L., Wagner, H.: Extracting oscillations: Neural coincidence detection with noise periodic spike input. Neural Comp **10**, 1987–2017 (1998)

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

[104] Krishna, B., Semple, M.: Auditory temporal processing: Response to sinusoidally amplitude-moulated tones in the inferior colliculus. J. Neurophysiol. **84**, 255–273 (2000)

[105] Kuttruff, H.: Room Acoustics, 3rd edn. Elsevier Applied Science, London and New York (1991)

[106] Landolfa, M., Barth, F.: Vibrations in the orb web of the spider *Nephilia clavipes*: Cues for discrimination and orientation. J Comp Physiol A **179**, 493–508 (1996)

[107] Lang, H.: Surface wave discrimination between prey and nonprey by the backswimmer *Notonecta glauca* L. (Hemiptera, Heteroptera). Beh Ecol Sociobiol **6**, 233–246 (1980)

[108] Langner, G.: Periodicity coding in the auditory system. Hearing Res. **60**, 115–142 (1992)

[109] Langner, G., Schreiner, C.: Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J. Neurophysiol. **60**, 1799–1822 (1988)

[110] Large, E.W., Crawford, J.D.: Auditory Temporal Computation: Interval Selectivity Based on Post-Inhibitory Rebound. J. Computational Neurosci. **13**, 125–142 (2002)

[111] Las, L., Stern, E.A., Nelken, I.: Representation of Tone in Fluctuating Maskers in the Ascending Auditory System. J. Neurosci. **25**(6), 1503–1513 (2005)

[112] Lebart, K., Boucher, J.M., Denbigh, P.N.: A new method based on spectral subtraction for speech dereverberation. Acta Acoustica **87**(3), 359–366 (2001)

[113] Leibold, C., van Hemmen, J.L.: Spiking neurons learning phase delays: how mammals may develop auditory time-difference sensitivity. Phys. Rev. Lett. **94**, 168102 (2005)

[114] Licklider, J.: A duplex theory of pitch perception. Experientia **7**, 128–134 (1951)

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

[116] Litovsky, R.Y., Colburn, H.S., Yost, W.A., Guzman, S.J.: The precedence effect. J. Acoust. Soc. Am. **106**(4), 1633–1654 (1999)

[117] Lomber, S.G., Malhotra, S.: Double dissociation of what and where processing in auditory cortex. Nat. Neurosci. **11**(5), 609–616 (2008)

[118] Ma, W., Beck, J., Latham, P., Pouget, A.: Bayesian inference with probabilistic population codes. Nat. Neurosci. **9**, 1432–1438 (2006)

[119] Mach, E.: Über die physiologische Wirkung räumlich vertheilter Lichtreize. Sitzungsber. Akad. Wiss. Wien II **54**, 393–408 (1866)

[120] Maeder, P.P., Meuli, R.A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J.P., Pittet, A., Clarke, S.: Distinct Pathways Involved in Sound Recognition and Localization: A Human fMRI Study. Neuroimage **14**, 802–816 (2001)

[121] Magal, C., Schöller, M., Tautz, J., Casas, J.: The role of leaf structure in vibration propagation. J. Acoust. Soc. Am. **108**, 2412–2418 (2000)

[122] Manley, G.A., Köppl, C., Konishi, M.: A neural map of interaural intensity differences in the brainstem of the barn owl. J. Neurosci. **8**, 2665–2676 (1988)

[123] Masters, W.: Vibrations in the orbwebs of *Nuctenea sclopetaria* (Araneidae). Beh Ecol Sociobiol **15**, 217–223 (1984)

[124] Masters, W., Markl, H., Moffat, A.: Transmission of vibration in a spider's web. In: W. Shear (ed.) Spiders: Webs, Behavior, and Evolution, chap. 3, p. 49 ff. Stanford University Press, Stanford, CA (1986)

[125] McAlpine, D., Grothe, B.: Sound localization and delay lines - do mammals fit the model? Trends Neurosci **26**(7), 347–350 (2003)

[126] Meddis, R., O'Mard, L.: A unitary model of pitch perception. J. Acoust. Soc. Am. **102**(3), 1811–1820 (1997)

[127] Meddis, R., O'Mard, L.: Virtual pitch in a computational physiological model. J. Acoust. Soc. Am. **120**(6), 3861–3869 (2006)

[128] Megela Simmons, A., Ferragamo, M.: Periodicity extraction in the anuran auditory nerve. J Comp Physiol A **172**, 57–69 (1993)

[129] Meyer, G.F., Wuerger, S.M., Röhrbein, F., Zetzsche, C.: Low-level integration of auditory and visual motion signals requires spatial co-localization. Experimental Brain Research **166**, 538–547 (2005)

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

[131] Miyoshi, M., Kaneda, Y.: Inverse Filtering of Room Acoustics. IEEE Trans. Acoust., Speech, Signal Process. **36**(2), 145–152 (1988)

[132] Monier, C., Chavane, F., Baudot, P., Graham, L.J., Frégnac, Y.: Orientation and Direction Selectivity of Synaptic Inputs in Visual Cortical Neurons: A Diversity of Combinations Produces Spike Tuning. Neuron **37**, 663–680 (2003)

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

[134] Neely, S.T., Allen, J.B.: Invertibility of a room impulse response. J. Acoust. Soc. Am. **66**(1), 165–169 (1979)

[135] Nelken, I., Rotman, Y., Bar Yosef, O.: Responses of auditory-cortex neurons to structural features of natural sounds. Nature **397**, 154–157 (1999)

[136] Nelson, P.C., Carney, L.H.: A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. J. Acoust. Soc. Am. **116**(4), 2173–2186 (2004)

[137] O'Craven, K.M., Downing, P.E., Kanwisher, N.: fMRI evidence for objects as the units of attentional selection. Nature **401**(7), 584–587 (1999)

[138] Oertel, D.: The role of timing in the brain stem auditory nuclei of vertebrates. Ann Rev Physiol **61**, 497–519 (1999)

[139] Ohm, G.S.: Ueber die definition des tones nebst daran geknüpfte theorie der sirene und ähnlicher tonbildender vorrichtungen. Ann Phys Chem **59**, 513–565 (1843)

[140] Okun, M., Lampl, I.: Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. Nat. Neurosci. **11**(5), 535–537 (2008)

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

[142] Oppenheim, A.V., Schafer, R.W., Stockham, T.G.: Nonlinear filtering of multiplied and convolved signals. Proc. IEEE **56**(8), 1264–1291 (1968)

[143] Otazu, G.H., Grothe, B., Leibold, C.: A corticothalamic circuit model for intensity invariant source identification in auditory scenes. Submitted to Nat. Neurosci. (2010)

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

[145] Paolini, A.G., Clarey, J.C., Needham, K., Clark, G.M.: Balanced inhibition and excitation underlies spike firing regularity in ventral cochlear nucleus chopper neurons . European Journal of Neuroscience **21**, 1236–1248 (2005)

[146] Pecka, M., Zahn, T.P., Saunier-Rebori, B., Siveke, I., Felmy, F., Wiegrebe, L., Klug, A., Pollak, G.D., Grothe, B.: Inhibiting the inhibition: a neuronal network for sound localization in reverberant environments. J. Neurosci. **27**(7), 1782–1790 (2007)

[147] Pickles, J.O.: An Introduction to the Physiology of hearing. Academic, 2nd ed., London (1988)

[148] Pouget, A., Dayan, P., Zemel, R.: Inference and computation with population codes. Ann. Rev. Neurosci. **26**, 381–410 (2003)

[149] Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: Numerical Recipes: The Art of Scientific Computing, 3rd edn. Cambridge University Press, Cambridge (2007)

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

[151] Qiu, F.T., Sugihara, T., von der Heydt, R.: Figure-ground mechanisms provide structure for selective attention. Nat. Neurosci. **10**, 1492–1499 (2007)

[152] Rauschecker, J.P., Tian, B.: Mechanisms and streams for processing of “what” and “where” in auditory cortex. Proc. Natl. Acad. Sci. USA **97**(22), 11800–11806 (2000)

[153] Rees, A., Møller, A.: Responses of neurons in the inferior colliculus of the rat to AM and FM tones. Hearing Res. **10**, 301–330 (1983)

[154] Rees, A., Møller, A.: Stimulus properties influencing the repsonses of inferior colliculus neurons to amplitude-modulated sounds. Hearing Res. **27**, 129–143 (1987)

[155] Rees, A., Palmer, A.: Neuronal responses to amplitude-modulated and puretone stimuli in the guinea pig inferior colliculus, and their modification by broadband noise. J. Acoust. Soc. Am. **85**, 1978–1994 (1989)

[156] de Rivaz, P., Kingsbury, N.: Bayesian image deconvolution and denoising using complex wavelets. IEEE International Conference on Image Processing **2**, 273–276 (2001)

[157] Roelfsema, P.R.: Cortical Algorithms for Perceptual Grouping. Annu. Rev. Neurosci. **29**, 203–227 (2006)

[158] Rosenfeld, D.: New Approach to Gridding Using Regularization. Magnetic Resonance in Medicine **48**, 193–202 (2002)

[159] Rossing, T.D.: The science of sound. Addison-Wesley, Reading, Massachusetts (1982)

[160] Rouiller, E., de Ribaupierre, Y., de Ribaupierre, F.: Phase-locked responses to low frequency tones in the medial geniculate body. Hearing Res. **1**, 213–226 (1979)

[161] Russel, B.: A History of Western Philosophy. Simon and Schuster, London (1945)

[162] Sarkar, T.K., Weiner, D.D., Jain, V.K.: Some mathematical Considerations in Dealing with the Inverse Problem. IEEE Trans. Antenn. Propag. **29**, 373–379 (1981)

[163] Schreiner, C., Langner, G.: Periodicity coding in the inferior colliculus of the cat. II. Topograhical organization. J Neurophysiol **60**, 1823–1840 (1988)

[164] Schuller, G.: Natural ultrasonic echoes from wing beating insects are encoded by collicular neurons in the CF-FM bat, *Rhinolophus ferrumequinum.* J Comp Physiol A **155**, 121 (1984)

[165] Seebeck, A.: Beobachtungen über einige Bedingungen der Entstehung von Tönen. Ann Phys Chem **53**, 417–436 (1841)

[166] Shannon, R., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M.: Speech recognition with primarily temporal cues. Science **270**, 303–304 (1995)

[167] Shinn-Cunningham, B.G., Wang, D.: Influences of auditory object formation on phonemic restoration. J. Acoust. Soc. Am. **123**(1), 295–301 (2008)

[168] Shirokuma, L.t.d.: Notre Dame de Budapest pipe organ samples; Full edition for GigaStudio 2 and 3. Handbook (2004). http://www.organa.org

[169] Shumway, C.A.: Multiple electrosensory maps in the medulla of weakly electric gymnotiform fish. I. Physiological differences. J. Neurosci. **9**, 4388–4399 (1989)

[170] Sichert, A.B., Friedel, P., van Hemmen, J.L.: Snake's perspective on heat: Reconstruction of input using an imperfect detection system. Phys. Rev. Lett. **97**, 68105 (2006)

[171] Smith, D.R.R., Patterson, R.D., Turner, R., Kawahara, H., Toshio, I.: The processing and perception of size information in speech sounds. J. Acoust. Soc. Am. **117**, 305–318 (2005)

[172] Smith, D.R.R., Walters, T.C., Patterson, R.D.: Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. J. Acoust. Soc. Am. **122**, 3628–3639 (2007)

[173] Smith, J., Marsh, J., Greenberg, S., Brown, W.: Human auditory frequency-following responses to a missing fundamental. Science **201**, 639–641 (1978)

[174] Smith, R., Zwislocki, J.: Short-term adaptation and incremental responses of single auditory-nerve fibers. Biol Cybern **17**, 169–182 (1975)

[175] Smith, Z., Delgutte, B., Oxenham, A.: Chimaeric sounds reveal dichotomies in auditory perception. Nature **416**, 87–90 (2002)

[176] Speck-Hergenröder, J., Barth, F.: Tuning of vibration sensitive neurons in the central nervous system of a wandering spider, *Cupiennius salei* Keys. J Comp Physiol A **160**, 467–475 (1987)

[177] Stanford, L.R., Hartline, P.H.: Spatial sharpening by second-order trigeminal neurons in crotaline infrared system. Brain Res. **185**, 115–123 (1980)

[178] Stavenga, D.G.: Reflections on colourful ommatidia of butterfly eyes. J. Exp. Biol. **205**, 1077–1085 (2002)

[179] Stecker, G.C.: Parallel Emergence of Spatial Tuning and Echo Suppression in the Auditory Midbrain? Focus on "A Neuronal Correlate of the Precedence Effect Is Associated With Spatial Selectivity in the Barn Owl's Auditory Midbrain". J Neurophysiol **92**(4), 1965–1966 (2004)

[180] Takahashi, T., Konishi, M.: Selectivity for interaural time difference in the owl's midbrain. J. Neurosci. **6**, 3413–3422 (1986)

[181] Theunissen, F.E., Doupe, A.J.: Temporal and Spectral Sensitivity of Complex Auditory Neurons in the Nucleus HVc of Male Zebra Finches. J. Neurosci. **18**(10), 3786–3802 (1998)

[182] Tian, B., Reser, D., Durham, A., Kustov, A., Rauschecker, J.P.: Functional Specialization in Rhesus Monkey Auditory Cortex. Science **292**, 290–293 (2001)

[183] Tikhonov, A., Goncharsky, A., Stepanov, V., Yagola, A.: Numerical Methods for the Solution of Ill-Posed Problems. Kluwer Academic Publishers, Dordrecht, Netherlands (1995). English translation of the original russian text

[184] Trussell, L.: Synaptic mechanisms for coding timing in auditory neurons. Ann Rev Physiol **61**, 477–496 (1999)

[185] Unoki, M., Furukawa, M., Sakata Keigo anf Akagi, M.: An improved method based on the MTF concept for restoring the power envelope from a reverberant signal. J. Acoustical Science and Technology **25**(4), 232–242 (2004)

[186] van der Waerden, B.L.: Mathematische Statistik. Springer, Berlin (1957)

[187] Wandell, B.A.: Foundations of Vision. Sinauer Associates, Sunderland, MA (1995)

[188] Wehr, M., Zador, A.M.: Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature **426**, 442–446 (2003)

[189] Westerman, L., Smith, R.: Rapid and short-term adaptation in auditory nerve responses. Hearing Res. **15**, 249–260 (1984)

[190] Wickesberg, R.E.: Rapid inhibition in the cochlear nuclear complex of the chinchilla. J. Acoust. Soc. Am. **100**(3), 1691–1702 (1996)

[191] Wickesberg, R.E., Oertel, D.: Delayed, frequency-specific inhibition in the cochlear nuclei of mice: A mechanism for monaural echo suppression. J. Neurosci. **10**, 1762–1768 (1990)

[192] Wu, M., Wang, D.: A two-stage algorithm for one-microphone reverberant speech enhancement. IEEE Trans. Audio, Speech, Language Process. **14**, 774–784 (2006)

[193] Yegnanarayana, B., Avendano, C., Hermansky, H., Satyanarayana Murthy, P.: Speech enhancement using linear prediction residual. Speech Communication **28**, 25–42 (1999)

[194] Yegnanarayana, B., Satyanarayana Murthy, P.: Enhancement of Reverberant Speech Using LP Residual. IEEE Trans. Speech Audio Process. **8**, 267–281 (2000)

[195] Yost, W.A.: Auditory image perception and analysis: The basis for hearing. Hear. Res. **56**(7), 8–18 (1991)

[196] Yost, W.A.: Fundamentals of Hearing: An Introduction, 3rd edn. Academic Press, San Diego, CA (1994)

[197] Zahn, T.P.: Neural achitecture for echo suppression during sound source localization based on spiking neural cell models. Ph.D. thesis, Technische Universität Ilmenau (2003)

[198] Zeil, J., Hemmi, J.M.: The visual ecology of fiddler crabs. J Comp Physiol A **192**, 1–25 (2006)

[199] Zhang, L.I., Tan, A.Y.Y., Schreiner, C.E., Merzenich, M.M.: Topography and synaptic shaping of direction selectivity in primary auditory cortex. Nature **424**(10), 201–205 (2003)

[200] Zhang, S., Trussell, L.: A characterization of excitatory postsynaptic potentials in the avian nucleus magnocellularis. J Neurophysiol **72**, 705–718 (1994)