How Nonlinear Dynamics in the Cochlea Shapes Human Sound Perception

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences (Dr. sc. ETH Zürich)

presented by

Florian Gomez
Master of Science in Physics, ETH Zurich
born 10. 11. 1987
citizen of Langrickenbach, TG

accepted on the recommendation of

Prof. Dr. Ruedi Stoop, Examiner, Supervisor
Prof. Dr. Hans Jürgen Herrmann, Co-Examiner
Dr. Tobias Reichenbach, Co-Examiner

2015
Acknowledgements

I am deeply indebted to my mentor and adviser Ruedi Stoop, who has guided me during this thesis in the best possible way. Over the course of the past four years, he has been not only a great teacher and a constant source of inspiration and good advice, but also a dear friend. I would like to express my sincere thanks to my co-examiners Hans Herrmann and Tobias Reichenbach for generously providing their time and support. Special thanks go to my predecessors and collaborators Albert Kern, Jan-Jan van der Vyver, Stefan Martignoli, Niko Buchheim and Victor Saase, whose works helped to lay the basis for the ongoing success of the Hopf cochlea project. I would like to thank the rest of the current and former Stoop group members, especially Tom Lorimer, for the many enjoyable discussions and the great time I have had with them. Last but not least, I wish to express my sincere gratitude to Mei Ying and my parents, who have always given me the support I needed.
Summary

In the past fifteen years, a detailed conceptual understanding of the mechanics of the mammalian cochlea has been achieved in terms of a model of forward-coupled subcritical Hopf oscillators. This ‘Hopf cochlea’ is based on the principle of small-signal amplification, exploiting the fact that systems close to a Hopf bifurcation show exactly the amplification characteristics observed in the mammalian ear. Even though the correspondence of model and experiment has been widely acknowledged, it has only been discussed to a limited extent how exactly the nonlinear characteristics of the cochlea shape the signal and what this implies for the perception of sound as a whole.

In this thesis, a thorough characterization and discussion of the nonlinear aspects of cochlear sound processing is presented. In addition to a detailed analysis of the key processes at work, there is a strong focus on the actual implications for perceptual hearing characteristics such as pitch. The work is formulated in the language of physics, but provides comparisons to biological concepts and psychoacoustical notions whenever needed and appropriate. Furthermore, most of the analysis is performed on the mesoscopic level, where one does not explicitly account for all the microscopic biochemical and biophysical details, but rather where one tries to build a model on the basis of robust generic principles, that simultaneously describe the essential aspects of the structure and dynamics of the system under investigation.

As a first step, the response of the Hopf cochlea is exhibited for different classes of signal input. Due to the lack of a superposition principle in nonlinear systems, already quite simple inputs can lead to complex excitation patterns. This is mostly a consequence of the emergence of combination tones, which are tones that are artificially generated within the cochlea. In a second step, these complex excitation patterns are studied by a characterization of the interactions of the signal components within the cochlea using a framework based on complex network theory. This not only allows for a detailed understanding of the nonlinear sound processing, but facilitates the description of the elicited complexity in the cochlea using paradigms from the study of biological neuronal networks. Specifically, the elicited responses in the cochlea are compared to avalanches in neuronal cell cultures. Based on a large number of inputs, the distribution of the size of the avalanches in the cochlea is found to follow a power-law. It is shown that this only holds when the cochlea is not biased towards certain inputs: if some of the Hopf oscillators are tuned away from bifurcation, mimicking a process of learning, the power-laws are lost. In correspondence with results from the thermodynamic formalism of dynamical systems, this suggests that power-laws characterize a ground state of general information processing systems.

In a third step, the consequences of the nonlinear interactions within the cochlea for the characterization of sounds by pitch is investigated. Through the accurate reproduction
of the well-known pitch-shift effects from psychoacoustics, it is assessed in what way the specific construction of the cochlea influences the perception of pitch. As the main result, it follows that not only the nonlinear amplification characteristics, but in a similar way also the dissipation processes in the cochlea affect pitch. Because the main contribution to dissipation is from viscous damping, the perceived pitch is found to be essentially shaped by the cochlear fluid. This highlights the importance of our Hopf-cochlea among the mesoscopic cochlea models. Together with the demonstration of the faithful transmission of the sound from the cochlea to the cortex via the auditory nerve, the results underline the fact that pitch is already fully present at the level of the cochlea and not, as is often assumed, generated higher up in the neural system.

Taking pitch as the guiding feature, it is then finally investigated to which extent the cochlea can be tuned towards signals of interest. Using a biologically motivated protocol to modify the bifurcation parameters of the Hopf cochlea, it can be seen that despite the complex nature of the cochlea, a simple tuning is possible and feasible. To a significant degree, this result is based on the fact that tuning not only allows the suppression of unwanted frequency components, but at the same time removes associated cross-source combination tones. This strongly suggests that cochlear tuning acts as a first step of active listening. Taking this feature into account is thus of utmost practical importance for the restoration of hearing in humans.
Zusammenfassung


mit Resultaten aus dem thermodynamischen Formalismus der dynamischen Systeme lässt sich daraus schliessen, dass Potenzgesetze einen Grundzustand der Informationsverarbeitung charakterisieren.


# Contents

Acknowledgements ........................................... ii
Summary ............................................................ iii
Zusammenfassung .................................................... vi

## Introduction
- Organization of this Thesis .................................................. 1
- Publication Statement ....................................................... 1

## 1 Key concepts and models
1.1 The principle of small-signal amplification ........................................ 5
  - Dynamical systems ........................................................... 5
  - Bifurcations ............................................................... 5
  - Small-signal amplification ............................................... 6
  - The Hopf amplifier ...................................................... 6
1.2 Small-signal amplification in vertebrates and insects ............................... 8
  - Vertebrates ............................................................... 9
  - Insects ................................................................. 9
1.3 The mammalian cochlea ................................................... 11
  - Physiology of the mammalian cochlea ................................... 11
  - Basilar membrane motion .............................................. 13
1.4 The Hopf cochlea model .................................................. 15
  - The model by Kern and Stoop ......................................... 15
  - Simplification and analog electronic implementation .............. 18
1.5 A full model of the peripheral auditory system ...................................... 19
  - Inner hair cells .......................................................... 20
  - Auditory nerve neurons .................................................. 20
  - Compound model ........................................................ 23

## 2 Cochlear complexity
2.1 Response to single tones .................................................... 27
  - Activation profiles ..................................................... 29
2.2 Response to two tones: combination tones ....................................... 29
  - Response of a single Hopf oscillator ................................ 30
  - Response of the full Hopf cochlea ................................... 33
2.2.3 Activation profiles ........................................... 36
2.3 Response to mixtures of complex tones ........................................... 37
  2.3.1 Cochlear interaction networks ........................................... 38
2.4 Statistical activation properties ........................................... 41
  2.4.1 Activation statistics ........................................... 41
  2.4.2 Network statistics ........................................... 42
  2.4.3 The effect of learning ........................................... 43
  2.4.4 Interpretation ........................................... 45

3 Pitch and the tuning of the cochlea ........................................... 49
  3.1 Pitch ........................................... 49
    3.1.1 Basic notions ........................................... 49
    3.1.2 The missing fundamental paradigm ........................................... 50
    3.1.3 Pitch-shift experiments ........................................... 50
    3.1.4 Discussion ........................................... 59
  3.2 Cochlear tuning ........................................... 60
    3.2.1 The medial olivocochlear (MOC) system ........................................... 60
    3.2.2 Simple MOC effects ........................................... 61
    3.2.3 A pitch-based measure for focusing on sounds ........................................... 62
    3.2.4 A full feedback system ........................................... 63
    3.2.5 Discussion ........................................... 64
  3.3 Conclusions ........................................... 67

Bibliography ........................................... 69
Curriculum Vitae ........................................... 78
List of Publications ........................................... 79
Introduction

The capability to hear the rustling of an insect as well as the rumble of a thunderstorm necessitates the ability to cope with sound intensities that can vary by more than 10 orders of magnitude. Most impressively, mammals are able perceive signals that have sound pressure levels just on the order of thermal noise. A major task of a good hearing organ or hearing device is therefore to amplify weak signals, while, at the same time, amplifying stronger signals to a lesser extent. This task is only feasible in nonlinear systems, and is exhibited in a generic and natural way by nonlinear systems close to a dynamical instability (bifurcation). Of special interest for the case of hearing, as well for many problems in neuroscience, is the occurrence of an Andronov-Hopf bifurcation, or simply, Hopf bifurcation. Here, two eigenvalues of the linear approximation at a fixed point cross the imaginary axis and lead the system from quiescence into the regime of stable self-sustained oscillatory behavior. Just below bifurcation, the system is still quiescent, but very sensitive to external perturbations. Past work has shown that this regime (the vicinity of a Hopf bifurcation) is very well-suited to describe not only the amplification characteristics of the mammalian cochlea, but also the behavior of other hearing-related organs such as insect antennae or hair cell bundles in amphibians.

Nonlinearity, however, comes at a price: it removes the superposition principle, with the consequence that the response of the system to superpositions of input may be hard to predict. Yet such superpositions of signals are the generic situation in the auditory world: we rarely hear sound from a single source, but typically from various (possibly very different) sources and their reverberations at the same time. This complex mixture of sounds enters the ear and leads to an oscillation of the eardrum, which is the basic information the inner ear and the corresponding neural regions receive from the outside. Different signals enter thus not in a spatially ordered way as is the case for the retina and the visual system, but as a single superposition. This leads to essential difficulties further on. Along its way through the cochlear duct, the signal is significantly changed by the nonlinear nature of the cochlea. New frequencies are generated (combination tones), and other frequencies are partially suppressed (two-tone suppression). In particular, already the presence of two or more similarly strong frequency components in the input signal can lead to whole cascades of artificially generated signal components. While one could dismiss these cascades as by-products of cochlear amplification, a careful study of them can reveal essential information about the system itself. Looking at the cochlea from a statistical perspective, we can compare the interactions in the cochlea to neuronal avalanches, the characteristics of which are often studied to infer the underlying state of a neural system. Screening a whole space of different inputs, the statistical distribution of the cochlear avalanche size should yield an overall measure of how the cochlea reacts to generic input. This not only allows for a comparison to other biological information-
processing systems, but additionally permits a description from a statistical mechanics point of view.

Seen from an application perspective, it is currently a major challenge to design hearing aids and cochlear implants which in terms of performance come close to normal human hearing. One main problem is that these devices still do not fully exhibit the complex nonlinear processes present in a normally functioning cochlea. In order to improve hearing aids and cochlear implants and to bring back hearing to people suffering from severe hearing loss, it is of key importance to understand the sound processing in the cochlea both on a conceptual, qualitative level, as well as in a detailed quantitative manner. Only in this way can one address the different aspects of hearing in the depth required for a significant technological step forward. Such an objective poses major challenges, some of which are simple to address, while others are more involved. Whereas e.g. the perception of speech can to a large extent be restored in cochlear implant users, a major shortcoming of current technology is that most users still have serious problems when listening to music. A natural starting point to tackle this problem is to try to understand the key components of music such as pitch, which is the perceptual attribute that enables the perception of melodies. Pitch is not a simple concept as it may seem at first, and is essentially shaped by the way the cochlea is built and how it works. In this sense, the detailed understanding of pitch may also pave the way for a further understanding of other perceptual characteristics in terms of nonlinear cochlear sound processing.

Finally, a decent artificial substitute for the functioning mammalian cochlea would not only relay signals to the auditory nerve, but would at the same time monitor the activity coming down from the brain. Such feedback is enabled through the medial olivocochlear fibers, which are neurons that target the outer hair cells in the cochlea. By means of controlling the gain in specific regions of the cochlea, this feedback can allow the suppression of unwanted signal components. Future cochlear implants would thus not only be able to mimic the processing of the signal in the cochlea and send it further on, but would at the same time incorporate the top-down information from the brain. This eventually would then combine the ability to hear with the ability to listen.

**Organization of this thesis**

In the first chapter of this thesis, the language of dynamical systems theory is used to discuss the essential properties of small-signal amplification in the vicinity of a bifurcation. Example systems which can be described in this way from the class of insects (*Drosophila*) and the class of vertebrates (mammals) are discussed. The mammalian cochlea is then presented in more detail, and the Hopf cochlea model with its extension to an auditory nerve model is introduced.

On the basis of the Hopf cochlea, the second chapter exhibits the cochlear response to different classes of input (comparisons to biological data will be shown wherever possible). Starting from the simplest case, single frequency tones, the response to two tones and then to complex input mixtures containing a number of frequency components is shown. A novel network-based view will be introduced to understand and investigate complex cochlear interactions and the resulting cochlear activation patterns, and a statistical description of these networks will be provided. This then leads to an interpretation of the cochlea in terms of the
thermodynamic formalism of dynamical systems.

In the third chapter, pitch will serve as the guiding line to relate input sound, cochlear excitations, and perception. The focus will be on two- and three-tone signals, where the well-investigated ‘pitch-shift effects’ from the field of psychoacoustics allow deep insights into the nature of cochlear sound processing. In the second part of the chapter, a pitch-based model will be used to investigate how much we can focus on desired sounds, by using a biologically-motivated tuning of the bifurcation parameters of the cochlea.

Publication statement

Parts of chapter 1 of this thesis have been published in [95]. The parts of chapter 2 related to combination tones have been published in [55] and its supplemental materials, and a manuscript on the network-based perspective and associated statistics is currently under review [56]. The results on the pitch-shift experiments (chapter 3) are published in [106] and [55], the results on cochlear tuning in [54].
Chapter 1

Key concepts and models

1.1 The principle of small-signal amplification

1.1.1 Dynamical systems

Without being too formal (cf. [121, 163] for a comprehensive treatment), we are interested in time-continuous dynamical systems of the form

\[ \frac{dx}{dt} = V(x, \mu, t), \tag{1.1} \]

where \( x \) is a (possibly complex-valued) \( N \)-dimensional vector in phase space, \( t \) is time, \( V(x, \mu, t) \) a differentiable vector field and \( \mu \) a vector of external parameters. This set of first-order differential equations determines the evolution of an initial point \( x_0 \), defining a trajectory (or orbit) \( x(t) \) with \( x(0) = x_0 \). If the system does not explicitly depend on \( t \), the system is autonomous, and we can write \( \frac{dx}{dt} = V(x, \mu) \).

To characterize the behavior of an autonomous dynamical system, one typically examines the fixed points \( x^\star \) of the system, which are determined by \( V(x^\star, \mu) = 0 \) (they do not change under time-evolution). These fixed points may be stable or unstable: if the fixed point is stable, orbits in the vicinity of \( x^\star \) are attracted towards \( x^\star \); otherwise, the fixed point repels its neighborhood. Given a fixed point \( x^\star \), the Hartman-Grobman theorem implies that if the real parts of the eigenvalues of the linearization (the Jacobian) \( DV(x^\star) \) are non-zero, the dynamics around \( x^\star \) are determined by \( DV \). If this holds, then the fixed point is stable (unstable) if all the real parts of the eigenvalues of \( DV(x^\star) \) are negative (positive). In addition to fixed points, there may exist other fixed objects such as limit cycles, which are stable, closed orbits \( x(t + T) = x(t) \) (stable oscillation of period \( T \)).

1.1.2 Bifurcations

An important role in Eq. (1.1) is taken by the parameter-vector \( \mu \). In the following, we restrict ourselves to the case of a single real-valued parameter \( \mu \). If, upon a small change in \( \mu \), the behavior of the dynamical system changes in a qualitative way, one speaks of a (codimension-1) bifurcation (cf. [58] for a formal characterization). Typically, bifurcations occur when new fixed points or limit cycles are born (or destroyed), or when the change in \( \mu \) induces a change in the stability of the fixed point / the limit cycle. A common situation is the case where
we have a quiescent system for $\mu$ smaller than some critical $\mu_c$, and an oscillating system for $\mu > \mu_c$. In this case, for $\mu < \mu_c$, there is a stable fixed point $x^*$. Increasing $\mu$ to the critical value $\mu = \mu_c$ leads to the loss of the fixed point or the loss of its stability. Beyond $\mu_c$, the trajectory is no longer attracted towards $x = x^*$, and moves e.g. to another stable fixed point, or to a pre-existing or newly generated limit cycle. The latter case is generic for excitable systems such as neurons: changing an external parameter (e.g. a dc-input current) brings the neuron into an oscillatory (spiking) state. Different types of bifurcations are associated with different properties of the eigenvalues of $DV(x^*)$ at bifurcation. Two prominent types of bifurcation may lead from quiescence to oscillations: saddle-node bifurcations and Hopf bifurcations. In the first case, a single eigenvalue becomes zero at bifurcation (when a stable and an unstable fixed point collide), and in the second case, two complex-conjugate eigenvalues cross the imaginary axis.

1.1.3 Small-signal amplification

If a system with a stable periodic orbit or a stable fixed point approaches a bifurcation by change of an external parameter, certain precursors of the approaching bifurcation become visible. Even though the system trajectory will theoretically stay on the stable orbit / fixed point until exactly at bifurcation, the stability properties and with them the reaction of the system when subject to external perturbations gradually change. Perhaps the first discussion of these facts was given in a theoretical contribution by Heldstab, Thomas, Geisel and Radons in 1983, where they addressed the response characteristics of time-discrete dynamical systems close to period-doubling bifurcations [66]. On the experimental level, in 1985 Wiesenfeld and colleagues considered pre-bifurcation power spectra for a driven p-n junction [76], and simultaneously analyzed and generalized their observations in a second paper [174]. While these studies primarily addressed the system’s response to noise, the potential for these systems to act as small-signal amplifiers for weak periodic signals of specific frequencies was recognized soon [175, 176], and a first practical implementation using a parametrically modulated laser was presented by Derighetti et al. [33]. The main idea behind the principle is that a system approaching bifurcation becomes more and more sensitive to external signals of frequencies close to the ones that the system exhibits post-bifurcation. This leads to prototypical, Lorentzian-shaped response peaks in the pre-bifurcation power spectrum. The closer to bifurcation, the more the external signal is amplified and the sharper the resonance. As a consequence, the distance to bifurcation $\mu_c - \mu$ serves as a natural control parameter for the gain and bandwidth of the small-signal amplifier.

1.1.4 The Hopf amplifier

In the year 2000, two groups independently related the small-signal amplification properties of the Hopf bifurcation to the functioning of the mammalian cochlea [16, 37] (even though they were seemingly unaware of previous work on small-signal amplification in bifurcating systems). While other bifurcation types may provide small-signal amplification as well, the Hopf bifurcation alone exhibits exactly the characteristics observed in our hearing organs.

The simplest equation exhibiting a Hopf bifurcation is of the form

$$\frac{dz}{dt} = V(z) = z(\mu + i\omega_{eh}) - |z|^2 z,$$  \hspace{1cm} (1.2)
1.1. THE PRINCIPLE OF SMALL-SIGNAL AMPLIFICATION

where $z$ is the complex-valued variable, $\mu$ is the bifurcation parameter and $\omega_{ch}$ denotes the oscillation frequency post-bifurcation (‘characteristic frequency’). In the case of $\omega_{ch} = 1$, Eq. (1.2) corresponds to the Hopf normal form. Rewriting Eq. (1.2) as a two-dimensional system by setting $z = x + iy$ and differentiating at the fixed point $(x, y) = (0, 0)$ yields the Jacobian

$$DV|_{(0,0)} = \begin{pmatrix} \mu & -\omega_{ch} \\ \omega_{ch} & \mu \end{pmatrix},$$

with corresponding eigenvalues $\mu \pm i\omega_{ch}$. For $\mu < 0$, the eigenvalues have a negative real part; for $\mu = 0$, the eigenvalues are located on the imaginary axis, and for $\mu > 0$, the real part is positive, which confirms the occurrence of a Hopf bifurcation at $\mu = 0$. For positive $\mu$, we find a limit cycle with angular frequency of oscillation $\omega_{ch}$ and amplitude $\sqrt{\mu}$.

To investigate the small-signal amplification property, we consider the system in the sub-critical phase ($\mu < 0$), but subject to an external forcing $F(t)$

$$\frac{dz}{dt} = z(\mu + i\omega_{ch}) - |z|^2z + F(t),$$

where we assume the forcing to be of the form $F(t) = F_0 e^{i\omega t}$, i.e. a sinusoidal forcing with amplitude $F_0$. In this case, the oscillator locks to the driving and we can write $z = ae^{i\omega t}$, where the complex amplitude $a$ includes a phase $\phi$. Inserting this into Eq. (1.4) yields (for all $t$ after some transient period)

$$a(i(\omega - \omega_{ch}) - \mu) + |a|^2a = F_0.$$  

Figure 1.1: Small-signal amplification characteristics for Eq. (1.4) with $\omega_{ch} = 10$, where the output (input) in dB is defined as $10 \log_{10} |a|^2 (10 \log_{10} |F_0|^2)$. a) Output curves as a function of the input frequency $\omega$ for different values of $|F_0|$ between $10^{-4}$ and 1 and $\mu = -0.01$. b) Output at resonance ($\omega = \omega_{ch}$) as a function of the input $F_0$ for different values of $\mu$. Dotted line: off-resonance curve ($\omega = 9.97, \mu = -0.01$) for comparison.
CHAPTER 1. KEY CONCEPTS AND MODELS

If we are interested in the magnitude of the response $|a|$, we can multiply both sides of Eq. (1.5) with the respective complex conjugates to obtain a cubic equation in $|a|^2$ that can be solved analytically [37, 86]. Directly from Eq. (1.5) one sees that close to bifurcation ($\mu \approx 0$) and for $\omega \approx \omega_{ch}$, the response goes as $|a| \propto |F_0|^{1/3}$, which yields a gain $\propto |F_0|^{-2/3}$. This embodies the small-signal amplification property, where weak signals are amplified more than strong signals. Away from bifurcation $\mu < 0$ but keeping $\omega \approx \omega_{ch}$, for weak stimuli, the linear term dominates and $|a| \propto -|F_0|/\mu$. The cubic term matters as $|F_0|$ gets stronger, which again leads to a decrease in the gain (we are entering the so-called ‘compressive regime’). Fig. 1.1 shows the full response properties of Eq. (1.4) for different stimulus levels $F$, different values of $\mu$ and different driving frequencies $\omega$. Note that in the form of Eq. (1.4), the amplification characteristics are not scale-invariant with respect to a logarithmic axis for the frequency (the bandwidth in linear $\omega$-space is constant). This can be remedied by using an $\omega_{ch}$-rescaled version of Eq. (1.5) as implemented in the Hopf cochlea model discussed later [168, 108].

1.2 Small-signal amplification in vertebrates and insects

The previous section described the nonlinear amplification characteristics of systems close to a bifurcation. Such systems amplify weak signals, exhibit a compressive regime, display narrow frequency tuning and, when crossing the bifurcation, may even account for certain types of spontaneous (otoacoustic) emissions. These are exactly the characteristics of the biological ‘active processes’ found in most hearing organs [101, 70]. Since 2000, there have been various experiments on different types of animals, which fully support the conjecture that biological sound amplification is well-described by a Hopf process [109, 110, 162, 71]. This demonstrates that despite the amazing diversity of hearing organs, on a macro- or mesoscopic scale, the key ingredients for sound detection seem to be similar [95].

Figure 1.2: Vertebrate hair cell. a) Sketch after [44], b) SEM image of an outer hair cell (s: stereocilia) from a guinea pig cochlea, from [42].
1.2. SMALL-SIGNAL AMPLIFICATION IN VERTEBRATES AND INSECTS

1.2.1 Vertebrates

The class of the vertebrates contains the fish, amphibians, reptiles, birds, and mammals. From these five families, the last four (the so-called tetrapods) evolved from the lobe-finned fishes almost 400 million years ago [102]. Common among all vertebrates is an evolutionarily very old type of cell as the sound-transducing element: the sensory hair cell [38] (Fig. 1.2; cf. [35] for a comprehensive overview on vertebrate hair cells). These cells are immersed in a fluid (the surrounding water for a fish or the cochlear fluid for hair cells in the mammalian cochlea) and their hair-like stereocilia are deflected by pressure waves. The stereocilia possess interadjacent ‘tip link’ connections, and sometimes they have a common overlying structure to which some of them are attached (often called the tectorial membrane). On the other side of the hair cell, afferent (“to the brain”) and efferent (“from the brain”) nerve connections form the communication interface of the hair cell. Hair cells do not only transduce signals; they can actively amplify them. This leads to a certain degree of hair cell specialization e.g. in birds and mammals: they are either primarily amplifiers (the outer hair cells in mammals) or receptors (inner hair cells in mammals). Consequently, this is reflected in either predominantly efferent (outer hair cells) or afferent (inner hair cells) nerve connections.

The specific properties of hair cells have been under investigation for many years, with a special focus on what constitutes the properties leading to active amplification. There seem to be two different mechanisms, termed active hair bundle motility (important in amphibians and reptiles) and electromotility (dominant in mammals), the interaction between the two still being under investigation [132, 118, 71]. A key role for the second process seems to be taken by the motor protein prestin in mammals (the mammalian cochlea is described in more detail in section 1.3), or prestin-homologous proteins in nonmammalian vertebrates and insects [92, 173, 26].

On a biophysical level, considerable efforts have been dedicated to the study of the mechanical properties of hair cell stereocilia, especially in amphibians and reptiles. A direct comparison to the Hopf amplifier concept was presented in [109, 110], where the active response of a hair bundle from the bullfrog’s sacculus was investigated. In compliance with a mechanical system close to a Hopf bifurcation (cf. Fig. 1.1b), the authors reported a response that grew linearly for weak stimuli and with a power of $1/3$ for stronger stimuli.

An alternative way to infer the characteristics of an active process is to focus on the spontaneous emissions, i.e. to concentrate on the state where the system produces self-sustained oscillations (see below and [162] for a detailed exploration of this principle). To this end, two active oscillator-based models were recently created to explain the characteristic spectrum of spontaneous emissions in certain lizard species [169, 45] (it must be mentioned that such an oscillator-based approach is, especially for mammalian spontaneous emissions, not undisputed [152, 11]). Both models are related to the Hopf bifurcation; whereas the first study [169] was based on an oscillator model that could directly be transformed into a Hopf normal form, the authors of [45] made use of the well-known van der Pol oscillator, which is another paradigmatic system exhibiting a Hopf bifurcation.

1.2.2 Insects

After a long tradition of research on evolutionary linkage and on physiological and genetical correspondences of species, it was proposed that a convergent evolution may have directed
insect, as well as jointly insect and mammalian audition ([95] and references therein). Insects possess a variety of sound detecting organs that are located all around the body. Differing in the types of sound they detect, there are two major building principles: flagellar organs (which serve as near-field detectors) and tympanal/chordotonal organs (for far-field sound detection) [68, 36]. Fig. 1.3a) provides a sketch of the locations of the hearing organs for the case of Drosophila. Although seemingly very different at first view, the human cochlear hair cells follow genetically closely the building principle of the chordotonal organs [75].

To investigate the biophysical nature of the amplification process in a system, one possibility is to make the system cross the bifurcation point so that it oscillates without additional external driving. This approach was taken in [162], where the antennal hearing organ of Drosophila could be brought into a state of self-sustained oscillations by injecting biochemical dimethyl sulfoxide (DMSO). From the observed velocity time series of the antenna oscillations, an underlying generalized van der Pol system

\[ \ddot{x} + P_n(x)\dot{x} + P_m(x) = 0, \]  

with polynomials \( P_n, P_m \) \((n = 2, m = 5)\) could be identified, which operates in the close vicinity of a Hopf bifurcation. For expressing the short-scale oscillations, a term \( A_0 \cos(2\pi ft) \) can be included into the equation \((A_0 = 70 \text{ and } f = 600 \text{ Hz}; \text{ this term does not compromise the nature of the bifurcation})\) [95]. An enlightening understanding of the amplification dynamics can be provided by the behavior around zero displacement position \( x = 0 \), where the nonlinear damping term \( P_n(x) < 0 \) implies that energy is injected into the system, indicating active amplification. Around \( x = 0 \), the nonlinear restoring force \( P_m(x) \), together with its first and second derivatives, are relatively small. This implies that for small receiver displacements, virtually no restoring force is present. By means of the negative damping term, the system is thus easily driven out to large amplitudes. The comparison between data and obtained trajectories reveals the close correspondence between the data and the model (Fig. 1.3b).
1.3. THE MAMMALIAN COCHLEA

In mammals, the central hearing organ is the coiled organ called cochlea. Only significantly differing in aspects such as size and frequency range covered (it spans a length of about 35 mm or 2 and 3/4 turns in humans), the cochlea is very similar for all mammals.

1.3.1 Physiology of the mammalian cochlea

When sound pressure waves arrive at our ears, the eardrum (tympanic membrane) transforms them into mechanical displacements that are transduced by the middle ear ossicles to a membrane-covered opening of the cochlea, the oval window. The middle ear, representing a sort of hydraulic apparatus, serves the essential purpose of “impedance-matching” (bridging the different impedances of air and water). A simple calculation shows that without it, only 0.1 % of the acoustic power would be transmitted into the cochlear fluid [84]. It is worth mentioning that quite recently, a similar type of apparatus was discovered in some insect (rainforest katydid) ears, suggesting a case of convergent evolution between insect and mammalian audition [114].

The cochlea is a snail-like structure with three distinct fluid-filled chambers, called scala
vestibuli, scala media, and scala tympani (see Fig. 1.4 for a cross-section). Two membranes, the thin Reissner’s membrane and the thicker basilar membrane, separate the chambers. The upper and the lower chamber, the scala vestibuli and the scala tympani, are filled with perilymph, a fluid with a low $K^+$ and a high $Na^+$ concentration similar to that of cerebrospinal fluid. The chambers meet at the end of the cochlear spiral (the apex), where they are connected through a hole called helicotrema. By contrast, the scala media is filled with a fluid called endolymph which has an ionic concentration similar to intracellular fluid (high $K^+$ and low $Na^+$) [7]. The gradient in concentrations is generated by an active transport in the stria vascularis, creating in this way an endocochlear potential of 80 to 100 mV.

When sound enters the cochlea at the oval window, a traveling wave on the main membrane of the cochlea, the basilar membrane, is generated. The basilar membrane has properties that change in an essential way along the longitudinal axis of the cochlea. While the width of the basilar membrane increases by about a factor of 5 from base to apex, its stiffness simultaneously decreases by a factor of about 100, a result already obtained by von Békésy’s pioneering experiments [170]. These two gradients result in a change of the resonance properties of the basilar membrane, which is the reason for the occasional comparison of the basilar membrane to a piano. Even though this is a problematic comparison (as we will see in chapter 2 of this thesis), the mechanical properties of the basilar membrane form the basis of the passive frequency-analysis characteristics of the cochlea.

On top of the basilar membrane sits the organ of Corti, which is the structure that contains both the outer and the inner hair cells, see Fig. 1.5. Organized in three (or more) rows, about 15,000 to 20,000 outer hair cells actively amplify the incoming signal. This happens in a complicated, cycle-by-cycle feedback loop, the details of which are still under current investigations [15, 3, 25, 71]. In a nutshell, it seems to work the following way: An incoming wave deflects (via shear between the basilar and the tectorial membrane) the hair cells’ stereocilia, which in turn leads to the opening of ion channels. The flux of ions and the resulting change in polarization of the hair cell then lead to an alteration of the hair cell’s length. This phenomenon, called electromotility, can be associated to piezoelectric properties of the motor protein prestin. When the hair cell changes its length, the movement is coupled back to the basilar membrane’s movement, in this way closing the feedback and generating...
amplification.

In addition to the outer hair cells, about 3,500 inner hair cells, organized in a single row, pick up the signal to send it further up the auditory pathway. This is mediated by synaptic contacts with the spiral ganglion nerve fibers: each inner hair cell is connected to several nerve fibers, but each nerve fiber receives input from exactly one inner hair cell. According to their spontaneous spike rate, auditory nerve fibers are often classified into three distinct categories: high-spontaneous, medium-spontaneous, and low-spontaneous auditory nerve fibers (the last two categories are sometimes merged into a single one) [90]. Differing in their axon terminal size, the classes project to distinct sides of the inner hair cells and seem to effectively transduce different sound level ranges (cf. Fig. 1.12).

In contrast to the inner hair cells, the outer hair cells’ neuronal innervation is primarily efferent. These efferent connections are part of a larger neural feedback system, where feedback loops throughout the brain provide efferent input to more peripheral parts of the auditory system. The olivocochlear bundle, the most peripheral loop, originates in the superior olivary complex (SOC) in the brainstem. From the medial part of the SOC, thick myelinated neurons (the so-called medial olivocochlear (MOC) efferents) synapse on the cochlea’s outer hair cells. Even though the functional role of these efferent connections is still not fully understood, experiments indicate that MOC neuron stimulation provokes a rapid reduction in cochlear amplification (see e.g. [59] for an overview). From this and other experimental studies, it has been suggested that the MOC connections provide a first step of the active listening process, a claim that will be discussed in detail in chapter 3 of this thesis.

1.3.2 Basilar membrane motion

Georg von Békésy was the first to perform real measurements on basilar membrane motion [170]. However, due to the fact that the measurements were carried out on dead cadavers, they could not reveal anything about the actual active process (which had had been postulated previously by Gold [49]; an indirect confirmation of the active process followed only in 1978 by Kemp, who recorded the first otoacoustic emissions [83]). In the 1960s and 1970s, the progress in measurement techniques marked the beginning of a new era in the understanding of cochlear mechanics. In 1967, Johnston and Boyle applied for the first time the Mössbauer technique to measure in vivo basilar motion [79]. Four years later, in 1971, Rhode could finally demonstrate the existence of a frequency-specific, compressive nonlinearity (Fig. 1.6a). Confirmations and extensions of these results have followed since, with the more recent measurements mainly making use of laser interferometry techniques [141, 137] (Fig. 1.6b).

Comparing the biological measurements to Fig. 1.1b), we can see the resemblance of basilar membrane motion to the simple Hopf system introduced in section 1.1. Stimulating the system with a frequency corresponding to the resonance frequency (characteristic frequency, ‘CF’) results in a compressive behavior with an exponent close to 1/3; stimulating it with frequencies off-resonance on the other hand yields linear behavior. A more complete picture of the properties of basilar membrane motion for different sound intensities and frequencies is shown in Fig. 1.6c). Again, the compressive nonlinearity becomes evident, here as the compression of the curves around CF. Moreover, another feature shows up: the place of maximal output (the peaks of the curves) shifts to lower frequencies when increasing the input level. This asymmetry has no correspondence in the simple Hopf system, and is due to the passive properties of the cochlea. A solution can be found in terms of a full cochlea model.
that not only embodies the local amplification properties, but also the passive characteristics of the cochlea. Such a model must contain a full set of amplifiers for the different cochlear frequencies, appropriately coupled, and a realistic set of equations modeling the dissipation processes in the cochlea. This has been achieved by the Hopf cochlea model proposed by Kern and Stoop [84, 86, 161], which will be discussed in next section.

![Figure 1.6](image)

Figure 1.6: Basilar membrane response (maximal output/velocity) to single-frequency input, for different frequencies (not normalized relative to middle ear filtering). a) Rhode’s original measurement of the compressive nonlinearity in a squirrel monkey cochlea (CF= 7.4 kHz) [134], b,c) more recent measurement in chinchilla (CF= 10 kHz) [141]. Red arrows indicate the compressive nonlinearity, which is only observed for stimulation with frequencies close to the characteristic frequency of the measurement location. The isointensity curves in c) correspond to the measurements in b); strong compression around CF is observed for 30-100 dB input level (red line).
1.4 The Hopf cochlea model

After the introduction of the Hopf amplifier concept in hearing in 2000 [37, 16], three groups in 2003 independently proposed a Hopf cochlea model (all in Physical Review Letters) [98, 34, 86]. Whereas the first two models had oscillators poised exactly at bifurcation, Kern and Stoop insisted on oscillators near, but below bifurcation, which is in line with the original small-signal amplification idea. This fact, together with an energy-based analysis of the passive cochlear properties, yields a model of basilar membrane motion that fully agrees with the biological observations exhibited in the previous section (Fig. 1.6), a correspondence not achieved by the other two approaches. Moreover, the description of the passive properties of the cochlea allows an interpretation of the dissipative processes, which in turn essentially influence perceptual characteristics such as pitch [55] (cf. chapter 3 of this thesis).

1.4.1 The model by Kern and Stoop

In the model of [86], amplification is governed by the Hopf equation as given in Eq. 1.4. Different Hopf oscillators provide amplification at different frequencies; for this, a tonotopic map of characteristic frequencies is defined (a geometric sequence due to the logarithmic character of the basilar membrane frequency scale in the biological template). For the passive components, a linear theory was found appropriate due to the small displacements of the basilar membrane. From this, the membrane-fluid behavior can be described in the framework of water surface waves (cf. [131, 133] for recent results on the fluid dynamics of the cochlea). Using \( x \) to denote the distance from the stapes along the unrolled cochlea, the water-surface wave is taken to have a fluid depth \( h \), a density \( \rho \), a surface mass density \( m \) and an exponentially decreasing transversal stiffness

\[
E(x) = E_0 e^{-\alpha x}.
\]  

Using \( e \) to describe the energy density and \( v_G \) the wave group velocity of the water wave, the general energy balance equation [84, 86, 85]

\[
\frac{\partial e}{\partial t} + \frac{\partial}{\partial x} (v_G \cdot e) = 0
\]  

leads with the ansatz \( \frac{\partial e}{\partial t} = -a + d \cdot e \) to the cochlea differential equation (CDE)

\[
\frac{\partial e}{\partial x} = -\frac{1}{v_G(x, \omega)} \left[ \frac{\partial v_G(x, \omega)}{\partial x} + d(x, \omega) \right] e + \frac{a(x, e, \omega)}{v_G(x, \omega)},
\]

where the local power \( a(\cdot) \) supplied by the local active amplification works against the viscous losses \( d(x, \omega) = 4\nu k(x, \omega)^2 + \sqrt{\frac{\nu}{2\omega} \frac{\rho \omega^3}{E(x)}} \). Around resonance, the second term describing the friction from the vibrating basilar membrane has a smaller effect compared to the fluid’s internal dissipation, due to \( k \)'s divergence (see below). \( \nu = \frac{\mu}{\rho} \) is the kinematic viscosity (\( \mu \): viscosity). The connection with the biologically observed basilar membrane movement is via the wave amplitude

\[
A_{BM}(x, \omega) = \sqrt{\frac{2e(x, \omega)}{E(x)}},
\]

where the equipartition theorem was used. To render the CDE useful, \( v_G \) and \( a \) need to be specified.
According to Patuzzi’s description of a two-dimensional shallow water wave [123], \( v_G, \omega \), and the wave number \( k \) are related by the following dispersion relation

\[
v_G = \frac{\partial \omega}{\partial k} = \frac{E(x) \rho}{2 \omega} \frac{kh + \sinh(kh) \cosh(kh)}{(mk \sinh(kh) + \rho \cosh(kh))^2}.
\] (1.11)

This shows that basilar membrane locations \( x = x_c(\omega) \) respond maximally when stimulated at characteristic frequencies \( \omega_c(x) = \sqrt{E(x)/m} \), which defines a tonotopic map for the passive elements. It can be shown that \( k(x, \omega) \) diverges as \( \omega \) approaches \( \omega_c(x) \) and that, as \( x \) approaches the point of passive resonance \( x_c(\omega) \) for fixed \( \omega \), the traveling wave stalls \( (v_G = 0) \). Due to dissipative losses, the wave amplitude reaches a maximum at \( x_{ch} < x_c(\omega) \).

An overview of the dissipative processes and their associated profiles is provided in Fig. 1.7 (cf. [84] and the supplemental material of [55]). Fig. 1.7a, b) compares dissipation in the cochlea due to two different processes: internal dissipation, i.e. viscous friction between the fluid elements; and friction of the moving fluid on the vibrating basilar membrane. It becomes apparent that internal dissipation causes the response to fall off at an earlier location than dissipation caused by friction on the vibrating basilar membrane alone, making the former the dominant contribution. Fig. 1.7c,d) shows the effect of introducing a membrane surface tension (with the dimensionless parameter \( \tilde{F} \) as the quotient of longitudinal and transversal stiffness) into the model. The incorporation of \( \tilde{F} \) has some influence on the response peak (because dissipation becomes more widely distributed), the decay of the response after resonance however stays comparable, indicating that tissue properties do not significantly contribute in this respect (but rather counteract the fluid’s dissipation). The combined effect of the internal viscous losses, the losses due to the friction on the vibrating basilar membrane and the basilar membrane surface tension has been shown to be well-modeled by a tailored 6th-order Butterworth filter [108] (see next section).

The active and passive components are finally linked by noting that by Hooke’s law, the tensile forces \( F \) that are the inputs to the Hopf system are proportional to their hair bundle displacements, which in turn are proportional to the wave amplitude. This yields, with \( z_0 \) solving the Hopf equation Eq. (1.4), the required connection in the following form

\[
a(e, x, \omega) = L \cdot z_0(\sqrt{\sigma e(x, \omega)})^2.
\] (1.12)

\( \sigma \) can be seen as the transfer parameter from \( A_{BM} \) to the tensile forces \( F \), and \( L \) is a proportionality constant. The full cochlea model is then described by Equations (1.4), (1.7), and (1.9) - (1.12). Even though the model provided excellent results when compared to experimental data, the complexity of the coupled equations posed a severe problem in terms of computation time. This then led to a simplified analog electronic implementation, which got finally patented in 2012 (see next section).
Figure 1.7: Passive Hopf cochlea properties (adapted from [84]). a) Internal dissipation $d_I(x, \omega)$ and dissipation due to friction on the vibrating basilar membrane $d_S(x, \omega)$ for stimulation with a fixed frequency $f = \frac{\omega}{2\pi} = 1000$ Hz, as a function of the location on the basilar membrane $x_{BM}$. Near the location of resonance, internal dissipation is dominant. b) Response amplitudes $A(x, \omega)$ ($f = 1000$ Hz) for internal dissipation (red), dissipation due to friction on the basilar membrane (blue) and combined effect (black). c,d) Typical passive response for different stimulus frequencies with surface tension $\tilde{F} = 10^{-5}$ (c), and without ($\tilde{F} = 0$, d)).
CHAPTER 1. KEY CONCEPTS AND MODELS

1.4.2 Simplification and analog electronic implementation

To achieve a fast Hopf cochlea implementation that could also deal with transient signals, a simplified analog electronic sensor was soon developed [168, 108, 160]. Based on the results of the full model described in the previous section, it was soon recognized that the passive properties of the cochlea could essentially be captured by tailored low-pass filters. For the active part, the amplification is governed by a more suitable $\omega_{ch}$-rescaled Hopf equation [108]

$$\frac{dz}{dt} = (\mu + i)\omega_{ch}z - \omega_{ch}|z|^2z - \omega_{ch}F(t),$$

(1.13)

which has the advantage that the amplification profile is now independent of $\omega_{ch}$ and solely controlled by $\mu$.

![Figure 1.8: Hopf cochlea section. A complex signal from the preceding section $j - 1$ stimulates the Hopf amplifier of section $j$ of characteristic frequency $\omega_{ch}^{(j)}$, the responsiveness of which is characterized by the Hopf parameter $\mu^{(j)}$. The two components of the signal are then filtered by Butterworth filters and passed to the next section.](image)

The cochlea model is organized as a feed-forward chain of discrete cochlea sections, where each section contains an active part (the Hopf oscillator) and two subsequent (parallel) 6th-order Butterworth filters (one each for the real / imaginary part of the signal), see Fig. 1.8. In each section, the Hopf amplifier precedes the low-pass filters, which follows the biological example and moreover avoids uncontrollable interactions of the phases of the passive and active components [108]. Given the characteristic frequency of the $j$-th Hopf amplifier $\omega_{ch}^{(j)}$, the cut-off frequency of the Butterworth filter $\omega_s^{(j)}$ is determined by a fixed relation $\omega_s^{(j)} = \omega_{ch}^{(j)}/1.05$.

The above described Hopf cochlea model was later implemented in software [146, 54], yielding almost identical (and moreover noiseless) results at comparable speed. Running roughly in real-time, the software implementation also provides the full control over the choice of the number of sections and the frequency range to be covered. Fig. 1.9 shows an overview of the Hopf cochlea discretization settings used in this thesis. In principle, the model is invariant with respect to shifts of the frequency range. For the purposes of this thesis, covering the lower frequencies ($< 100$ Hz) was not found necessary (very little experimental data available) and a ‘discretization density’ of approximately 4 sections/octave proved sufficient and convenient.
1.5 A FULL MODEL OF THE PERIPHERAL AUDITORY SYSTEM

Chapter 2 of this thesis will provide a detailed description of the response of the model, with comparisons to experimental data.

Figure 1.9: Hopf cochlea discretizations. a) Approximate human hearing range, b) the four discretization settings used in this thesis. A Setting equivalent to electronic hardware implementation (4 sections/octave), B extended setting with exactly 20 sections (3.8 sections/octave), C setting used when higher frequencies needed to be taken into account (3.8 sections/octave), D cochlea with extended range for statistical descriptions (4 sections/octave).

1.5 A full model of the peripheral auditory system

A full model of the peripheral auditory system in principle requires three components: a middle ear, a cochlea with a basilar membrane, inner and outer hair cells and a description of the hydrodynamics, and auditory nerve neurons. The middle ear is essentially just an impedance-matching component with bandpass characteristics and does not significantly contribute to the peripheral sound processing. However, to transform the Hopf cochlea’s output (representing basilar membrane motion) into neuronal spikes, we still need a model for the inner hair cells and the auditory nerve neurons. Building upon an already established model of inner hair cells [94], a detailed nonlinear model of the auditory nerve was developed in [105], which we extended and applied in [106]. The model exhibited what was already known
CHAPTER 1. KEY CONCEPTS AND MODELS

from physiological experiments: the main share of peripheral auditory processing is achieved by the cochlea and its nonlinear characteristics. Even though the interface of the inner hair cells to the auditory nerve has proved to be nontrivial, the auditory nerve essentially just transduces the signals to stations higher up the pathway.

1.5.1 Inner hair cells

The dynamical role of IHC is to half-wave rectify and slightly compress the cochlear signal: on top of a frequency-dependent DC-component, the output has a slightly low-pass filtered AC-component [94]. To obtain a detailed model of the inner hair cells, the task is essentially to define a set of differential equations that models the behavior of the relevant ion channels and their dependence on stereocilia deflection. The output variable of the model is a voltage variable \( V \), which is then taken as the input for a model of the auditory nerve neurons. Different models of inner hair cells exist (see [116] for a review); in the following, we briefly discuss a rather recent approach by Lopez-Poveda and Eustaquio-Martín, which we used in our compound system [106]. Building on the framework of an electrical-circuit equivalent, the model contains variables for a mechanically driven \( K^+ \) conductance (fast \( g_{mf} \) and slow \( g_{ms} \)). The mechanically driven \( g_m \) is described as a three-state Boltzmann function

\[
g_m(u(t)) = \frac{G_M}{1 + \exp((u_0 - u(t))/s_0)(1 + \exp((u_1 - u(t))/s_1))},
\]

where \( u(t) \) denotes the stereocilia deflection. For the membrane potential \( V_{IC} \), the following equation with the kinetics for the fast and slow conductances was proposed:

\[
\frac{dV_{IC}}{dt} = \frac{1}{C_A + C_B}[(E_l - V_{IC})(g_m(u(t)) + g_l) + (E_{k_f}' - V_{IC})g_{k_f}(V_{IC}) + (E_{k_s}' - V_{IC})g_{k_s}(V_{IC})],
\]

\[
\frac{d^2g_{k_f}}{dt^2} = -\left(\frac{1}{\tau_{1f}} + \frac{1}{\tau_{2f}}\right)\frac{dg_{k_f}}{dt} - g_{k_f} + g_{\infty,f}(V_{IC}),
\]

\[
\frac{d^2g_{k_s}}{dt^2} = -\left(\frac{1}{\tau_{1s}} + \frac{1}{\tau_{2s}}\right)\frac{dg_{k_s}}{dt} - g_{k_s} + g_{\infty,s}(V_{IC}),
\]

where \( g_{\infty,f,s}(V_{IC}) = G_m((1 + \exp((V_{1,f,s} - V_{IC})/S_{1,f,s})(1 + \exp((V_{2,f,s} - V_{IC})/S_{2,f,s})))^{-1} \) and where \( g_l \) is the constant leak conductance. The model was fitted to in vitro data, which yielded a full set of parameter values then used to run in vivo simulations. Fig. 1.10 shows a comparison of the model to experimental data from guinea pig [122], exhibiting the membrane potential’s locking to the mechanical driving. When the driving frequency is increased, the oscillation amplitude (AC-component) progressively decreases. In contrast, a DC-offset becomes evident, which for high frequencies gets ever more dominant.

1.5.2 Auditory nerve neurons

Modeling neurons has been a topic of research for decades. From the simplest, spin-based models to the most sophisticated multi-compartment models, there exists every desired level of abstraction. In the present framework, a suitable model is a model that is as simple as
possible, but still possesses the nonlinearities required to exhibit the main neuronal firing characteristics. While other approaches (e.g. Izhikevich’s approach [74]) may be suitable as well, the model of choice here is a version of the Rulkov neuron model. The general approach of Rulkov to modeling neurons is based on two-dimensional iterated maps, instead of differential equations. This has both the advantage of a reduced dimensionality (the “uninteresting” time-dimension is singled out), and increased computational speed (no Runge-Kutta integration or similar required). Different versions of Rulkov neurons exist; the first one was presented in 2001 [142] in the context of synchronization of neuronal bursts. Replacing the differentiable first component of the map with a piecewise differentiable function, the map was then generalized in order to obtain a model that could reproduce the main types of typical neuronal firing behavior [143, 154]: quiescence, periodic spiking, irregular spiking, periodic bursting, or irregular bursting. This model has the following form

\[
x_{n+1} = f(x_n, y_n + \beta_n),
\]

\[
y_{n+1} = y_n - \mu(x_n + 1) + \mu\sigma_n,
\]

where the index \( n \) indicates the iteration step, and where the function \( f \) is given by

\[
f(x_n, y) = \begin{cases} 
\frac{\alpha}{1 - x_n} + y, & x_n \leq 0 \\
\alpha + y, & 0 < x_n < \alpha + y \land x_{n-1} \leq 0 \\
-1, & x_n \geq \alpha + y \lor x_{n-1} > 0.
\end{cases}
\]

In this model, \( x_n \) represents the membrane voltage at discrete time steps \( n = 0, 1, 2, ... \) and \( y_n \) is the slow recovery or adaptation variable. The parameter \( \mu \), typically set to \( \mu = 0.001 \), establishes the fast-slow correspondence of the two variables (\( x \) and \( y \) respectively). The basic shape of \( f \) and the resulting dynamics of the map are depicted in Fig. 1.11. Map \( f \) has three components: an ascending branch, a top branch (the spike), and a reset branch. Depending on the second variable \( y \) and the parameters of the map, the ascending branch may lie above the diagonal (solid line), or cross it (dashed line). The first case, for fixed \( y \), results in a superstable periodic solution (see blue orbit), with the \( x \)-variable tracing out a spike-shaped

Figure 1.10: IHC characteristics. Comparison of a) experimental results [122] to b) model simulations [94] for increasing driving frequency.
trajectory (see inset). The second case however, exhibits a stable fixed point (circled) and thus yields a quiescent neuron. The switching between these two basic states is then established by the dynamics of the \(y\)-variable, which in this way enables the neuron to burst. Setting the parameters \(\alpha\) and \(\sigma\) in an appropriate way (the latter being something like an external DC-current), the map can be brought into any desired working state. However, due to the map’s intrinsic nonlinearity, periodic and irregular behavior are found very close together (see [57] for details and a discussion of potential implications for further data processing).

Figure 1.11: The Rulkov neuron model. a) Map as defined in Eqs. (1.18)-(1.20) for \(\alpha = 5\) and \(\sigma = 0.8\). Solid curve: \(f(x)\) for fixed \(y = -3.37\), with corresponding superstable periodic orbit (blue). Dashed curve: \(f(x)\) for a hypothetically decreased \(y = -3.6\), which yields a stable fixed point (circled). Inset: \(x\)- and \(y\)-component of the map as a function of time; the \(y\)-variable only oscillates mildly around \(y = -3.37\) (vertically stretched in figure). b) Critical avalanches in a network of Rulkov neurons (from [81]). The distribution of the avalanche size \(S\) follows a power law of exponent \(\alpha \approx 2.4\). Inset: Raster plot of spiking activity.

Modifications and extensions of the map to bring it closer to biology, have been presented by Rulkov and co-workers in [144]. Based on a modification that includes the dynamics of the synapses, Rulkov neurons have recently been used to study neuronal avalanches, where an avalanche size distribution following a power-law can be observed [81] (Fig. 1.11b); for a discussion of avalanches, cf. chapter 2). Another proposition in [144] was to include spike-afterhyperpolarization, which can be achieved by replacing the second variable \(y\) by a phenomenological hyperpolarizing current \(y^{hp}\) so that

\[
y^{hp}_{n+1} = \gamma^{hp} y^{hp}_n - \begin{cases} g^{hp}, & \text{if the n-th iteration spiked,} \\ 0, & \text{otherwise,} \end{cases}
\]

where the parameters \(\gamma^{hp}\) and \(g^{hp}\) control the duration and amplitude of the hyperpolarizing...
1.5. A FULL MODEL OF THE PERIPHERAL AUDITORY SYSTEM

Using $\nu := y^{rs} + \beta^{hp} y^{hp} + \beta^{e} I_{n}$, the map for the first variable then reads

$$x_{n+1} = \begin{cases} \frac{a}{1-x_n} + \nu, & x_n \leq 0, \\ \alpha + \nu, & 0 < x_n < \alpha + \nu, \quad x_{n-1} \leq 0, \\ -1, & x_n \geq \alpha + \nu, \text{ or } x_{n-1} > 0. \end{cases}$$

(1.22)

In our model of the peripheral auditory system (chapter 3 and [106]), Eqs. (1.21) and (1.22) together with $\alpha = 3.8$, $y^{rs} = -2.9$, $b^{hp} = 0.5$, $g^{hp} = 0.1$ and $b^{e} = 0.1$ [144] were used. Timescale correspondence was established by setting 1 iteration = 0.05 ms, which yielded a natural sampling rate of 20 kHz.

1.5.3 Compound model

To assemble the different parts of the compound model, it is assumed that the inner hair cell stereocilia deflection is proportional to the output of the Hopf cochlea (i.e. basilar membrane motion). Using $V_{Co}$ to denote the real part of the Hopf cochlea output $V_{Co} = \text{Re}(z)$, the chosen correspondence reads $u(t) = 20 \cdot 10^{-9} \cdot V_{Co}(t)$. For the synaptic transmission from the inner hair cells to the auditory nerve neurons, it was found necessary to take the omnipresent synaptic noise into account (especially for pitch perception; cf. chapter 3). This was achieved by relating the IHC-voltage to the neuron’s input as

$$I_{n} = I(t) = A + B \cdot 20 \cdot (V_{IC}(t) - V_{IC,rest}) + \sigma \cdot \xi(t),$$

(1.23)

where constant $A$ has the effect of a firing threshold and where constant $B$ scales the IHC voltage to the evoked nerve current. $\xi$ is exponentially correlated noise of intensity $\sigma$ (from the algorithm of [41]), independent for each transmission channel. The correlation time was determined by comparing the model to Kiang’s experimental data on spontaneous firing [87], which yielded an optimal correlation time of $\tau = 3$ ms. To finally match the model to the three distinct classes of auditory nerve neurons [90], we adjusted the parameters $A$, $B$, and Rulkov neuron’s “speed constant” $\gamma^{hp}$ appropriately (see Table 1.1 and Fig. 1.12). Variations of the parameters (see colored lines in Fig. 1.12) underline the fact that these parameters are sufficient to reproduce any type of input-output-behavior observed in auditory nerve neurons.

<table>
<thead>
<tr>
<th>Neuron type</th>
<th>A</th>
<th>B</th>
<th>$\gamma^{hp}$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hi-spontaneous</td>
<td>0 (-0.02)</td>
<td>1 (0.8)</td>
<td>0.97</td>
<td>0.1 (0.2)</td>
</tr>
<tr>
<td>Me-spontaneous</td>
<td>-0.2 (-0.25)</td>
<td>1.25 (1.15)</td>
<td>0.5</td>
<td>0.06 (0.04)</td>
</tr>
<tr>
<td>Lo-spontaneous</td>
<td>-0.2 (-0.25)</td>
<td>1.05 (1.15)</td>
<td>0.5</td>
<td>0.04 (0.06)</td>
</tr>
</tbody>
</table>

Table 1.1: Parameter values of the three auditory nerve neuron classes. Values in brackets correspond to variations of the parameters, see Fig. 1.12.

The full pathway for an auditory signal (exemplified by an amplitude-modulated sine-wave) is exhibited in Fig. 1.13. A real-valued input signal is first Hilbert-transformed to make it complex-valued, and then enters the Hopf cochlea, whose output (here at section 9) serves as input for the inner hair cell model. After the addition of noise as described in Eq. (1.23), the IHC’s output signal is transduced to the auditory nerve neuron, the spiking of which preserves the essential periodicities of the signal (cf. chapter 3).
24

CHAPTER 1. KEY CONCEPTS AND MODELS

Figure 1.12: Classes of auditory nerve neurons (SPR: spontaneous rate). a) Experimental data from [117], b-d) compound model simulations with parameters from Table 1.1 (black; frequency 1.76 kHz and Hopf cochlea setting A from Fig. 1.9, output at section 5). Colored lines are from the bracketed values of parameter \( A \) (red), \( B \) (green), \( \sigma \) (purple) and \( \tau \) (orange). Spontaneous rates are given by \( B = 0 \) (blue).
1.5. A FULL MODEL OF THE PERIPHERAL AUDITORY SYSTEM

Figure 1.13: Signal pathway in the compound model of the peripheral auditory system. Amplitude-modulated sound with carrier frequency $f_{\text{car}} = 850$ Hz and modulation frequency $f_{\text{mod}} = 200$ Hz. Responses evoked at a place corresponding to $f_{\text{ch}} = 880$ Hz. Top row: physical signal, bottom row: Fourier transform (histograms of inverse interspike intervals, respectively). The auditory nerve firing (AN) represents the high-spontaneous (top), and the medium-spontaneous case (bottom; several neurons pooled together for the histogram).
Chapter 2

Cochlear complexity

2.1 Response to single tones

In a first step, we describe the response of the Hopf cochlea model from section 1.4.2 to stationary stimulation with single frequency tones (pure tones); earlier results are described in [108] and the supplemental material of [107]. To permit an optimal comparison to biological measurements which were mostly performed on the apical part of the cochlea (CF of several kHz), we monitor the Hopf cochlea’s signal at a section with a CF roughly one octave below the model’s highest CF. For the cochlea discretization setting \( C \) (cf. Fig. 1.9; frequency range 14.08 - 0.44 kHz covered by 20 sections), Fig. 2.1a) shows the output of the fifth section in terms of isointensity curves (various fixed input levels). The bifurcation parameters of the first five sections were set to \( \mu = -0.1 \); afterwards, they decrease with -0.025/section (last section: \( \mu_{20} = -0.475 \)). This choice, which will always be used together with discretization \( C \), leads to amplification and tuning curves that closely match those measured experimentally, preventing at the more apical sections excess amplification because of the amplification cascade.

In contrast to a single Hopf oscillator (Fig. 1.1), the response of the compound cochlea is no longer symmetrical and shows the desired left-shift of the peak output value as a function of increasing input level. This observation is fully consistent with biological data (Fig. 1.6) and originates in the passive behavior of the cochlea and the subsequent coupling of the cochlea sections (cf. [86] and [108] for the corresponding observations in the original Kern-Stoop-model and the analog electronic implementation, respectively). Fig. 2.1b) replots the data in terms of the gain (output divided by input); comparing the peak values for two curves 70 dB apart yields a value that is consistent with observations in animals [141]. Fig. 2.1c) shows so-called tuning curves: here, the input amplitude required to generate a response of a predefined output level is plotted for different frequencies. The sharpness of these curves is predominantly determined by the \( \mu \)-values of the oscillators (going closer to bifurcation leads to sharper tuning curves). In Fig. 2.1d), the phase of the response of section 3 for different input levels is plotted relative to the phase generated by an input of -30 dB. Here, an essential characteristic consistent with experimental data [141] is observed: low-level inputs show phase leads for frequencies smaller than CF, but phase lags above CF. High-level inputs, on the contrary, show phase lags below CF and phase leads above CF. This phase-flipping can essentially be traced back to the response of a single Hopf oscillator, where the curves cross zero exactly at CF. Coupling the oscillators and adding the passive filters enhances the
Figure 2.1: Hopf cochlea response to pure tones. Cochlea discretization setting C with 20 sections covering 14.08 - 0.44 kHz; output at section 5 (CF = 6.79 kHz) for a)-c) and section 3 (CF = 9.78 kHz) for d). Small numbers denote input levels in dB. a) Response in dB, b) gain in dB; a difference of 33 dB in peak gain for two input levels differing by 70 dB corresponds to observations in chinchilla (32.5 dB, or slightly higher, between 20 and 90 dB SPL curves [141]). c) Tuning curves for fixed output levels, d) Phase for different input levels relative to -30 dB.

effect, but smears out the sharp flip so that the zero-crossings are not exactly at CF anymore.
2.2. RESPONSE TO TWO TONES: COMBINATION TONES

Figure 2.2: Cochlear activation from single frequency stimulation. Hopf cochlea setting C, input frequencies are 8, 3, 1 kHz in a), b) and c), respectively. Upper panel: Approximate extrapolated continuous excitation along the cochlea, lower panel: activation patterns on the discrete level of the cochlea sections.

2.1.1 Activation profiles

So far, the behavior of the model was exhibited from a ‘local’ point of view, expressing how a single cochlea section (or correspondingly a single location on the basilar membrane) sees and processes an incoming signal. It is now interesting to see how this translates into an excitation profile (or ‘activation profile’) of the whole cochlea, as shown in Fig. 2.2.

The resulting activation profiles are simple and comparable for all input frequencies. A tone of frequency $f$ enters the cochlea, gets more and more amplified towards the location where $f = \omega_{ch}/2\pi$, and shortly after gets dissipated by internal friction (implemented by the low-pass filters). Apart from high input frequencies which activate only the first few sections of the cochlea, also the ‘width’ of the cochlear activation profile roughly becomes a constant: if we define a section to be activated when its output exceeds $-50$ dB (a value that is motivated by comparisons to psychoacoustic experiments, see chapter 3), then a single-frequency tone activates about 9-10 sections (cf. Fig. 2.2b,c).

2.2 Response to two tones: combination tones

For nonlinear systems, the superposition principle does not hold. This implies that in general, the response to the sum of two inputs is not the sum of the individual responses to each input. In the cochlea, this manifests in a variety of effects, the most prominent one being the generation of additional frequencies, so-called combination tones. These combination tones (CT) were already known to 18th century musicians Georg Sorge and Giuseppe Tartini, and can easily be heard also by non-trained listeners. CT are thus not by-products that are filtered
out on the way to the brain. Indeed, they travel on and can be measured at all stages of the auditory pathway (e.g. in the inner hair cell response [119], in the auditory nerve [52, 140], or in the inferior colliculus [1]). Many of the CT generated in the cochlea are not dissipated immediately, but propagate along the cochlear duct and get amplified, and interact with other frequencies to create additional CT. From this, the cochlear activation profiles often get highly nontrivial. First, we will present the basic principles of combination tone generation using the example of a single Hopf oscillator. A detailed analysis however shows then that only the full Hopf cochlea model, can exhibit combination tone responses comparable to the ones observed in biological and psychoacoustic experiments.

2.2.1 Response of a single Hopf oscillator

The following analysis exhibits in detail how a single Hopf oscillator generates CT. These CT are a product of the cubic nonlinearity acting in the Hopf equation, with the result that from the stimulation with two frequencies, a well-defined set of CT (the ‘cubic CT’) is generated. The cubic CT show, as a function of their order, exponentially decaying amplitudes, a fact that has been mentioned in the context of early psychoacoustic studies as well as more recent investigations of cochlear mechanics [138, 50] (see also [80, 161] for comparable but less detailed discussions related to CT in the Hopf cochlea approach).

We start with the $\omega_{ch}$-rescaled Hopf equation from section 1.4.2

$$\frac{dz}{dt} = (\mu + i)\omega_{ch}z - \omega_{ch}|z|^2z - \omega_{ch}F(t),$$

(2.1)

where $\omega_{ch}$ is the characteristic frequency of the oscillator, $F(t)$ is the (complex) external forcing and $\mu$ is the bifurcation parameter. Given a harmonic two-tone forcing

$$F(t) = F_1 e^{i\omega_1 t} + F_2 e^{i\omega_2 t},$$

(2.2)

with $\omega_1 = k\omega_0$ and $\omega_2 = (k+1)\omega_0$ s.t. all CT are multiples of $\omega_0$, we can expand the response in a Fourier series

$$z(t) = \sum_j a_j e^{i\omega_0 t},$$

(2.3)

where the $a_j$ are complex (i.e. include a phase). For a given frequency $\omega_l = l\omega_0$ we then obtain by inserting Eq. (2.3) into Eq. (2.1)

$$i(\omega_l - \omega_{ch})a_l - \mu\omega_{ch}a_l + c.i.t. = -\omega_{ch}F_l,$$

(2.4)

where c.i.t. denote terms from cubic interaction ($\propto \omega_{ch}a_{k'}a_{k''}a_{k'''}$ so that $k' + k'' - k''' = l$). More explicitly, we may calculate the responses at $\omega_1$, $\omega_2$ and all combination tones. We now assume low to moderate sound levels and an $\omega_{ch}$ close to the forcing frequencies. Inserting the Fourier ansatz into the Hopf equation, one obtains to lowest order for $\omega_1$

$$i\omega_1 a_1 = (\mu + i)\omega_{ch}a_1 - \omega_{ch}|a_1|^2a_1 - 2\omega_{ch}|a_2|^2a_1 - \omega_{ch}F_1,$$

(2.5)

Similarly, we get at $\omega_2$

$$i\omega_2 a_2 = (\mu + i)\omega_{ch}a_2 - \omega_{ch}|a_2|^2a_2 - 2\omega_{ch}|a_1|^2a_2 - \omega_{ch}F_2.$$

(2.6)
2.2. RESPONSE TO TWO TONES: COMBINATION TONES

Solving Eqs. (2.5) and (2.6) simultaneously yields complex-valued expressions for $a_1$ and $a_2$. For the first combination tone CT1 at $\omega_1 - \omega_0$, we obtain to lowest order (appropriate for low sound levels)

$$i(\omega_1 - \omega_0)a_{CT1} = (\mu + i)\omega_{ch}a_{CT1} - \omega_{ch}a_1a_2^*.$$  

(2.7)

For higher sound levels, one would have to add to the r.h.s. the cubic interaction terms

$$-2\omega_{ch}a_{CT1}|a_1|^2 - 2\omega_{ch}a_{CT1}|a_2|^2.$$  

(2.8)

This, however, keeps the equation linear in $a_{CT1}$, since the $|a_{CT1}|^2a_{CT1}$-term is of higher order.

The next combination tone CT2 at $\omega_1 - 2\omega_0$ follows to lowest order ($\propto (a_1)^3(a_2)^2$) from

$$i(\omega_1 - 2\omega_0)a_{CT2} = (\mu + i)\omega_{ch}a_{CT2} - 2\omega_{ch}a_{CT2}a_1a_2^*.$$  

(2.9)

and is thus mainly a result of the interaction of the the first CT and both primary responses (and not $a_{CT1}$ and $a_1$ alone, as was seemingly the assumption of Ref. [80]). The latter term and two other interaction terms, i.e.

$$-\omega_{ch}a_{CT1}^2a_1^* - 2\omega_{ch}a_{CT1}a_2|a_1|^2 - 2\omega_{ch}a_{CT1}a_2|a_2|^2.$$  

(2.10)

need to be considered when including higher order contributions. Close to bifurcation ($\mu \approx 0$) and for $\omega_{ch} \approx \omega_1$, Eq. (2.9) yields

$$a_{CT2} \approx \frac{\omega_{ch}a_1a_2^*}{i\omega_0}a_{CT1}.$$  

(2.11)

For the third combination tone CT3 at $\omega_1 - 3\omega_0$, we proceed correspondingly and get to lowest order

$$i(\omega_1 - 3\omega_0)a_{CT3} = (\mu + i)\omega_{ch}a_{CT3} - 2\omega_{ch}a_{CT3}a_1a_2^* - \omega_{ch}a_{CT1}^2a_2.$$  

(2.12)

Here, the last two terms are of the same order ($\propto (a_1)^4(a_2)^3$). Using Eqs. (2.7) and (2.11), one approximately gets for $\omega_{ch} \approx \omega_1$ and $\mu \approx 0$

$$a_{CT3} \approx \frac{\omega_{ch}a_1a_2^*}{i\omega_0}a_{CT2},$$  

(2.13)

which is exactly of the same form as Eq. (2.11). Carefully handling the interaction terms of lowest order, the same law holds for all subsequent combination tones CT4, CT5, ..., which leads to an approximate exponential decay of CT amplitudes satisfying

$$a_{CTk} \approx \frac{\omega_{ch}a_1a_2^*}{i\omega_0}a_{CT(k-1)} = \kappa a_{CT(k-1)},$$  

(2.14)

with $\kappa := (\omega_{ch}a_1a_2^*)/(i\omega_0)$. Approximate exponential decays are still found off-bifurcation (e.g. $\mu = -0.1$) and for $\omega_{ch}$ not exactly at $\omega_1$, see below.

Two examples will now illustrate the above results. In the first example we choose a two-tone stimulus of $\omega_1 = 2\pi \cdot 2000$ rad/s and $\omega_2 = 2\pi \cdot 2200$ rad/s with amplitudes $F_1 = F_2 = 0.01$. In Fig. 2.3a), we compare the numerical integration of Eq. (2.1) with our above-derived analytical calculations for $\omega_{ch} \approx \omega_1$ and $\mu = 0$. For the theoretical approach, $a_1$ and $a_2$ follow from solving Eqs. (2.5) and (2.6) simultaneously, which yields $-14.1$ and $-22.3$ dB
Figure 2.3: CT in a single Hopf oscillator. a) Fourier transform of the response of a single Hopf oscillator ($\omega_{ch} = 2\pi \cdot 2000$ rad/s, $\mu = 0$) when subject to two-tone forcing ($\omega_1 = 2\pi \cdot 2000$ rad/s and $\omega_2 = 2\pi \cdot 2200$ rad/s with amplitudes $F_1 = F_2 = 0.01$); spectrum: numerical integration, crosses: analytical results; exponential decay factor $|\kappa| = |\frac{\omega_{ch} \alpha_1 \alpha_2}{i\omega_0}|$ from Eq. (2.14). b,c) Same as a), but $\omega_{ch} = 2\pi \cdot 1900$ rad/s, $\omega_{1,2} = 2\pi \cdot (2000, 2100)$ rad/s with amplitudes $F_1 = F_2 = 0.01$, and b) $\mu = -0.1$ and c) $\mu = 0$ (oscillator at criticality).

respectively. $a_{CT1}$ follows from Eq. (2.7) together with Eq. (2.8), yielding a response of $a_{CT1} = -33.06$ dB. Using this, we obtain (to lowest order) from Eq. (2.9) $a_{CT2}$ and from Eq. (2.12) $a_{CT3}$ (and correspondingly the other CT). For the numerical integration, we chose a sample rate $SR = 80$ kHz and a fourth-order Runge-Kutta scheme with integration step $h = SR^{-1}$ (discarding the first 0.25s).

The second example considers the case where $\omega_2/\omega_1 = 1.05$ and $\omega_{ch} = 2\omega_1 - \omega_2$, which corresponds to the biological experiment of [138]. We use $\omega_1 = 2\pi \cdot 2000$ rad/s and $\omega_2 = 2\pi \cdot 2100$ rad/s for simplicity ($\Delta f := (\omega_2 - \omega_1)/2\pi = 100$ Hz), and amplitudes $|F| = 0.01$ (-40 dB, moderate to high sound level). Fig. 2.3 shows the obtained response for b), $\mu = 0$, and c), a biologically more appropriate value of $\mu = -0.1$. Both settings produce exponential CT amplitude decays, but with significantly too high decay exponents for corresponding sound pressure levels (Biology: 5-6 dB/$\Delta f$ for the lower CT [138]). For the whole set of biological measurements from 30 to 80 dB SPL made in [138], a single Hopf oscillator underestimates
2.2. RESPONSE TO TWO TONES: COMBINATION TONES

CT levels substantially. This is, however, not the case for the compound Hopf cochlea, as will be exhibited below.

2.2.2 Response of the full Hopf cochlea

In contrast to single Hopf elements, in biology, CT of frequencies lower than stimulus propagate down the cochlea until the waves are amplified and stopped where their frequency matches the characteristic frequency $\omega_{ch}$. This leads to an asymmetric (low-pass) and generally slower CT decay. In a first step, we can exhibit the response corresponding to the first cubic combination tone ‘CT1’, i.e. the $2f_1 - f_2$-tone. For this, we focus on a location with $\omega_{ch} = \omega_{CT1}$ to see how strong a two-tone signal from two single pure tones of equal strength would need to be to generate the same effect as from a direct stimulation by a tone of frequency $f_{ch}$, for a single Hopf oscillator, and for a compound cochlea. We then compare these results to the biological measurements. A conventional quantification of this difference is the ‘relative CT1-tone strength’, obtained by measuring at $x_{ch}$ how much stronger a two-tone input (inputs of equal strength) having a CT1 with $\omega_{ch}$ would need to be to generate the same response as a given pure tone with $\omega_{ch}$ [138]. The horizontal distances between the black and the green lines in Fig. 2.4a,b) illustrate this measure; the results show that CT1-amplitudes vary nontrivially with stimulation level, and that this depends on whether we consider a Hopf amplifier alone or a section in the compound cochlea. Only the compound cochlea reproduces the effect correctly, and is also consistent with measurements from the apical part of the cochlea [24].

We now shift our focus toward the full set of cubic combination tones. Fig. 2.5 provides a detailed comparison of the CT response in the full Hopf cochlea (setting C) to measurements from chinchilla [138]. From Fig. 2.5, we see that the CT are stronger for higher input levels; nevertheless, the first CT to the left of the stimulus frequencies $f_1$ and $f_2$ (at frequency $2f_1 - f_2$) is even for an input level as low as 30 dB SPL (-84 dB) still clearly visible. CT are thus not only relevant at high sound levels, as was for a long time assumed (from which the alias ‘distortion products’ emerged). As we will see below, the CT to the left of $f_1$, $f_2$ propagate further down the cochlea and become ever more dominant in the response. This has far-reaching consequences e.g. for pitch perception (chapter 3 of this thesis). Comparing Fig. 2.5 to Fig. 2.3, we can also note that single Hopf oscillators are fully inadequate to reproduce biological CT response. The feed-forward coupling of the oscillators and the low-pass filtering change CT response in an essential way: CT amplitudes decay in general more mildly, and moreover asymmetrically. Last but not least, Fig. 2.5 provides an excellent opportunity to match the sound level scales of biological measurements (dB SPL) to our simulations. For the experiment shown, an input level of -74 dB in the simulation is comparable to an input sound pressure level of 40 dB SPL. From this reference point, all other sound levels can be deduced (e.g. -64 dB $\cong$ 50 dB SPL). This matching has provided a detailed reproduction of the well-known pitch shift effects (see later), but also seems to hold for other comparisons to biological data.

In order to characterize the full cochlear response to two-tone stimulation, we need to track the signal along its way through the cochlear sections. Fig. 2.6 shows the propagation of a two-tone signal (waveform and spectrum) as it travels down the cochlea. The first cochlea sections mildly amplify the two tones, without any further particular effect on the signals’s waveform or spectrum. When the frequencies $f_1$, $f_2$ get close to their respective place of
Figure 2.4: The behavior of the first cubic CT in a single Hopf element, in the compound cochlea and in biology. a,b) Response amplitude to a pure tone of frequency $f_{ch}$ (black), response to a two-tone input (equal strength of components) with $f_{ch} = f_{CT1}$ (green). The difference (arrows) is the ‘relative strength of CT1’. a) Single Hopf amplifier (no fluid comprised), and b) cochlea section 6 where $f_{ch} = f_{CT1}$. c) Relative strength of CT1 for two $f_2/f_1$-frequency ratios. Red: cochlea section 6, black: biological data [138] ($f_{ch} = 9000$ Hz). The blue arrows in b) and c) describe the same experimental result.

Resonance (section 10), the first CT start to appear. Further on (section 12), the lower CT ($f < f_1, f_2$) get stronger, and new CT appear. At section 14, the waveform and the spectrum have changed completely: the higher frequencies (including $f_1, f_2$) have been dissipated, and the remaining frequencies are grouped around the local characteristic frequency $f_{ch} = 1.31$ kHz. Towards the end of the cochlea, also the remaining lower CT start to gradually vanish.
2.2. RESPONSE TO TWO TONES: COMBINATION TONES

Figure 2.5: CT in the cochlea. a) Basilar membrane response spectrograms for two-tone stimulation of amplitudes 30, 40, 50 dB sound pressure level (SPL) (frequencies $f_2/f_1 = 1.05$ and $2f_1 - f_2 = f_{ch}$) and biological data [138] ($f_{ch} = 7.5$ kHz). b) Full Hopf cochlea, sixth section ($f_{ch} = 5.66$ kHz). Dashed lines: exponential amplitude scaling ($\Delta f = f_2 - f_1$).
Figure 2.6: Two-tone propagation in the cochlea. Input tones were $f_1 = 2200$ and $f_2 = 2400$ Hz, each at -74 dB sound level.

### 2.2.3 Activation profiles

Fig. 2.7 shows the activation profiles of the cochlea for two different two-tone inputs. The activation pattern is primarily determined by two factors: the input amplitude and the distance between the two tones $\Delta f = f_2 - f_1$. For an input of two tones at 2800 and 3500 Hz, a particular excitation becomes visible at $f_0 = 700$ Hz; this is the so-called missing fundamental. It is introduced through a series of CT, first $2f_1 - f_2 = 2100$ Hz, then $3f_1 - 2f_2 = 1400$ Hz, and ultimately $f_0 = 4f_1 - 3f_2 = 700$ Hz, with $f_0$ being the greatest common divisor of $f_1$ and $f_2$. In the case here, $f_1$ and $f_2$ are harmonics number four and five of $f_0$. \(^1\)

We now clearly see that the comparison of the cochlea to a set of tuned strings (a piano) is fundamentally misleading: excitation of the cochlea does not only occur at the locations corresponding to the input frequencies, but due to nonlinearity, also at many other locations. While this is already true for two input frequencies, the picture gets much more complex in the case of three or more frequencies, or mixtures of realistic complex tones. This will be the topic of the next section.

\(^1\)Even though $f_0$ here in principle also corresponds to the perceived pitch of a listener, the pitch is in general not related to the activation of the cochlea at $f_0$. A more detailed discussion of the (missing) fundamental and its relation to pitch will be provided in chapter 3.
2.3. RESPONSE TO MIXTURES OF COMPLEX TONES

The sounds we usually deal with in the real world are not pure tones. Whatever the generator of a sound, a spectrum containing more than one significant frequency is usually observed. Typically (especially for instruments), the observed frequencies correspond to a fundamental frequency and its higher harmonics (the fundamental is the first harmonic). Depending on the tone-generating element (pipe, string, ..), the higher harmonics may be stronger or weaker, and their exact characteristics partially determine the timbre of the tone. In the following, we focus on stationary mixtures of two or three complex tones (fundamental + higher harmonics) as cochlear input. Such mixtures are simple enough to still understand the interactions occurring in the cochlea, but at the same time general enough to explore the full complexity of the cochlea. We will not deal with non-stationary sounds (which for the Hopf cochlea would in principal pose no problems) because we do not expect the extra complexity and difficulties to pay out in additional insight. Furthermore, it is well-known that people with cochlear implants (i.e. without a functional natural cochlea) have much fewer problems in understanding speech, which is very transient, than in enjoying music (which is, on an appropriate timescale, fairly stationary).

As the basic element of our tone mixtures, we chose a ‘typical’ complex tone of a fundamental frequency and four additional harmonics \( a_n = 1/n \). Four higher harmonics have turned out to provide enough additional input (using 10 harmonics does not essentially change the results), and the decay of the relative amplitudes should mimic what is typically found in musical instruments. If we use a single complex tone as our cochlear input, we essentially only activate the cochlear locations corresponding to the input frequencies, because the generated CT come to lie exactly on top of the input frequencies \( n f_0 \). This results in a change in relative weighting of the harmonics along the cochlea, but not in additional complexity. Using a mixture of two complex tones with fundamentals \( f_0^{(1)} \), \( f_0^{(2)} \) as input, this no longer holds. The harmonics of both tones start to interact in a complicated way that becomes difficult to track. This is illustrated in Fig. 2.8, which shows the output of four cochlea sections to stimulation with two complex tones of varying \( \{f_0^{(1)}, f_0^{(2)}\} \) (using cochlea discretization setting \( D \) with \( \mu = -0.2 \) for all sections).
CHAPTER 2. COCHLEAR COMPLEXITY

Figure 2.8: Activation profiles for stimulation with mixtures of two complex tones of fundamental frequencies \{f_0^{(1)}, f_0^{(2)}\} (input level -60 dB per tone, and fundamental frequencies from (0, 15) kHz. Cochlea setting D with 29 sections (all at \(\mu = -0.2\)), characteristic frequencies are 2.96 kHz (section 10), 1.25 kHz (section 15), 0.52 kHz (section 20) and 0.22 kHz (section 25).

The nonlinear interaction of the frequencies generates complex ‘receptive fields’ (a term we borrowed from neuroscience [153, 69]) which differ from section to section. Specifically, the contributions from fundamental combination tones of increasing order can be identified as non-vertical/non-horizontal structures of activation.

2.3.1 Cochlear interaction networks

In order to get a handle on the nonlinear frequency interactions in the cochlea, we developed an algorithm to visualize them as a complex network (cf. [2, 12] for overviews on complex networks). In the network picture, the nodes of the network refer to the cochlea sections and the links denote interactions. The idea is the following: a cochlea section \(j\) is activated (defined as: output > −50 dB) either by a frequency contained in the input (amplified by the cochlear cascade), or by CT generated within the cochlea. In order to visualize the activation
of a section, we check for every cochlea section $j$ if there is a previous section $i$, $i < j$ that contributed to the activation of section $j$. This is done in the following way: first, we search at section $j$ (via Fourier transform) for active frequencies ($> -50$ dB) that lie within the (logarithmic) frequency band covered by the section (a narrow band centered around $f_{ch}^{(j)}$). Suppose there is such a frequency and denote it by $f$. We now go through all the previous sections and check if at a previous section $i$, there is a set of active frequencies $f'$, $f''$, $f'''$ so that $f' + f'' - f''' = f$. If this holds, then section $i$ contributes (via the cubic interaction term) to the activation of section $j$, which results in an edge $i \rightarrow j$. If $f' = f''$, this corresponds e.g. to the typical case of a $2f_1 - f_2$-CT generation. Moreover, we do not rule out the possibility that $f' = f'''$ or $f'' = f'''$, which would correspond to the simple amplification of a single frequency. In this way, we also account for the pre-amplification of frequencies before they reach the respective place of maximal excitation (where $f \approx f_{ch}$). We checked the case where this is explicitly forbidden and only real CT generation is represented by network edges; the networks are then in general slightly smaller, but the larger networks and the distributions (see later) are very similar.

Fig. 2.9 shows the resulting networks for two cases: a simple two-tone input (see previous section), and an input mixture of two complex tones. In the context of networks and statistical descriptions, we have always made use of the extended cochlea discretization setting $\mathbf{D}$, with 29 sections covering 7 octaves. To keep the description as general as possible, all bifurcation parameters were set to the same value (‘flat tuning’).
CHAPTER 2. COCHLEAR COMPLEXITY

Figure 2.9: Two examples of cochlear activation networks ($\mu = -0.25$). a) Two-tone input $f_1, f_2 = 375, 500$ Hz, input level -60 dB each; b) input mixture of two complex tones ($f_0^{(1,2)} = 2000, 3350$ Hz) at -60 dB each. Top panels: networks in typical graph layout. Middle panels: Same networks, symbolically plotted as cochlear spirals (arrow: direction of propagation). Lower panels: corresponding activation profiles.
2.4 Statistical activation properties

2.4.1 Activation statistics

Based on the activation profiles introduced earlier, we may now ask what statistical activation distributions are generated along the cochlear duct. To address this question, we stimulated the cochlea with two-tone or three-tone mixtures of random fundamental frequencies and random amplitudes, or, for a comparison, at fixed amplitudes. We adhere to the activation threshold of -50 dB, but checked that differing choices, e.g. -40 or -60 dB, do not compromise any of the following results.

For obtaining a statistics of activity along the cochlear duct generated by more complex input, we define the activity $A(j)$ of a section $j = 1, ..., 29$ by

$$A(j) = \frac{1}{N} \sum_{i=1}^{N} \Theta(f_i, j),$$

where $f_i$ denotes the input frequency (or frequencies) of stimulation experiment $i$, $N$ is the total number of stimulations, and $\Theta(f_i, j)$ is 1 if the output at section $j$ exceeds the threshold and 0 otherwise. For pure-tone input uniformly sampled from a (non-logarithmic) frequency interval, the average activity $A$ follows a power-law of exponent one (each section is activated by a proportion of frequencies corresponding to the ‘bandwidth’ of the section, which $\propto f$). For two- or three-tone inputs, however, CT lead to additional activity (cf. Fig. 2.8). Fig. 2.10 shows the results over all cochlea sections on log-log-scale, for two-tone and for three-tone mixtures, for different $\mu$-settings, and for fixed or random input strengths. In all cases, we observe power-laws $A \propto f^\beta$ with exponents $0 < \beta < 1$ (if we sample the random fundamental frequencies from a distribution such as $\propto 1/f$, we also obtain power-laws, but this time with negative exponents $\beta$).

Figure 2.10: Cochlear activation distributions. Average activity $A$ from two complex tones from $N = 10,000$ random input frequencies. a) Random sound levels (-80,-40) dB (rms) per tone, b) fixed sound levels (-60 dB). Inset: Results of mixtures of three complex tones. Dashed lines: power-law guidelines (exponents $\beta$ from maximum likelihood estimation (bottom to top): for a), $\beta = 0.6, 0.44, 0.3, 0.2, 0.13$, for b), $\beta = 0.64, 0.43, 0.29, 0.2, 0.12$).
CHAPTER 2. COCHLEAR COMPLEXITY

Figure 2.11: Network size distributions, generated by inputs consisting of two complex tones with fundamental frequencies from (0,15) kHz (≥ 30,000 stimulations per dataset). a,b) Random sound levels and \( \mu = -0.2, -0.25, -0.3 \), c,d) Fixed sound levels (-50,-60,-70 dB) and \( \mu = -0.25 \). a,c) Raw data \( (N(s) = \text{Number of occurrences, rescaled to 30,000 trials}) \), b,d) sliding-window-averaged binned (bin-size 4) probability distribution \( P(s) \) for highlighted area. Thin lines: power-law guidelines of exponent \( \beta = 1.5 \).

2.4.2 Network statistics

We now move our investigations from the statistical activation toward the cochlear interaction networks and their statistical properties. In neuroscience, a recently well-investigated paradigm is the concept of neuronal avalanches [8, 9] (cf. [28, 27, 89] for results on simple neuronal avalanche modeling). To measure such avalanches, a neuronal culture or slice is monitored by recording its electrical activity through multi-electrode arrays. Depending on the developmental state of the culture, the neurons can show highly synchronized spontaneous bursting activity [172]. If these bursts are examined on a finer scale (where mostly the time binning is of crucial importance), it is possible to define ‘avalanches’ as the network’s spiking activity between two time-bins of quiescence (no spiking). Creating a histogram of avalanche sizes, distributions following a power-law are observed, where the size of an avalanche is usually determined by the number of electrode channels activated in an avalanche, or alternatively by the sum of the amplitudes of the local field potentials [8, 9].

In the cochlea, we can follow a similar approach by looking at the size distribution of the cochlear interaction networks. Depending on the input frequencies, networks of arbitrary size
may be generated (only limited by the size and discretization of the cochlea); simple inputs (e.g. two pure tones separated by an octave) will elicit a small network, more complex inputs (e.g. two complex tones close together) will elicit a larger network. Feeding the cochlea with a large set of inputs, we can then measure the distribution of the size of these networks, where due to the small size of our cochlea (here: 29 nodes) we measure the network size by the number of links (the actual interactions). The results for different sound levels (random or fixed) and different \( \mu \)-settings are exhibited in Fig. 2.11, where we stimulated the cochlea by at least 30,000 two-complex-tone inputs per dataset. Typically, we obtain distributions which roughly follow a first power-law until around size 30, and then fall off as a second power-law afterward. For medium sound levels and typical \( \mu \)-values, the exponent of the second power-law is in the vicinity of \( \beta = 1.5 \), which is the exponent observed in neuronal avalanches and theoretical percolation models (or critical branching) \[177, 2\]. Going to higher (lower) sound levels or with \( \mu \) closer to (further away from) bifurcation, the distribution starts to appear more supercritical (subcritical). The goodness of a power-law fit can be measured by comparing the Kolmogorov-Smirnov (KS) distance measure to surrogate data \[23, 32\]; the proportion of cases where the actual data is better fitted by a power-law (measured by KS) than the surrogate data defines the \( p \)-value. Typically, \( p = 0 \) follows from data which is clearly not in correspondence with a power-law, whereas \( p \geq 0.1 \) can already be considered a good power-law fit \[23\]. Applying this method to the distributions in Fig. 2.11 on a reasonable interval (we typically have to discard the section before the break around \( s = 30 \)), we observe \( p \)-values often exceeding 0.5. For the specific example of the blue line (random sound levels, \( \mu = -0.25 \)), the test yields a \( p \)-value of 0.9 (exponent 1.5) on the interval [90,220].

In general, the network size distributions depend to some extent on the input space chosen (where a certain class of inputs has to be assumed). In addition to the class of mixtures of two complex tones with five harmonics each, we also performed experiments with mixtures of three complex tones, mixtures of four random pure tones, or mixtures of two tones with less/more harmonics as input. The general behavior is similar, although the response to mixtures of three complex tones appears more supercritical than usual. This seems to represent a case of overstimulation, where the abundance of different frequencies in the input (fifteen frequencies) leads to a bias towards large networks. On the other hand, stimulation with two complex tones of three harmonics each yields a distribution that even more consistently follows a power-law than the example in Fig. 2.11d). The choice of input thus represents one of the difficulties of the approach, although the consistency of the results suggests the conclusion that the observed effects are generic.

In addition to network size, we also looked at other characteristics of cochlear networks. Typical degree properties are displayed in Fig. 2.12, which exhibits the average in- and out-degree for each cochlea section and the degree distributions for the networks ordered according to their size (based on the number of links, we define three classes). In particular, when we consider large networks (> 150 links), a ‘humpy’ degree distribution resembling the degree distributions found in the \textit{Drosophila} courtship networks is observed \[96\].

### 2.4.3 The effect of learning

In chapter 3, we will specifically address the tuning of the Hopf cochlea’s \( \mu \)-parameters as a means to focus on desired sounds (a first step of ‘listening’). By detuning the \( \mu \) of some sections, the response to tones in the vicinity of these sections’ characteristic frequencies can
be suppressed. Based on such a tuning process, we can now investigate the tuning’s effect on the activity and network size distributions exhibited before (where $\mu$ was always kept fixed and the same for all cochlea sections). The choice of the particular sections to be detuned is in this context not relevant; we follow, however, the results from our tuning experiments (chapter 3) by detuning not single sections, but rather collections of sections covering certain frequency bands.

The effect of detuning a single frequency band on the cochlear activation distribution is exhibited in Fig. 2.13a). The detuning results in a clear reduction of activity in the corresponding frequency band; afterwards, the ‘ground-state’ power-law behavior seems to be partially restored (at slightly lower levels, however). More interesting is the effect of tuning on the activation networks. Fig. 2.13b) exhibits the impact of detuning sections 11 and 12 on a single activation network, and how this affects the size of the network (Fig. 2.13c) and the ‘small-worldness’ $S := \gamma / \lambda$. Here, $\gamma$ is the clustering coefficient of the network divided by the average clustering coefficient of random networks of same size (nodes and links), while $\lambda$ is the average shortest path length of the network divided by the corresponding quantity from random networks [72]. By definition, a network is small-world if $S > 1$. Note however, that $S$ scales roughly linearly with the size of the network [72]. In our case, this is corroborated ($S$ scales approximately as $n^{1.2}$, where $n$ is the network size in terms of number of nodes); the reduction in network size from tuning therefore leads to a corresponding decrease in $S$. In that sense, we can conclude that tuning works against the small-worldness property.

Finally, we consider the effect of a general learning (tuning) on the ‘avalanche distributions’ (network size distributions) shown earlier. Fig. 2.14a) shows the effect resulting from detuning two frequency bands, spanning 970-1360 Hz (sections 15 and 16) and 400-680 Hz (sections 19-21). Such a $\mu$-setting would e.g. correspond to a situation where a sound of fundamental 550 Hz and its second harmonic would be suppressed (cf. Fig. 3.14 in chapter 3).
2.4. STATISTICAL ACTIVATION PROPERTIES

Figure 2.13: Effect of tuning (learning). a) Activation distributions (fixed -60 dB input level, cf. Fig. 2.10) after detuning Hopf sections 19,20,21, from $\mu_{19,20,21} = -0.2$ (dashed), to $\mu_{19,20,21} = -0.3, -0.5, -1, -2$ (magenta to black, respectively). b-d) Effect of detuning sections 11,12 (from $\mu_{11,12} = -0.25$ to $\mu_{11,12} = -1.0$) on a single activation network (input: two complex tones with $f_0 = 1331, 2120$ Hz, five harmonics, at -70 dB rms each). Upon tuning, the network from the unbiased cochlea (blue) is gradually transformed into a distinctly smaller sub-network (red). c) Size of the network as a function of detuning ($\mu_{11,12} \in (-2, -0.25)$), d) Small-worldness $S$ as a function of node number.

2.4.4 Interpretation

In statistical physics, power-laws are fingerprints of systems near a second order phase transition. These phase transitions are characterized by an order parameter $m$ that is zero in one phase, and nonzero in the other phase. In the vicinity of the phase transition, $m$ grows as $m \propto (x - x_c)^\beta$, where $\beta$ is a critical exponent related to the type of second-order phase transition (i.e., to which universality class it belongs), and $x$ is a parameter (critical point at $x = x_c$). Well-studied model systems are e.g. the Ising model [73], where the order parameter is given by the mean magnetization (which is zero above the critical temperature, and non-zero below), the Potts spin model [127], stochastic lattice gas systems (which often show a nonequilibrium phase transition into an absorbing state [67]), or various percolation models [158]. In addition to the order parameter, also other observables follow power-laws near the phase transition, each with its own critical exponent. These exponents are not independent, however, but related through scaling relations [159].

On a mathematical level, a power-law $f(x) \propto x^\beta$ expresses the lack of a typical scale. This
Figure 2.14: Effect of learning on power law distributions for the network sizes in the cochlea, and in the framework of the thermodynamic formalism, based on invariant measures (signals). a) Network size distributions (40,000 stimulations with two complex tones, random amplitudes) for flat tuning (blue), and after detuning two frequency bands (sections 15-16, 19-21) from $\mu = -0.25$ to $\mu = -1.0$ (red crosses), and $\mu = -2.0$ (red stars). b) In the thermodynamic formalism, the observability $O$ of an invariant measure $\epsilon$ decays with time $t$ as $O(\epsilon, t) \sim e^{t(\epsilon - S(\epsilon))}$ (red and blue arrows). Points of the graph on the diagonal $\epsilon = S(\epsilon)$ indicate measures that do not experience a temporal decay. Blue: Entropy function $S(\lambda)$ of intermittent systems associated with power-law characteristics. Red: Entropy function associated with non-power law behavior, as the result of focusing on a particular measure. The green circle and the green arrows symbolize the result of the learning process.

is reflected in the fact that $f(\lambda x) \propto (\lambda x)^\beta \propto \lambda^\beta x^\beta$, i.e. rescaling the system with a factor $\lambda$ does not change the form of the function, but simply adds a pre-factor. This is contrary to an exponential, for example, where the function $f(x) \propto e^{ax}$ is essentially governed by the parameter $a$ and rescaling $x$ by $\lambda$ affects the growth/decay of the function. In physical systems, the lack of scale or the so-called ‘scale-freeness’ near a phase transition is typically expressed by the divergence of correlation functions, which is related to the divergence of a susceptibility-like quantity (through the fluctuation-dissipation theorem). This expresses that near a phase transition, the system is not governed by its microscopic details, but essentially behaves in a macroscopic way.

We can apply these ideas to systems exhibiting avalanches, as was first done by Bak and colleagues for the sandpile model in the context of the so-called ‘self-organized criticality’ [5, 6, 78]. Here, the lack of scale in the system (i.e. the ‘criticality’ of the system) is expressed by the observation of power-law distributed avalanche sizes. Power-law distributed avalanches have since then been observed in a variety of models and contexts, such as in studies on granular flow [103], investigations of solar flares [97], or the behavior of lung inflation [164]. In our context, the cochlear avalanches are elicited by sound input of varying frequency components. Being based on a wide enough screening of the input space, the observation of power-law distributed sizes thus seems to reflect that the system can react to any kind of input, in this sense expressing the highest possible adaptability. However, if we want to
actively listen to something, we need to enhance certain sounds and suppress others, which then distinguishes a scale of the system. Upon learning, the power-law must therefore be lost, which is exactly what we observe.

The ideas can be brought onto a more rigorous level within the thermodynamic formalism of dynamical systems [139, 62, 165, 166, 124]. From ergodic theory, we know that the intermittent maps are the simplest model for producing power-law distributed invariant density measures. These maps are equipped with an infinity of ‘observable invariant measures’ (related to phase-transition like behavior), which is in strong contrast to the ‘normally behaved’ systems that only have one observable invariant measure or state. From a continuum of observable invariant measures, upon learning the system puts a focus on a particular measure, which destroys the power law (Fig. 2.14b). Whereas the predictability of the reaction of an unbiased system upon arbitrary input is extremely difficult, for the tuned system this task is much simpler. Accordingly, unbiased power-law systems perform a much smaller amount of computation, if compared to the tuned system (where computation is defined as the reduction of the complexity of prediction imposed by the system).
Chapter 3

Pitch and the tuning of the cochlea

3.1 Pitch

3.1.1 Basic notions

Pitch, or *Tonhöhe* in German ("tone height"), is one of the most salient perceptual characterizations of sound. Despite the omnipresence of pitch in daily life (singing, musical instruments, or the buzzing of a machine), it has turned out to be difficult to give a proper definition. The defining feature of pitch is that it relates to an ordering of sounds on a one-dimensional scale related to frequency, which allows to compare two pitches and say that one of them is "higher". The close relation of pitch and musical melodies is expressed by the American Standards Association’s definition of pitch in 1960, where pitch is “that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale” (Acoustical Terminology SI, American Standards Association) [125].

The simplest sounds eliciting a clear pitch are pure tones, where we find (within some reasonable frequency range) a simple one-to-one mapping of stimulus frequency and pitch. When several frequencies are involved however, the story gets more complicated and the relation of the signal’s frequency content and pitch becomes nontrivial. From here, two lines of possible pitch perception models open up: 1) models that are still based on the frequency content despite the difficulties, and 2) models based on the autocorrelation (ACF) of the signal. In spite of the actual mathematical correspondence between the two (the Wiener-Khinchin theorem relating the ACF to the power spectrum), the use of autocorrelation offers a simpler and cleaner way to model pitch. The reason for this is that pitch is often closely related to the intrinsic periodicity (or repetition rate) of a sound, a characteristic that is usually well-visible in the ACF, but sometimes hard to spot in the Fourier transform (a simple example is provided by two- or three-tone signals, see below).

Historically, the proper investigation of tones and pitch was enabled by the mathematical framework by Joseph Fourier (1768-1830) and the development of the siren, which was an acoustic instrument that permitted the generation of tones and signals with well-determined physical properties [30]. Georg Simon Ohm (1789-1854) formulated a law relating a tone to its Fourier content ("Ohm’s acoustic law"), expressing that a tone of frequency $m$ must contain a partial of the form $\sim \sin(2\pi mt)$ [120]. Two years earlier, however, the young physicist August Seebeck (1805-1849) had already disproved this statement [151]. In a seminal experiment, Seebeck had produced a sound that contained (besides the higher harmonics) only a very
weak fundamental, which however still elicited a clear pitch corresponding to frequency \( m \). He thus concluded that the perception of partial frequencies is not one-to-one related to their strength in the signal, and that pitch and its salience are not necessarily determined by the lowest Fourier component. Ohm, however, did not accept any of Seebeck’s objections, and the disagreements among the two scientists became later known as the “Ohm-Seebeck dispute” [167]. Several years later, Hermann von Helmholtz (1821-1894) extended Ohm’s law and related the timbre of a sound to the distribution of Fourier components [171, 30] (a point that seems to have already been made by Seebeck, however [167]). Moreover, Helmholtz was the first to specifically address the ear and its role in sound perception, describing the tuning of the organ of Corti, and discussing the key role of nonlinear distortions (combination tones).

3.1.2 The missing fundamental paradigm

When we listen to a typical (harmonic) complex tone of fundamental \( f_0 \) and higher harmonics \( 2f_0, 3f_0, \ldots \), we hear a pitch corresponding to the fundamental \( f_0 \) (i.e. the pitch is not higher or lower than the pitch elicited by a pure tone of frequency \( f_0 \)). The phenomenon of the ‘missing fundamental’ refers to the fact that as we remove the fundamental \( f_0 \) from the signal, the elicited pitch is, under ‘generic’ conditions, still the same. This perception of a pitch \( f_0 \) that is not part of the signal itself but generated by the higher harmonics, is also known as residue pitch perception. First described by Seebeck (see above; in his case, the fundamental was present in the signal, but very weak), the residue pitch phenomenon has been a topic of research for more than 150 years.

Fig. 3.1 displays three different signals which all elicit the same pitch of 200 Hz. The ‘missing fundamental’ paradigm is represented in part c), where the signal consists of two harmonics \( 3f_0, 4f_0 \) only, which is enough to generate a clear pitch at \( f_0 \). To explain this perception of the residue, historically two lines of argumentation have been followed. The first one is that through nonlinearity, the fundamental is re-introduced as a combination tone (the original line of argumentation by Seebeck and Helmholtz). Even though this is true for this specific case, the basic reasoning has later been disproved by the use of inharmonic sounds (see next section), or the masking of the lower frequencies with noise. The other (and better) explanation assumes that for the perception of a pitch at \( f_0 \), the presence of \( f_0 \) in the spectrum is not required in any case, neither in the input signal nor in the cochlea. To still infer the pitch at the fundamental however, one can deduce \( f_0 \) from the other partials in the spectrum and their spacing (pattern-matching models [51]), or read it off from the waveform’s periodicity (directly or through autocorrelation [93, 31]), cf. the arrows in Fig. 3.1.

3.1.3 Pitch-shift experiments

After Helmholtz, it took almost a century to make further progress on the topic of residue pitch perception. The first modern experiments were conducted around 1940 by Marten Schouten, who used an optical siren that allowed him to accurately cancel partials in the acoustic signal. This enabled him to provide further evidence that the perception of the residue is unrelated to the physical presence of \( f_0 \) in the spectrum [148]. In a seminal experiment, Schouten then created a signal where all partials were shifted by the same amount \( \Delta f \) [149, 150, 30]. These so-called ‘inharmonic’ sounds have some interesting properties. Take e.g. a three-tone signal with frequencies \( f_1, f_2, f_3=400, 600, 800 \) Hz, which corresponds to an amplitude-modulated
3.1. PITCH

signal of carrier frequency 600 Hz and modulation frequency 200 Hz, and shift the partials by
\( \Delta f = 50 \) Hz to obtain a signal with partials 450, 650, 850 Hz. This signal still has a modulation
frequency and a difference combination tone of 200 Hz; Schouten, however, found that the
residue pitch perceived shifted as well, approximately linearly with \( \Delta f \). This is the so-called
‘first pitch-shift effect’. Confirmations and extensions of Schouten’s experiments followed in the
1950s and 1960s mainly by the Dutch scientists de Boer [29, 30], Plomp [126], Ritsma
[136], and Smoorenburg [156]. Whereas these studies essentially confirmed the existence of
the first pitch-shift effect, they found a consistent and regular deviation from a simple model
of the first pitch-shift effect (de Boer’s formula). This deviation was termed the ‘second pitch-
shift effect’, and was attributed to the generation of combination tones in the cochlea [156].
As we will see later, the second pitch-shift effect can be accurately reproduced and explained
not by the generation of combination tones alone, but by a subtle interplay of CT generation
and the low-pass filtering provided by the cochlear fluid.

The first pitch-shift effect: de Boer’s formula

To explain the roughly linear shift of the perceived pitch resulting from a shift of the partials,
a simple argument can be given, which became known as de Boer’s formula. To derive
it, a single assumption is sufficient: the inverse of the perceived pitch \( 1/f_p \) corresponds to

Figure 3.1: Three signals which elicit the same pitch \( f_p = 200 \) Hz (left panels: waveform, middle
panels: Fourier transform, right panels: autocorrelation ACF). a) Pure-tone of frequency
\( f_0 = 200 \) Hz, b) Complex tone with five equal-strength harmonics \( f_0, 2f_0, ..., 5f_0 \), c) two-tone
of frequencies \( f_1, f_2 = 600, 800 \) Hz. All signals are normalized to the same rms-power.
Figure 3.2: The first pitch-shift effect. Pitch-shift experiment for $N = 2$, $k = 3$, $f_0 = 200$ Hz and $\Delta f = 0, 50, 100$ Hz (a), (b), (c), respectively). The perceived pitch $f_p$ or $f_p'$ is visible as the distance between the main peaks in the waveform (l.h.s.), or as the main peak in the autocorrelation function (ACF, r.h.s.), and can be simply calculated by de Boer’s formula (see text). $\Delta t$ is the period of the fast-changing signal (inverse of the carrier frequency).

the distance between the main peaks in the waveform, cf. Fig. 3.2. If there are several comparably salient distances, this leads to an ambiguity in pitch perception. In that case, the perceived pitch depends on the history of the psychoacoustic experiment, and a hysteretic effect is observed.

We can now derive de Boer’s formula (cf. [156, 30, 20]) for a general signal composed of $N$ subsequent harmonics of fundamental frequency $f_0$. In the case of a harmonic signal with lowest component $kf_0$, the signal takes the form

$$A(t) = A_0 (\sin(2\pi kf_0 t) + \sin(2\pi (k+1)f_0 t) + \ldots + \sin(2\pi (k+1-N)f_0 t)),$$

(3.1)

where $A_0$ is the amplitude. Using trigonometric identities, these signals can always be rewrit-

---

1Note that de Boer’s formula could also be derived in the spectral picture, where the formula provides the missing fundamental whose harmonics would be closest (‘best match’) to the given spectrum [156].
3.1. PITCH

ten as a product of a carrier term $\propto \sin(2\pi k' f_0 t)$ and several modulation terms, where $k' = k + (N - 1)/2$. Pictorially, $k' f_0$ is the ‘center of gravity’ of the signal (for $N = 2$, $k' = k + 1/2$, or for $N = 3$, $k' = k + 1$). Independent of the number of components $N$ or a shift of the partials by $\Delta f$, the periodicity of the envelope is always $f_0$. In contrast, a shift of the partials by $\Delta f$ shifts the center of gravity to $k' f_0 + \Delta f$, which decreases the carrier frequency. These two facts can now be used to determine the distance between the main peaks in the signal’s waveform for arbitrary $\Delta f$.

In the harmonic case ($\Delta f = 0$), the location of the waveform’s main peaks within the envelope is always the same (there are $k'$ carrier periods $\Delta t = 1/(k' f_0)$ of the signal within a period of the envelope), which yields a distance between two major peaks of the form $1/f_0 = 1/f_p$. The perceived pitch is thus simply the (missing) fundamental $f_0$. This changes if we now shift the partials by a small amount $\Delta f$; in this case, the distance between two subsequent (carrier) peaks of the waveform changes to

$$\Delta t = \frac{1}{k' f_0 + \Delta f}, \quad (3.2)$$

being slightly smaller than in the harmonic case. Within one period of the envelope, still $k'$ periods of the fast-changing waveform separate two main peaks. The perceived pitch (the inverse of this so-called ‘pseudo-period’) is thus given by

$$f_p = \frac{1}{k' \Delta t} = \frac{k' f_0 + \Delta f}{k'} = f_0 + \frac{\Delta f}{k'}, \quad (3.3)$$

which is de Boer’s formula. Specifically, for $N = 2$, $f_p = f_0 + \Delta f/(k + 1/2)$, and for $N = 3$, $f_p = f_0 + \Delta f/(k + 1)$.

A specific example is provided by Fig. 3.2, which shows the case of $N = 2$, $k = 3$ and $f_0 = 200$ Hz for three values of $\Delta f = 0, 50, 100$ Hz. For $\Delta f = 100$ Hz, the situation can equivalently be considered a pitch-shift experiment with $k = 4$ and $\Delta f = -100$ Hz. This yields a second possible perceived pitch $f_p'$ which is equally salient because the distance between the main peaks in the waveform alternates between $1/f_p$ and $1/f_p'$. In between, at $\Delta f = 50$ Hz, this second interpretation is possible as well ($k = 4$ and $\Delta f = -150$ Hz), but the corresponding distance $1/f_p'$ occurs less often in the waveform. Note that in all cases, the autocorrelation function provides the correct perceived pitches (r.h.s. panels in Fig. 3.2). Moreover, the autocorrelation picture gives an indicator of the salience of the perceived pitch through the size of the corresponding peak in the ACF.

In the case of $N = 2$, the predicted pitch $f_p = f_0 + \Delta f/(k + 1/2)$ can also be interpreted as the main three-frequency resonance formed by the two signal frequencies [19]. For general $N$, the formula moreover correctly predicts the resonances for a simple neuron driven by noise and several sub-threshold periodic signals [20]. This is intuitively clear, since the equation merely provides the distances between the main peaks in the signal, which in a case of simple sub-threshold stochastic resonance will automatically be reflected in the resulting interspike intervals.

The second pitch-shift effect

The actual pitch perceived by a listener in a pitch-shift experiment does only to a first approximation follow de Boer’s formula. As shown in Fig. 3.3, the slope of the psychoacoustically
Figure 3.3: The second pitch-shift effect. Perceived pitch $f_p$ upon stimulation with three-tone inputs and center frequency $f_c = (k + 1)f_0 + \Delta f$ and frequency spacing $f_0 = 200$ Hz. Black stars and lines: psychoacoustic data from [150]. Grey lines: de Boer’s formula with $k' = k + 1$, and $k' = k + 1/2$ (dashed). Red circles: $f_p$ in the Hopf cochlea model from peaks in the autocorrelation function (input amplitude -77 dB; cochlea setting C; read-out sections indicated in red).

evaluated lines of $f_p$ is constantly higher than predicted by de Boer’s equation. This deviation from de Boer’s formula is called the second pitch-shift effect. Bearing in mind the results in chapter 2 of this thesis, this comes as no surprise: in the cochlea, combination tones emerge, and tend to dominate the signal more and more towards the end of the cochlea. When the pitch is read out at a cochlea section beyond the location corresponding to the signal frequencies, the whole frequency spectrum of the signal shifts towards the lower frequencies, which lowers $k'$ and hence increases the pitch-shift $f_p(\Delta f)$ (cf. the propagation of a two-tone signal in Fig. 2.6). As a first correction of de Boer’s formula for $N = 3$, one could thus use $k' = k + 1/2$ (the center of the lower two frequencies) instead of $k' = k + 1$, which is the basis of the match in [19]. As demonstrated in Fig. 3.3, this is however not sufficient, and works even less well in the case of $N = 2$ (see below).

In the following, we describe an exact quantitative reproduction of the second pitch-shift effect for $N = 2$ and $N = 3$ on the basis of the Hopf cochlea. A reproduction of the second pitch-shift effect for $N = 3$ is exhibited in Fig. 3.3 (red circles; see later for details). In this case, the second pitch-shift is rather weak compared to the first pitch shift. Stronger pitch-shift effects are observed if two tones are used instead of three (i.e. $N = 2$); this will be the basis of the following investigations. The most elaborate psychoacoustic study related to the second pitch-shift effect was performed by Smoorenburg [156], who tested two subject using two-tone signals at 40 dB SPL. Smoorenburg not only measured the shift of the perceived pitch, but at the same time investigated the audibility of combination tones to assess how CT shape the second pitch-shift effect. This combined investigation allows us to deduce a full pitch extraction model.

In the Hopf cochlea as well as in biology, two key parameters influence the amount of pitch-shift: the stimulus amplitudes that determine the CT decay, and the cochlear pitch

3.1. PITCH

Figure 3.4: Pitch extraction. a) Spectra for two-tone stimulation (-74 dB, $f_1, f_2 = 2200, 2400$ Hz) at three cochlea sections. The lowest audible combination tone (CT) (hearing threshold: -53 dB, blue line) is the response at 1400 Hz (section 14, circled). The perceived pitch is the residue pitch (red arrow) associated with the spectrum at this location. b) Psychoacoustical lower hearing frequency limit of CTs [156] (dashed black line). Simulation: Lowest CTs above the implemented amplitude threshold (solid red line) and highest CTs below the limit (unconnected red circles).

read-out place that keeps track of the fluid’s amplification and low-pass filtering history. Detailed comparisons to experimental data allow us to assess both parameters in a natural way, rendering any parameter fine-tuning unnecessary. The comparisons of cochlear response (especially CT response) to biological data in chapter 2 provided us with a scale to relate input amplitudes in the experiment (in dB SPL) to our simulations. This fixes the input amplitude to use (40 dB SPL corresponds to -74 dB). The detailed investigation of CT audibility by Smoorenburg then provides us with a model for the place of pitch extraction (illustrated in Fig. 3.4). We posit that pitch is extracted at the last cochlea section that has a CT just over hearing threshold (the lowest audible CT). The value of the hearing threshold (-53 dB) follows from the comparison to Smoorenburg’s results [156] (Fig. 3.4b). The procedure yields a read-out place that shifts monotonically with stimulation frequency (for our cochlea covering a frequency range from 440 to 14,080 Hz, over ranges from around 700 to 1,600 Hz), which leads to the exact correspondence of the second pitch-shift values exhibited in Fig. 3.5a).

A corroboration of our results can be obtained by a comparison to data from a place on the auditory pathway between cochlea and cortex. Such data is available through Rhode’s investigations of the cat cochlear nucleus [135], where the ‘pitch’ can be read-off from interspike interval (ISI) histograms ($1/f_p$ is given by the peaks in the ISI histograms). In the experiment exhibited in Fig. 3.5b), the measurement is based on a single neuron having its own preferred frequency, which already sets the read-out place in our cochlea (section 15). The exact agreement of the model and the experimental data confirms our main assertions without relying on a particular read-out mechanism.

To uncover the actual biophysical principles underlying the second pitch-shift effect, we
may consider the active and the passive properties of the cochlea independently. The failure in the reproduction of the second pitch-shift effect by single Hopf oscillators [80] confirms the improperness of modeling hearing with single Hopf elements (which was already exhibited in the context of CT response in chapter 2). CT behavior and correspondingly the second pitch-shift in the full Hopf cochlea and in biology are thus fundamentally different from the case of isolated Hopf amplifier elements. From the model, the origin of this must be seen either in the feed-forward coupling of Hopf amplifiers, or directly in the viscous damping by the cochlear fluid. Both effects are due to the presence of the fluid. Qualitative insight into the respective importance of the two effects can be obtained simply by running the pitch-shift experiments in a cochlea lacking the viscous damping term (i.e. without filters). As a result, further down the cochlea, the signal frequencies $f_1$ and $f_2$ remain dominant, so that de Boer’s formula would continue to hold and the second pitch-shift would not be observed.
3.1. PITCH

In biology, tissue-related effects may be suspected to contribute as well (cf. the description of the original Kern-Stoop model in chapter 1). We expect that those effects will contribute similarly to surface tension, which only smooths (and therefore lowers) the effect of the fluid viscous losses. As the presence of the cochlear fluid is undisputed, these observations are strong support for the conclusion that human-perceived pitch is essentially shaped by the fluid physics of the cochlea [55].

From the cochlea to the auditory nerve

Up to now, we have always directly compared cochlear waveforms to results from psychoacoustics, or to electrophysiological measurements in the brainstem (cochlear nucleus). The exact correspondence between pitch in the Hopf cochlea, and pitch in the brainstem or from psychoacoustic experiments suggests the following conclusion: the cochlear waveform’s periodicities are one-to-one transduced into the auditory nerve, where they manifest as peaks in the histogram of interspike intervals. Using the compound model cochlea-inner hair cell-auditory nerve exhibited in chapter 1 (see there for the parameter values), we can investigate in detail which ingredients are necessary for such a faithful transmission of the pitch [106].

![Figure 3.6: Stochastic resonance of high/low-medium spontaneous ANN (upper/lower panel). a) Spike trains of one/four neuron(s), b) interspike interval distribution at the indicated noise level, c) probability p for the instantaneous frequency to coincide with frequency of the perceived pitch $f_p$, for variable noise levels $\sigma$.](image)

The key result is exhibited in Fig. 3.6. For a three-tone pitch-shift experiment with $f_c = 850$ Hz and a modulation frequency of 200 Hz (in this case, de Boer’s formula roughly holds), the figure shows the ISI histograms of the different classes of auditory nerve neurons (for the low-medium class, four neurons are pooled together, which is the so-called ‘volley principle’), and the probability $p$ that the instantaneous frequency $(1/\text{ISI})$ coincides with the frequency of the perceived pitch $f_p = 212.5$ Hz as a function of the noise strength $\sigma$. The
clear peak of $p$ at a non-zero location highlights the essential role played by the noise in the system and exemplifies a ‘real-world’ case of stochastic resonance. The mechanism is detailed in Fig. 3.7, showing the ISI distribution for different levels of noise. For the high spontaneous rate neuron, a vanishing amount of noise leads to a locking to the modulation frequency (200 Hz) and renders the neuron incapable of resolving the fine-structure within the envelope. Only when a certain amount of noise is added, the neuron can correctly transduce the pitch (212.5 Hz). For the low-medium spontaneous class of neurons, the stochastic resonance has a different origin. Without noise, the neurons do not fire at all (threshold not crossed). Once the firing is enabled by the addition of noise, they however immediately start to fire at the correct pitch frequency (this is the more classical case of stochastic resonance).

We can now test whether the correct pitch transmission is also possible for varying shifts $\Delta f$, and check whether also Smoorenburg’s second pitch-shifts are correctly transduced. This is exhibited in Fig. 3.8, and confirms the assertion that the essential factors influencing the
3.1. PITCH

perceived pitch lie within the cochlea, and not in the neural system.

Figure 3.8: Perceived pitch from the peak in the ISI-histogram of high-spontaneous auditory nerve neurons (full dots; solid lines are linear fits). a) Three-tone signals (600 + ∆f, 800 + ∆f, 1000 + ∆f Hz) with varying ∆f (noise level σ = 0.07); dotted line indicates de Boer’s formula for k’ = 4. b) Two-tone stimulation f_2 = f_1 + 200 Hz (input amplitude A_0 = −25 dB_1V, scaling factor B = 3.0, cochlea output at f_ch = 622 Hz). Crosses: Smoorenburg’s psychoacoustic data [156]. Inset: ISI-histogram for f_1 = 900 Hz exhibits pitch ambiguity in terms of a doubly-humped distribution.

3.1.4 Discussion

The previous investigations have exhibited how the perceived pitch f_p is shaped by the cochlea and its nonlinearities, and how pitch can be successfully extracted by the auditory nerve. From there, pitch travels onwards toward the cortex, and does not seem to be essentially altered in any further way (as the comparisons between the cochlea and the psychoacoustical experiments suggest). The only thing we expect on the way to cortex is some mechanism which maps the most frequent ISIs (corresponding to 1/f_p) onto a corresponding place on a cortical ‘pitch map’ (assuming that Bendor and Wang’s observation of such a pitch map hold [10]).

One may then wonder what part in pitch perception will finally be left to the brain? In the presence of complex sound inputs, the complex interactions in the cochlea may lead to several possibilities for where to extract the pitch. This multivaluedness of the pitch is even exhibited for relatively simple inputs, and would be even stronger in the presence of complex sound mixtures. At the end of the cochlear periphery, the phenomenon is reflected by multi-peak spike distributions, for one single physical condition (as exhibited in the inset in Fig. 3.8b). The existence of pitch could then reflect the necessity of an overall characterization of complex cochlear activation (cf. chapter 2 and [95]). Moreover, cascades of feedback loops from the cortex to the cochlea are known to provide efferent input to the cochlea, which modifies the Hopf parameters. Given a mixture of sounds, this permits focusing on individual

---

²In our discussion, we have excluded the phenomenon of dichotic pitch or binaural pitch; this type of pitch, which is formed by information from both ears and generated somehow in the neural system, is very special and usually “extremely weak” (comment by Greenberg in [88]).
CHAPTER 3. PITCH AND THE TUNING OF THE COCHLEA

auditory objects that have a characteristic pitch of their own. This will be the topic of the next section.

3.2 Cochlear tuning

3.2.1 The medial olivocochlear (MOC) system

The physiological basis for an active tuning of the cochlea is provided by the existence of the medial olivocochlear nerve fibers, which originate in the superior olivary complex in the brain stem and innervate the cochlea in a tonotopic arrangement [91]. Fig. 3.9 provides an overview of the whole olivocochlear system (cf. [14] for a detailed discussion of MOC anatomy, and [59, 60] for reviews on MOC physiology).

![Figure 3.9: The olivocochlear system (from Guinan [59]). Left: Brainstem cross-section, exhibiting where the MOC neurons are located (blue/red reflect ipsilateral/contralateral feedback). CN: Cochlear Nucleus, LOC neurons: Lateral olivocochlear neurons. Right: The terminations of the olivocochlear fibers in the organ of Corti; the thick MOC fibers synapse on the outer hair cells.

In contrast to the effect by activation of the (here not further discussed) lateral olivocochlear (LOC) system, the MOC effect is fast. Experiments indicate that MOC neuron stimulation provokes a rapid reduction in cochlear amplification, which was first indirectly discovered by Galambos (who looked at the activity of the auditory nerve) [43], and then corroborated by measurements directly assessing cochlear mechanics [115, 155]. Despite the efforts to study the MOC system at different levels and from different points of view (cf. e.g. [100, 48] for a biochemical approach, or [4, 104] for general results on the cochlear amplifier), a full biophysical description of how efferent MOC activity eventually leads to a reduction in cochlear amplification is still missing.

Concerning the function of the MOC system in the general context of hearing, basically three roles have been suggested [59]: it may protect the ear from acoustic overstimulation [129], enhance the detection of signals in noise [82], and help in selective auditory attention [128, 47, 99]. Of these three points, the last one is the most interesting and the one we are going to focus on. In this context, especially a clinical study by Scharf et al. [147] is worth mentioning. In patients whose olivocochlear bundle had been cut due to health reasons (Ménière’s disease), the main effect of the removal of MOC connections was that patients detected signals at unexpected frequencies better than before, implying an impaired attentional ability in the frequency domain. Otherwise, their hearing seemed to be normal, which strongly suggests that
3.2. COCHLEAR TUNING

MOC activity is not a prerequisite for proper cochlear functioning, but rather an additional ‘tool’ to suppress out undesired/unexpected signals.

---

![Diagram of cochlear system](image)

Figure 3.10: Cochlea with efferent feedback. a) Mammalian example, b) our model, where efferent inhibitory input to the cochlea’s outer hair cells is realized by tuning the sectional Hopf parameters away from criticality (“µ-tuning”).

On the modeling side, several studies have taken reverse-engineered models of the auditory periphery as their starting point to study the MOC system (circumventing the problems and the complexity that emerge from a more concise biophysical description) [77, 21, 40, 13, 22, 46, 113]. In these approaches, a close-to-biology implementation of efferent control of the cochlea is inherently difficult, and in most approaches, the frequency specificity of efferent innervation is not considered. Using the Hopf cochlea as our basis, we now provide a quantitative test to determine to what extent the MOC system can contribute to the ability of mammals to follow signals of interest within sound mixtures. In the Hopf cochlea, µ provides the natural parameter for efferent gain control. The correspondence between our model of efferent gain control and biology is exhibited in Fig. 3.10.

### 3.2.2 Simple MOC effects

In a first step, we compare the effect of detuning single Hopf parameters with experimental data resulting from electrical stimulation of the MOC fibers. Over a wide range of input levels, the response at the intermediate section 5 of the Hopf cochlea is shown in Fig. 3.11a). To model frequency-specific MOC control, we compare flat tuning (where all µ parameters have the same value, µ = −0.1) to the case where the section’s µ is pushed further away to µ₅ = −1.0. A substantial effect is observed, even for frequencies well below a section’s CF. It was found in biological efferent stimulations [145] that the cochlear level shifts [61] were largest for stimulus frequencies below CF at low to moderate sound levels, whereas for frequencies above CF, the salient level shifts were obtained at higher levels. The corresponding modeling results are exhibited in Fig. 3.11b) and fully consistent with these observations.
CHAPTER 3. PITCH AND THE TUNING OF THE COCHLEA

Figure 3.11: The basic effects of frequency-specific cochlear tuning. a) Gain isointensity curves at section 5 ($f_{ch} = 1.42$ kHz) without (solid lines) and with (dashed lines) MOC efferent input. Flat tuning ($\mu = -0.1$) for the non-MOC case. MOC stimulation is implemented by shifting $\mu_5$ to $\mu_5 = -1.0$ (curves for -80 and -100 dB SPL collapse). b) Corresponding phase shift at section 5. Phase delays result for frequencies below CF, phase leads above CF. c) BM level shifts (arrows) at section 2, $f_{ch} = 16.99$ kHz, when stimulated by 16 and 19 kHz (left and right) pure tones. Open circles: Flat tuning ($\mu = -0.05$). Filled circles: MOC stimulation; $\mu_2$ is shifted to $\mu_2 = -0.5$. Insets: Corresponding animal data [145].

In experiments, efferently low-frequency stimulated, high-CF auditory nerve fibers show substantial phase lags if compared to the unstimulated case ([157] auditory nerve measurements; corresponding cochlear evidence is yet missing). The phase changes that we obtain in our model for different $\mu$-values and frequencies below and above CF are shown in Fig. 3.11c). Tuning the oscillator from $\mu = -0.1$ away from bifurcation generates phase delays for frequencies below CF and phase leads for frequencies above CF, which is fully consistent with biology.

3.2.3 A pitch-based measure for focusing on sounds

After having checked that our efferent tuning replicates the properties known from the biological example, we now measure to what extent this may be helpful for extracting target signals from a mixture of sounds. Taking pitch as the guiding feature for listening suggests taking the signal autocorrelation function (ACF) as the target. In our cochlea, the most general
way is to take a summation of the signal autocorrelations over the different sections, which changes the ACF into the so-called summary autocorrelation function (SACF; in this way, we also avoid the use of a particular read-out mechanism) [111, 112, 17, 18]. The first prominent SACF peak again indicates the pitch evoked by the stimulus. However, we maintain the full normalized summary autocorrelation function (NSACF) that accounts for sound characteristics other than pitch as well (e.g., timbre). To measure how strongly a mixture of two input sounds $x, y$ is biased (“tuned”) towards signal component $x$, we use the Euclidean distance from the mixture NSACF to the normalized autocorrelation function (NACF) of the target signal $x$ divided by the Euclidean distance from the mixture NSACF to the NACF of the undesired signal $y$. The tuning error (TE) - the measure we develop to assess how close we arrive to the target - has the expression

$$TE(x, y) := \frac{||\text{Normalized}(\sum_i ACF(f_i(x + y))) - \text{NACF}(x)||_2}{||\text{Normalized}(\sum_i ACF(f_i(x + y))) - \text{NACF}(y)||_2},$$

where $f_i$ denotes the output at section $i$ of the cochlea and where the summation extends over the $N$ sections. TE values are between 0 and $\infty$. $TE = 0$ indicates a perfect focus, and a larger TE is a less perfect focus on the target signal. If one source dominates the mixture, then TE values below unity may be observed even before tuning.

### 3.2.4 A full feedback system

Whereas the amplification of a pure tone can simply be monitored via the change of the Hopf parameter of a section (“local tuning”), for complex sounds, due to emergent combination tones, the monitoring of a whole pattern of Hopf parameters is required. On the computational level, the implementation of active listening by cochlear tuning amounts to finding the patterns of $\mu$-values that minimize the target’s TE. For the following results, we used a state-of-the-art genetic algorithm [63], where we always start from a flat tuning of $\mu = -0.1$. The sounds

![Figure 3.12: Frequency spectra of the two instruments used (here at -40 dB rms-power). Left: Flute (Flöte 8’, fundamental $f_0 = C#$ at 277 Hz), right: Cornett (Zinke 8’, same fundamental). The cornett’s frequency spectrum contains strong higher harmonics (which gives it the metallic sound). Insets: waveforms (20 ms).](image-url)
CHAPTER 3. PITCH AND THE TUNING OF THE COCHLEA

are from two church organ pipes from Zurich Paulus Church (Fig. 3.12): a flute and a cornett having distinctly different sound timbres, recorded at a sampling frequency of 44.1 kHz and equalized in loudness with respect to the sum of squared wave (.wav) coefficients. The autocorrelation is performed over 300 time steps. Since the frequencies in our tuning experiments are in the range of 270 Hz to 3 kHz, we used the cochlea discretization setting B with 20 sections covering a range from 220 Hz to 7.040 kHz. From time series containing 8192 measurements (185.8 ms) at each section, we discard the first 4096 elements to avoid, for simplicity, the computationally more demanding transient responses.

Stationary sounds

When we disturb a cornett target signal with a flute at high pitch (interacting with the higher harmonics of the cornett), µ-tuning not only suppresses the responses at the frequencies of the flute, it, moreover, cleans the signal from two-signal combination tones (Fig. 3.13). We emphasize that the target sound is enhanced via the suppression of undesired frequencies. For all tested pitch combinations of the two instruments, we observe a substantial decrease in TE at all input sound levels (Fig. 3.14b). This is nontrivial, as the strength of the generated combination tones depends in a nonlinear fashion on the stimulation amplitude. Moreover, the results are independent of the initial tuning (µ-level).

Dynamical sounds

A convincing tuning strategy should be able to follow sounds that change in amplitude and frequency via simple enough tuning patterns. That this is possible is not obvious in light of the combination tones and suppression mechanisms inherent in our setting. We report here on experiments where a frequency-variable sound is to be separated from a stationary disturber. The result is the emergence of remarkably simple tuning patterns (Fig. 3.14a) that provide excellent signal separation results (Fig. 3.14b,c).

3.2.5 Discussion

Regarding biological plausibility, the question remains whether feedback loops are strong enough to achieve the predicted effect and on what temporal scale the tuning would have to change. Whereas our results indicate that Hopf cochlea’s µ-control parallels the biological experiment closely, the objection may arise that the electrical stimulations generally used in the biological experiments are not too close to the natural conditions. To be able to decisively answer this question, we lack dedicated biological data. The time scale of the tuning we report here (approximately 100 ms, the fast MOC effect) is slightly above the attention escape time, but it appears to be appropriate for tuning in on a quickly changing signal. Realistically, the temporal scale of µ-tuning will be determined by the acoustic properties of the monitored signal according to which appropriate feedback loops will be chosen (SOC, other auditory nuclei, auditory cortex, neocortex, or combinations thereof). The correlation method that we use is independent of the time scale (coarse graining, subsampling, and sliding window smearing will yield a desired time scale); this is a main reason for monitoring the entire autocorrelation of the signal.

In our tuning approach, the cochlea’s SACF is guided by the target signal’s ACF. In real-world applications, such a guidance may be based on past experience (we know what an
3.2. **COCHLEAR TUNING**

Figure 3.13: TE improvement by $\mu$-tuning. a) Frequency spectrum at section 8 (CF = 1964 Hz). Blue: Flat tuning (-80 dB, target cornett $f_0 = 392$ Hz, disturber flute $f = 2216$ Hz). Cross-combination tones (CT, two explicitly labeled) between the flute fundamental $f$ and higher harmonics of the cornett. Red: Optimized tuning. $f$ (flute) and cross-combination frequencies are suppressed, leaving a harmonic series of the target (small arrows). b) Averaged TE over 13 different fundamental target frequencies (steps of 1 semitone) demonstrates input amplitude independence. Blue lines: flat tuning. Red lines: optimized $\mu$-tuning. Left panel: (full lines) target sound cornett (277 to 554 Hz), disturbing sound flute (at 277 Hz); (dashed lines) same target but flute at 2216 Hz. Right panel: same experiment but target and disturber are interchanged. TE improvements: arrows in b).

Instrument and speaker should sound like), may exploit particular information acquired at the beginning of the listening process, or may be a self-enhancing process. We often invest considerable effort “tuning in to” a target sound before we are able to follow it. Even without a biophysical justification of all modeling details, the efferent tuning of the cochlea could be of great technological interest. Humanlike sound separation in cocktail-party environments and attending to a selected sound source are key abilities for artificial intelligence and for robotics [65]. By starting the sound separation and selection process already in the cochlea, the effort in dealing with a breadth of combination products among undesired and desired components is drastically reduced.
3.3 Conclusions

In this thesis, we have investigated the sound processing in the mammalian cochlea from a mesoscopic perspective, addressing the specific characteristics of the system’s response in a detailed qualitative and quantitative manner. At the core of the problem is the necessity of the cochlea to be essentially nonlinear, which leads to the generation of combination tones that systematically influence all subsequent processing steps including the perception of pitch. Whereas most of the quantitative results in this thesis are specific for the mammalian cochlea and its unique design, many of the underlying principles will be transferable to other biological systems. Without much hesitation, it can be said that all interesting biological systems are nonlinear, given the fact that even a model of a stable oscillation requires a nonlinear equation. Furthermore, it will most often be the case that the system is not only nonlinear as a whole, but that even its isolated components are nonlinear themselves. In particular, this holds true for all models in neuroscience when we require the single neurons to be able to exhibit the biologically observed fundamental types of firing patterns. In view of the composed system, such local nonlinearities then also shape the macroscopic observables of the system (see e.g. [39] for the example of frequency synchronization in neural networks).

The specific properties of the relationship between the microscopic and the macroscopic level depend on several factors such as the order of the local nonlinearity, the topology or architecture of the system, and the type of coupling. In a first step, the type of local nonlinearity determines the response properties of a single node, which includes characteristics such as the class of interaction products (the cubic combination tones in the Hopf cochlea). Which type of nonlinearity is implemented will mainly depend on the task the network node is expected perform: in the cochlea, this is the amplification of weak sounds, which leads to the description with a Hopf equation, being of cubic order. Given the dynamics of a single node, the topology of the system together with the coupling then prescribes how the nodes talk to each other. In the cochlea, the specific feed-forward design and the associated low-pass filtering leads to the propagation and amplification of the low-frequency combination tones and the gradual disappearance of higher frequencies. In other biological networks, this may be different, due to a different order of nonlinearity, a different coupling, or a different topology. Nevertheless, we expect processes like the generation and propagation of interaction products to be present in a generic way. These effects may however not always be as visible and accessible as in the mammalian cochlea, with the consequence that they may often be overlooked.

From the perspective of complex network theory, the class of networks represented by the cochlear networks seems to be a quite general and to-date not yet well-investigated system class. The topology of the system is at first sight a simple feed-forward connectivity, where each section impacts on every subsequent section. The effective topology, however, is input-specific, and can range from very simple to very complex. Moreover, due to the location in parameter space below the bifurcation point, the system is silent if there is no input (apart from noise). In this sense, the mammalian cochlea (and with it probably many other biological systems) are quite the opposite of the well-studied networks of self-sustained oscillators, where the topology is a priori defined and the nodes oscillate even in the absence of input. To explore this class of networks with an input-driven topology in more detail, it could be interesting to study the networks and the associated avalanche characteristics under different assumptions than those prescribed by the design of the mammalian cochlea. This could be implemented
e.g. by letting go of the feed-forward restriction, or by choosing different types of input and stimulation paradigms.

One further important aspect of cochlear networks is that the effective topology is not simply accessible through an observation of the system’s output. The correct understanding of the system necessitates the knowledge of the generating mechanism, which in the case of the Hopf cochlea is provided by the Hopf equation. This puts a warning sign against a too simple extraction of so-called ‘functional networks’ from timeseries data. In the context of the Hopf cochlea, a simple thresholding of e.g. the correlations of the output (as is often done when electrophysical recordings from neural networks are analyzed) would yield entirely misleading results. This suggests the conclusion that it is essential to try to understand the main dynamical principles underlying the system’s components first, and only then draw conclusions about higher-order and more implicit observables.

Finally, biological systems are rarely isolated. Even if a specific system is understood on all possible levels and in all possible aspects, it will be necessary to study the system in the context of the system’s neighbors and the respective interaction between them. In the cochlea, a feedback is enabled through the connections of the olivocochlear system, the activity of which can push parts of the systems towards a more quiescent behavior. Bearing in mind the striking importance of the $\mu$-parameter in the context of hearing, we expect that also in many other systems, the parameter controlling the excitability of the system’s nodes will be the main target of feedback. Yet, given the nonlinear nature of the system, it will in general not be straight-forward to predict the eventual effect of such feedback. By a change of the excitability of only a few oscillators, the system’s response may change as a whole. This seems to be a drawback at first sight, but it must not be so: targeting the right nodes, controlling the system may become much simpler than expected. This is the case for the Hopf cochlea, and it will most likely also be the case for many other systems.

To conclude, one may say that the understanding of the mammalian cochlea and biological information processing is in general still far from complete. We are currently looking at some of the above mentioned generalizations and extensions of coupled Hopf-type systems and have already seen a variety of interesting effects [53]. The natural questions arising are: Are there prototypical network topologies enhancing Hopf-type neural network information processing? What properties would these networks have? The answers will hopefully be provided by future work.
Bibliography


Curriculum Vitae

Florian Gomez
Gubelstrasse 50
8050 Zürich
Born on 10.11.1987 in Thun, BE
Citizen of Langrickenbach, TG

2011 - 2015  PhD at the Institute of Neuroinformatics,
             University of Zurich and ETH Zurich,
             and the Department of Physics, ETH Zurich.
2006 - 2011  Master of Science in Physics, ETH Zurich.
Fall 2009    Exchange Semester, The Chinese University of Hong Kong.
2003 - 2006  Matura, Gymnasium Thun-Schadau.
List of Publications

Journal Articles


Selected Conference Proceedings

