Diss. ETH No. 21204

Sparse Coding and Its Application to Modeling the Zebra Finch's Auditory System

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
FLORIAN BLÄTTLER
Dipl. Phys. ETH

born June 21, 1978
citizen of Hergiswil NW

accepted on the recommendation of
Prof. Dr. Richard H.R. Hahnloser, examiner
Prof. Dr. Walter Senn, co-examiner

2013

My life's course is guided
Decided by limits drawn
On charts of my past ways
And pathways since I was born.

Pictures at an Exhibition: Promenade
Greg Lake and Richard Fraser

Acknowledgements

First and above all, I have to thank the giants of the past: the great philosophers and scientist of our history. I shall look up to them and bow humbly to them. They privileged me by putting me on their shoulders and they let me see what they already saw long time ago. None of my work would have been possible without their legacy. But most of what they taught us has become common knowledge and is no longer cited separately and they are no longer mentioned personally. Nevertheless, we should never forget to thank each one of them!

I want to express my gratitude to my supervisor and mentor Richard Hahnloser. He convinced me to take up my doctoral studies and guided me through this process with his broad knowledge and his intuition. Further I want to thank my co-examiner Walter Senn. His positive attitude towards my work and his discerning eye were an enrichment.

My fellow physicist Alexandros Guekos should be mentioned and thanked specially, not only for proofreading the dissertation regarding both, language and mathematics. But he should also be thanked for providing me new insights into my work from his view throughout my whole studies.

A lot of friends from our institute I have to give sincere thanks: Andreas Steimer who was a good friend to me and who was always challenging me and my ideas. During our coffee breaks I had more insights into science and the world than during any lecture or while reading, thinking, and working at my desk. Alessandro Canopoli and Moritz Kirschmann who helped me by proofreading parts of my dissertation and preparing my defense. Joshua Herbst, Georg Keller and Alexei Vyssotski who provided me with song data. Janie Ondracek whom I could deliver simulation data and who compared it neuronal data of the zebra finch. Michael Graber, Daniele

Oberti and Claude Wang who I had many fruitful discussions with. Many more, especially the undergraduate students I had projects with and the people sitting in the same open-plan office with me. I know, it is not easy.

And last but not least I should thank Aniela Kremser, for always believing in me and my work, even when I did not.

Abstract

In the first part of this thesis I discuss characterizations of neuronal functionalities especially in response to sensory stimuli, followed by an overview over the knowledge about the zebra finch's auditory system as far as it is known today.

The main part of the thesis is based on and partly reproduced[1] from our publication (Blättler and Hahnloser, 2011). I present a new nonsymmetric sparse coding algorithm and its application to modeling the zebra finch's neuronal activity in response to auditory stimuli. In contrast to other, symmetric sparse coding algorithms it is adapted to neuronal modeling as biological neurons themselves are only able to relay unsigned messages (action potential). However, models based on sparse coding schemes have successfully been applied in the past to model low-level sensory systems, mainly the primary visual cortex. But whether such models will be successful in explaining the more complex behavior of neurons in higher sensory brain areas is unknown. I show that applying our nonsymmetric sparse coding algorithm on zebra finch vocalizations we are able not only to model neuronal behavior in low-level brain areas such as Field L, but also in high-level areas such as HVC (used as a proper name). In our model one single parameter controls the transition between these behaviors: the firing threshold of the neurons.

In the last part of the thesis four possible applications in machine learning based on our nonsymmetric sparse coding algorithm and inspired by the zebra finch's auditory system will be presented: The first application is a simple method to automatically identify sound files containing subsong. As a second application a direct way to track development of song learning

---

[1] sections 4.3, 5.2, 5.3, 6.3, and chapter 7

during development will be shown. The third application will demonstrate the possibility of smart noise suppression and as a fourth application an approximation algorithm for underdetermined blind source separation will be presented.

In a nutshell, I present a new algorithm that gives new insight into sensory processing of the brain and could serve as a tool for machine learning algorithms.

Zusammenfassung

Im ersten Teil dieser Dissertation diskutiere ich quantitative Beschreibungen neuronaler Funktionen, im Speziellen als Antwort auf sensorische Stimuli, gefolgt von einer Übersicht des Wissen über das auditorische System von Zebrafinken, wie es zum heutigen Tag bekannt ist.

Der Hauptteil dieser Dissertation basiert auf und ist teilweise übertragen[2] von unserer Publikation (Blättler and Hahnloser, 2011). Ich präsentiere einen neuen nichtsymmetrischen Sparse-Coding-Algorithmus und seine Anwendung zur Modellierung neuronaler Aktivität im Zebrafinken als Antwort auf auditorische Stimuli. Im Gegensatz zu anderen, symmetrischen Sparse-Coding-Algorithmen ist ein solcher für neuronale Modellierung geeignet, da biologische Neuronen gleichfalls nur vorzeichenlose Botschaften übermitteln (Aktionspotential). Nichtsdestotrotz wurden Modelle, welche auf Sparse-Coding-Algorithmen basieren, erfolgreich angewendet, um primäre sensorische Systeme zu modellieren, allen voran den primären visuellen Cortex. Ob solche Modelle auch fähig sind, das komplexere Verhalten von Neuronen in höheren sensorischen Hirnarealen zu erklären, ist unbekannt. Ich zeige, dass wenn wir unseren nichtsymmetrischen Sparse-Coding-Algorithmus auf die Vokalisierung von Zebrafinken anwenden, wir das Verhalten von Neuronen nicht nur in tieferen Hirnarealen wie Field L modellieren können, sondern auch in höheren Arealen wie HVC (Eigenname). In unserem Modell wird der Übergang durch einen einzigen Parameter kontrolliert: vom Schwellenwert, ab welchem die Neuronen feuern.

Im letzten Teil der Dissertation werden vier mögliche Anwendungen des nichtsymmetrischen Sparse-Coding-Algorithmus im Bereich des Machine-

---

[2] Unterkapitel 4.3, 5.2, 5.3, 6.3, und Kapitel 7

Learnings vorgestellt, welche inspiriert sind vom auditorischen System der Zebrafinken: Die erste Anwendung ist eine einfache Methode, um automatisch Sounddateien zu identifizieren, welche Subsong enthalten. Als zweite Anwendung wird eine Möglichkeit gezeigt, um das Lernen des Gesangs während der Entwicklung zu verfolgen. Die dritte Anwedung zeigt die Möglichkeit, intelligent Lärm zu unterdrücken, und die vierte Anwendung ist ein Approximations-Algorithmus für die unterbestimmte, blinde Quellentrennung.

Zusammengefasst präsentiere ich einen neuen Algorithmus, welcher neue Einsichten gibt in die Verarbeitung sensorischer Daten im Hirn und welcher als Werkzeug für Machine-Learning-Algorithmen dienen kann.

Contents

Acknowledgements i

Abstract iii

Zusammenfassung v

1. Introduction 5

1.1 What Is the Goal of Sensory Coding 6

1.2 The Zebra Finch 8

1.2.1 The Vocalization of a Zebra Finch 9

2. Characterization of a Neuronal Sensory System 15

2.1 Response Strength, z-Score and Selectivity 15

2.2 Receptive Fields 17

2.2.1 Time-Frequency Representation of Sound 17

2.2.2 Receptive Field Estimation 22

2.2.3 Spectral-Temporal Receptive Fields 24

2.2.4 Ensemble Modulation Transfer Function 25

3. The Auditory System of Songbirds 27
3.1 From the Mechanosensory System to the Thalamus 29
3.1.1 Middle Ear 29
3.1.2 Cochlear Duct 30
3.1.3 Cochlear Nerve 30
3.1.4 Hindbrain Nuclei 31
3.1.5 Midbrain - MLd 33
3.1.6 Thalamus - Nucleus Ovoidalis 34
3.2 Primary Auditory Cortical Nucleus - Field L 35
3.3 Secondary Auditory Cortical Nuclei 38
3.3.1 NCM 38
3.3.2 CM 39
3.4 Tertiary Auditory (Premotor) Cortical Nuclei 41
3.4.1 NIf 41
3.4.2 HVC 42
3.4.3 Av 46
3.5 Auditory Responses in Other Areas 46
3.5.1 Premotor Nuclei and the Anterior Forbrain Pathway 46
3.5.2 HVC Shelf and RA Cup 48
3.5.3 MMAN and VTA 48

4. A New Nonsymmetric Sparse Coding Algorithm 51
4.1 Independent Component Analysis 52
4.1.1 Contrast Function 53
4.1.2 Optimization Algorithms 54
4.2 Sparse Coding 56
4.3 A New Nonsymmetric Sparse Coding Algorithm 58

4.3.1 Zero-Threshold Algorithm 61
4.3.2 Non-Zero Threshold 65
4.3.3 Overcomplete Representation 66
4.3.4 Low-Density Receptive Fields 67
4.3.5 Inclusion of Temporal Features 68
4.3.6 Reconstruction Error and Decoding 68

5. Sensory Modeling Using Nonsymmetric Sparse Coding . 73
5.1 An Overview over Sensory Modeling 73
5.2 The Model 76
5.3 Results 78
5.3.1 Spectral-Temporal Receptive Fields and Ensemble Modulation Transfer Functions 78
5.3.2 Synaptic Currents and Firing Rates 86
5.3.3 What Has Been Encoded 89
5.3.4 Selectivity and Sparseness 92
5.3.5 Multilayer Networks 95

6. Nonsymmetric Sparse Coding as a Computational Tool 97
6.1 Sparse Coding Used for Subsong Detection 98
6.1.1 Training of the Detector 98
6.2 Tracking Song Development 102
6.3 Smart Noise Suppression by Selective Neuron Exclusion . . 106
6.4 Underdetermined Blind Source Separation of Zebra Finch Songs 107

7. Discussion 113
7.1 A New Model? 113

7.1.1 Song Selective Neurons in the Auditory Forebrain . 114
7.1.2 STRFs and Their Relation to Spike Responses . . . 115
7.1.3 Function of Selectivity Reversal 116
7.1.4 Applicability to Other Sensory Modalities 117
7.1.5 HVC and Song Learning 118
7.1.6 The Algorithm from an Engineering Viewpoint . . 119
7.2 Conclusion 120

Appendix A. List of Variables and Functions 121

Appendix B. List of Abbreviations 127

Bibliography 131

Chapter 1

Introduction

> If the brain were so simple we could understand it, we would be so simple we couldn't.
>
> Lyall Watson

Every moment of our life millions of rods and cones on our retina tell us what the world around us looks like (Jonas et al., 1992), while thousands of hair cells in our cochlea tell us how it sounds (lehlov et al., 1987). We are probably not aware all the time of the olfactory receptor neurons and taste buds. But we will be the next time we sit down and eat. And there is more, somatosensory receptors send one million bit of information each second to your central nervous system (Schmidt and Altner, 1978). But regardless of these innumerable, atomically small particles of information impinging on all our sensory systems we normally only perceive one cohesive environment. All our senses have limited resolution in all dimensions, be it time, space, frequency, or any other. But we never get the impression of a grainy or jerky environment. And with a retina containing only three different types of cones, each with its own absorption spectrum (centered around red, green, and blue), we are able to distinguish roughly 10 million different colors (Judd and Wyszecki, 1975).

Another astonishing feature of our perceptive system is its failure tolerance. Defects or damages of the sensory systems often remain unnoticed, unless directly verified. Examples are the blind spot of the eye (Durgin, 1995) or the reduced range in aging humans.

These three features, cohesion, continuity, and failure tolerance, all are rendered possible by highly redundant sensory information, be it in time, space, or any other dimension. If we would live in a world made of independent noise, such features would be impossible. Or in simple terms, all this robustness of perception boils down to the fact that our brain stores a continuously updated model of our environment and projects all sensation onto to this model (Fiser et al., 2010).[1] Anyone can get an idea of this internal model by examining stimuli that do not fit the model, everyone has encountered such stimuli: visual and auditory illusions (Eagleman, 2001).

1.1. What Is the Goal of Sensory Coding

Knowledge about sensory coding of single neurons has been gathered for several modalities and for different species (Cariani, 2001).

However, the knowledge about the operation performed by a single or few subunits, even if all subunits are akin, does not explain the function of a entire system. If we want to know more about a system - with access only to subunit operation - we have to assume a possible goal of the system. If we know the goal, we can search the space of functions for the ones leading to this goal. And of these functions one now has to choose the ones that fit the operation of the subunits.

On a first, swift glance one could say that sensory coding as a whole has to subserve survival and reproduction. This approach is not as useless as it might seem, it gives as some rough ideas.

In different species the nervous system has developed to completely different levels. Some animals like the jellyfish do not posses a brain, some like the sponge completely lack a nervous system, whereas higher vertebrates developed brains embodying up to dozens of billions of neurons (Herculano-Houzel, 2010). What are the advantages for developing smaller or bigger brains? The most dominant downside of a big brain is its high demand of energy resources. The adult human brain consumes roughly

---

[1] One could argue that this sentence is an endorsement of indirect perception and therefore controversial. However, as I will not discuss conscious awareness of the environment, the sentence is unaffected by this controversy.

20% of the body's total energy, in children even up to 50% (Kennedy and Sokoloff, 1957), while most of the vertebrates' brains demand only a single-digit percentage (Mink et al., 1981). This high energetic cost is therefore thought as the upper limit of brain size, as it reduces the available energy for other indispensable organs such as the digestive tract (Aiello and Wheeler, 1995). This limitation of energy consumption for the brain is intrinsically implying a limitation of the sensory systems and sensory coding.

Brains allow animals to learn from success and failure and adapt their behavior accordingly (van der Helden et al., 2010). Such abilities highly facilitate survival rates. In the case of females choosing their mating partner they have to evaluate the fitness of possible partners in order to optimize the survival rate of their offsprings, thus of their genes. And for animals in which part of the courtship is learned, males have to be able to refine their strategies to outperform rivals.[2] And as a third exemplary benefit some of the brained animals are able to plan future actions. Several mammals and birds stash food (Smith and Reichman, 1984), apes are able to prepare tools for future use (Mulcahy and Call, 2006), and humans try to plan most of their lives in advance. Whether this planing is consciously made by the animal or just an outcome of a genetically encoded program has no bearing on our argument.

In all the discussed cases there are some decisions and/or actions performed. These decisions and actions are based on current sensory input (choosing of mate, notion of success or failure, memorizing stash) or on memorized sensory inputs (stash retrieval, tool preparation). Therefore to enable the animal to perform the optimal action the sensory systems have to acquire and (pre-)process all *relevant* information about the environment[3] for later stages of processing.

---

[2] For a discussion about sensory systems, courtship, and their evolution, see Ryan (1990) and Endler and Basolo (1998)

[3] The term 'environment' includes the animal itself as the most important constituent

1.2. The Zebra Finch

The zebra finch (Taeniopygia guttata) belongs to the family of the sparrows (Passeridae), subfamily Estrildidae. Its natural habitat are the Lesser Sundas (eastern Indonesia) and the Australian continent. The lesser Sundas zebra finch and the Australian zebra finch both form an own subspecies, the Taeniopygia guttata guttata and the Taeniopygia guttata castanotis. Even though there is a dimorphism between the two subspecies, most research results will not state which subspecies was subject, but because they are much more often kept as cage birds one can assume them to be Australian zebra finches.[4]

On the Australian continent the zebra finch is omnipresent. In all four climatic regions of Australia (tropical, subtropical, transitional zone, and warm temperate (Heinrich Walter, 1975)) and in 16 of 18 faunal regions (Blakers et al., 1984) zebra finches were found breeding. The only regions avoided by zebra finches are forests. Their preferred fauna consist of bushes and single trees to build nests on and vast areas of grass to feed upon the seeds. Most of these regions are arid, but differ greatly in temperature. While some regions have constant high temperature throughout the year, others show seasonal changes. The zebra finch is a highly robust bird that will survive in high temperature but is also known to breed down to 4 °C (Zann, 1996). Essential for survival is a sufficient supply of grass seeds. Wild zebra finches drink roughly 1/3 of their body weight per day, depending on temperature. However they would survive for a undetermined amount of time in a conditioned laboratory environment even when water deprived and fed with dry seeds (Cade et al., 1965; Lee and Schmidt-Nielsen, 1971; Sossinka, 1972).

Zebra finch breed also in captivity. Even tough zebra finches are highly monogamic birds they will reproduce with any new partner as soon as the old partner disappears.

Since the mid-19th century the zebra finch is one of the most popular cage birds in Europe, and because of their exalted breeding no new birds needed to be imported since the beginning of the 20th century. People love the zebra finch for his plumage and song.

---

[4] If not mentioned otherwise this thesis discusses the Australian zebra finch

The zebra finch's song repertoire and breeding ability made the bird a model animal in (neuro-)biology. It appeared in the focus of research 1959 when Immelmann published his doctoral thesis, the first scientific study on the zebra finch (Immelmann, 1959).

1.2.1. The Vocalization of a Zebra Finch

Vocalizations of songbirds fall in two categories: calls and songs. Calls are short, single tones without any gaps in between. Calls are dominantly innate and produced by both sexes. Songs consist of any number of tones and are sung only by male birds. The song of an adult bird is the result of learning process.

The zebra finch produces a series of calls. The most prominent call is the 'short call' or 'tet'. These calls are gentle and of very short duration, roughly 50 ms. Zebra finches emit them regularly while moving, in states of excitement, or total isolation. But due to its softness 'short calls' can only be heard by nearby birds.

The 'long call or 'distance call' is the loudest call of the zebra finch and serves multiple purposes. It is used for greeting conspecific birds, especially family members, but also new flock members. It is regularly exchanged during flight, but also part of the mating ritual and emitted while feeding hatchlings. When a bird loses visual contact to the other birds, it will try to reach them by 'long calls'. The 'long call' is produced by both male and female birds. However, it is sexually dimorphic. The female 'long call' is completely innate and consist in a simple harmonic stack, while the male 'long call' is learned. Every adult male zebra finch has its own individual call which he developed during adolescence. From lesion studies it is known that in the male brain the innate 'long call' is overridden by the song production circuitry (Simpson and Vicario, 1990). When this circuit is lesioned, male zebra finches will produce the same call as females. Cross-fostering experiments with Bengalese finches showed that the male 'long call' is actually learned from its parents, as the zebra finches produced calls similar to their foster-father's (Zann, 1985).

The third regular call is the 'medium call'. It is shorter and fainter than the 'long call', but longer and louder than the 'short call'. Wild zebra finches use it mainly to indicate a take-off or a landing. When held in isolation, 'medium calls' will be the most common call of the zebra finch, more often than 'long calls' or 'short calls'. There are roughly nine more calls emitted by the zebra finch, but only on specific occasions, such as begging for food, copulation, pain or warning (Zann, 1996).

The vocalization that distinguishes songbirds are their songs. In the whole animal kingdom only very few species are able the learn a vocalization apart from humans. Several birds posses this ability (parrots, songbirds, and hummingbirds) as well as few mammals (marine mammals, probably also bats and elephants) (Janik and Slater, 1997; Poole et al., 2005). In the order of primates we are most probably the only species displaying vocal learning.

A song is classified on 3 levels: The bottom level is formed by tones, syllables, or elements (the nomenclature is not standardized). Syllables are similar to calls. For the zebra finch their duration is in the order of 100 ms. Normally a syllable is defined by a minimum duration of silence in the beginning and end. However, single syllables might consist of 2 or 4 completely different sounds, sometimes called sub-syllables or tones. A categorization of the sub-syllables is given by Zann (1993). On the intermediate level is the phrase or the motive (sometimes also called song). A motive is a more or less fixed sequences of syllables and intervals, depending on the species. A zebra finch will only produce one single motive (with some errors)[5]. Typically a zebra finch motive will consist of 3 to 8 syllables, which do not have to be unique and can be repeated within the motive. On the top level finally we have the song or (song-)bout. The zebra finch song starts off with a variable number of introductory notes which are similar to the short call. The introductory notes are then followed by the motive that can be repeated a few times. Birds with more than just one motive might mix up the different motives. However, there is no evidence that the semantic information exceeds 'mate me', and conveying some information about the sexual fitness (Berwick et al., 2011). Or how Ian Anderson wrote in the introduction to 'The Secret Language Of Birds, Pt. II': "*Semantic set-aside. You with me?*" (Anderson, 2000)

However, an interesting feature is the variability of the song. Zebra finches

---

[5] in contrast to other songbird species that may have several motives in their repertoire (Devoogd et al., 1993a). This simplicity makes the zebra finch an optimal model animal

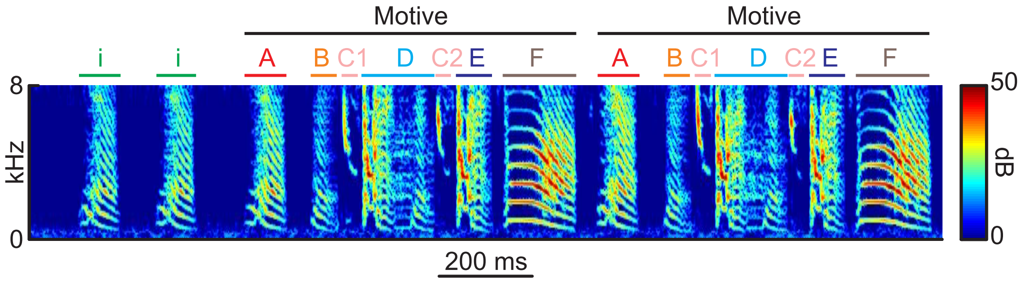


Fig. 1.1: Example of Zebra Finch Song. The song consists of two introductory notes i followed by two identical motives. The motives are rather rich for a zebra finch song and consist in 7 syllables. Syllable A actually is an identical copy of the introductory notes but it is a real syllable as it is repeated in each motive. Syllables C1 and C2 are very similar and in most cases would be classified as the same, however, there are tiny differences that are consistent over different motives. Syllable D is a very complex note consisting in 4 sub-syllables: noise-stack-stack-downsweep. Syllable E again is a fragment of syllable D, consisting in the first two sub-syllables of the latter. Syllables such as F are very often found in zebra finches: constant harmonic stacks that end with a downsweep.

will produce two different types of song depending on the situation: directed song and undirected song. The defining feature of the directed song is its very low variability in both spectral and temporal structure, has more introductory notes and more repetitions of the motif, and is sung faster (Sossinka and Böhner, 1980; Zann, 1996). It is mainly used when courting a female (Sossinka and Böhner, 1980; Zann, 1996). The undirected song in contrast shows a certain degree of variability and is mainly sung towards no particular bird (Immelmann, 1962). Fathers will sing undirected songs more often when having offsprings in the sensory phase (Ten Cate, 1982) and during nest building and breeding (Zann, 1996). There is some evidence that undirected song could play a role in song maintenance in the adult bird (Jarvis et al., 1998).

1.2.1.1. Acquisition of Song

In their early life male zebra finches will learn their song. The template of their song will be provided by the tutor. In captive breeding the chosen tutor is normally the rearing father or foster-father of the young bird (Böhner, 1990). So when raised by heterospecific birds they will try to imitate their supposed father (Immelmann, 1965). However, in studies with more natural settings or field studies the choice of tutor seems more complex. Most studies still show a preference for the father as tutor, with about every second bird copying his father (Mann and Slater, 1995; Zann, 1990).

When raised together with a (foster-)father young zebra finches are able to memorize a tutor song during the so-called sensory phase lasting more or less till day 35 (Arnold, 1975). However, this first template of a tutor song is not final. If the tutor is exchanged during their sensory-motor phase before their song finally crystallizes around day 90 to 120, when they reach sexual maturity, the birds may completely change the tutor song or mix up the two songs and come up with a hybrid song (Eales, 1985).

The development of the song is most often divided into three phases: subsong, plastic song, and crystallized song. The subsong is best compared to the babbling of children (Goldberg and Fee, 2011). It is a phase of vocal exploration around day 35 to day 45 with no identifiable repeated pattern neither temporally nor harmonically (Goldberg and Fee, 2011; Veit et al., 2011). The plastic song in contrast starts to express repeated syllables which are gradually or abruptly assimilated from the tutor song (Tchernichovski et al., 2001). Around day 120 the song does not further evolve and will stay unchanged in the birds further life unless they are seriously perturbed (Leonardo and Konishi, 1999). Between this three song types no fixed boundaries have been reported, they are stages of a fluid song-development.

A big question is song acquisition in the absence of a tutor. The male zebra finch will start singing subsong normally, but as they do not have a template they will start to repeat calls and random sounds as their song which sounds unlike natural zebra finch songs(Williams et al., 1993). An interesting study has been performed by Feher et al. (2009): They formed a whole colony with only untutored males and females. Even tough their offsprings copied the stunted songs, slight changes where introduced, so that after 3 to 4 generations the songs evolved towards wild-type songs. "*These birds behave as though they possess extensive innate knowledge about their species song*" (Marler, 1997)[6].

If a bird is given a late tutoring there are some evidences that the critical phase of song learning can be prolonged (Slater et al., 1992).

---

[6] We will discuss this statement later in chapter 6.2.

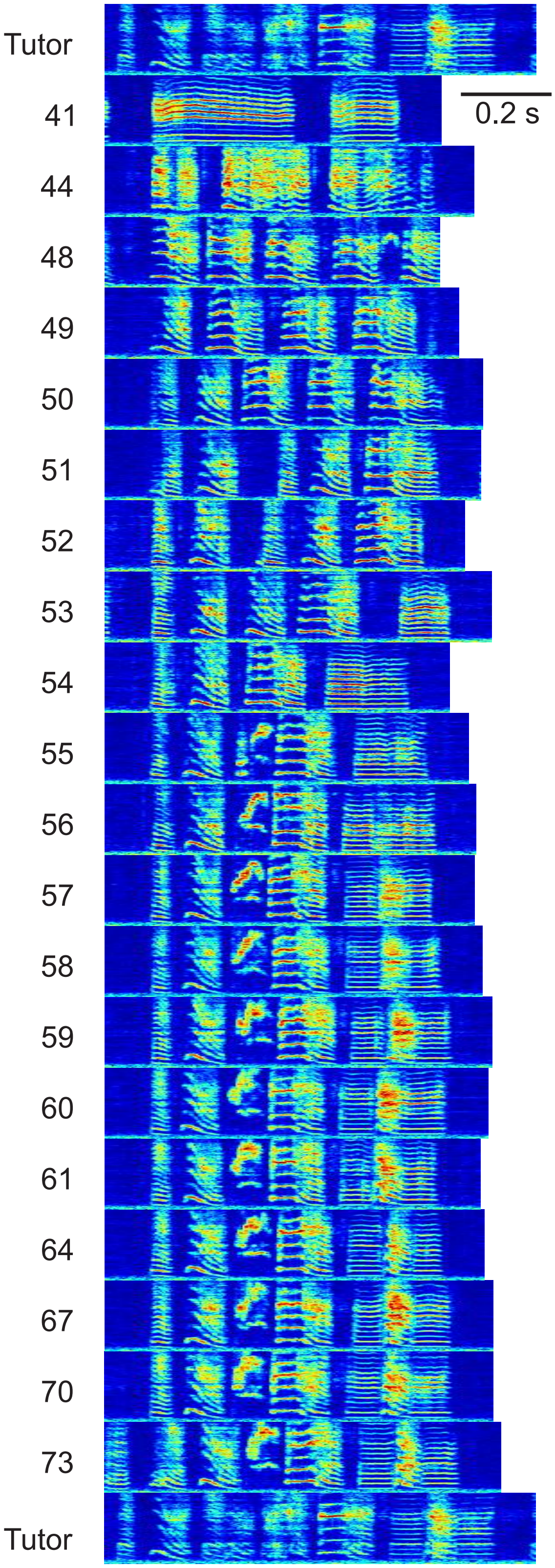


Figure 1.2: Example of song development. On top and bottom row an examples of the tutor song is shown. This bird was not exposed to male conspecifics until day 39 and was then given access to tutor song. Each time the bird pressed a button, he would be exposed to a zebra finch song (Tutor), until the daily maximum was reached. On day 44 one can see the bird producing classical subsong, where no clear spectral pattern can be seen and temporal patterns are not repeated. With day 48 the bird starts to sing plastic song and already by day 49 the second last syllable of the tutor song has been copied, but will still be further refined on later days. From day 53 on, the bird tries to copy the last syllable of the tutor song which he gradually prolongs element by element till day 57. In parallel on day 55 suddenly a new syllable appears in the center of the song which tries to mimic the noisy center syllable of the tutor song. However he will not succeed. So from day 57 on all syllables of the tutor song are copied and further refined to match the tutor song. The spectral range of the spectrograms spans from 0 to 8 kHz. The data was provided by Joshua Herbst.

Chapter 2

Characterization of a Neuronal Sensory System

> Est modus in rebus. - There is measure in all things.
>
> Horace, Sermones

If th raw data about the neural system we want to investigate is limited, and even if we were able to describe the system in every detail of its operations, we will not gain any insight to its functionality. We will have but a mere neural painting of colors and shapes, but with no identified content. We have to search for patterns and regularities and quantify them. Only with these characteristic numbers, we can put it on a level with different (sub-)systems, search for underlying principles, and compare models of the systems to the limited data available.

In this chapter I will explain the measures most often used to describe neural activity in sensory systems and I will discuss the strength and limitations of each measure.

2.1. Response Strength, z-Score and Selectivity

The measures in this section are probably the most simple and direct ways to describe the main response properties of neurons. The response strength

$RS_A = \bar{r}_A/\bar{r}_{BG}$ of a neuron to a stimulus A describes the ratio of mean firing rate $\bar{r}_A$ during stimulus presentation compared to the so called mean baseline firing rate $\bar{r}_{BG}$ of the neuron when no stimulus is presented. So a $RS_A < 1$ represents a neuron that is inhibited by stimulus A on average, while a $RS_A > 1$ represents a neuron excited by the stimulus.

The response strength tells you the sign of the change in firing rate in response to the stimulus. However, it says nothing about the reliability of the change and, given a neuron with a very low baseline firing rate $\bar{r}_{BG}$, the response strength can achieve absurdly high values. Therefor, the z-score is introduced, sometimes also called the normalized response strength. The idea is to normalize the difference in firing rate by standard deviation of the mean firing rates during different stimulus presentations and baseline,

$$z_A = \frac{\bar{r}_A - \bar{r}_{BG}}{\sqrt{\sigma_A^2 + \sigma_{BG}^2 - 2 \cdot \mathrm{covar}\left(r_A, r_{BG}\right)}}, \tag{2.1}$$

where covar (., .) denotes the covariance of the mean firing rates. Compared to the RS, the z-score tells us not only, whether a neuron boosts its firing rate in response to a stimulus or whether it is suppressed. It also tells us, how reliable the change is. A z-score of somewhere between -1 and 1 tells us that from the response of this neuron, we will most of the time not be able to tell whether or not the stimulus was presented. But when we are faced with a z-score of e.g. 10, there will never be an uncertainty about whether or not the stimulus is presented just by looking at the response of the neuron. The firing rates of baseline firing and during stimulus presentation are nonoverlappingly distributed.

The third measure from this group is the selectivity. It is very closely related to the z-score, but does not compare stimulus evoked response to baseline, but the responses to two different stimuli A and B:

$$d' = \frac{2\left(\bar{r}_A - \bar{r}_B\right)}{\sqrt{\sigma_A^2 + \sigma_B^2}}. \tag{2.2}$$

This measure was first proposed by Green and Swets (1966) and has since then found its way into neuroscience. The two differences to the z-score are the factor 2 and the missing covariance term. The factor 2 should compensate for the division by the square root of two variances and is only a matter of taste in my eyes. A bigger issue is the missing covariance which should reflect the variance that does not originate in noise or stimulus, but in global changes of the system e.g. the state of mind, the wakefulness of the subject, or simply a deterioration of your electrode. It therefore should have been included in the selectivity measure, but history has chosen not to. However, it is difficult to estimate and thus often ignored, also when calculating the z-score.

The name ”selectivity” also is misleading as a high value does not mean that a neuron is selective towards a stimulus. Nor does a zero value indicate a nonselective neuron. The value does not even reflect the preference of stimuli, as a positive value could mean a suppression of firing in response to the second stimulus, while the first stimulus will not drive firing away from baseline. Therefore sometimes the more appropriate term ”discriminability” is used, as this measure describes how well two stimuli can be told apart by looking at the mean response of this single neuron.

The main drawback of these three measures are their stationarity assumption. They all just work on mean responses which may be a good assumption using static stimuli such as pictures. However, as we will discussion in the next section, sound is seldom stationary and neural responses are not simply up- or down-regulated by stimuli, but show a high degree of inner variability. Nevertheless, this measures proved to be useful as a first description of neural behavior and help us to get a first insight into the songbird’s brain.

2.2. Receptive Fields

2.2.1. Time-Frequency Representation of Sound

Sound itself is a mechanical pressure wave in a medium such as air. This one-dimensional signal is taken up by the auditory system (see chapter 3). However, cortical (and most subcortical) auditory neurons will not respond to the rapid changes in air pressure (hearing range goes as high as 120 kHz in bats, Neuweiler (1984)). Much more will they respond to temporal and spectral features within the sounds. Already in the first step of auditory processing the sound signal will be decomposed into different frequency bands (3.1.3). In the following I will present a few computational methods to decompose a sound signal into its temporal and spectral components.

2.2.1.1. Log-Power Spectrogram

One of the most often used methods of time-frequency representations is the log-power spectrogram, often just called spectrogram. The big advantage of spectrograms is their mathematical comprehensibility and easy computability. By Fourier transform (FT) short excerpts of the waveform (Figure 2.1A) are converted into frequency domain (short-time Fourier transform, STFT). FT is energy-conservative, so the square amplitude of a single Fourier component represents exactly the amount of energy in the corresponding frequency band. Phase information is discarded[1] and only the square amplitude (power) is preserved. However, it turns out that this representation is more useful in the log-domain (in decibel).

If an event happens within such an excerpt of length δt we will be able to locate it temporally with a precision of δt and spectrally with a precision of $\delta f = \frac{1}{\delta t}$. By overlapping the excerpts, the signal will temporally be smeared, as the temporal distance between two excerpts is smaller than the temporal resolution of the signal. But a further problem arises from the fact, that the amplitudes of Fourier-series are invariant under circular permutation, i.e. the STFT of the excerpt is equivalent to an normal FT of a signal that consists in an infinite repetition of the excerpt. But the excerpts in general will not end at the same phase as they started, leading to a discontinuity in the waveform (Figure 2.1B). The FT of a jump however will lead to energies in the high frequency domain, even tough the underlying signal might be a constant low-frequency sine-wave (Figure 2.1E). In order to suppress this erroneous high frequency portion the excerpts get multiplied by a windowing function w, such as Hamming windows or Gaussian windows (Figure 2.1F), that flattens the boarders of the excerpts (Figures 2.1G and H). However, this windowing functions do not only suppress the high frequency portions in the FT, they in fact

---

[1] the original signal is recoverable from the amplitude of STFTs with temporal overlap of more than 50% (Griffin and Lim, 1984).

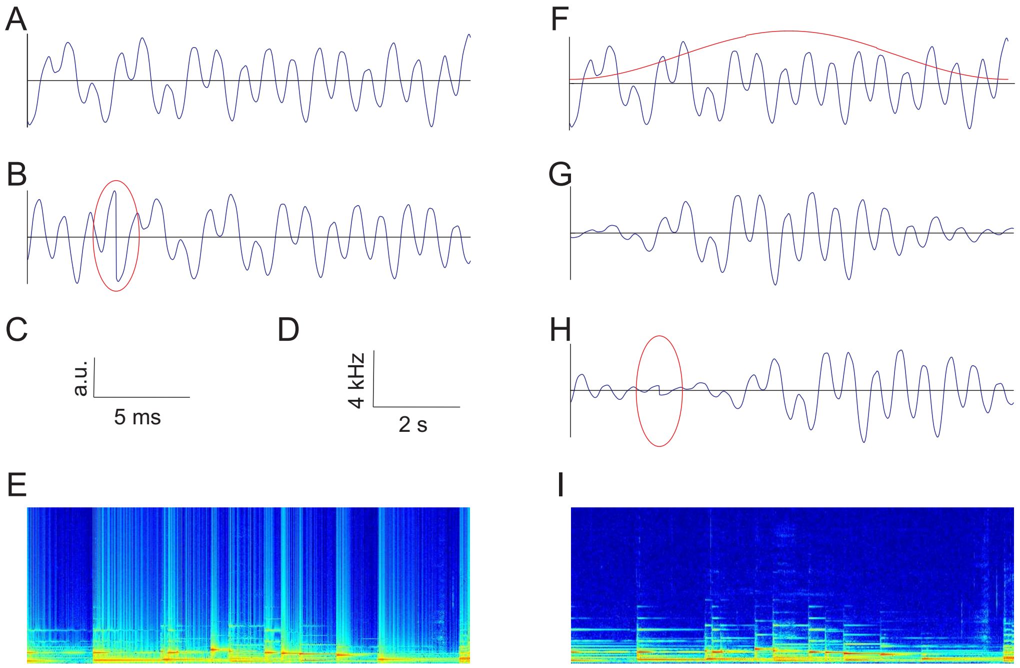


Fig. 2.1: Calculation of a spectrogram. (A) Excerpt of the pressure wave of a piano. (B) Same excerpt, but phase-shifted. In the red ellipse the waveform has a discontinuity. (C) Scale of the waveforms. The amplitude is in arbitrary units. (D) Scale of the spectrograms. (E) Spectrogram outtake of the piano piece without windowing. Due to the discontinuity depicted in B, the energy is distributed over all frequency bands and higher harmonics tend to dissolve. The logarithm of the energy is color-coded over a range 100 dB. (F) Excerpt of the waveform plus a Hamming window of the same size (in red). (G) Windowed excerpt. The windowed excerpt is created by pointwise multiplication of the excerpt with the window. (H) Phase-shifted version of the windowed excerpt. The discontinuity in the red ellipse is highly reduced, leading to less energy in the high frequency band due to it. (I) Spectrogram outtake of the piano piece with windowing. High frequencies which are not due to the piano piece itself are highly suppressed and the harmonics are clearly visible, in comparison to E. The logarithm of the energy is color-coded over a range 100 dB.

also change the resolution of the spectrogram: frequency resolution gets worse, and the signal gets smeared over adjacent frequency bands (spectral leakage)[2] and the temporal resolution gets slightly better, sharpening the spectrogram temporally.

2.2.1.2. Log-Spaced Spectrogram

When dealing with sounds, a linear frequency scale is often not preferable as perception of sounds does not depend on absolute values, but on their relative frequency[3]. Therefore a logarithmic frequency scaling seems more natural. One could imagine different ways to calculate such a spectrogram, but I want to stay as close as possible to the original idea of the spectrogram. The discrete FT can also be seen as matrix multiplication of a matrix containing complex sine waves of frequencies zero to the sampling frequency whereby the second half of the waves (above the Nyquist frequency) matches exactly the complex conjugate of the first, except for the zero frequency. If the excerpt is windowed, the windowing function can also be drawn into the matrix by windowing each complex sine wave instead of windowing the excerpt. To get a logarithmical frequency scale the complex sine waves need no longer to be linearly scaled but logarithmically. So the frequency of the first component should be $f_0 = f_{min}$ and the following frequencies $f_n = f_{min} \cdot 2^{(n/s)}$ where s is the number of frequency bands per octave. Due to the different spectral spacing of the spectrogram we are also able to adapt the temporal resolution of it. Equivalent to the normal spectrogram where the temporal resolution is $\delta t = \frac{1}{\delta f}$ we can define $\delta t_n = \frac{1}{\delta f_n} = \frac{2}{f_{n+1} - f_{n-1}} = \frac{2}{f_n \cdot \left(2^{1/s} - 2^{-1/s}\right)}$. What we see is a temporal resolution that is inverse proportional to the frequency, so we can achieve higher temporal resolution in the high frequency domain and at the same time retain the high spectral resolution in the low frequency domain. However, again the real spectral resolution is somewhat worse due to the windowing.

---

[2] The multiplication of the excerpt with the windowing function is equivalent to a convolution in the Fourier-domain. Therefore the only windowing functions w that do not smear need to satisfy $\mathcal{F}(w)(n \cdot \Delta f) = 0$ for all $n \in \mathbb{Z} \setminus \{0\}$. The above equation is only satisfied for rectangular windows.

[3] Only very few people have access to absolute pitch, Levitin and Rogers (2005).

2.2.1.3. Chirplet Transformation

A further shortcoming of the STFT is the implicit assumption that sounds will have constant pitch over the duration the window. This assumption might be valid for pianos as in the example of Figure 2.1. However, many natural sounds tend to gradually shift their pitch. Such a shift leads to a further spectral broadening of the representation as well as to the emergence of sidebands. An interesting approximation to this problem was proposed by Mann and Haykin (1991). The general idea is to take shifts into account. Good candidates for a first approximation are linear shifts and exponential shifts. Similar as in Section 2.2.1.2 FT is replaced by multiplication with a matrix F containing complex sine waves. However, the frequency $f(\tau) = \frac{\partial\phi(\tau)}{\partial\tau}$ within each wave $w(\tau) = e^{i\cdot 2\pi\cdot\phi(\tau)}$ is either going linearly from $f_c\cdot(1-\Delta)$ to $f_c\cdot(1+\Delta)$ or exponentially from $f_c\cdot e^{\Delta}$ to $f_c\cdot e^{-\Delta}$, where f_c is the center frequency[4]. So for a set of center frequencies (either linearly or logarithmically spaced) a transformation matrix $W(\Delta)$ is calculated.

The most difficult point is the choice of the correct chirp Δ. A simple way is to create a set of transformation matrices for a different Δ each and apply them on excerpts s of the sound. The correct chirp Δ for an excerpt will then be the one that minimizes the spectral broadening and the sidebands. A possible solution is the Δ that produces the smallest normalized 1-norm in amplitude space: $\Delta = \operatorname{argmin}_{\tilde{\Delta}} \frac{\left\|abs\left(W(\tilde{\Delta})\cdot s\right)\right\|_1}{\left\|W(\tilde{\Delta})\cdot s\right\|_2}$.

Similar algorithms have been proposed, using chirplet-atoms in matching pursuit (Bultan, 1999). The idea is to have an overcomplete set of chirplet-atoms with different spectral and temporal resolutions and different chirps (similar to the transformation matrices $W(\Delta)$ in the algorithm presented above). The original signal then is approximated as good as possible using a minimal amount of chirplet-atoms.

---

[4] one can imagine an infinite number of different chirplets (Mann and Haykin, 1995), however, for sound applications those two seem the most promising.

2.2.2. Receptive Field Estimation

In order to understand the functionality of sensory systems we have to describe how stimuli are encoded by neurons. In auditory systems this is often done by calculating the spectral-temporal receptive field (STRF) which is a linear approximation of the relations between stimulus and output of auditory neurons and has first been described by Aertsen and Johannesma (1981). However, the concept of receptive fields (RF) in visual system was established long before by Hartline (1938). He defines the receptive field as follows:

> *No description of the optic responses in single fibers would be complete without a description of the region of the retina which must be illuminated in order to obtain a response in any given fiber. This region will be termed the receptive field of the fiber.*

This definition however is insufficient as neurons do not only experience excitatory stimuli, but very often also depressing stimuli which could go unnoticed, if not paired with an exciting stimulus, and which are not described separately. Hubel and Wiesel (1959) therefore divided RFs into "*excitatory and inhibitory ('on' and 'off') areas*". This definition was further refined. Areas of the RF not only have a sign but also an amplitude and work as a linear filter on the stimulus.

Calculation of RF is normally done by reverse correlation. We define the RF as the linear filter h whose prediction $\tilde{r^t} = h^T * X^t$ minimizes the summed squared prediction error $F = \sum_t \left(r^t - \tilde{r^t}\right)^2$ of the neurons response r^t over a set of stimuli X^t. The stimuli X^t and the response r^t are mean-subtracted.

To optimize it, we can set the derivation of F to zero and get:

$$\frac{\partial F}{\partial h} = \sum_t -2X^t r^t + 2X^t {X^t}^T h = 0 \tag{2.3}$$

$$h = \left(\sum_t X^t {X^t}^T\right)^{-1} \cdot \sum_t X^t r^t = C_{SS}^{-1} C_{SR}. \tag{2.4}$$

C_{SS} simply describes the stimulus covariance, while C_{SR} is the cross-covariance between stimulus and response and is closely related to the spike triggered average. To estimate the RF, white noise stimuli are often chosen, as the covariance degenerates to a simple scalar factor and the RF is identical to the cross-covariance. It is a good choice for neurons with very linear response properties, but for cortical neurons white noise often leads to a poor response, making an estimation of the RF impossible (Theunissen et al., 2000; Cohen et al., 2007). White noise therefore is replaced by natural stimuli that drive the neurons well. The RF is no longer the cross-covariance itself, but it is normalized by the stimulus covariance. The problem that arises is that natural stimuli often occupy only a low dimensional subspace of all possible stimuli. This low dimensionality manifests itself by a covariance matrix which is not of full rank, or differently said, the covariance matrix has eigenvalues which are virtually zero. Calculating the inverse, dimensions with eigenvalues close to zero will be blown up monstrously, even though such dimensions most likely just represent noise. Predictions on new stimuli are destined to fail badly, as noise will be the main driving force given such model. It is a classical example of overfitting.

To avoid overfitting several methods have been proposed. Theunissen et al. (2001) used only eigenvectors of the stimulus covariance that are associated with an eigenvalue not smaller than a certain fraction of the biggest eigenvalue for the matrix inversion. To enhance this method the cross-covariance was low-pass filtered by checking the frequencies for significance using a jackknife (Theunissen et al., 2000). Alternatively a jackknife was applied on the raw RF directly to test for significant regions, without previous manipulations (Sen et al., 2001).

A more direct way would be a Tichonov-regularization of the RF by weighting it with a matrix Γ and adding it to the cost function F:

$$F = \sum_t \left(r^t - \tilde{r^t} \right)^2 + \|\Gamma h\|_2^2. \tag{2.5}$$

The open question is the choice of Γ. Often it is chosen as a constant factor $\Gamma = \mu \cdot \mathbf{I}$. However, in my experience one achieves better results by scaling the factors according to the root of the stimulus variance $\Gamma =$

$\mu\sqrt{\operatorname{diag}(C_{SS})}$, where diag (.) is a diagonal matrix with the same elements on the diagonal as the argument. The RF can still be directly calculated by

$$h = \left(C_{SS} + \Gamma\Gamma^{T}\right)^{-1} C_{SR} = \left(C_{SS} + \mu \cdot \operatorname{diag}(C_{SS})\right)^{-1} C_{SR} \qquad (2.6)$$

The optimum factor μ can be estimated by calculating the predictive power of RFs on validation data with different μ. This method is equivalent to assuming to have uncorrelated noise on the signal with a local signal to noise ratio of $1/\mu$. By experience, the optimum μ lays between 0.1 and 1, depending on the amount and quality of the data.

One of the most elegant methods in my eyes of RF estimation is maximally informative dimensions (MID) by Sharpee and Bialek (2007). While the input-output relationship of neurons described by RFs is purely linear, sometimes an additional nonlinearity is introduced (Calabrese et al., 2011). However, this methods have to make assumptions about the nonlinearity. MID on the other hand searches the stimulus-space for the stimulus dimensions (or produces a simple RF in the case of one stimulus dimension) that show the highest mutual information with the neurons output. If the output is defined by just one RF, this method will find it regardless of any nonlinearity. Drawbacks of this method are its need for data to reasonably estimate the density of the stimulus space and its high computational cost.

2.2.3. Spectral-Temporal Receptive Fields

In the visual domain most often static stimuli are used to determine the RF of neurons, such that the response r^t is only depending on stimulus X^t. However, RF estimation from static stimuli (tones) often predict the behavior of neurons inferiorly to RF estimation from (naturally) modulated stimuli (Theunissen et al., 2000; Machens et al., 2004). But also in the visual domain natural, non static stimuli yield better RF estimations (David et al., 2004).

To catch the dynamic features to which the neurons respond, I will not only look at the actual stimulus but at the stimulus history of length τ preceding the response. The response r^t is now depending on the stimulus $X^{t-\tau:t}$

which we write as a single vector and we have $C_{SS} = \sum_t X^{t-\tau:t} X^{t-\tau:tT}$ and $C_{SR} = \sum_t X^{t-\tau:t} r^t$.

An important question remaining is the representation of the stimulus. A short discussion about representation of sound in the spectral-temporal domain has been given at the beginning of this section 2.2.1. In my experience, estimations based on the simple log-power spectrogram yield results qualitatively similar to estimations based on more elaborate representations and are therefor my preferred representations. A comparison of power spectrogram, log-power spectrogram and the Lyon's model (Lyon, 1982) used for estimation of the STRF and their predictive power applied on zebra finches is given by Gill et al. (2006).

2.2.4. Ensemble Modulation Transfer Function

An interesting feature was introduced into songbird research by Singh and Theunissen (2003). The idea is to summarize the spectral-temporal features the neurons are coding for. In simple words, STRFs are calculated for auditory neurons and on these STRFs a two-dimensional FT is performed (dropping the phase). We now have a two dimensional map called Modulation Transfer Function (MTF) which tells us how many cycles per kHz and how many cycles per s (or Hz) the stimulus must have (without the phase) to optimally drive the neuron. For a single STRF this measure might not be of great value, but if we sum over the whole population, we get the ensemble MTF (eMTF). The eMTF can then directly be compared to the FT of different stimuli and we can determine which elements of the stimulus get overrepresented in the neuronal code and which elements are suppressed. However, only few studies (Theunissen et al., 2004; Woolley et al., 2005, 2009; Amin et al., 2010) estimated eMTFs in songbirds by now, as a lot of data is needed to estimate the RF for each neuron and a large number of neurons to make statements about the ensemble.

Chapter 3

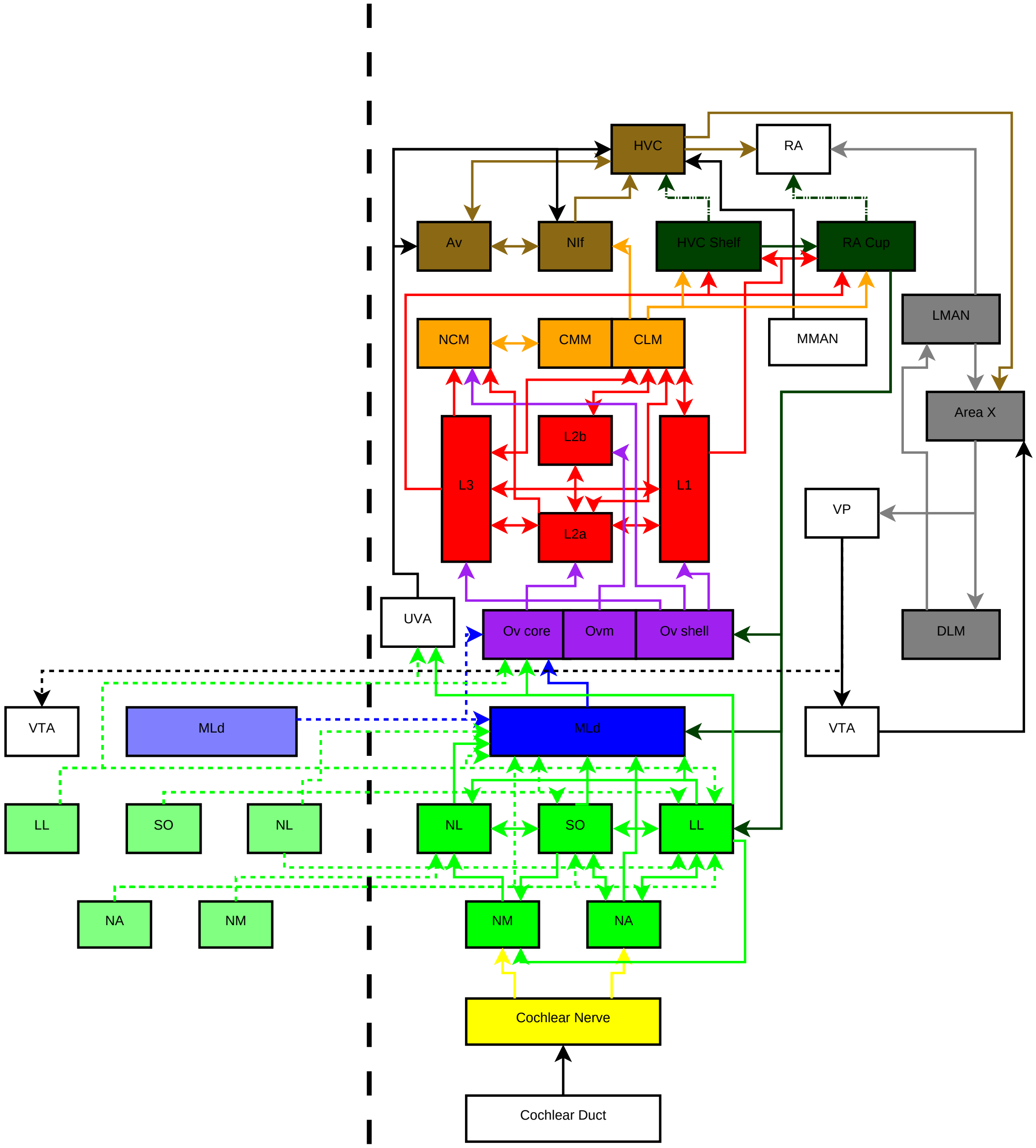
The Auditory System of Songbirds

> Everything you see or hear or experience in any way at all is specific to you. You create a universe by perceiving it, so everything in the universe you perceive is specific to you.
>
> Douglas Adams, Mostly Harmless

In this chapter I shall to some extent try to unravel the auditory system of songbirds in general and of the zebra finch in particular. We can not draw any conclusions, because - despite the overwhelming amount of published material - the knowledge about the system consists in relatively few and tiny fragments. It is like trying to see the dolphins in a 10'000-pieces jigsaw puzzle of the maritime wildlife with only a dozen pieces at hand. But by analyzing these pieces carefully, one may detect the maritime environment.

One problem which we will not discuss any further is the definition of auditory. Do you call an area auditory if it changes its state solely (but

Fig. 3.1 ***(facing page)*:** Schematic drawing of the auditory nuclei of one hemisphere and their connections. The colors of the areas correspond to the following structures: yellow - peripheral nervous system, green - hindbrain, blue - midbrain, purple - thalamus, red - primary auditory cortical nuclei, orange - secondary auditory cortical nuclei, brown - tertiary auditory cortical nuclei, dark green - shelf and cup, grey - anterior forebrain pathway, white - other (auditory) areas. Only known and verified connections are drawn. The solid lines denote ipsilateral connections and the dotted lines contralateral connections to or from the faintly painted contralateral nuclei.


HVC
RA
Av
NIf
HVC Shelf
RA Cup
NCM
CMM
CLM
MMAN
LMAN
Area X
L2b
L3
L1
L2a
VP
UVA
Ov core
Ovm
Ov shell
DLM
VTA
MLd
MLd
VTA
LL
SO
NL
NL
SO
LL
NA
NM
NM
NA
Cochlear Nerve
Cochlear Duct

always) in response to auditory stimuli and is independent of other influences? Probably no brain area would satisfy this definition. Or would you call any area auditory that changes its state in response to auditory stimuli, regardless of other influences? Then nearly any area could be called auditory. A useful definition lies somewhere in between these extreme examples. However, in this thesis I do not want to assess the definition, but rather stick to generally accepted terms.[1]

The data presented in these sections is not from the zebra finch only. Especially the first section 3.1 deals with other avian species: barn owl, chicken, sparrow, or canary. But from the evolutionary point of view, ear and hindbrain are old parts of the auditory system which did not change much, even when compared to reptiles. It is therefore safe to assume that there exists a certain homology among birds, except for size and quantities.

Naming of the different nuclei in the avian brain has been very inconsistent, but since the 2002 Nomenclature Forum there has been a standardization of the nomenclature (Reiner et al., 2004a,b,c; Jarvis et al., 2005).

3.1. From the Mechanosensory System to the Thalamus

3.1.1. Middle Ear

The middle ear of a bird has a very similar function to the mammalian one. In both animal classes the inner ear mechanically transfers sound arriving at the tympanic membrane to the oval window. In order to facilitate the transfer of sound from air to the fluid of the inner ear (a huge transition regarding the acoustic impedance, i.e. with no direct sound transfer from one medium to the other) a leverage system is used. But as the bird possesses evolution's oldest ossicle, the columella - homologous of our stapes - the lever is generated by the flexible extracolumella. This system of leverage creates the disadvantage of being a low-pass filter. Frequencies above

---

[1] To quote Whitfield (1967): "*the 'auditory cortex' is hard to define, either anatomically or functionally. While certain areas are so predominantly connected with the rest of the auditory system as to be unequivocally auditory areas, surrounding regions become less and less certainly so, as we move from the primary projection area.*"

4 kHz are cumulatively damped and the theoretic limit of transmission lies at 12 kHz. For highly specialized birds such as the barn owl evolution pushed the hearing ability up to this limit (Manley, 1981).[2]

3.1.2. Cochlear Duct

The auditory organ of the inner ear in birds is often called cochlea as for mammals. This naming might be misleading as the word 'cochlea' is derived from the Greek word *kokhlias* (snail). But the cochlear duct of birds is not coiled. Its shape resembles a straight, twisted cylinder (Schwartzkopff and Winter, 1960). The inner structure of the cochlear duct is similar to the mammalian one, consisting of the liquid-filled scala tympani and scala vestibuli, which are separated by the basilar membrane and by the tegmentum vascolosum from the scala media. The basilar papilla is connected to the basilar membrane and consists of roughly 5'800 hair cells in starlings (Gleich and Manley, 1988), less than for most mammals. Most of these hair cells belong to two populations: the tall hair cells located mainly on the neural side of the basilar papilla and the short hair cells mainly on the abneural side. A striking difference between these two populations (except for their morphology) is the complete lack of afferent innervation of the short hair cells, though they receive efferent input (Fischer, 1994). There is evidence that the function of the short hair cells is to change the mechanics of the system in order to amplify low-intensity sounds and dampen high-intensity sounds (Yates et al., 2000). Such a nonlinear amplification is a well-studied phenomenon in the mammalian cochlea (Brownell et al., 1985).[3]

3.1.3. Cochlear Nerve

In consistency with the shorter cochlear duct and the smaller number of hair cells, the number of auditory nerve fibers in birds is smaller than in mammals. The highly specialized barn owl which has roughly 31'000

---

[2] See Manley and Gleich (1992) for a more detailed description of the avian middle ear.

[3] For a good overview of the avian cochlear duct, see Köppl et al. (2000a) and Manley and Gleich (1992).

afferent auditory nerve fibers (Köppl, 1997a) is once again an exception. In songbirds this number is significantly smaller: reported numbers of auditory nerve fibers are between 6'000 in the canary (Gleich et al., 1998) and 9'000 in the starling (Köppl et al., 2000b).

Neural activity of hair cells is rarely measured at the hair cells themselves, but at the cochlear nerve. As with most vertebrates, each auditory nerve fiber shows one distinct sound frequency to which it responds at lowest sound pressure level, the so called characteristic frequency (CF). Measurements of the CF and the minimal sound pressure level that elicits a response in the fibers reveal two organizing principles of the basilar papilla: a tonotopic organization and an audibility organization. The connections made by auditory nerve fibers at the apical end of the papilla show the lowest characteristic frequencies, while the ones connecting to the basal end the highest. On the other dimension, fibers connecting to the neural end respond to low sound pressure levels, while the ones connecting to the abneural end respond only at very high sound pressure levels (Manley et al., 1985).

When stimulated with a single tone at the CF, auditory nerve fibers will respond with spikes that are phase-locked up to roughly 1 kHz. Interestingly, when measuring spontaneous activity in these fibers, half of them show regular interspike intervals which are roughly 15% longer than a period of the CF (Manley et al., 1985). This seems consistent with the fact that the electrical properties of the fibers modeled as a RLC-circuit (proper name) have a resonance frequency which is roughly 20% lower than their CF.

3.1.4. Hindbrain Nuclei

The recipients of the auditory nerve are the ipsilateral nucleus magnocellularis (NM) and nucleus angularis (NA) (Correia et al., 1982). In case of the redwing blackbird neurons in NM show quite a high spontaneous firing rate between 16 and 237 spikes per second with a mean of 116 (Sachs and Sinnott, 1978). The organization of the neurons is tonotopic, but interestingly only a few neurons have CFs above 5 kHz (Sachs and Sinnott, 1978; Konishi, 1970). For the barn owl these neurons show responses that are phase-locked to a stimulus at the CF up to nearly 10 kHz (Köppl, 1997b). This phase-locking is important as NM projects tonotopically to both ipsilateral and contralateral nucleus laminaris (NL) as its only efferent projection (Krützfeldt et al., 2010a; Parks and Rubel, 1975; Young and Rubel, 1983; Takahashi and Konishi, 1988). NL is thought to be the nucleus that calculates the interaural time difference (ITD) and thereby locates the sound sources (Carr et al., 1989; Hyson, 2005). Neurons in NA show a very rigid tonotopic organization. Lowest CF have been measured between 135 Hz and 655 Hz (depending on the species) and highest CF were between 6.2 kHz and 9.4 kHz (Konishi, 1970). These values are considered to define the limit of hearing in the bird, as the other recipient of the auditory nerve, NM, can only represent a more limited bandwitdh. An interesting feature of a subpopulation of NA neurons is the presence of not only a CF, but also of a characteristic loudness. These neurons are inhibited when the loudness exceeds a certain threshold (Sachs and Sinnott, 1978).

NL and NA both provide the main input to the superior olive (SO) and the lateral lemniscal nuclei (LL) on both, the ispi- and contralateral side (Krützfeldt et al., 2010a,b; Wild et al., 2010; Correia et al., 1982; Takahashi and Konishi, 1988; Yang et al., 1999). SO and LL themselves project back to all other four ipsilateral hindbrain nuclei in the bird (Wild et al., 2010; Yang et al., 1999). Further they are both projecting to their respective contralateral counterpart (Wild et al., 2010). Additionally LL receives input from the RA Cup (Martin Wild et al., 1993; Mello et al., 1998). The exact functionality of SO and LL is mostly unknown. However, part of LL is known for interaural level difference (ILD) estimation in case of the barn owl (Takahashi and Keller, 1992). For small birds the ITD becomes too small for reliable source localization and ILD becomes more important. Additionally, ITD and ILD are not identical. Therefore not only the azimuth but also the elevation can be estimated (Moiseff, 1989).

LL form two interesting connections, bypassing the midbrain: firstly, they bilaterally connect directly to the auditory thalamus, the nucleus ovoidalis (Ov), and secondly, they provide auditory input to the thalamic nucleus uvaeformis (Uva) bilaterally (Wild et al., 2010), which is gating auditory input to the premotor nucleus HVC (Coleman et al., 2007).

A very thorough study of auditory hindbrain connectivity in the zebra finch was published by the group of Prof. JM Wild (Krützfeldt et al., 2010a,b; Wild et al., 2010).

3.1.5. Midbrain - MLd

The auditory nucleus of the avian midbrain is the mesencephalicus lateralis dorsalis (MLd). MLd receives auditory input from both, ipsilateral hindbrain nuclei and contralateral hindbrain nuclei SO, LL, NL, and NA (Krützfeldt et al., 2010a,b; Wild et al., 2010; Correia et al., 1982; Takahashi and Konishi, 1988). Except for a few bypasses, all auditory information to the cortex passes through MLd. It also forms a feedback connection with the RA Cup (Martin Wild et al., 1993; Mello et al., 1998). In addition, it is connected to its contralateral counterpart (Karten, 1967; Akesson et al., 1987). One should keep in mind that all nuclei further upstream do no longer have contralateral connections.

Woolley and Casseday (2004) examined the auditory behavior of neurons in the MLd of zebra finches. A striking difference to the hindbrain neurons is the nearly complete absence of spontaneous firing. The encoding of auditory information in MLd is unsigned, only in a positive change of firing rate, not in suppression of an ongoing firing. In response to pure tones roughly half of the neurons respond only to the onset, while the other half respond with sustained firing (Woolley and Casseday (2004) further broke down these response patterns into 5 subgroups). The onset-neurons are responding slightly faster with a first-spike latency of 10.31 ± 0.44 ms while the neurons with ongoing activity show a first-spike latency of 13.8 ± 1.05 ms (Amin et al. (2010) report a mean latency of 7.1 ms using conspecific songs and estimations of STRFs). The CFs and the response thresholds of all neurons are broadly distributed over the whole frequency and loudness range. Also tuning curves range from narrow (< 1.5 kHz) to broad, regardless of any other feature of the neurons.

In a second study Woolley and Casseday (2005) researched the response of MLd neurons to more complex sound, namely white noise, band-limited noise, frequency modulated sweeps, and sinusoidal amplitude-modulated tones. Their main result was that the response to noise of most cells can be well predicted from the pure tone responses, contrary to responses in cortical cells (see sections 3.2 and following). However, a small fraction of neurons (13%) showed a selective response for up- or downsweeps. These responses can not be explained with the simple features obtained from pure-tone responses. The common feature in these neurons was that the gained frequency-tuning curve was asymmetric, a feature unseen in neurons without selectivity for sweeps. For the amplitude-modulated tones, most neurons, except for a big fraction of onset neurons, did respond to the tone modulation. The best modulation frequencies were found to lie between 20 Hz and 200 Hz (or 140 Hz, depending on the method), a range "*well suited for encoding the temporal modulations that characterize zebra finch song*". These findings were generally confirmed calculating STRFs using natural birdsong (Woolley et al., 2009). However, the neurons in MLd are able to change their response property depending on the stimulus: (Woolley et al., 2006) found that the STRFs had spectrally narrower excitatory and inhibitory subfields when using modulation-limited noise stimuli than when using CON. But the STRFs had temporally narrower and faster subfields when using CON than when using noise stimuli. This change leads to a more precise and faster coding of CON - a behaviorally relevant stimulus - compared to noise in MLd.

3.1.6. Thalamus - Nucleus Ovoidalis

The auditory relay between midbrain and cortex is formed by the thalamic nucleus ovoidalis (Ov). It is sometimes functionally further subdivided into its core (Ov core), nuclues ovoidalis medialis (Ovm) and the surrounding shell (Ov shell) (Vates et al., 1996). Ov receives auditory input from both ipsi- and contralateral MLd (Karten, 1967; Akesson et al., 1987), as well as from ipsi- and contralateral LL (Wild et al., 2010) and gets feedback from the RA Cup (Martin Wild et al., 1993; Mello et al., 1998).

The functional role of Ov is not well understood and only few studies have been published. According to Amin et al. (2010) "*because it has been difficult to record single unit activity from this small ovalshaped nucleus deep in the brain*". The lack of knowledge is unfortunate as Brauth et al. (2007) showed gene-expression-based evidence for neural plasticity in Ov during the process of familiarization and habituation (in budgerigars). An early electrophysiological study was conducted by Bigalke-Kunz et al. (1987) in head-restrained European starlings. Their main findings where a tonotopic organization of Ov as well as a high spontaneous firing rate (mean 61 Hz) compared to MLd (see section 3.1.5). A more recent study was performed by Amin et al. (2010) in anesthetized zebra finches. Playing back conspecific songs and white noise they estimated STRFs. In contrast to Bigalke-Kunz et al. (1987) they found a much lower background firing rate of 3.25 ± 4.05 Hz, yet still significantly higher than in MLd (0.15 ± 0.46 Hz). The mean latency derived from the STRFs was 10.2 ms, responding 3.1 ms later than MLd on average and 2.2 ms before subfield L2a. Their main result was a much closer similarity of STRFs in Ov to the more complex STRFs in Field L than to the simpler STRFs in MLd. This similarity and the high differential latency to MLd urge us to drop the idea of Ov being a simple relay between midbrain and cortex. Further studies are needed to understand the computational role of Ov and whether traces of feedback signals from higher areas can encountered.

3.2. Primary Auditory Cortical Nucleus - Field L

Field L in birds was first defined by Rose (1914) by Nissl staining. It was redefined by Karten (1968) as the region where axons from Ov terminate in the avian cortex. Field L lies in the birds nidopallium and as auditory thalamorecipient zone in the bird's cortex. It is the putative homologue of the primary auditory cortex in mammals (Jarvis et al., 2005). It is subdivided into four (Bonke et al., 1979) or sometimes into five subfields (Fortune and Margoliash, 1992), L1, L2a, L2b, and L3. However, the two definitions (Rose, 1914; Karten, 1968) of Field L are not completely congruent. When defined, the fifth subfield is the discrepancy of the two above definitions and is, irritatingly, again called (field) L. The subdivisions into the four subfields are very well-defined by staining studies (Fortune and Margoliash, 1992) as well as by tracer studies (Vates et al., 1996).

Main input to Field L is provided to L2a by the Ov core (Vates et al., 1996). It is thereby the main auditory thalamorecipient area. This finding is supported by the time to peak response measured by Sen et al. (2001). Neurons in L2a show a peak response at 14 ± 1 ms after optimal stimulus onset, while neurons in the other subfields are significantly slower (L2b: 20 ± 1 ms, L1: 21 ± 3 ms, L3: 22 ± 3 ms). Furthermore L2a is reciprocally connected to all other subfields of Field L as well as to CLM (Vates et al., 1996). L1 and L3 both receive thalamic input from Ov shell. They are reciprocally connected to L2a and to each other. L2b receives thalamic input from Ovm. It is reciprocally connected to L2a. All subfields are further reciprocally connected to CLM (Vates et al., 1996).

The internal organization of these subfields is not completely understood. In guinea fowls (Bonke et al., 1979) and mynah birds (Langner et al., 1981) tonotopic maps have been found. In zebra finches Gehr et al. (1999) found several tonotopic gradients which surprisingly were not congruent to the subfields.

The measured figures for the background firing rates of neurons in Field L differ greatly : Grace et al. (2003) reported a mean background firing rate of 2.8 ± 2.7 spikes per second in anesthetized zebra finches. Woolley et al. (2009) on the other hand reported a mean background firing rate of 0.6 spikes per second in zebra finches anesthetized with only a slightly higher dose of urethane.

One of the first studies on the functionality of Field L was performed by Leppelsack and Vogt (1976) in immobilized starlings. They played natural sounds, mainly conspecific songs and calls. They analyzed and categorized the responses by their predominance for certain features. Most of their neurons responded to simple spectral features. In conclusion they stated that "*there remains a fundamental doubt whether it will be possible to assign it* [Field L] *a distinct and clearly defined function*".

Margoliash (1986) investigated the preferences of neurons in Field L of anesthetized white-crown sparrows. However, he found no preference for BOS over REV or other modifications of BOS. Lewicki and Arthur (1996) looked for selectivity in Field L and HVC of anesthetized zebra finches. They compared BOS versus REV, the syllables of BOS in reverse order, and the subsyllables of BOS in reverse order. They found a slight BOS-selectivity, however much less than in HVC. This finding is supported by newer studies finding slightly negative (Janata and Margoliash, 1999, in L1) and zero to slightly positive (Amin et al., 2004, for the different subfields) mean BOS-REV selectivity. When comparing the responses for BOS and CON in anesthetized zebra finches a negative mean selectivity throughout all subfields of Field L is found (Janata and Margoliash, 1999; Amin et al., 2004). On the other hand, for the acoustically similar TUT they measured a negative mean TUT-CON selectivity only in L2a, while in the other subfields they measured mean selectivities close to zero. This non-selectivity was later also found in juveniles (Amin et al., 2007).

Neurons in Field L tend to prefer natural stimuli over artificial stimuli, except white noise due to the big onset response (Grace et al., 2003; Theunissen et al., 2004). However, these preferences do not seem to be congenitally given. Comparing responses in juvenile (roughly 35 dph) and in adult birds, Amin et al. (2007) mostly found significantly lower selectivities for CON over the different artificial stimuli in the juvenile birds. The one big exception was white noise: while neurons in L2a were selective for CON against white noise, the mean selectivity disappeared in the adult bird. Cross-fostering experiments between different finches reveal that this development of neural responses is not purely hard coded but at least partly experience dependent (Woolley et al., 2010a; Woolley, 2012; Hauber et al., 2013).

One question arising in these studies was the influence of the anesthesia on the auditory responses. However, Cardin and Schmidt (2003) stated that "[A]*uditory responses in Field L were unaffected by arousal in both acute and chronic recordings. In fact, Field L auditory responses in chronically implanted birds are very similar during sleep, light sedation, deep anesthesia, and wakefulness*".

When estimating STRFs in Field L many - however, less than in MLd or Ov - neurons are found to respond to onsets regardless of frequency (Woolley et al., 2009; Amin et al., 2010). The second largest group in Field L is formed by neurons responding mainly to a certain frequency (Woolley et al., 2009; Amin et al., 2010). The rest of the neurons respond to features such as harmonic stacks (Sen et al., 2001; Woolley et al., 2009), offsets (Woolley et al., 2009; Amin et al., 2010), changing frequencies (Sen et al., 2001; Nagel and Doupe, 2008; Woolley et al., 2009), as well as to complex, non-categorizable stimuli (Sen et al., 2001; Nagel and Doupe, 2008; Woolley et al., 2009; Amin et al., 2010). Neurons in L2(a) respond faster (Sen et al., 2001; Nagel and Doupe, 2008) and integrate the stimulus input over shorter periods Nagel and Doupe (2008) than neurons in L1 and L3. Also, responses of neurons in L2a are more linear, i.e. the STRFs explain more of the response (L1: 0.48 ± 0.05 , L2a: 0.63 ± 0.02 , L2b: 0.44 ± 0.04, L3: 0.37 ± 0.05) (Sen et al., 2001). One possible source of nonlinearity is given by Nagel and Doupe (2008): they propose that excitatory and inhibitory regions in the STRFs have different threshold, leading to intensity-depending STRFs. A summary of STRFs found in Field L is given as eMTF: most of the energy is distributed between 5 and

60 Hz temporal modulation and between 0 and 1 cycle/kHz spectral modulation (Woolley et al., 2005). However, the upper limitations are given by the noise-induced necessity for smoothing the STRFs[4]. Newer data indicates that the upper limit for both temporal and spectral modulation might be higher (Woolley et al., 2009).

In contrast to the clearly auditory functions of Field L Keller and Hahnloser (2009) found subsets of neurons in Field L and CLM showing selective response to auditory perturbations only during singing or stereotyped responses during singing even when perturbed, but not during playback. These neurons might indicate the existence of an efference copy of the song's motor command as well as an error sensitivity in these areas.

3.3. Secondary Auditory Cortical Nuclei

3.3.1. NCM

The caudomedial nidopallium (NCM) is an anatomically structure adjacent to L3 (Mello and Clayton, 1994). It receives auditory input from L2a and L3 as well as from the Ov shell and is reciprocally connected to CMM (Vates et al., 1996). It further gets input from ParaHVC (Foster and Bottjer, 1998).

NCM is thought to be a key area in formation and storage of auditory memory, especially the memory of TUT (Hahnloser and Kotowicz, 2010). A first hint was given by Mello et al. (1992). They found an increased expression of ZENK - an immediate early gene involved in memory formation (Goelet et al., 1986) - in response to CON, much less in response to heterospecific songs, and none after tone bursts. However, there is habituation: the increase in ZENK-expression rapidly declines back to background levels within 30 minutes, if a single CON is presented repeatedly. But as soon as a new CON is introduced, ZENK-expression is increased again (Mello et al., 1995; Gentner, 2004).

The necessity of ZENK-expression in the auditory system for tutor memory formation was shown by London and Clayton (2008): By temporary pharmacological obstruction of ZENK-expression in primary and sec-

---

[4] see Section 2.2.2

ondary auditory areas before any tutoring session they successfully suppressed the capability of copying TUT in zebra finches.

The habituation seen in ZENK-expression is also supported by measurements of spike responses: a fast reduction of multiunit responses to repeated presentation of the same CON is found, especially between the first few presentations (Chew et al., 1995; Stripling et al., 1997). However, after several hundred presentations of a single song, responses to previously presented CON are similarly strong as when first presented (Stripling et al., 1997). But when training starlings in a go/no-go operant-conditioning procedure, Thompson and Gentner (2010) did not find a general habituation to CON due to simple exposure, but only when the song was coupled to behavior. Additionally they found a steep gradient of habituation along the dorsal-ventral axis: while the most dorsal quarter of neurons of NCM showed no or even negative habituation, the most ventral neurons in NCM showed significantly lower responses to the trained CON than to novel CON.

Strong evidence for NCM being the location of TUT memory was provided by Gobes and Bolhuis (2007): they measured the behavioral preference of adult male zebra finches for the TUT compared to novel CON and then lesioned NCM. They showed that preference was significantly reduced, but the birds still were able to sing and to distinguish male from female calls. However, they did not test, whether the effect was TUT vs. CON or familiar vs. novel songs.

3.3.2. CM

The caudal mesopallium (CM) is an area adjacent to L1, separated from Field L by the lamina hyperstriatica. It is subdivided into a medial (CMM) and a lateral part (CLM) which are highly interconnected. CLM is reciprocally connected to all subfields in L and to CMM. It projects to NIf, HVC shelf and RA Cup. CMM is reciprocally connected to NCM and CLM (Vates et al., 1996). There is some indication that CM also projects directly to HVC (Bauer et al., 2008). There is some further evidence that CM receives input from UVA and Ov shell (Martin Wild et al., 1993; Fortune and Margoliash, 1995; Vates et al., 1996).

In accordance with the hierarchical connectivity found, neurons in CM will mostly respond later to stimuli than neurons found in Field L (Sen et al., 2001). In parallel the neurons tend to respond less linearly to natural stimuli, i.e. STRFs explain less of the response variability than in L1, L2a, and L2b (0.37 ± 0.06, similar to L3) (Sen et al., 2001). However, this lower number could also result from a smaller number of spikes. Generally responses to CON in CM are smaller than in Field L (Sen et al., 2001) and have lower z-score (Grace et al., 2003; Amin et al., 2004). Nevertheless, the neurons show slightly positive selectivities for BOS vs. CON and for BOS vs. REV (Amin et al., 2004; Bauer et al., 2008), as well as a clear selectivity for CON over artificial stimuli (Grace et al., 2003; Bauer et al., 2008). In contrast to Field L this selectivity is already present in young birds (Amin et al., 2007).

The weaker CON-responses in CM however are not global, i.e. in response to all CON. Neurons in CM often are very selective for certain (complex) features. Interestingly, this selectivity for certain features is negatively correlated to the background firing rate of the neurons in CMM of starlings (Meliza et al., 2010). The features for which the neurons are selective are not just random features but trained by experience and behavioral relevance. When trained on a standard go/no-go operant-conditioning procedure starlings are able to distinguish two classes of CON motives. Jeanne et al. (2011) measured responses to these motives in CMM and CLM. In both subdivisions the entropy of firing rates was higher in response to rewarded motives than to unrewarded motives and this entropy was again higher than to novel motives. Identically was the mutual information between firing rate and motives highest for the rewarded songs and higher for unrewarded than novel motives. Mutual information between motives and firing rates was generally higher in CMM than CLM, and neurons in CMM had a higher coefficient of variation. Also, neurons in CMM encoded significantly more information about the rewarded/unrewarded categories than neurons in CLM and more information about the rewarded/unrewarded categories than about random categories.

Similar to NCM also CMM shows an elevated ZENK-expression after the playback of novel songs. But in contrast to NCM the number of ZENK-positive cells was also above baseline after the playback of familiar songs, although lower than after novel songs (Gentner, 2004).

However, similar to Keller and Hahnloser (2009) in Field L, Bauer et al.

(2008) found a discrepancy in responses while listening to BOS and activity during singing in many neurons in CM. Where this discrepancy arises from remains unclear.

3.4. Tertiary Auditory (Premotor) Cortical Nuclei

3.4.1. NIf

The interfacial nucleus of the nidopallium (NIf) is a small nucleus located within Field L, between L1 and L2a (Fortune and Margoliash, 1995). NIf gets input from CLM (Vates et al., 1996; Bauer et al., 2008) and UVA (Nottebohm et al., 1982; Akutagawa and Konishi, 2005) and is reciprocally connected to Av (Akutagawa and Konishi, 2010). NIf shows clear premotor activity related to singing (McCasland, 1987; Lewandowski and Schmidt, 2011). However, the role of NIf in song production is not well understood. It seems that in adult birds NIF is not necessary for song production and that lesions have only minor impact on song quality, but influence syllable and motive sequence (Hosino and Okanoya, 2000; Cardin et al., 2005; Naie and Hahnloser, 2011). In older juveniles in contrast NIf lesions significantly reduce song quality (Naie and Hahnloser, 2011).

But neurons in NIf also show auditory responses. Main auditory input to NIf is provided through CLM, as auditory responses are abolished while CM is inactivated (Bauer et al., 2008). Similar to HVC, auditory response (as well as baseline firing) is modulated by the behavioral state, although not completely abolished: when aroused, baseline firing rates are higher and response strength is reduced (Cardin and Schmidt, 2004a).

Auditory neurons in NIf present themselves as BOS selective. In anesthetized zebra finches Janata and Margoliash (1999) found 13 out of 14 neurons to be selective for BOS vs. CON (mean selectivity $d' = 1.50$) and 15 out of 16 to be selective for BOS vs. REV (mean selectivity $d' = 1.39$). Similar values were obtained by Coleman and Mooney (2004) (BOS vs. CON $d' = 1.26 \pm 0.23$, BOS vs. REV $d' = 1.29 \pm 0.23$). This clear BOS-selectivity arises in NIf, as neurons in CM - the source of auditory input to NIf (Bauer et al., 2008) - does not respond with such clear selectivity.

3.4.2. HVC

HVC (used as a proper letter-based name[5]) is probably the best studied nucleus in the songbird's brain. Reasons for this scientific attention are that HVC is uniquely found in songbirds (Lovell et al., 2008) and is considered to be the control center of song production, as discussed below.

HVC is located in the caudal nidopallium (Reiner et al., 2004c) and gets input from UVA, NIf, and MMAN (Nottebohm et al., 1982; Fortune and Margoliash, 1995; Vates et al., 1997; Akutagawa and Konishi, 2005) and is reciprocally connected to Av (Nottebohm et al., 1982; Akutagawa and Konishi, 2010). There is some evidence that the connection to RA is also reciprocal (Roberts et al., 2008). Whether or not it also gets input from the HVC shelf is unsure. Few axons from the shelf to HVC have been found (Mello et al., 1998), as well as dendrites of HVC neurons in the shelf (Fortune and Margoliash, 1995; Vates et al., 1996). But there is no clear evidence whether these connections are functional (Wang et al., 2001; Shaevitz and Theunissen, 2007).

HVC is often considered to be the central nucleus in song production. It is necessary for song production in older juveniles and adults: when HVC is bilaterally lesioned birds will restart singing subsong. However, the production of subsong as well as calls is not impaired by this lesion (Aronov et al., 2008). Further, by cooling of HVC the song can be slowed down by up to 45%, while cooling of downstream nucleus RA has no influence on speed (Long and Fee, 2008).

HVC shows great sexual dimorphism. In zebra finches the volume of HVC differs by a factor of 8 to 10 between adult males and females (Nottebohm and Arnold, 1976; Gahr and Metzdorf, 1999). In general there seems to be a positive correlation between the size of the song repertoire of a bird and the size of its HVC, across species (Devoogd et al., 1993b) and sometimes within species (Nottebohm et al., 1981; Airey et al., 2000), but not for all (Brenowitz et al., 1991; MacDougall-Shackleton et al., 1998). In contrast, the size of HVC is not constant, in open-ended learners -

---

[5] Originally the nucleus was called hyperstriatum ventrale, pars caudalis (Nottebohm and Arnold, 1976) which turned out to be anatomically incorrect. The acronym was so prominent by then that it was preserved as a proper name (Reiner et al., 2004c). In between it was sometimes referred to as higher vocal center to match the acronym.

birds that relearn a new song each year such as canaries - the size of HVC varies with the season, it is bigger in spring than in autumn (Nottebohm, 1981; Nottebohm et al., 1986; Alvarez-Buylla and Kirn, 1997; Vellema et al., 2010). This example of (adult) neurogenesis is not limited to open-ended learners but has also been observed in closed-ended learners, during development as well as in adults where the new neurons are incorporated in HVC and replacing dead neurons without altering the song (Nordeen and Nordeen, 1990; Tramontin and Brenowitz, 1999; Lipkind et al., 2002).

Generally three categories of neurons in HVC are distinguished: projecting to RA ($\mathrm{HVC_{RA}}$), projecting to Area X ($\mathrm{HVC_X}$), and interneurons ($\mathrm{HVC_{int}}$). $\mathrm{HVC_{RA}}$ and $\mathrm{HVC_X}$ neurons are known to fire very sparse and reliable bursts of action potentials while singing and while listening to BOS. In singing zebra finches $\mathrm{HVC_{RA}}$ fire zero or one burst per motive rendition at very high precision (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007), while $\mathrm{HVC_X}$ neurons will fire zero to four bursts per motive rendition at a slightly lower precision (still, for most neurons the root-mean-square jitter of the first spike is < 2 ms) (Kozhevnikov and Fee, 2007). In contrast, $\mathrm{HVC_{int}}$ fire throughout the whole motive with a modulated firing rate and without a precise single spikes (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007).

Between all three types of HVC neurons direct or indirect connections were found, but the main connectivity was $\mathrm{HVC_{RA}}$ neurons inhibiting $\mathrm{HVC_X}$ neurons, most likely disynaptically through $\mathrm{HVC_{int}}$ neurons. But also $\mathrm{HVC_X}$ neurons inhibiting other $\mathrm{HVC_X}$ neurons and $\mathrm{HVC_X}$ neurons exciting $\mathrm{HVC_{RA}}$ neurons were prominently found (Mooney and Prather, 2005).

An interesting feature was discovered by Wang et al. (2008): as mentioned above, the cortex of birds consists of two hemispheres that are only connected through mid- and hindbrain, as birds do not possess a corpus callosum. They addressed the question which hemisphere was dominantly producing the song. What they found was a rapid switch of dominance between left and right HVC every few dozens of ms throughout the song and surprisingly not synchronized with the syllable structure of the song. The underlying mechanism of the switch however remains unknown.

The auditory responses of neurons in HVC is gated and depending on the behavioral state of the bird: in zebra finches neurons in HVC respond

neuron type	mean d' BOS vs. REV	mean d' BOS vs. CON
HVC_{RA}	1.3 - 1.53	1.72
HVC_{X}	0.59 - 1.65	0.68
HVC_{int}	2.54 - 3.0	0.27
unspecified multiunit	1.7	2.3

Tab. 3.1: Mean d'-values reported for HVC-neurons in anesthetized zebra finches (Compiled from: Theunissen and Doupe, 1998; Mooney, 2000; Rosen and Mooney, 2003; Coleman and Mooney, 2004).

to auditory stimuli during anesthesia, sleep, and probably quiet resting, but not when aroused or singing (Schmidt and Konishi, 1998; Cardin and Schmidt, 2003, 2004a). In swamp sparrows and bengalese finches auditory responses in HVC_{X} neurons are only gated off while singing (Prather et al., 2008;contrariwise: Sakata and Brainard, 2008). The mechanism of gating is not completely understood, but it is known that UVA and probably NIf play an important role (Cardin and Schmidt, 2004b,a; Coleman et al., 2007). Main auditory input to HVC is provided through CM and NIf as deactivation of CM shuts down nearly all auditory response in HVC and NIf and deactivation of NIf auditory response in HVC (Coleman and Mooney, 2004; Cardin et al., 2005; Bauer et al., 2008). However, also MMAN provides auditory input to HVC (Vates et al., 1997; Williams et al., 2012).

When HVC shows auditory response (see above), neurons of all three categories in HVC are known to respond very selectively for BOS. Janata and Margoliash (1999) measured a highly significant preference BOS vs. CON or REV in anesthetized zebra finches, without specifying the type of neuron in HVC. Figures reported in other studies supported this finding (see Table 3.1). In contrast, Cardin and Schmidt (2003) reported that most neurons in HVC are unselective in awake zebra finches and that responses were very unstable. However, a more thorough study found an overall BOS vs. CON selectivity of $d' = 0.640 \pm 0.075$ for putative HVC_{int} in awake zebra finches (Raksin et al., 2012). This value is significant and in the range of values reported for anesthetized birds. Even higher values have been reported in awake bengalese finches (Sakata and Brainard (2008): unidentified neurons, BOS vs. REV $d' = 3.01$, BOS vs. CON

$d' = 3.18$). Furthermore, Sakata and Brainard (2008) detected for the first time auditory feedback during singing.

A set of studies has focused on the source of BOS-selectivity in HVC neurons. For all three categories neurons were found to receive already selective input from NIf (Rosen and Mooney, 2006). Interestingly, even though all three categories show a selectivity BOS, their auditory responses are completely different. In anesthetized birds both types of projection neurons have a very low baseline firing rate (Mooney (2000): $\mathrm{HVC_{RA}}$: 0.6 ± 0.4 Hz, $\mathrm{HVC_X}$:0.6 ± 1.5 Hz) and in response to BOS (roughly 1 Hz above baseline) and fire with a high sparseness (Coleman and Mooney, 2004). In contrast $\mathrm{HVC_{int}}$ neurons fire with a baseline rate of 12.0 ± 4.3 Hz (Mooney, 2000) and with roughly 11 Hz above baseline in response to BOS (Coleman and Mooney, 2004). The projection neurons do not differ much in the firing rates, but more in their subthreshold behavior in response to BOS: while $\mathrm{HVC_{RA}}$ are generally depolarized throughout the whole BOS presentation, $\mathrm{HVC_X}$ are mainly hyperpolarized during BOS presentation (Mooney, 2000). This difference in subthreshold behavior is unmasked when (further) depolarizing the neurons by current injection. While $\mathrm{HVC_{RA}}$ neurons stay BOS selective, $\mathrm{HVC_X}$ lose their selectivity or the selectivity is even reversed (Mooney, 2000). However, when local inhibition was shut down the response of $\mathrm{HVC_X}$ neurons became similar to the one of $\mathrm{HVC_{RA}}$ neurons (Rosen and Mooney, 2006).

Prather et al. (2008) addressed the question of how auditory response to BOS and premotor activity of HVC neurons are related in swamp sparrows. They actually found a very precise and robust mirroring in $\mathrm{HVC_X}$ neurons, firing at the same point in the song both while singing and listening to BOS. This finding means that if $\mathrm{HVC_X}$ have influence on sound production, they will not respond to the same sound they produce. A sound produced due to the firing of a neuron will be produced after the firing of that neuron. However while listening the neuron can only fire in response to sound that has already passed. So either $\mathrm{HVC_X}$ neurons are not premotor in the sense that they influence the motor output, or $\mathrm{HVC_X}$ neurons do not respond to the sound they influence themselves. Whether or not this mirroring also happens in zebra finches remains to be tested.

3.4.3. Av

The nucleus avalanche (Av) is a small nucleus in the middle of CM (Nottebohm et al., 1982; Akutagawa and Konishi, 2010). Av gets input from UVA and is reciprocally connected to NIf and HVC (Akutagawa and Konishi, 2010). Even though this nucleus has been known for a long time, it was mostly ignored and and only recently moved back into focus of research. It is known that while singing Av has an amplified ZENK-expression (Jarvis and Nottebohm, 1997; Feenders et al., 2008). Additionally Akutagawa and Konishi (2010) showed - as expected from its connectivity - that neurons in Av are clearly BOS-selective, in contrast to the surrounding CM. However, up to date no further electrophysiological study is known to me measuring the neural activity in Av, neither the premotor activity during singing, nor the auditory responses. But due to its neglect and its position within CM, there is a certain risk that studies on CM might have unintentionally incorporated Av.

3.5. Auditory Responses in Other Areas

3.5.1. Premotor Nuclei and the Anterior Forbrain Pathway

The nucleus uvaeformis (UVA) is a thalamic nucleus that amongst other inputs gets bilateral auditory input from LL (Wild et al., 2010). It also integrates different sensory cues, such as visual or somatosensory (Wild, 1994). It projects to the premotor nuclei NIf, Av, and HVC (Nottebohm et al., 1982; Akutagawa and Konishi, 2010). The activity of neurons in UVA is depending on the behavioral state of the bird (Hahnloser et al., 2008). By changing the state or by electrical stimulation of the neurons, auditory responses in HVC is gated on or off (Coleman et al., 2007; Hahnloser et al., 2008). The neurons in UVA also show clear auditory responses under anesthesia. But in contrast to the other premotor nuclei, most neurons in UVA are unselective (Coleman et al., 2007).

The robust nucleus of the archopallium (RA) is a nucleus that gets direct input from HVC as well as indirect through the anterior forebrain pathway (AFP) (Nottebohm and Arnold, 1976; Bottjer et al., 1989) and probably from the RA cup (Mello et al., 1998). It projects ipsilaterally directly and indirectly to the hypoglossal nucleus, a motor nucleus that innervates the muscles in the syrinx and ipsilaterally directly and indirectly to brainstem nuclei controlling respiration (Nottebohm and Arnold, 1976; Reinke and Wild, 1998; Roberts et al., 2008). RA is absolutely needed for song production: while unilateral RA lesions will lead to a distortion of the song, bilateral lesion will stop any song production. But the bird will still be producing calls (Nottebohm and Arnold, 1976; Aronov et al., 2008). Neurons in RA show a clear selectivity for BOS (Doupe and Konishi, 1991; Vicario and Yohay, 1993). However, as HVC is the main direct or indirect auditory input to RA, responses are also gated (Dave et al., 1998).

The indirect pathway from HVC to RA is formed by the anterior forebrain pathway (AFP) a cortico-basal-thalamic loop. HVC is projecting to Area X which is projection to the medial nucleus of the dorsolateral thalamus (DLM). DLM is projecting to the lateral magnocellular nucleus of the anterior nidopallium (LMAN) which projects back to Area X as well as to RA (Vates and Nottebohm, 1995). Area X is part of the songbird's basal ganglia and further projects bilaterally to VTA via the ventral pallidum. Dopaminergic neurons in VTA in return mainly innervate ipsilaterally the striatal part of Area X (Gale et al., 2008). These dopamineric neurons might be an indication of an involvement of the AFP in learning (Schultz, 1998). The AFP is not needed for song production, but during song development the AFP is driving the song (Scharff and Nottebohm, 1991; Aronov et al., 2008). Although it is not necessary for song production in adult birds, the output of LMAN to RA induces slight variations in the song and might be involved in song maintenance (Brainard and Doupe, 2000; Kao et al., 2005).

As in the other premotor nuclei neurons in both Area X and in LMAN are responding very selectively for BOS vs. REV or CON. But this selectivity for BOS is much lower in juvenile birds singing plastic song and is raised while developing the adult song (Doupe and Konishi, 1991; Doupe, 1997). Unlike in HVC premotor activity in Area X and in LMAN is completely different from auditory response to BOS (Hessler and Doupe, 1999). It remains an open question of how the AFP is involved in song learning.

3.5.2. HVC Shelf and RA Cup

The HVC shelf and RA cup are two nuclei bordering the premotor nuclei HVC and RA. Both areas get auditory input from L1, L3, and CLM and the HVC shelf projects to the RA cup (Fortune and Margoliash, 1995; Vates et al., 1996; Mello et al., 1998). Despite their physical proximity there is disputed evidence that they provide (auditory) input to the respective premotor nuclei (Fortune and Margoliash, 1995; Vates et al., 1996; Mello et al., 1998; Wang et al., 2001; Shaevitz and Theunissen, 2007). The RA cup projects down to OV shell, MLd, and LL (Martin Wild et al., 1993; Mello et al., 1998). The two nuclei are both considered auditory: after song playback both areas show an increased ZENK-expression (Mello and Clayton, 1994; Jarvis et al., 1998). Further weak evidence of auditory response is given by functional magnetic resonance imaging (fMRI) where a response in the blood oxygenation level-dependent (BOLD) signal could be seen in the region of HVC/HVC shelf and RA/RA cup after auditory stimulation (Voss et al., 2007). However, electrophysiological confirmation is still missing. Whatever the use of such an auditory loop in parallel to the premotor system is, still is unclear. Farries (2004) speculates that it could be the homologous of a loop found in nonoscine birds and that the song system evolved out of this loop.

3.5.3. MMAN and VTA

The medial magnocellular nucleus of the anterior nidopallium (MMAN) receives input from the dorsomedial nucleus of the posterior thalamus (DMP) and projects to HVC and paraHVC (Vates et al., 1997; Foster et al., 1997). The role of MMAN is not understood, however, DMP gets input from the hypothalamus and could convey information about the internal state of the bird and control social and sexual behavior (Foster et al., 1997). Furthermore DMP could also receive input from RA and thereby forming a loop (Williams et al., 2012). Interestingly, neurons in MMAN show auditory response and are highly selective for BOS (Vates et al., 1997; Williams et al., 2012). The source of the auditory input is unknown. But as auditory response in MMAN lacks behind HVC and is similar to the response in HVC (Williams et al., 2012) there is a high chance that auditory input is delivered directly or indirectly through HVC.

The ventral tegmental area (VTA) gets input from Area X via the ventral pallidum (VP) and projects by dopaminergic neurons to most of the songbird's songsystem, but mainly to the striatal part of Area X (Gale et al., 2008). Both, VP and VTA, show clear auditory responses and neurons in both areas are highly selective for the birds own song (Gale and Perkel, 2010). During singing, responses in VTA are highly depending on social context: most neurons fire with a significantly high rate while singing directed song (Yanagihara and Hessler, 2006). Dopamin is often associated with learning and expectation (Schultz, 1998), it is however not yet clear what role it plays in song learning. There are some indication that it influences the context dependency of song variability (Leblois and Perkel, 2012).

Chapter 4

A New Nonsymmetric Sparse Coding Algorithm

> EM/3/Green: We'll all die here!
> Mr. Spock: A statistical probability.
> Lara: You ever quote anything besides statistics, Vulcan?
> Mr. Spock: Yes. But philosophy and poetry are not appropriate here.
>
> Star Trek: The Jihad

Mathematically speaking encoding is the mapping of sequences of source alphabet symbols onto new sequences of target alphabet symbols. This encoding is called lossless or lossy depending on whether the mapping is injective or not. So if we have a thought and vocalize it, it is the encoding of a thought as sound. If we write down what has been said, it is the encoding of sound as letters. And if a third person reads these letters, they will again be encoded as neural activity, forming a new thought. But as you can see, each encoding looks different. Biological encoding schemes, as the ones presented, are optimized to their field of application.

A possible goal of encoding is compression of the data in order to minimize the amount of transferred symbols. Maybe the channel of transmission is noisy, so the goal a robust encoding. Or the channel is unsave, so we need an encoding to ensure that the sequence is hard to interpret. Or exactly the opposite, we want an encoding that can be interpreted without heavy computation. So, if we artificially create an encoding scheme, we have to know what we want.

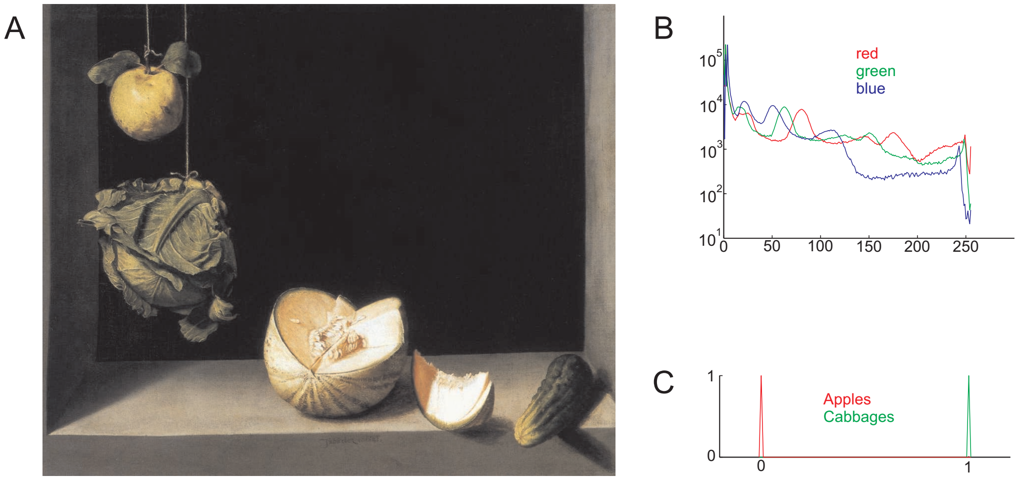


Fig. 4.1: Quince, Cabbage, Melon and Cucumber by Juan Sánchez Cotán. (A) The painting itself. (B) Histogram of the pixel values in each color channel. Very precise, but not informative about the content of the picture. (C) Histogram of the fruit/vegetable count in the cabbage and apple channel. Even tough not as exact as the histogram of pixel values, we have a better understanding of picture content.

4.1. Independent Component Analysis

The name 'Independent Component Analysis' (ICA) was coined in 1994, when Pierre Comon published 'Independent component analysis, a new concept?' (Comon, 1994), however other concepts existed already before. The basic idea of ICA is that the data is a sum of a lot of independent features. Imagine your data being the painting 'Quince, Cabbage, Melon and Cucumber' by Juan Sánchez Cotán (Figure 4.1A). Looking at the pixel values in Figure 4.1B, we know how reddish, greenish, and blueish our painting is. And we can be pretty sure that Juan Sánchez Cotán did not plan to paint a picture with such a color histogram. He was actually thinking of painting a quince, a cabbage, a melon, and a cucumber. These vegetables are the independent components of the picture. If we move one vegetable as a whole a few pixels, the content of the painting would still be the same, but if we move random patches of the painting by the same amount, we would no longer consider it a baroque painting, but rather something more modern.

Mathematically speaking we have a generative model. The observed data X was created by (unknown) independent underlying causes S being linearly mapped onto the space of observation by a mixing matrix A:

$$X = A \cdot S, \tag{4.1}$$

where X is a vector representing the picture in pixel space, S is a vector of the underlying causes (1 cabbage, 0 apples, 1 quince, ...) and A is a matrix with the picture (and the position) in pixel space of a cabbage in the first column, of an apple in the second column, and so on.

The goal of an ICA algorithm is to find an unmixing matrix W so that the vector

$$Y = W \cdot X = W \cdot A \cdot S \tag{4.2}$$

is equal to to the underlying causes S, except for scaling and permutation ($W \cdot A$ is a square matrix with exactly one non-zero entry in each row and column). The problem is, that normally both, the underlying causes S and the mixing matrix A are unknown. But assuming that the elements of the underlying causes S are independent of each other, we only need to find a matrix W, so that the elements of Y are independent ($p(Y) = \prod_i p_i(Y_i)$) over all our data (e.g. all paintings in all galleries). Unless more than one element S_i is Gaussian distributed, the matrix W is the can be found.

One problem is that in real life nothing is ever independent. If we have a painting with already a cabbage on it, there is an increased probability, that we will see more vegetables on the same painting. A second problem is the assumption of linearity. In the picture overlying elements are not adding up, but they conceal one another. Three-dimensional elements will create shadows that have shape depending nonlinearly on other elements. We therefore have to assume that no matrix W exist that projects our data X onto truly independent components. The goal of an ICA algorithm thus is not to find independent components, but a projection W that maximizes the independence of the components.

4.1.1. Contrast Function

One faces the problem of how to define the level of independence. Typical contrast functions to approximate independence of the elements Y_i are:

- Maximization of negentropy (Hyvärinen, 1998)
- Minimization of mutual information or minimization of the Kullback-Leibler-divergence between distribution and product of the marginal distributions (Comon, 1994; Hyvarinen, 1999)
- Maximization of the likelihood given a distribution of the underlying causes (including the infomax principle) (Pham et al., 1992; Bell and Sejnowski, 1995)
- Maximization of the kurtosis or other cumulant-based contrast function (Hyvarinen and Oja, 1997)

However, the differences are not are not as big as it might seem. Minimization of the mutual information is equivalent to the minimization of the Kullback-Leibler divergence (Hyvärinen, 1999) and asymptotically identical to a maximum likelihood estimation (Cardoso, 2000). On the other hand cumulant-based contrast function can be used to approximate both, the negentropy and mutual information (Hyvärinen, 1999). So finally it all comes down to choosing between single dimension contrast functions and multidimensional contrast functions which are actually equivalent to the sum of single dimensional functions with a decorrelating term.

4.1.2. Optimization Algorithms

4.1.2.1. Preprocessing

After choosing the contrast function there still remains the question of the optimization to be chosen accordingly. Most often the data is whitened before the optimization process itself, i.e. we center our data and linearly transform it, so that the preprocessed data's covariance matrix is the unity matrix:

$$X_P = P \cdot (X - \langle X \rangle) \tag{4.3}$$

$$\left\langle X_P^T \cdot X_P \right\rangle = \mathbf{I} \tag{4.4}$$

Normally this is done using principal component analysis (PCA) which in the same step can be used to reduce the dimensionality of our data. E and Λ being the matrix of eigenvectors and the matrix of eigenvalues of the covariance matrix of our data the projection matrix P becomes

$$P = \Lambda^{-1/2} \cdot E^T. \tag{4.5}$$

However any additional rotation on P fulfills equation 4.4.

4.1.2.2. Gradient Ascent/Descent

Gradient ascent/descent is probably the most straight forward method for optimization. Bell and Sejnowski (1995) tried to maximize the output entropy of a two-layer network with nonlinear output units ($\tanh(Y)$). However as mentioned above, it was shown that this is equivalent to a maximum likelihood estimation (Cardoso, 1997). The gradient ascent lead to the following update rule for the projection matrix:

$$\Delta W \propto W^{T^{-1}} - 2\tanh(Y) \cdot X^T. \tag{4.6}$$

Their algorithm does not need a prewhitening of data, but it performs better when applied. For further speed increase we replaced the stochastic gradient by the natural gradient (Amari, 1997)

$$\Delta W \propto (\mathbf{I} - 2\tanh(Y) \cdot Y^T) \cdot W. \tag{4.7}$$

The great advantage of the natural gradient is not its slightly faster convergence but that it does not need an explicit matrix inversion (however, for the prove the inverse has to exist).

A gradient descent algorithm for single components was first presented by Delfosse and Loubaton (1995), where they extract one component after another by searching for the one-dimensional subspace in which data has the highest kurtosis. An algorithm for a broad number of cost functions was later presented by Hyvärinen and Oja (1998).

The advantage of gradient ascent/descent algorithms is that they can handle both, on-line learning and batch learning. However, when using a gradient ascent/descent algorithm, it should be combined with a line search algorithm instead of fixed step sizes.

4.1.2.3. Fixed-Point Algorithm

Hyvarinen and Oja (1997) presented a fixed-point algorithm that maximized the kurtosis of the components. The algorithm, called FastICA, normally shows good convergence and can be used for estimating the independent components at once or one component after the other. Later the algorithm was generalized for any contrast function (Hyvarinen, 1999). However, the algorithm is restricted to orthogonal components whether they are learned at once or separately.

4.1.2.4. Neural Network-Inspired Algorithm

The oldest class of algorithms were inspired by neural networks. Jutten and Herault (1991) published a two-layer network algorithm that decorrelated the output of a non-linear output layer:

$$Y = (\mathbf{I} + W)^{-1} \cdot X \tag{4.8}$$

$$\Delta W \propto g_1(Y) \cdot g_2(Y)^T \text{ with } \Delta W_{ii} = 0. \tag{4.9}$$

If the network decorrelates the input for any nonlinearities $g(.)$, the outputs Y_i would be independent. However, this algorithm converges only under restrictions. Further improvement have been made to make it computationally more stable and faster (Laheld and Cardoso, 1994).

4.2. Sparse Coding

Sparse coding looks at the whole problem from the opposite point of view. Sparse coding is indifferent to the nature of the data, be it cabbages or

coding	+	-
dense, non-redundant	few channels, low energy demands	no error correction
dense, redundant	error correction	many channels, high energy demands, error correction only over several channels
sparse	error correction within single channel	many channels

Tab. 4.1: Advantages and disadvantages of coding schemes

apples. Sparse coding's only objective is to answer the question: What would be an efficient way (for a brain) to encode all this data?

When for example a painting as in Figure 4.1 is projected onto our retina, the encoding of it is highly redundant: neighboring pixels often share the same color and most of the background is completely black. If we look at all the paintings in the world, we will see similar redundancies in most of them. This redundancy means that, even though every channel itself carries a lot of information about the picture, i.e. has a high Shannon entropy (dense coding), the collective information of all channels is just slightly higher than the information in a single channel. An easy reduction scheme is PCA, reduction of the channels to a few decorrelated, dense-coding channels. The information in each channel is similar to the information in the original channels, but the total information is roughly the sum of the information in the single channels. On the downside, the brain represents a very noisy environment, and the redundancy gave us some kind of error correction. A third encoding method, sparse coding, combines the advantages. The goal of sparse coding is not to get rid of the redundancy, but to constrict the redundancy over many channels into redundancy within one single channel, so that each channel has a built-in error correction (See Table 4.1).

A first sparse coding algorithm was proposed by Olshausen and Field (1996), where they are trying to maximize the following cost (the variables have been adapted to fit the nomenclature of the thesis):

$$F(X,Y|A) = \sum_i (x_i - A_i \cdot Y)^2 + c \cdot \sum_j R\left(\frac{y_j}{\sigma}\right). \tag{4.10}$$

The first term $\sum_i (x_i - A_i \cdot Y)^2$ is a measure of the mean square reconstruction error of the original stimulus X by a representation Y and a given decoding matrix A. The second term $\sum_j R\left(\frac{y_j}{\sigma}\right)$ with σ^2 as the mean variance of all y_j, is the cost of the representation Y, where the elementwise function $R(.)$ is sparseness-enforcing. They tested the functions $R(y) = -e^{-y^2}, log(1+y^2), |y|$, which all lead to a qualitatively similar result. The factor c is the trade-off between reconstruction error and sparseness.

The learning algorithm used was again a gradient descent algorithm. Alternatingly they optimized the decoding matrix A and the representation Y. But even after the training is finished, the representation Y of a new stimulus X has to be optimized by gradient descent.

Note the similarity to the formulation of ICA. If we add a term of Gaussian white noise[1] to the ICA $X = A \cdot S + \eta$, the cost function of sparse coding is equivalent to a log-likelihood formulation with the distribution of S being proportional to $e^{-R(\cdot)}$.

4.3. A New Nonsymmetric Sparse Coding Algorithm

The following section is partly reproduced from the publication Blättler and Hahnloser (2011).

As seen in the previous section, many data-driven coding algorithm already exist. But all the algorithms presented so far are symmetric. Symmetric means that finding an encoding matrix W is equal to finding an encoding matrix $-W$, which would just exchange Y with $-Y$. But if we look again at the example of Figure 4.1A, there is 1 cabbage, 1 quince, and 0 apples. And we might look at a million more paintings (before artists started to play with negatives), and we will always find a non-negative number of cabbages. There are already a number of algorithms trying to implement this restriction. The probably best known is non-negative matrix factorization by Lee and Seung (1999). The goal of this method is to approximate a non-negative stimulus $X \approx A \cdot Y$ by a product of two

---

[1] Therefore, sparse coding is sometimes called noisy ICA

non-negative matrices. However, the two major drawbacks of this method are that no negative features are allowed[2] and that the encoding Y has to be optimized iteratively. A second algorithm was proposed by Hoyer (2002), called non-negative sparse coding. The cost function is actually identical to equation 4.10, it just imposes a non-negativity restriction on Y. But again, the drawback of the algorithm is its inability to encode new stimuli without iterative optimization. A third algorithm that should be mentioned here is non-negative ICA by Plumbley (2003). This algorithm minimizes the Euclidean norm of all negative elements of the encoding Y. But the algorithm does not care about the exact span of the subspace of positive values and the encoding W is restricted to rotations.

I therefore shall present a new algorithm that does not suffer the drawbacks of the algorithms mentioned so far.

The model we use is the generative model of noisy ICA, under the restriction that the hidden causes S are all non-negative and most of the time equal to zero:

$$X = A \cdot S + \eta. \tag{4.11}$$

The data vectors X^t are of dimensionality N_0. But before we feed them into our algorithm itself, we will whiten them and reduce the dimensionality to N_P by a standard PCA algorithm (see equations 4.4 and 4.5). In case of on-line learning we would have to replace the standard PCA by an on-line version, e.g. Rao and Principe (2002).

We will then project the Data using a square encoding matrix W

$$Y = W \cdot X_P. \tag{4.12}$$

The complete projection including the whitening will be called $\xi = W \cdot P$. If $\xi \cdot A$ is the unity matrix, we have

$$Y^t = S^t + \xi \cdot \eta^t + K, \tag{4.13}$$

---

[2] How would you represent shadows without being allowed to subtract something from the background?

where K is a constant offset due to the centering in the PCA step. η^t is N_0-dimensional uncorrelated Gaussian white noise with zero mean.

We will now make the following assumptions:

(I) the variance in all channels has been normalized, i.e. $\left\langle \left(A_i \cdot S^t\right)^2 \right\rangle_t = 1$

(II) the noise in all channels is equal, i.e. $\left\langle \eta_i^{t2} \right\rangle_t = \sigma_X^2$

(III) each source S_i has an equally sparse activity, i.e. $\left\langle s_i^t \neq 0 \right\rangle_t = n_a \ll 1$

(IV) each source's activity is independent of all other sources

(V) each source has the same influence on the signal, i.e. $\left\langle \left(\mathbf{A}_i \cdot s_i^t\right)^2 \right\rangle_t = \sigma_{S_i}^2 \cdot \|A_i\|_2^2 = k$

(VI) the projection matrix A is dense.

From this assumptions we can derive the following

(VII) if the number of sources N is big enough the elements of X will converge towards a Gaussian distribution by the central limit theorem.

(VIII) the influence k is the ratio between source and channels $k = \frac{N_0}{N} = \sigma_{S_i}^2 \cdot \|A_i\|_2^2 = \frac{\sigma_{S_i}^2}{\|\xi_i\|_2^2}$.

(IX) the noise variance on Y_i is $\sigma_{Y_i}^2 = \left\langle \left(\xi_i \cdot \eta^t\right)^2 \right\rangle_t = \|\xi_i\|_2^2 \cdot \sigma_X^2$.

(X) the SNR on X is $SNR_X = \frac{1}{\sigma_X^2}$.

(XI) the SNR on Y_i $SNR_Y = \frac{\sigma_{S_i}^2}{\|\xi_i\|_2^2 \cdot \sigma_X^2} = \frac{N_0}{N \cdot \sigma_X^2}$ is independent of the source i.

(XII) the total SNR therefore gets enhanced by a factor $\frac{SNR_Y}{SNR_X} = \frac{N_0}{N}$.

(XIII) if we only consider nonzero entries in the source S the SNR will be enhanced by a factor $\frac{N_0}{N \cdot n_a}$.

From (XIII) we see that the correct matrix ξ yields an enhancement of the SNR by the factor of $\frac{N_0}{N \cdot n_a}$. So for very sparse sources we can reach very high SNRs for nonzero entries.[3] It is therefore save to assume that any value y_i^t below a certain threshold θ stems from a source $s_i^t = 0$. We therefore introduce two new matrix: Y_+ as the matrix suprathreshold value of Y and Y_- as the matrix of subthreshold values:

$$y_{i,+}^t = \begin{cases} y_i^t & y_i^t > \theta \\ 0 & \text{else} \end{cases} \tag{4.14}$$

$$Y_- = Y - Y_+. \tag{4.15}$$

As a third matrix we define the firing rate R as the matrix of the threshold excess of Y:

$$r_i^t = \begin{cases} y_i^t - \theta & y_i^t > \theta \\ 0 & \text{else} \end{cases} \tag{4.16}$$

4.3.1. Zero-Threshold Algorithm

Given the matrices in the last section we can now formulate our cost function

$$F(Y) = \sum_{i,t} f\left(y_i^t\right) \tag{4.17}$$

with

$$f\left(y_i^t\right) = \frac{1}{2}\left(y_{i,-}^t - y_0\right)^2 + c \cdot r_i^t, \tag{4.18}$$

graphically shown in Figure 4.2. This cost function is similar to equation 4.10, if we set the threshold $\theta = 0$, the subthreshold minimum $y_0 = 0$, and restrict the coding matrix W to rotations. However, we do not restrict W

---

[3] this remains valid if we drop assumptions (I), (II), and (III). It should be an easy exercise for the experienced reader to proof it.

just to rotations. If not mentioned otherwise, the restriction on projection W was that the basis vectors of the projection had to be of length 1. Mathematically speaking the columns in the left-side inverse J of W are of length one:

$$J \cdot W = \mathbf{I} \tag{4.19}$$

$$\mathrm{diag}(J^T \cdot J) = \mathbb{1} \tag{4.20}$$

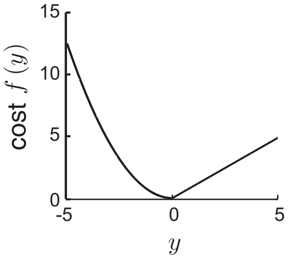


Fig. 4.2: Cost function for a single unit. The constant c is set to $c = 1$ and the threshold $\theta = 0$ and subthreshold minimum $y_0 = 0$.

However, we also performed some simulations where we restricted $\det(W) = 1$ and some simulation where we restricted W to rotations, $W^T \cdot W = \mathbf{I}$. But the results were not qualitatively different. To ensure the restriction in equation 4.20 we parametrized the inverse projection J by

$$j_{mn} = \frac{\sin b_{mn}}{M_n \left|\cos b_{mn}\right|} \tag{4.21}$$

$$\tag{4.22}$$

with the normalization

$$M_n = \sqrt{\sum_l \tan^2 b_{ln}}. \tag{4.23}$$

This parametrization was chosen because it showed superior convergence compared to others. The optimization is now done by gradient descent.[4]

The gradient $-A^T$ is calculated replacing the gradient by the directional derivative:

$$\mathrm{trace}(-A^T \cdot A) = \mathrm{trace}\left(\frac{\partial F(B)}{\partial B} \cdot A\right) = \nabla_A F(B) = \left.\frac{\partial}{\partial \tau} F(B + \tau A)\right|_{\tau=0} \tag{4.24}$$

---

[4] For $\det(W) = 1$ the parametrization was $W = B \cdot \det\left(B^{-1/N}\right)$ and for $W^T \cdot W = 1$ it was $W = e^{B - B^T}$.

where we can formulate the cost function

$$F = \operatorname{trace}\left(\frac{1}{2} Y \cdot Y_-^T + c \cdot Y \cdot \operatorname{sign}(R)^T - \theta \cdot \operatorname{sign}(R) \cdot \operatorname{sign}(R)^T\right). \quad (4.25)$$

The derivation of the signum function is zero everywhere but at zero. However, in physics there is no zero, just smaller than measurable. The last term can therefore be omitted in the derivation:

$$\begin{aligned} &\frac{\partial}{\partial \tau} \operatorname{trace}\left(\frac{1}{2} Y \cdot Y_-^T + c \cdot Y \cdot \operatorname{sign}(R)^T\right)\bigg|_{\tau=0} \\ &= \operatorname{trace}\left(\frac{\partial W(B+\tau A)}{\partial \tau}\bigg|_{\tau=0} \cdot X_P \cdot (Y_- + c \cdot \operatorname{sign}(R))^T\right). \end{aligned} \quad (4.26)$$

The deviation can be further resolved

$$\frac{\partial W(B+\tau A)}{\partial \tau}\bigg|_{\tau=0} = -W \cdot \frac{\partial J(B+\tau A)}{\partial \tau}\bigg|_{\tau=0} \cdot W, \quad (4.27)$$

so that the directional derivative becomes

$$\frac{\partial F}{\partial \tau}\bigg|_{\tau=0} = -\operatorname{trace}\left(\frac{\partial J(B+\tau A)}{\partial \tau}\bigg|_{\tau=0} \cdot Z\right) \quad (4.28)$$

with

$$Z = Y \cdot (Y_- + c \cdot \operatorname{sign}(R))^T \cdot W. \quad (4.29)$$

The derivation of a single element is

$$\frac{\partial j_{mn}}{\partial \tau}\bigg|_{\tau=0} = a_{mn} \frac{\operatorname{sign}(\cos b_{mn})}{\cos^2 b_{mn} \cdot M_n} - \sum_l a_{ln} \frac{\tan b_{ln} \cdot j_{mn}}{M_n^3 \cdot \cos^2 b_{ln}} \quad (4.30)$$

If we resolve equation 4.24 we get

$$a_{mn} = \frac{\sin b_{mn}}{M_n^2 \cos^3 b_{mn}} \left(\frac{z_{nm}}{j_{mn}} - \frac{\sum_l z_{nl} j_{ln}}{M_n} \right). \tag{4.31}$$

The encoding matrix W is calculated according to algorithm 1. At each step n a random subset of our data is chosen to calculate the local gradient A^n. We then search for the optimum τ^n by estimating a maximal step size based on the last τ^{n-1} and so loosely circling in on the optimum value.

$$\tau^n = \text{argmin}\ {}_\tau F(B^{n-1} + \tau \cdot A^n) \tag{4.32}$$

The randomization of the subset taken at each optimization step makes it harder to estimate the convergence, but at the same time it overcomes saddle points and local minima very fast.

Algorithm 1 Learning algorithm

```
Initialize random matrix $B^0$, $n = 0$
repeat
    $n$++
    Calculate $W(B^{n-1})$
    Chose random subset of the data $X_P$
    Calculate for the subset $Y = W \cdot X_P$
    Calculate $A^n$
    Perform a line search for the optimal step $\tau^n$
    Update $B^n = B^{n-1} + \tau^n \cdot A^n$
until Convergence
```

However, as mentioned above, it is possible to run this algorithm on-line. The changes would be to use a on-line PCA algorithm and to chose an appropriate update rule for the training data. However, it is costly as for each update step the inverse of J has to be calculated.

4.3.2. Non-Zero Threshold

With a training threshold $\theta \neq 0$ or a subthreshold minimum $y_0 \neq 0$, we encounter a problem: our cost function is no longer a continuous function. It has a discontinuity at the threshold in each dimension. Therefore the gradient descent algorithm will fail. A possible solution to this problem is a smoothing of the discontinuity by a sigmoid function $U_k(y)$. The cost function is therefore changed to

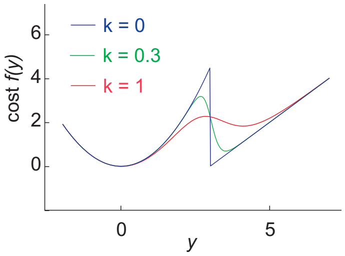


Fig. 4.3: Non-zero threshold cost function for a single unit. The constant c is set to one. the subthreshold minimum to zero, the threshold θ is 3, and the subthreshold minimum y_0 is 0. The factor k takes the values 0 (blue), 0.3 (green), and 1 (red).

$$f\left(y_i^t\right) = \frac{1}{2} U_k\left(y_i^t\right) \cdot \left(y_i^t - y_0\right)^2 \\ + c\left(1 - U_k\left(y_i^t\right)\right) \cdot \left(y_i^t - \theta\right) \tag{4.33}$$

with

$$U_k(y) = \frac{1}{2}\left(\tanh\left(\frac{1}{k}(\theta - y)\right) + 1\right). \tag{4.34}$$

The constant k has to be chosen of the same order as the distance between neighboring data points in each dimension around the threshold. The influence of k on the cost function can be seen in Figure 4.3.

If we now use the same algorithm to solve it we can just change the definition of the matrix Z:

$$Z = Y \cdot \left((Y - y_0) \circ U_k(Y) + \frac{1}{2}(Y - y_0)^{\circ 2} \circ V_k(Y) \right. \\ \left. + c\left(1 - U_k(Y)\right) - c\left(Y - \Theta\right) \circ V_k(Y)\right)^T \cdot W \tag{4.35}$$

where $\circ$ is the elementwise Hadamard multiplication and $^{\circ 2}$ the elementwise Hadamard sqare. We have further the function $V_k(y)$ being the elementwise derivative of sigmoid function $U_k(y)$

$$V_k\left(y\right) = \frac{dU_k}{dy} = -\frac{1}{2k}\cosh^{-2}\left(\frac{1}{k}\left(\theta - y\right)\right) \tag{4.36}$$

which both act elementwise. After this redefinition, we can use equation 4.31 to calculate the derivative A and applay the same algorithm.

4.3.3. Overcomplete Representation

Several data-driven learning algorithms have been extended by different authors to comply with overcomplete basis (Olshausen and Field, 1997; Lewicki and Sejnowski, 2000; Hyvärinen and Inki, 2002; Delgado et al., 2003). In the painting of Juan Sànchez Cotàn (Figure 4.1), he painted a quince, a cabbage, a melon, and a cucumber. But he could have painted thousands of other vegetables . Or fruits. Or animals. Or humans. One easily could imagine many more objects for a painting than the dimensionality of subspace spanned by all paintings. If we assume that the subspace spanned by our stimuli is of lower dimension than the number of possible hidden sources, we should work with overcomplete representations.

In the section 4.3.1 we assumed that the dimensions N_P and N are the same, i.e. the matrix W is quadratic. However, by definition this restriction is not needed, we only have to restrict $J \cdot W = \mathbf{I}$. But for a non-quadratic NxN_P matrix W with $N > N_P$ there is an infinite number of left-side inverse J and vice versa. Which one to chose? The solution is to let the algorithm also. We define two additional matrices J_0 and W_0, as folows:

$$\begin{bmatrix} J \\ J_0 \end{bmatrix} = \begin{bmatrix} W & W_0 \end{bmatrix}^{-1}. \tag{4.37}$$

Actually we are not interested in W_0 at all. What we need is J_0 as it defines the orthogonal complement to W and therefore W itself for a given J. The parameter matrix B now is a square matrix of size NxN that parametrizes the matrix $\begin{bmatrix} J \\ J_0 \end{bmatrix}$ by

$$j_{mn} = \frac{\sin b_{mn}}{M_n \left|\cos b_{mn}\right|} \tag{4.38}$$

$$j_{0mn} = \frac{\sin b_{(m+N_P)n}}{M_n \left|\cos b_{(m+N_P)n}\right|} \tag{4.39}$$

with the same normalization as equation 4.23,

$$M_n = \sqrt{\sum_l \tan^2 b_{ln}}. \tag{4.40}$$

When calculating the derivative (equation 4.24) we can define the matrix Z as in equations 4.29 (for $\theta = 0$ and $y_0 = 0$) or 4.35 (else). The derivative finally gets

$$a_{mn} = \begin{cases} \frac{\sin b_{mn}}{M_n^2 \cos^3 b_{mn}} \left(\frac{z_{nm}}{j_{mn}} - \frac{\sum_l z_{nl} j_{ln}}{M_n} \right) & |m \leq N_p \\ \frac{\sin b_{mn}}{M_n^2 \cos^3 b_{mn}} \frac{z_{nm}}{j_{mn}} & |else. \end{cases} \tag{4.41}$$

4.3.4. Low-Density Receptive Fields

To master the density of STRFs (sharply tuned versus broadly tuned neurons) we run a few simulations with an additional term in the cost function, corresponding to a linear cost on absolute synaptic weights,

$$W = \arg\min{}_{W'} \sum_{i,t} f(y_i^t) + c_s \frac{\sqrt{N} s_b}{\sqrt{N_0} \left\|P\right\|_F} \left\|\xi\right\|_1, \tag{4.42}$$

where s_b denotes the batch size (number of cochlear input samples per weight update), $\|.\|_F$ the elementwise Frobenius-norm, and $\|.\|_1$ an elementwise 1-norm and c_s is the relative weight of the new term.

To implement the low-density RF we have to adjust the matrix Z in equation 4.29 (or equation 4.35 accordingly) to

$$Z = \left(Y \cdot (Y_{-} + c \cdot \mathrm{sign}(R))^T + c_s \frac{\sqrt{N} s_b}{\sqrt{N_0} \left\| P \right\|_F} \xi \cdot \mathrm{sign}(\xi)^T \right) \cdot W. \quad (4.43)$$

It is possible to implement further optimization targets into the cost function without great efforts. Possible targets would be to minimize correlations between different neurons for a certain time shift, as the algorithm right now does not care about temporal continuity and only cares about single frames. Each additive cost adds one term to the matrix Z and can be calculated with ease as in the aforementioned examples.

4.3.5. Inclusion of Temporal Features

What we left out until now are temporal features, any equation presented so far works on the input matrix X where single data points X^t are represented as columns and any permutation of columns will lead to exactly the same result, except that the columns of Y are equally permuted. But natural stimuli are not a sequence of static stimuli that are randomly switched in zero time. One possible approach of taking into account temporal statistics is by replacing data points X^t with sliding windows $X^{t:t+\tau}$ of length τ. If we whiten our data first we then get

$$X_p^t = P \cdot X^{t-\tau:t} \quad (4.44)$$

and have reduced the window to a single time step again. However, be aware that even though the windows might overlap, the algorithm will treat them as no more related than any two windows in the training set.

4.3.6. Reconstruction Error and Decoding

The first term of the cost function in Equation 4.17 imposes a linear cost on suprathreshold synaptic currents. Because suprathreshold synaptic currents are equivalent to instantaneous firing rates at zero noise, the first term enforces firing sparseness across the population (for the training threshold). The second term imposes a quadratic cost on subthreshold synaptic currents, which is equivalent to minimizing an error bound on decoded cochlear inputs. To see this, consider the following estimate $\hat{X}_p^t$ of whitened cochlear inputs $X_p^t = PX^{t-\tau:t}$ from suprathreshold synaptic currents at time t:

$$\hat{X}_p^t = J\left(Y_+^t + Y_{E-}^t\right), \tag{4.45}$$

where Y_{E-}^t is the vector with the expected subthreshold currents: $y_{i,E-}^t = y_{E-}$ for $y_i^t < \theta$ and $y_{i,E-}^t = 0$ otherwise.

The decoding error associated with the decoding $\hat{X}_p^t$ is given by the mean square Euclidean norm between $\hat{X}_p^t$ and X_p^t. This error is related to the distribution of subthreshold currents as follows:

$$\begin{aligned}
\sum_t \left\| X_p^t - \hat{X}_p^t \right\|_2^2 &= \sum_t \left\| JY^t - J\left(Y_+^t + Y_{E-}^t\right) \right\|_2^2 \\
&= \sum_t \left(Y_-^t - Y_{E-}^t\right)^T J^T J \left(Y_-^t - Y_{E-}^t\right) \\
&= \sum_t \left(Y_-^t - Y_{E-}^t\right)^T (\mathbf{I} + C) \left(Y_-^t - Y_{E-}^t\right) \\
&\simeq \sum_t \left(Y_-^t - Y_{E-}^t\right)^T \left(Y_-^t - Y_{E-}^t\right) \\
&= \sum_{y_i^t < \theta} \left(y_i^t - y_{E-}\right)^2
\end{aligned} \tag{4.46}$$

where in the first line we have used that $X_p^t = PX^{t-\tau:t} = JY^t$, and in the third line $\mathbf{I}$ represents the identity matrix and C a matrix with zeros on the diagonal. The approximation in the third line is based on the assumption of equally distributed and mutually independent subthreshold currents: $p(y_i, y_k) = p(y_i)p(y_k)$. Note that the approximation in Equation 4.46 is exact for rotations ($J^T J = \mathbf{I}$), whereas for the constraint $\mathrm{diag}(J^T J) = \mathbb{1}$, we found the approximation to be within 4% of the reconstruction error for $\theta = 0$, and to be even closer for higher θ.

The key insight is that for the training threshold θ, the approximated reconstruction error is proportional to the subthreshold term of our cost function if we choose $y_0 = y_{E-}$ in Equation 4.17. Hence, by minimizing the subthreshold term in our cost function, we minimize the approximation of the decoding error. The final term in Equation 4.46 shows that the reconstruction error is small when the subthreshold currents are rare ($p\left(y<\theta\right)$ is small) and their variance is small ($\left\langle\left(y-y_{E-}\right)^2\right\rangle_{y<\theta}$ is small).

Note that the decoding scheme defined in Equation 4.45 may not be globally optimal, but it is motivated by our assumption that only suprathreshold events carry meaningful information. Also, a benefit of this decoding scheme is its asymptotic robustness (for a threshold of minus infinity the decoding error vanishes).

Our algorithm minimizes an approximation of the reconstruction error, but not the reconstruction error itself (Equation 4.46). Exploratively, we have adapted the algorithm to directly minimize the reconstruction error (defined in Equation 4.46) for a given threshold $\theta > 0$. The resulting reconstruction error for BOS at the given threshold was only marginally better than with our sparse-coding algorithm. Synaptic current distributions were nearly bimodal with a first peak at zero and a second peak slightly above θ; interestingly, reconstructions became worse than in Figure 5.10E when performed using a different threshold from the one used during learning.

4.3.6.1. Reconstructed Spectrograms

From the decoded whitened inputs $\hat{X}_p^t$, the spectrograms (Figure 5.10) are reconstructed using the pseudoinverse $P^{-1} = E\Lambda^{1/2}$. The element t of the reconstructed spectrogram was defined by

$$X_{rec}^t = E\Lambda^{1/2}\hat{X}_p^t. \tag{4.47}$$

From these elements, the fully reconstructed spectrograms are computed by averaging over all overlapping regions in the sequence X_{rec}^t, X_{rec}^{t+1}, $\ldots$, i.e. the overlapping time slices.

It is also possible to reconstruct the spectrogram probabilisticly from the firing rates R:

$$E(X|R) = \int X \cdot p(X|R)dX = \frac{\int X \cdot p(X)\prod_t \frac{p(X|R^t)}{p(X)}dX}{\int p(X)\prod_t \frac{p(X|R^t)}{p(X)}dX} \tag{4.48}$$

given that $p(R|X) = \prod_t p(R^t|X)$, which is fulfilled by $p(R^t|X)$ being delta peaks in the noiseless case and independent Gaussian distributions with variance k^2 in the noisy case, thresholded at θ. We can further make the approximation that the posteriors $p(X|R^t)$ are much sharper defined than the prior $p(X)$: $p(X)\prod_t \frac{p(X|R^t)}{p(X)} \approx \alpha \cdot \prod_t p(X^{t-\tau:t}|R^t)$, where α is a proportionality factor. The pointwise approximation then becomes:

$$E(x_i^t|R) = \frac{\int x_i^t \prod_{u=0}^{\tau} p(x_i^t|R^{t+u})dx_i^t}{\int \prod_{u=0}^{\tau} p(x_i^t|R^{t+u})dx_i^t} \tag{4.49}$$

In the case of a threshold $\theta \to -\infty$ and no noise, the expected value $E(x_i^t|R^{t+u})$ is $(\xi^{-1})_i^u \cdot Y^{t+u}$ with the variance of projections onto the orthogonal complement of the PCA-space ${\sigma_i^{t|u}}^2 = \left(C - E^T \cdot \Lambda \cdot E\right)_{ii}^{uu}$, under the assumption that projections onto PCA-space and projections onto its orthogonal complement are independent. If we have imperfect information, i.e. a finite threshold and noise, the expected value becomes $(\xi^{-1})_i^u \cdot \left(Y_+^{t+u} + Y_{E-}^{t+u}\right)$ (see section 4.3.6) and the variance is augmented by variance of the unknown elements of Y^{t+u} and the noise, ${\sigma_i^{t|u}}^2 = \left(C - E^T \cdot \Lambda \cdot E\right)_{ii}^{uu} + (\xi^{-1})_i^u \cdot \Sigma(Y^{t+u}|R^{t+u}) \cdot {(\xi^{-1})_i^u}^T$, with $\Sigma(Y^{t+u}|R^{t+u}) = k^2 \cdot \mathbf{I} + \Sigma_-^{t+u}$ being the covariance matrix of the thresholded synaptic currents given the firing rate, with Σ_-^{t+u} the covariance matrix of the subthreshold currents. The expected value of x_i^t under this assumptions becomes

$$E(x_i^t|R) = \frac{\sum_{u=0}^{\tau} \frac{E(x_i^t|R^{t+u})}{{\sigma_i^{t|u}}^2}}{\sum_{u=0}^{\tau} \frac{1}{{\sigma_i^{t|u}}^2}}. \tag{4.50}$$

Compared to the averaging of the windows proposed above, we end up with a weighted averaging, weighted by the inverse of the variance. However, reconstructions based on this weighted averaging are only marginally better than when simply averaged.

Now that we have written down the equations and the algorithm, we shall discuss in the next chapter the implementation and the characteristic behavior when applying the method to natural data.

Chapter 5

Sensory Modeling Using Nonsymmetric Sparse Coding

Il y a aussi deux sortes de vérités, celles de Raisonnement et celle de Fait. Les vérités de Raisonnement sont nécessaires et leur opposé est impossible, et celles de Fait sont contingentes et leur opposé est possible.

Gottfried Wilhelm Leibniz, La monadologie

5.1. An Overview over Sensory Modeling

It is an ongoing discussion on how to filter for relevance and in what code to present the data to later stages. One of the earliest proposals was made by James (1890) who suggested the existence of a "pontifical cell". In his model all neurons would have to report (directly or indirectly) to this one cell, to which also our consciousness is bound to. Information about the whole world would be needed to fit in a single sequential code. A new idea was introduced by Barlow (1972) replacing the one "pontifical cell" by a number of "cardinal cells". No longer should one single neuron encode all information, but a set of neurons should each encode for a specific percept:

> *Among the many cardinals only a few speak at once; each makes a complicated statement, but not, of course, as complicated as that of the pontif if he were to express the whole of perception in one utterance.*

As an example Barlow mentions the infamous “grandmother cell”, a cell that response to all views of a grandmother’s face. But he does not think of this cell responding solely, but simultaneously with other “cardinal cells” that respond to position, surrounding etc. The number of these “cardinal cells” is estimated very high:

> [The collage of these cardinals] *...must include a substantial fraction of the* 10^{10} *cells of the human brain.*

The new idea was to have a set of neurons each one representing a cause underlying a sensory stimulation, very similar to how someone would describe a situation, and not just single pixels or elements. But how should the sensory system be built to lead to such sensations?

Todays models are still far away from answering this complex question. They rather try to solve small pieces of this puzzling question. The design of a model can coarsely be attributed to one of two groups: bottom-up design and top-down design. The biologically inspired bottom-up design starts at the very ground level. Depending on how and what to model, this is often a single neuron (such as Hodgkin and Huxley, 1952) or even single axons and dendrites (Bressloff, 1995). These models try to explain the biophysics of a system as precise as necessary and, by aggregating them to larger systems, search for the tasks these systems can perform.

The second approach, top-down, is rather mathematically inspired. The starting point for such models is the functionality of the system. Its creator defines the tasks/output of the system and a learning rule for the system, that guarantees him success. Then we look into the system analyzing the behavior of subunits (e.g. neurons) and compare it to the behavior of its biological counterpart.

Hubel and Wiesel (1959) were the first to describe the functionality of sensory neurons by receptive fields, focusing on simple cells in the primary visual cortex of cats. One of the first approaches to model these receptive fields was by Bossomaier and Snyder (1986). They were philosophizing about how the brain could best decorrelate a picture in order to reduce redundancy. In the case of infinite size pictures with shift-invariant statistics, PCA (the optimal decorrelator) is equivalent to the Fourier transformation. However, it turns out that real pictures do not follow exactly these statistics and a better decorrelation can be obtained by localized Fourier transformation, in their case by Gabor functions, leading to the maximal space-frequency resolution. This model of simple cells in the primary visual cortex as Gabor filters holds still today as a good approximation.

In 1992 Hancock et al. (1992) really calculated the first few principal components of a set of natural 64x64 grayscale pixel images. The resulting receptive fields were not exactly Gabor filters, they could better be described as (zeroth to n-th order) derivatives of two-dimensional Gaussians. As they discuss in the conclusions, PCA may not exactly be what the visual cortex does.

The same year Atick (1992) published a very interesting review paper, where he proposed neural coding strategies based on information theory. Instead of decorrelation he introduced a maximization of the code entropy, ideally with independent symbols. He applies this idea on the example of retinal output. However, he did not use real pictures, but made statistical assumptions about them.

Five years later Bell and Sejnowski (1997) took up the idea of maximizing the entropy for sensory modeling. They applied their infomax algorithm (Bell and Sejnowski, 1995) to real images and found that most of their units were edge detectors at different positions and orientations, much like simple cells in primary visual cortex of cats and monkeys.

In parallel there was the work by Olshausen and Field (1996). Their approach was a different one. Neurons should maximize the entropy, they should maximize the sparseness of their output while maximizing the information about the original stimulus. The outcome was very similar to Bell and Sejnowski (1997), which is not so surprising, as the two algorithms are closely related[1]. A direct advantage of this approach is the free number of dimensions. In contrast to many ICA algorithms that need a square projection matrix, their number of outputs has no limitations, which makes it preferable for neural modeling. For example primary visual cortex has much more neurons than its input LGN and its coding can be modeled as an overcomplete representation of the input (Olshausen and Field, 1997).

---

[1] see chapter 4.2

However, sensory processing is not a one-layer story, but includes many, complexly wired areas[2]. Hyvarinen and Hoyer (2001) created a two-layer model with two layers of identical size. The cells in the layers were topographically arranged as 25x25 torus, and the connections between layer one and two were fixed by layer-two cells getting input from the 25 nearest layer-one cells. Layer-one cells performed a rotation on the whitened input and rectified their output. The learning algorithm optimized this rotation in order to maximize the output sparseness of layer-two cells. The results were topographically arranged layer-one cells. Cells tuned to similar orientation sat next to each other. Layer-two cells thus were phase insensitive, very much like the complex cells found in V1.

A new interesting approach was made by Smith and Lewicki (2006). Using a matching pursuit algorithm they decomposed waveforms of natural sounds and speech. The resulting kernels after training highly resembled the transfer function obtained from cat auditory nerve fibers. However, it remains an open question whether this method would also be successful in predicting behavior of cortical neurons.

5.2. The Model

The following two section is mainly reproduced from the publication Blättler and Hahnloser (2011).

We model the auditory pathway of the zebra finch as a feedforward network that receives auditory input from the cochlea in the form of spectral temporal sound patterns[3]. These patterns are multiplied by synaptic weights and summed, to result in the total synaptic current impinging onto neurons. Mathematically, the set of synaptic weights onto a neuron represent its spectral temporal receptive field (STRF). We devised an algorithm that optimizes synaptic weights for their propensity to decorrelate (whiten) and sparsify cochlear inputs: First, we whitened cochlear inputs using a projection matrix P (principal component analysis, PCA), and then we sparseness-transformed the inputs using a matrix W that minimizes an asymmetric cost imposed on total synaptic currents, equation

---

[2] see chapter 3 for an example

[3] The cochlear input is approximated by a log-power spectrogram. It is a very rough, however useful approximation (Gill et al., 2006)

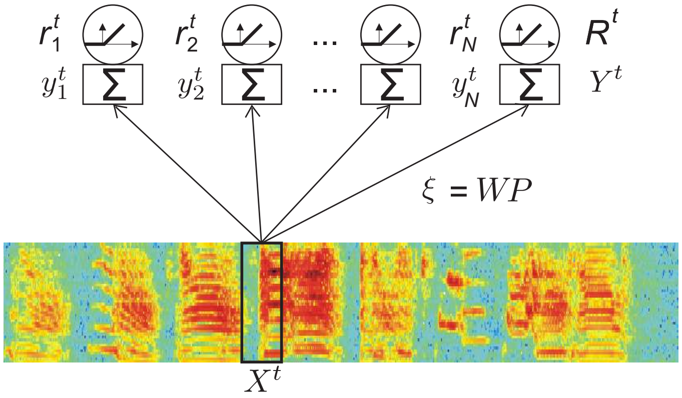


Fig. 5.1: Schematics of the model. At time t, the auditory input to the network is a 50-ms window $X^{t-\tau:t}$ of the sound spectrogram. This input is multiplied by synaptic weights $\xi = WP$ to result in total synaptic currents $Y^t = (y_1^t, ..., y_N^t)$ onto N neurons. P stands for whitening and dimensionality reduction (principal component analysis), and W stands for a sparseness transformation. Neural firing rates $R^t = (r_1^t, ..., r_N^t)$ are given by rectified synaptic currents.

4.17 and Figure 4.2. In this cost, hyperpolarizing synaptic currents are punished quadratically, whereas depolarizing currents are punished linearly. Intuitively, to minimize the cost, weak and frequent cochlear inputs must be hyperpolarizing (such that the quadratic cost is smaller), whereas strong and rare inputs must be depolarizing (such that the linear cost is smaller). We defined firing rates by simple rectification of total synaptic currents at variable firing thresholds. The linear cost of depolarizing currents is in effect a cost on the average population firing rate (at the learning threshold); and, the quadratic cost of hyperpolarizing currents is a cost on a reconstruction error associated with simple decoding of the original cochlear inputs from firing rates. In simple words, the algorithm tries to maximally sparsify population responses without discarding any relevant sensory information. We minimized the cost function over training data consisting mostly of renditions of a particular zebra finch song (BOS) and a few CONs. After training, we evaluated the network for a wide range of firing thresholds.

Most of the data we are going to present in this chapter is produced by a training set of 34 files containing BOS and 12 files with containing CON, produced by two different birds, 6 files each bird or a training set of 34 BOS-files (different bird) and 44 CON-files of 22 birds. The spectrograms produced from this files had a temporal resolution of 0.7 ms and a spectral resolution of 86 Hz. The sliding windows used had a temporal span of $\tau =$ 50 ms (64 samples) and a spectral span from 0 to 11 kHz (128 frequency bands), leading to a total dimensionality of $N_0 = 8192$ of the input space. If nothing else is mentioned the threshold θ, the subthreshold minimum y_0 as well as the density weight c_s where all set to zero, and sparseness factor c was set to one.

5.3. Results

5.3.1. Spectral-Temporal Receptive Fields and Ensemble Modulation Transfer Functions

In the contemporary literature we find a variety of publications describing measured STRFs in the zebra finch's areas DLM, Ov, Field L, and CM, mainly by the groups of Theunissen and Woolley (Theunissen et al., 2000; Sen et al., 2001; Theunissen et al., 2001; Hsu et al., 2004; Theunissen et al., 2004; Woolley et al., 2005; Gill et al., 2006; Woolley et al., 2006; Nagel and Doupe, 2008; Woolley et al., 2009; Amin et al., 2010). After training, our model neurons displayed a large diversity of STRFs. Typically, STRFs were patchy and had multiple adjacent inhibitory and excitatory spectral/temporal subfields. In many neurons, STRFs were regularly arranged into horizontal or vertical stripes (Figure 5.2A), similar to receptive fields in field-L neurons that encode elementary spectro-temporal sound features such as sound onsets or a particular sound pitch.

Model STRFs had excitatory and inhibitory subfields that together covered the entire spectro-temporal window of the STRF. Typically STRFs in field L are of considerably lower density in that they mostly possess only two or three subfields instead of more than six. We therefore explored whether STRFs in our model would be of lower density when we added a third term to the cost function, a term corresponding to a linear cost on absolute synaptic weights. We found indeed low-density STRFs if the parameter c_s weighing this third term exceeded roughly 0.1, Figure 5.2B. Density of STRFs could be controlled independently of the sparseness of model responses.

We explored the correspondence between STRFs and the stimulus features to which neurons responded most. In most cells, presentation of

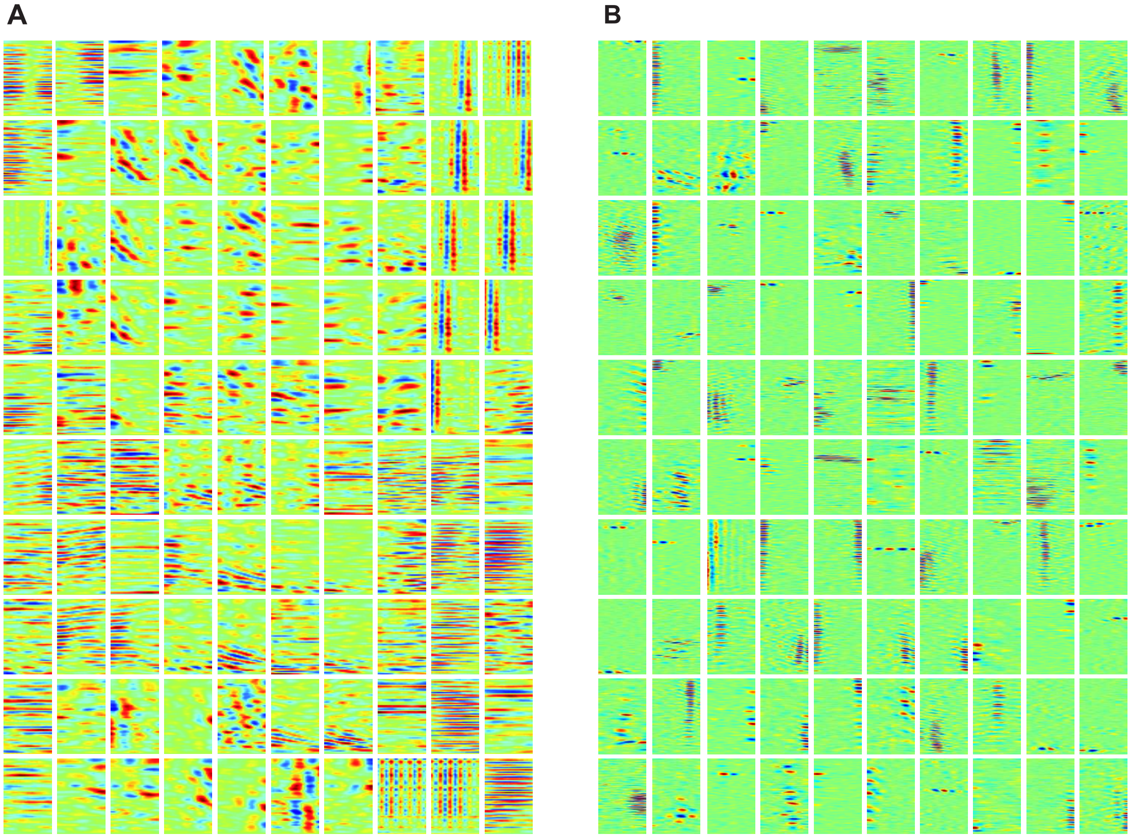


Fig. 5.2: Spectral temporal receptive fields. (A) Spectral temporal receptive fields (STRFs) of $N = 100$ neurons, arranged by nearest-neighbor similarity (circular boundary conditions). Neurons tend to be either temporally tuned (vertical stripes, top right), spectrally tuned (horizontal stripes, middle rows), or display more complex spectro-temporal patterns. Spectral resolution is 172 Hz, offset between subsequent cochlear inputs is 1.5 ms. (B) STRFs obtained with a linear cost on synaptic weight magnitudes. The linear cost forces many synaptic weights to be close to zero (green), leading to low-density STRFs most of which contain a smaller number of excitatory and inhibitory subfields than in A. Interestingly, excitatory and inhibitory subfields tend to be close to each other and aligned horizontally or vertically, similar to observations in field L neurons. The 100 presented STRFs were randomly chosen out of the total 800. $c_s = 0.2$.

different BOS versions elicited reliable responses to specific song notes, Figure 5.3A-C. For example, the total synaptic current in neuron 10 with a checkerboard-like STRF reliably peaked after the down sweep of the introductory note and to a lesser extent it also peaked at the offsets of some other syllables. Neuron 23 with a narrow and slanted STRF responded most strongly to the down-sweep of the harmonic stack in Syllable A1. The STRF of Neuron 88 had sharp vertical subfields, the synaptic current to this neuron peaked during rapid increases of sound intensity such as during the onsets of Syllables C and D. Another neuron with a vertically dominated STRF (Neuron 131) responded a few milliseconds after the onsets of Syllables E and F. This neuron was able to respond to different syllable onsets than Neuron 88 by virtue of its sensitivity to a low-frequency tone immediately followed by a high-frequency tone, which is a common characteristic of both Syllables E and F. Very particular was Neuron 106. Its receptive field and that of several other neurons were centered on a single frequency band close to 7 kHz. It turned out that this cell responded to electrical noise by which our recordings were affected; during BOS presentation, the total synaptic current to this cell was small and increased mainly during syllable gaps where no signal except the noise was present. The structure of the sparsely checkered STRF of Neuron 121 was particularly well adapted to Syllable E, the neuron responded almost exclusively during the transition between sub-Syllables E1 and E2. Finally, Neurons 55 and 147 did not show either robust or strong responses to BOS; a more thorough analysis revealed that they responded strongly to a CON in the training set. When we ordered all neurons by the time at which their synaptic currents peaked in response to a particular version of BOS, we found that the resulting stack plot exhibited a staircase-like shape: peak currents were widely distributed across cells with many more peaks seen during syllables than during syllable gaps, Figure 5.3D.

Some neurons were not just tuned to a particular syllable within the motif, but had even more specific tuning to particular subsets of that syllable. Finches often produce harmonic stack syllables and can subtly vary the pitch of these syllables in a well-controlled and goal-directed manner (Tumer and Brainard, 2007; Andalman and Fee, 2009). When we trained a network of 196 neurons on the songs of a bird produced during an entire day, we found that synaptic currents of two neurons peaked during a harmonic-stack syllable produced by that bird (neurons with such harmonic-stack receptive fields have been illustrated in Amin et al. (2010)). Interestingly, for any given stack syllable, the synaptic current in only one of the neurons peaked, but not in both, Figure 5.4. The two neurons divided the representation of that syllable between each other, one represented the high-pitch version of the stack, the other the low-pitch

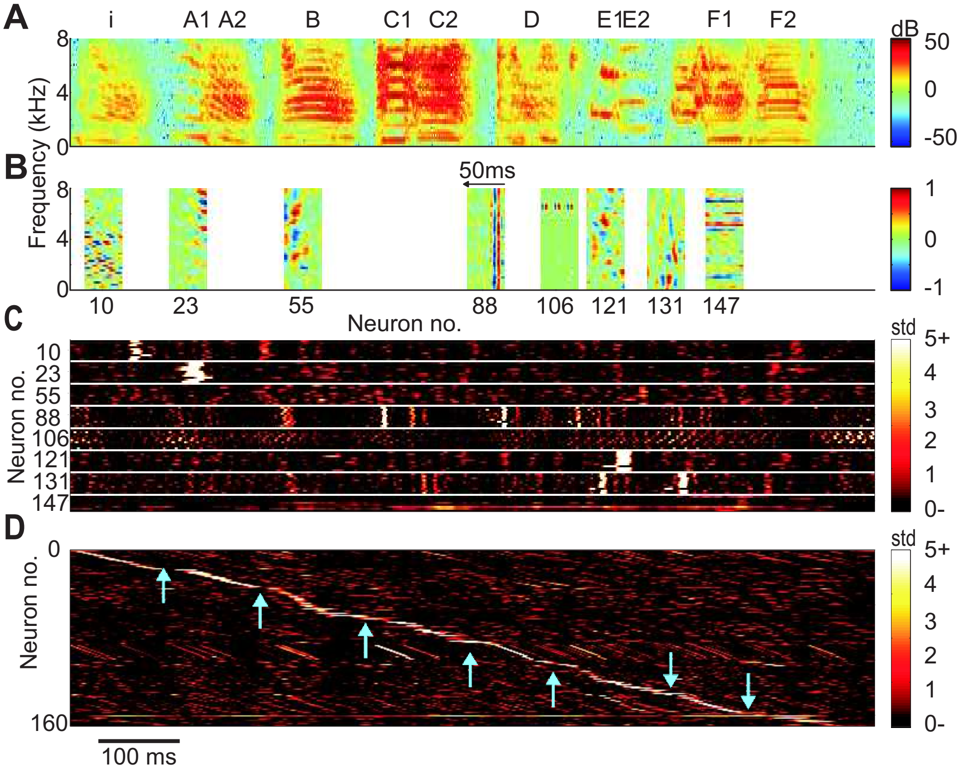
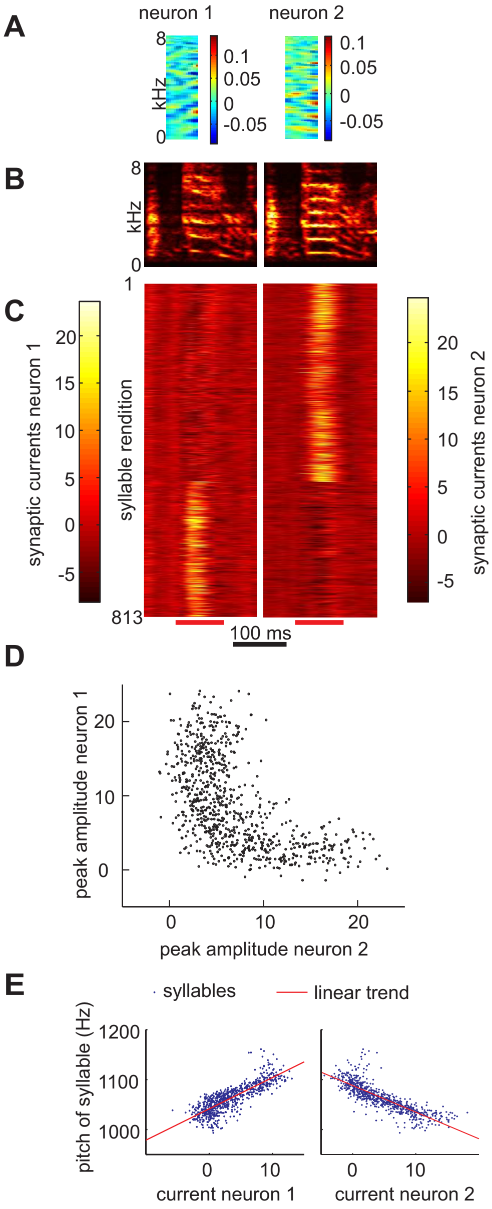


Fig. 5.3: Receptive fields and neurogram. (A) Power spectrogram of a bird's own song (BOS). (B) STRFs ξ_i of 8 representative neurons (i = 10, 23, 55, 88, 106, 121, 131, 147). The horizontal alignment of STRFs with the spectrogram in A is such that the trailing edges of the STRFs correspond to the respective peak times of synaptic currents. The temporal axis of the STRFs is inverted for better comparison with the BOS spectrogram. (C) Synaptic currents of representative neurons in B in response to 10 different versions of BOS, vertically aligned to A. (D) Neurogram of synaptic currents in response to the BOS in A. The $N = 160$ neurons are sorted according to the peak times of their synaptic currents. Fewer neurons display synaptic current peaks during syllable gaps (blue arrows) than during syllables.

versions, Figure 5.4E. Hence, our algorithm is able to 'allocate' more than a single neuron to a song feature, depending on the extent of its variability.

STRFs are optimal models of the linear part of neural responses. Can we

Figure 5.4: Neurons encode behavioral variability, for example song pitch. (A) Two receptive fields formed by training a network on all songs produced by a bird on a single day. (B) Spectrograms of a song syllable containing a harmonic stack. The left version has median pitch 1024 Hz, the right version 1138 Hz. (C) Stack plot of synaptic currents in the two neurons elicited by 813 syllable renditions. The stack plots have been sorted identically to reveal that for a given syllable rendition either the left or right neuron exhibits a peak in synaptic current, but not both. Peaks in synaptic currents are computed in intervals indicated by red bars on the bottom. (D) Scatter plot of peak synaptic currents in the two neurons. The distribution is sparse ('L'-shaped). (E) Median synaptic current in same intervals versus median song pitch of the harmonic stack. The two neurons are detectors of low and high pitch versions of the stack, respectively. Red and blue lines are linear regressions (Neuron 1: $R^2 = 0.68$, $p < 10^{-170}$, Neuron 2: $R^2 = 0.69$, $p < 10^{-185}$), $N = 196$, $y_0 = 0$.



recover the STRFs found by our algorithm in simulated neural responses which contain a threshold nonlinearity? To this end, we estimated STRFs from nonlinear responses to birdsong stimuli for a range of firing thresholds θ. We estimated STRFs using reverse correlation (equation 2.6). Mostly, we found strong resemblance between estimated and actual STRFs, Figure 5.5. Strong resemblance was seen in all cases in which the correlation coefficient cc between estimated and actual responses was above 0.1, i.e. in cases in which the linear model was reasonably good. Moreover, estimated STRFs did not change much with increasing firing threshold θ, except that with increasing θ the STRFs had a small tendency to extend over larger time-frequency regions (Figure 5.5B) than the original low-density STRFs (Figure 5.5A). We found similarly satisfying results when estimating high-density STRFs (not shown). Hence, estimated STRFs were quite insensitive to the nonlinearity and to changes in firing threshold.

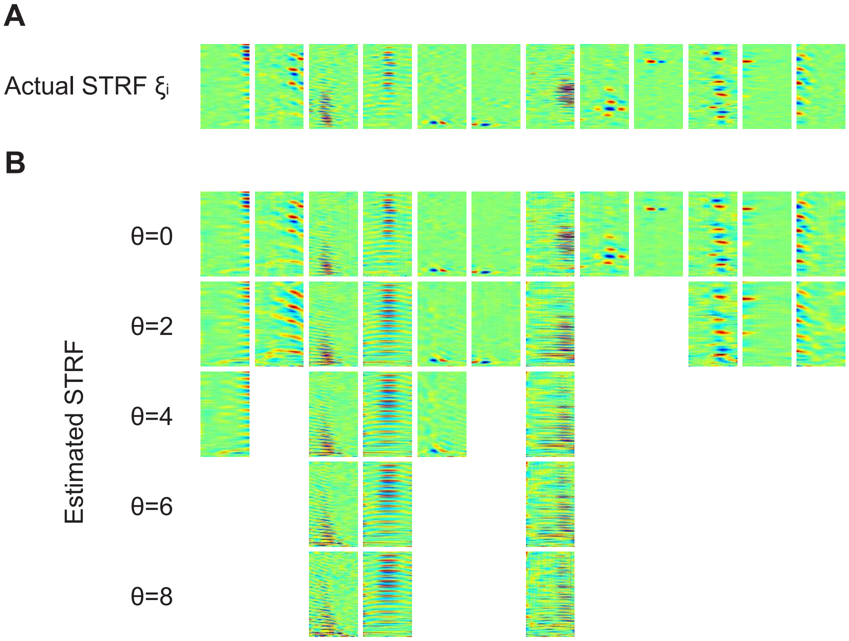


Fig. 5.5: Actual STRFs and STRFs estimated using reverse correlation. (A) A selection of twelve STRFs ξ_i obtained after convergence of the algorithm ($N = 800$). (B) Estimated STRFs (reverse correlation) based on the predicted firing rates r_i^t. Shown are only estimated STRFs for neurons associated with a correlation coefficient cc between predicted and actual firing rates of $cc > 0.1$. $\mu = 1$. $c_s = 0.2$.

An interesting feature of a sensory system as a total is its eMTF (see section 2.2.4). This measure describes the modulations of the stimulus

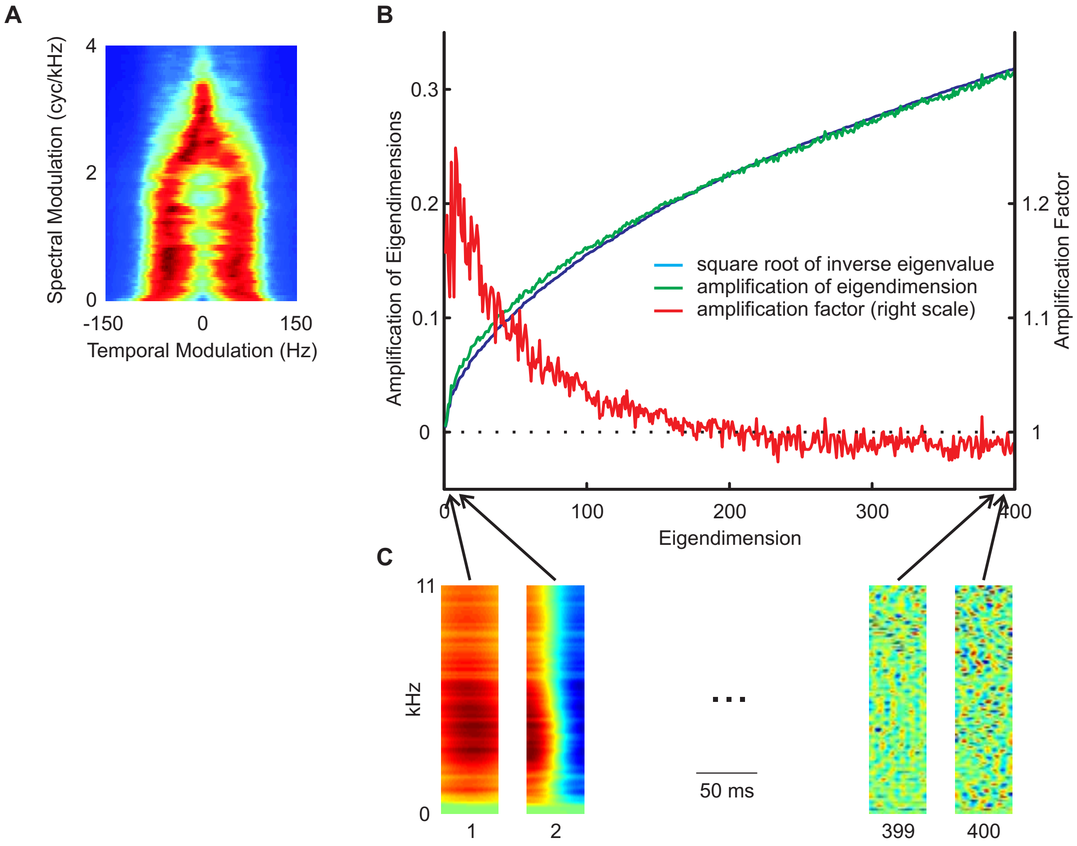


Fig. 5.6: Ensemble transfer modulation function and modulation amplification. (A) Example of an eMTF gained from a simulation of 400 neurons. The amplitude is color coded and in arbitrary units. The main energy lies between -80 Hz and +80 Hz with a small gap around 0 Hz temporal modulation and up to 2 cycles/kHz spectral modulation. Between 2 and 3 cycles/kHz a lot of energy is around 0 Hz temporal modulation. (B) Example of the amplification of eigendimensions of the birdsong. (blue, left scale) Due to whitening the different eigendimensions get amplified by the inverse square root of the corresponding eigenvalue (equation 4.4. (red, right scale) The sparsifying matrix W is not restricted to rotations and each eigendimension is further amplified. Eigendimensions with high eigenvalue thereby are amplified, while eigendimensions with low eigenvalue are slightly suppressed. (green, left scale) The total amplification due to PCA and sparsification. (C) Examples of eigenvectors. The eigenvectors resemble Fourier components. The first eigenvalue is similar to the zero-component (on-off), with the color-coded amplitude displaying the variance of the pixels. The second eigenvectors resembles the zero-one (spectral-temporal) component. In contrast, the 399th and 400th eigenvectors show high spectral and temporal modulations.

mostly encoded by the system as a total, independent of the absolute temporal or spectral position. Woolley et al. (2005) found similar eMTFs for MLd, field L and CM: all of them had their main power between −60 Hz and +60 Hz for the temporal modulation, with a clear gap between −5 Hz and +5 Hz, slightly more pronounced upstream. The temporal modulation was very limited for MLd and CM and went only up to roughly 0.2 cycles/kHz, while for field L it goes up to roughly 1 cycle/kHz.

The eMTF estimations gained from our models (Figure 5.6A) are very similar: the temporal modulation is slightly wider from −80 Hz to +80 Hz, but shares the similar gap around zero. One difference is the spectral span up to 2 cycles/kHz. But the main difference is a purely spectral modulation (around 0 Hz temporal modulation) between 2 and 3 cycles/kHz which is unseen in experimental literature.

A possible explanation would be insufficient data for the estimation of STRFs in the experiments. As explained in section 2.2.2, overfitting is a serious problem in the estimation processes. The methods used to avoid it smooth the STRFs and thereby selectively remove high modulation frequencies. When this is done using a jackknife filtering will be more sever when data is noisy and/or little. This is partly supported by the fact that the field L shows higher spectral modulations as the other nuclei[4], in correspondence to their respective firing rates and therefor the available data.

The question however remains how the model ends up with such eMTF. If we assume that the stimuli are scale (in time and frequency) invariant than the PCA step of the algorithm would be identical to the Fourier transform. But because we multiply it with the inverse of the squarer root of the eigenvalues, we would get an eMTF of the whitening filters that grows from the origin to the boarder. And as we reduce the dimensionality, there would be a hard cut at some value and everything further from the origin would be zero. And under any rotation in the filter space does not change the eMTF. Now are the stimuli not scale invariant, but even so eigenvectors with high eigenvalues tend to represent very low frequency modulations while low eigenvalue eigenvectors tend to be more complex (Figure 5.6C. The interesting part is that the sparsening matrix

---

[4] How should high modulation frequency components arrive and be represented in field L when they where not present in the downstream nucleus MLd

W is not restricted to rotations. So the total amplitude with which an eigendimension is encoded can be changed by sparsification. This change is represented by the square root of the diagonal of the matrix W^TW, depicted in Figure 5.6B. Low-frequency components get amplified (maximum amplification around the 10th eigenvector) and the high-frequency components get suppressed, or in total less amplified than by whitening alone. This suppression/amplification is in practice not unlimited. However in the overcomplete case, where this limitation is looser, the qualitative form of the red curve does not change.

5.3.2. Synaptic Currents and Firing Rates

The distribution of total synaptic currents over all neurons and over all training stimuli was highly asymmetric and contained many positive but few negative outliers, Figure 5.7A. The distribution of BOS-evoked currents could be reasonably well approximated by a unit Gaussian on the negative side and a long-tail exponential on the positive side. This combined Gaussian-exponential behavior follows from the fact that minimization of the quadratic-linear cost function is equivalent to locally maximizing the log-likelihood of a Gaussian model density below the threshold and of an exponential model density above the threshold, under the global restriction of zero mean and fixed variance. Interestingly, large synaptic currents were mostly elicited by the BOS rather than by other stimuli, illustrating that neurons were best tuned to the features of the most prominent stimulus in the training set, which was the BOS. The same finding was true for low-density STRFs (when $c_s > 0.1$ instead of zero): BOS-elicited currents exhibited a heavier positive tail than currents elicited by CON and REV (Figure 5.7B). Hence, model responses were robustly tuned for the BOS.

We also explored the influence of other model parameters such as y_0, which sets the location of minimal quadratic cost. When we changed y_0 to nonzero values different from the firing threshold θ during training, we found that BOS tuning of synaptic currents was qualitatively unchanged. The only effect of changing y_0 was to slightly increase the distribution of synaptic currents around y_0 (where cost is minimal) and to slightly decrease it around θ (not shown).

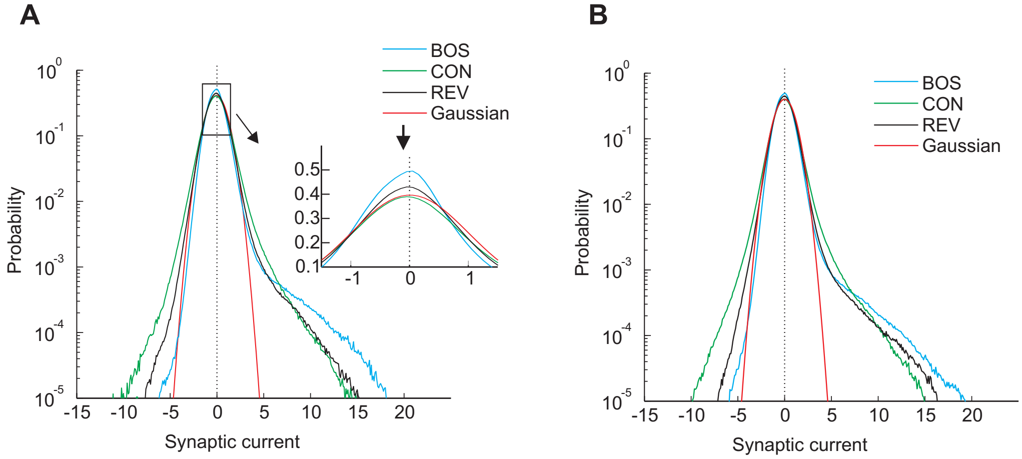


Fig. 5.7: Probability density of total synaptic currents. (A) The probability density of total synaptic currents y averaged over all neurons has a heavy tail on the positive side. Shown are the densities for BOS (blue), CON (green), and REV (black). Near zero synaptic currents, the curves are approximatively unit Gaussian (red), though their excessive peaks are slightly shifted to the negative side (inset, arrow). The curves cross each other such that large positive synaptic currents are preferentially elicited by the BOS and small positive currents by REV and CON. $N = 400$, $y_0 = y_{E-}$. (B) The distributions of synaptic currents for sparse STRFs (Figure 5.2B) are qualitatively similar to (A). The only noticeable difference is that the distribution for REV is closer to BOS, reflecting a lower selectivity for temporal order. $N = 400$, $y_0 = y_{E-}$, $c_s = 0.2$.

5.3.2.1. Distribution of Firing Rates

We computed firing rates in model neurons by thresholding total synaptic currents. A recent analysis of sparsely firing cells in primary auditory cortex of unanesthetized rats has revealed that both background and stimulus-evoked firing rates are well fit by log-normal distributions (Hromdka et al., 2008). We speculated that log-normal firing may be a corollary of efficient coding that could be reproduced in simulations. We inspected the distributions of firing rates across all neurons for all BOS and CON stimuli. Indeed, we found that the density of firing rates was best fit by a log-normal distribution, which was especially true for low firing thresholds, Figure 5.8. Note that recently published firing rate distributions of field L neurons in zebra finch were fit by Gamma distributions (Woolley et al., 2010b). However, the published data suggests that a fit with a log-normal distribution should be equally good.

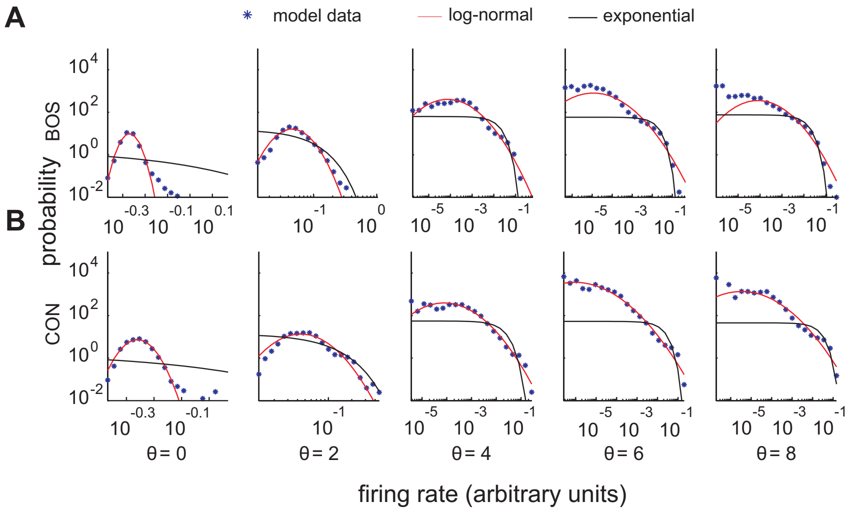


Fig. 5.8: Probability densities of mean firing rates. Mean firing rates in response to (A) a BOS stimulus and (B) a CON stimulus. For each cell we computed the mean firing rate to one stimulus trial. Our simulation data (blue asterisks) are better fit by log-normal densities (red) than by exponential densities (black). Firing rates are plotted in arbitrary units. Fit parameters for log-normal densities were determined by the mean and variance of logarithmic firing rates, and for exponential densities they were determined by the mean firing rates. Thresholds varied from $\theta = 0$ to $\theta = 8$. Noise amplitude $k = 1$, $N = 400$.

5.3.2.2. Independence of Neural Responses

Training the network increased the independence of neural responses. If responses were perfectly independent among neurons, the size distribution of coactive neurons would be binomial (the size distribution is the probability that a given number of neurons fire synchronously). The sole parameter of this binomial model is the single-neuron firing density that we estimated in terms of the fraction of suprathreshold events observed in the entire neuron population and for all training stimuli. In comparison to this binomial model, the observed size distribution elicited by whitened cochlear inputs was substantially wider, illustrating strong firing dependencies. The sparseness transformation significantly narrowed the observed size distribution towards the binomial case, Figure 5.9A. This increase of independence (decrease in Kullback-Leibler divergence to the binomial model) was true for both high and low firing densities, and true for nearly all firing thresholds tested, Figure 5.9B, revealing that the sparseness transforma-

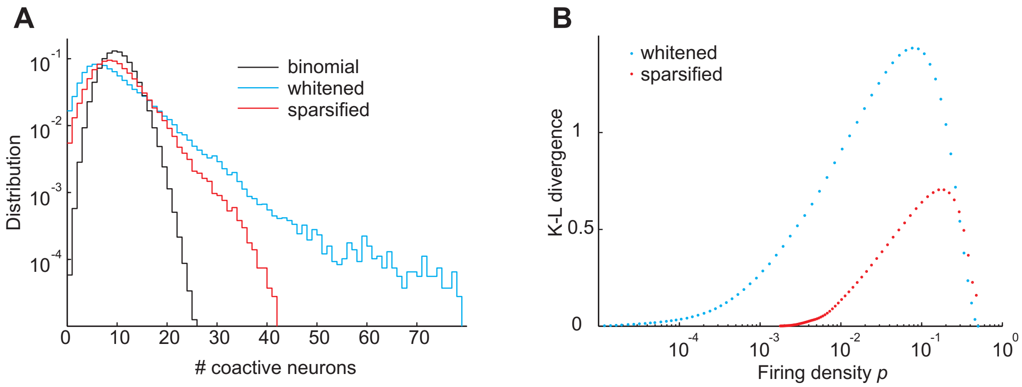


Fig. 5.9: Sparsification reduces firing dependences. (A) The sparseness transformation renders the size distribution of coactive neuron groups (whitening+sparseness) closer to binomial. The probability p of the binomial distribution (that a neuron is active per unit time) was estimated in terms of the firing density (the fraction of suprathreshold events over all neurons and training stimuli). Firing densities p were nearly identical for whitening and whitening+sparseness when $\theta = 2$ ($\theta = 0$ and $y_0 = y_{E-}$ during learning). (B) The Kullback-Leibler divergence between size distributions is smaller when comparing the whitening+sparseness model to the binomial model than when comparing the whitening model to the binomial model, for nearly all firing densities tested. $N = 400$.

tion is a robust mechanism to increase independence of neural responses. Qualitatively, this behavior of the network to render responses more independent applied to all firing thresholds used during training (we tested thresholds up to $\theta = 5$).

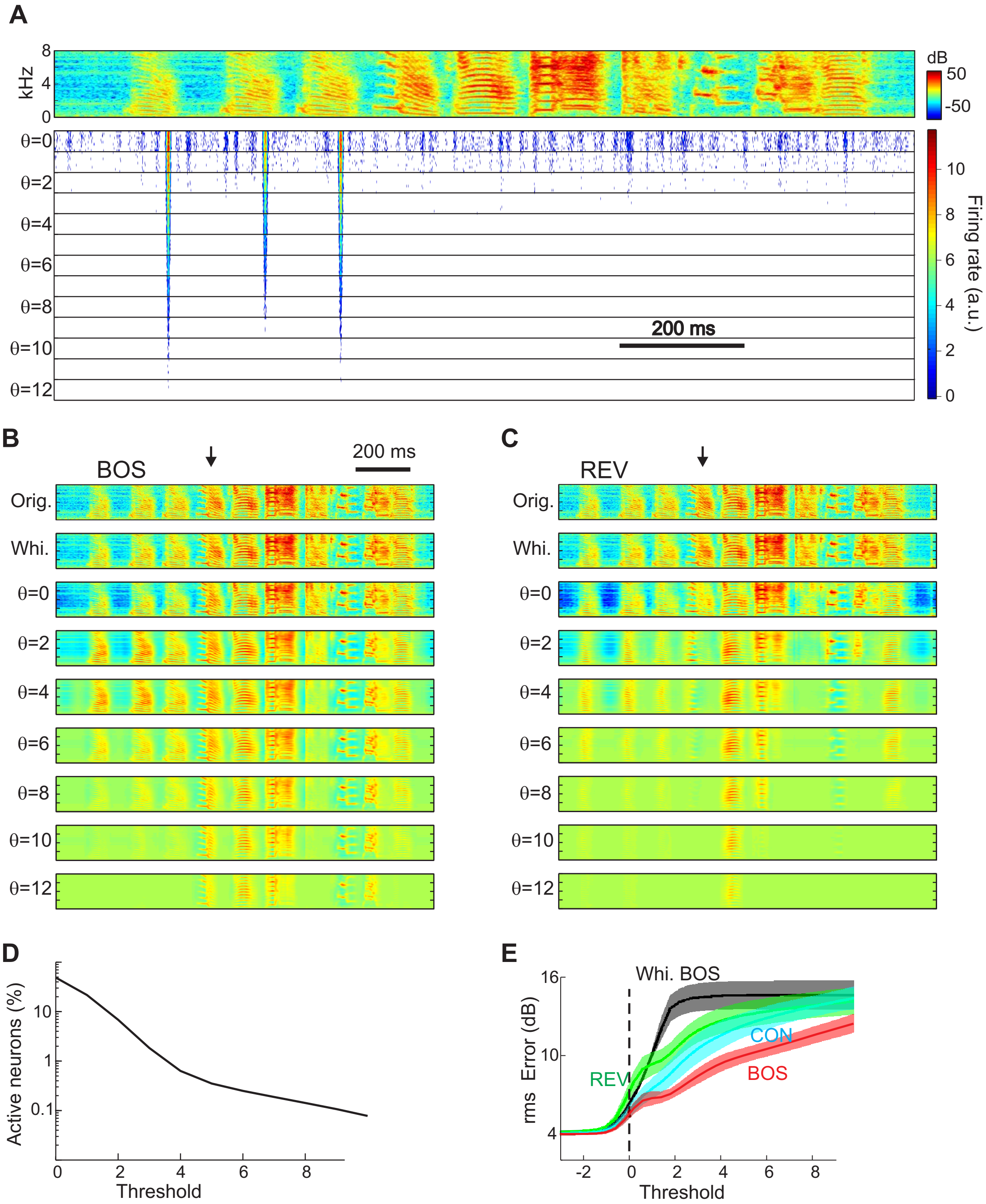
5.3.3. What Has Been Encoded

Our network allowed us to decode firing rates and reconstruct the spectro-temporal sound patterns that elicited them, using the pseudo-inverse of the sparseness transformation W (see equation 4.45). We evaluated the reconstructions for various firing thresholds (after training at a fixed threshold of zero), Figure 5.10A-C. For a threshold of zero, neurons produced dense firing patterns in response to BOS, with roughly 50 percent of neurons active at any time, Figure 5.10D. The percent active neurons decreased from 20% for $\theta = 1$, to 1-2% for $\theta = 3$, and down to 0.4% for $\theta = 5$. At thresholds higher than roughly three, reconstruction errors associated with non-BOS stimuli were often due to missed syllables because none of the neurons fired in response to these syllables.

In all cases, reconstruction errors increased with increasing firing threshold in a monotonic manner, Figure 5.10E. For a given threshold, reconstructions from sparseness-transformed cochlear inputs were much better than reconstructions from merely whitened inputs. This superiority was true even though for thresholds up to approximately 1.3, mean firing rates were lower for sparseness-transformed inputs than for merely whitened inputs.

Moreover, for a given threshold, reconstruction errors tended to be larger for the BOS played back in reverse (REV) than for BOS or CON, illustrating that reconstructions were optimized for stimulus ensembles experienced during training but not for novel ensembles. In the Methods we show that the reconstruction error is approximately equal to a term that grows not only with the number of subthreshold events, but also with their variance; hence, stimuli that induce narrow subthreshold distributions (such as the BOS, Figure 5.7) lead to smaller reconstruction errors.

Fig. 5.10 ***(facing page)*:** Reconstructing the cochlear spectrograms from firing rates. (A) Firing-rate of one example neuron in response to BOS for increasing firing thresholds ($\theta = 0$ to 12). The BOS spectrogram is shown on top. This neuron is tuned to a feature present in introductory notes and responds to it up to thresholds higher than seven. For each threshold, ten different responses are plotted, corresponding to ten different instantiations of synaptic noise. $k = 1$. (B) The reconstruction of a BOS spectrogram (orig., top) using all neurons, based on a firing threshold of minus infinity (whi., 2nd from top) is fairly complete with little information loss (arising from dimensionality reduction). With increasing thresholds (below), more and more syllables are lost in the reconstruction, but the reconstructed spectro-temporal patterns remain clearly recognizable. The arrow points to a down-sweep syllable. (C) Reconstructions of REV (flipped horizontally for comparison with B) are worse than reconstructions of BOS at the same threshold; for example the down-sweep syllable is not well reconstructed (arrow), presumably because zebra finches produce almost no up-sweeps. (D) The fraction of active neurons (averaged over all BOS stimuli) decreases with increasing threshold such that at $\theta = 3$ about 1% of neurons are active on average. This fraction decreases to 0.1% at about $\theta = 9$. (E) The reconstruction errors averaged over different stimulus ensembles are monotonic functions of the firing threshold. For a given positive threshold, reconstruction errors increase from BOS to CON to REV. $N = 400$.


A
8
4
0
kHz
dB
50
-50
θ=0
θ=2
θ=4
θ=6
θ=8
θ=10
θ=12
200 ms
10
8
6
4
2
0
Firing rate (a.u.)
B
BOS
200 ms
Orig.
Whi.
θ=0
θ=2
θ=4
θ=6
θ=8
θ=10
θ=12
C
REV
Orig.
Whi.
θ=0
θ=2
θ=4
θ=6
θ=8
θ=10
θ=12
D
Active neurons (%)
10
1
0.1
0
2
4
6
8
Threshold
E
Whi. BOS
16
10
4
rms Error (dB)
REV
CON
BOS
-2
0
2
4
6
8
Threshold

5.3.4. Selectivity and Sparseness

We explored the response selectivity of model neurons using the psychophysical d' measure (Green and Swets, 1966) that is routinely applied in birdsong studies. According to this measure, the selectivity for a stimulus over another is given by the difference in mean firing rates elicited by these stimuli, normalized by their standard deviations (see section 2.1). We assessed the selectivity of neurons to BOS versus matched spectro-temporal stimuli such as CON and REV. Variability of responses to a fixed stimulus was generated by a white-noise current source.

We found a wide range of selectivity behaviors. Many neurons responded more strongly to CON than to BOS, but this CON preference often reversed to BOS preference at high firing thresholds, Figure 5.11B. For a firing threshold of zero, the median d' selectivity for the BOS was negative across the population, both with respect to REV and to CON, Figure 5.11B. Hence, at this low threshold, the majority of neurons preferred REV and CON over BOS. BOS anti-preference remained true for a range of firing thresholds θ up to three. However, all of the median and mean BOS-REV and BOS-CON selectivities became positive at thresholds $\theta \geq 5$, Figure 5.11C. Thus, the response selectivity of the network was non trivial in that it reversed at higher thresholds. From the point of view of stimulus selectivity, the low-threshold regime of our network is a model of densely firing field-L neurons and the high-threshold regime is a model of sparsely firing HVC neurons.

Our training set contained many versions of two different CONs. For BOS-CON selectivity reversal it did not matter whether selectivity was tested on three novel CONS as in Figure 5.11, or on twelve novel CONs, or on the two trained CONs, because for all these cases the median and mean BOS selectivities reversed at around $\theta = 3 - 4$. However, when CONs from many more birds (> 20) were in the training set, then the median BOS-CON selectivity became positive only at very high thresholds ($\theta \geq 8$ for 22 CONs), whereas the mean selectivity became positive already at $\theta \geq 3$. Thus, when responses to many different songs are sparsified, then increasing numbers of neurons develop a feature preference that best matches a CON in the training set rather than the BOS; however, this match is not particularly good as illustrated by BOS that is preferred on average already at relatively low thresholds.

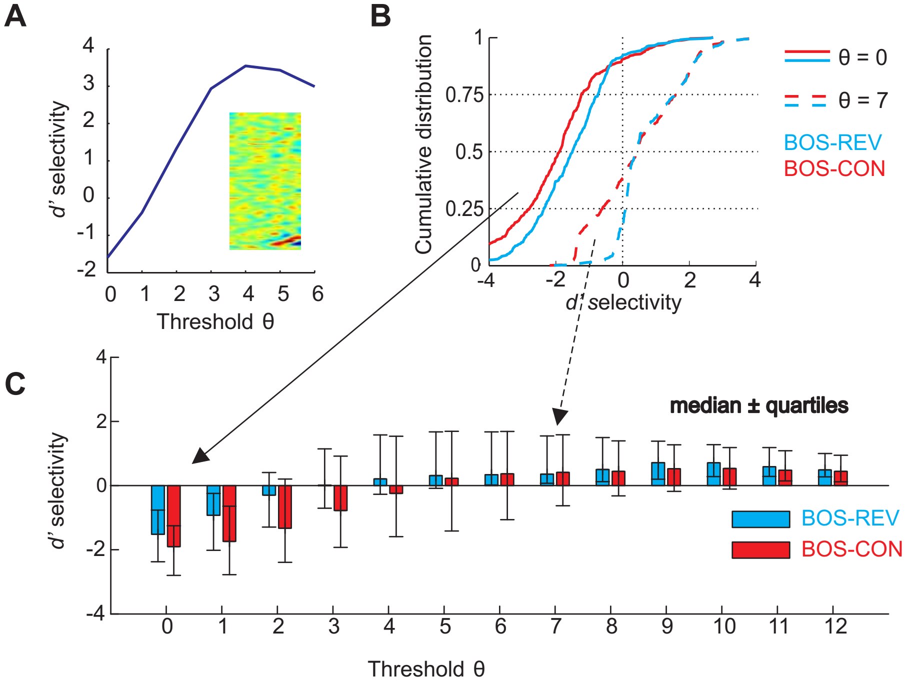


Fig. 5.11: d' selectivity for BOS reverses at high firing thresholds. (A) Example model neuron with reversing BOS-CON selectivity. This neuron's STRF (inlay) codes for an up-sweep from 500 to 800 Hz over 20 ms. The resulting BOS-CON d' selectivity is negative for low thresholds $\theta = 0, 1$ and turns positive for thresholds $\theta \geq 2$. (B) Example cumulative distributions of BOS-REV (blue) and BOS-CON (red) d' selectivities across $N = 400$ neurons for $\theta = 0$ (solid lines) and $\theta = 7$ (dashed lines). For $\theta = 0$ the selectivities are biased towards negative values, whereas for $\theta = 7$ the distributions are biased towards positive values. (C) Bar plot summarizing BOS-REV (blue) and BOS-CON (red) selectivities for a wide range of firing thresholds. The colored bars indicate the median d' selectivity and the error bars delimit the first and third quartiles. Selectivity reverses at around $\theta = 3$ (REV) and $\theta = 5$ (CON).

For any given threshold θ, the median d' selectivity (be it positive or negative) depends on the noise level. When increasing the noise level, the median d' selectivity goes toward zero, and, when decreasing the noise level, the median d' selectivity diverges from zero. d' magnitudes are also influenced by the number of different song renditions used to probe selectivity. When selectivity is probed with a single BOS and a single CON file and noise is small, d' values can become arbitrarily large. Hence, our model allows an arbitrary scaling of d' values by manipulating the intrinsic noise.

We tested the dependence of BOS selectivity on the temporal summation window and found that our results did not depend critically on STRF width. Model STRFs were 50 ms wide. For 100-ms wide STRFs, BOS-CON and BOS-REV selectivities reversed at around $\theta = 2-3$; and, for 25-ms wide STRFs, selectivity reversal was seen at around $\theta = 4$. Hence, with increasing temporal summation window, selectivity reversal was seen at lower thresholds. Note that 25-ms STRFs are much shorter than estimated integration times in BOS-selective neurons (Lewicki and Arthur, 1996; Sen et al., 2001).

We also tested the effect of neuron number on d' selectivity. Doubling that number from $N = 400$ to $N = 800$, or reducing it to $N = 200$ or $N = 100$ preserved selectivity reversal in the range $\theta = 2-5$, for all CON ensembles tested and for both firing-rate models. Also, we found that the value of the threshold θ during learning has little influence on selectivity reversal after learning. For $\theta = 0$, $\theta = 2$, and $\theta = 5$ during learning, selectivity for BOS reversed to positive values at around $\theta = 3 - 4$ after learning. In summary, selectivity reversal at high thresholds was very robust and did not depend on model details.

We assessed whether our model neurons preferred CON over artificial stimuli, as has been reported in field L (Theunissen et al., 2004; Grace et al., 2003). We found high median selectivity for CON versus tone pips, tone stacks (ripples), and white noise (Figure 5.12). This CON preference was true for all thresholds $\theta \geq 0$ examined. Solely tones (sparse colored noise) were preferred over CON for thresholds up to $\theta = 1$. For higher thresholds $\theta \geq 2$, CON-tones selectivity reversed and CON was strongly preferred.

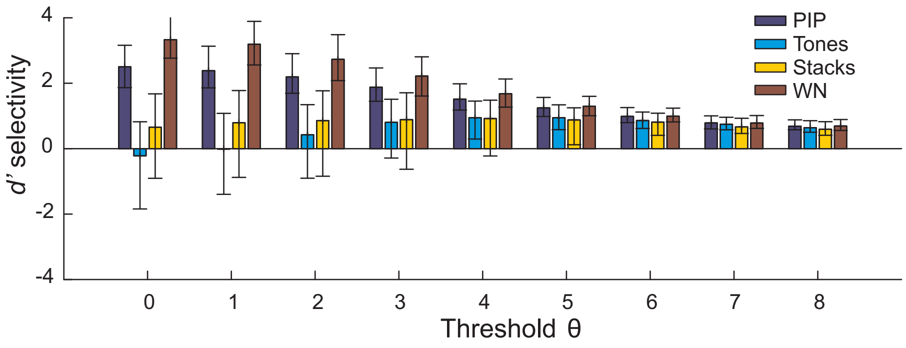


Fig. 5.12: Selectivity for CON vs. different artificial stimuli. Depicted are median selectivities ± quartiles. PIP=tone-pip stimuli, WN=white noise. $N = 400$.

5.3.4.1. Selectivity Reversal in Different ICA Algorithms

The key element in our model seemed to be the positive tail of synaptic currents. Because this tail has a non-Gaussian shape, our model can be seen as part of a broader class of ICA algorithms that extract maximally non-Gaussian components from data (Hyvärinen and Oja, 2000). To test whether BOS preference at high thresholds arises also in other ICA algorithms, we trained an identical network using the classical ICA algorithm by Bell and Sejnowski (results not shown) (Bell and Sejnowski, 1995). In this algorithm, as in most similar algorithms, the final distribution of synaptic currents is symmetric, with heavy tails on both sides. For this reason we applied the firing threshold to absolute synaptic currents. We computed BOS selectivities for different firing thresholds θ and found that BOS-CON and BOS-REV selectivities reversed as in our model, but at higher thresholds: The median and mean BOS-REV selectivities reversed at $\theta = 4$, whereas the mean and median BOS-CON selectivities reversed at around $\theta = 7$. Thus, the emergence of BOS preference in the ultrasparse regime of simple networks did not depend on how efficient coding was enforced, but appears to represent a generic consequence of non-Gaussian statistics and the choice of the training set.

5.3.5. Multilayer Networks

To account for the layered architecture of the auditory pathway, we also explored a two-layer network, in which the second layer was trained on thresholded first-layer outputs. In simulations, first-layer outputs were first summed over consecutive time bins (to extend receptive field widths in the second layer) and were then subjected to whitening and sparseness transformations (as we did for the first layer, see Figure 5.13). In simulated networks in which the first layer was a universal encoder (small θ), we found that response selectivity in the second layer reversed at high firing thresholds in preference of BOS. Hence, our high-threshold model of response selectivity in HVC can also be expressed in an architecture that is consistent with the feedforward organization of the auditory pathway.

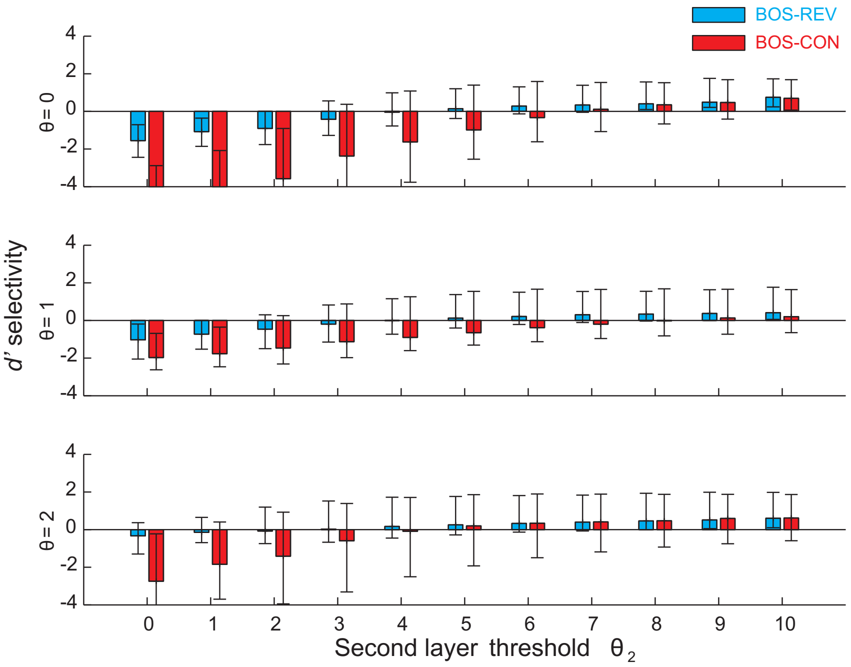


Fig. 5.13: Selectivity in two layers. Median and quartile selectivities in a network of two layers for various firing thresholds θ in the first layer and θ_2 in the second layer. In each simulation, second-layer responses were evaluated using the first-layer threshold applied during training. As can be seen, BOS preference in the second layer is restricted to the high-sparseness regime there (right part of the three subplots). $k = 1$, $N = N_2 = 400$.

Chapter 6

Nonsymmetric Sparse Coding as a Computational Tool

> Die schlechten ins Kröpfchen
> Die guten ins Töpfchen
>
> Gebrüder Grimm, Aschenputtel

In chapter 5 I showed how nonsymmetric sparse coding may be a valuable approach to explain the computational dynamics in the auditory processing of the zebra finch. We established that such an encoding scheme would be desirable from a metabolic point of view. But as stressed in the introduction sensory coding has to present the environment to the animal in a fashion that the animal is able to perform adequate actions or take correct decisions. I already mentioned the possibility of song learning based on the sparse representation of tutor song and bird's own song in section 5.2. In this chapter I want to go step further and present four computational tools for four problems where nonsymmetric sparse coding is the key element. While classical approaches to these problems are often heuristic and involve mathematics difficult to predict, the methods we will discuss are entirely data-driven and mostly linear. All results are elaborated for birdsong but can easily be applied to more complex problems.

6.1. Sparse Coding Used for Subsong Detection

The scientific interest in the zebra finch is mainly focused on the bird's ability to imitate vocalization of a tutor as mentioned in the introduction. The main phase of song learning is the so called sensory-motor phase of the juvenile bird. During this time the bird's vocalization shifts gradually from subsong (similar to human babbling) with no defined syllables, no clear temporal or spectral structure, to plastic song where syllables are present, but still lack the spectral smoothness and a static song. And finally to the crystallized adult song, with very low variability (see section 1.2.1 for a detailed description of zebra finch song and it's development).

Many experiments with zebra finches, be it lesion experiments, electrophysiological experiments, or observational studies, rely on the analysis of changes or on the development of the bird's (sub-)song . When songs are recorded over several days it is desirable to automate the recording process. Computers do a good job in recording the songs of adult birds, as they can easily be detected on-line based on heuristic features, such as typical rhythm or spectral features. However, when dealing with subsong, they will fail. Lacking any typical rhythmicity, no repeated syllables, and a misty spectrum subsongs are not the preferred target of such detectors. When these detectors are running at the same settings as for adult song, most of the subsongs will be missed, and the ones recorded will be highly biased. If the detection criterion is loosened in order to produce less false negative detections, the recorder will start to detect all the calls as well as a lot of (cage-)noise. In our lab we recorded for one bird up to 10'000 files per day and (i.e. a nearly continuous recording) of which only several hundred contained subsongs. Sorting this amount of files by hand would take roughly the same amount of time as their recording.

6.1.1. Training of the Detector

The method is completely data-driven and has to be trained off-line prior to application. This training consists in two steps: in a first step we will produce a set of linear filters using nonsymmetric sparse coding.

But first we will have to define the training set. If we just want to sort the files of a day's recording, it is best to use all the files of this day. If we want

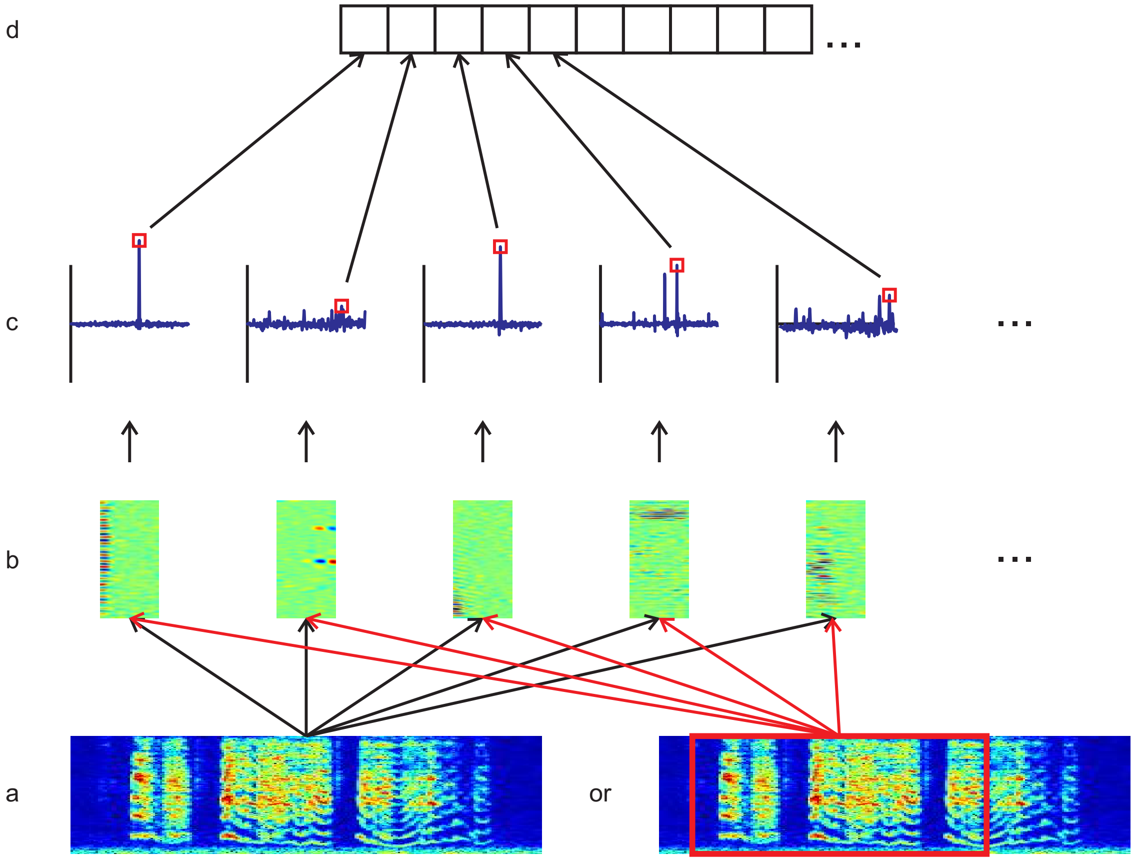


Fig. 6.1: Diagram of feature vector calculation. The spectrogram of a potential subsong file, or a window containing a potential subsong (a) are convolved with the set of filters, previously trained by our nonsymmetric sparse coding algorithm (b). Of the output (c) only the maximum is taken and written into the feature vector (d).

to have a general detector however, it would be best to randomly choose recordings of different juveniles at different ages. It is important to be aware, that such a data-driven detector will only be able to reliably work on a stimulus space which it was trained on. Any stimulus perpendicular to the training set will randomly be classified as subsong or not. The number of dimensions n needed is highly dependent on the diversity of the stimuli. Recordings from a single bird and day without complex background noise might be completely represented by even less than 100 filters, while a general detector might need more than 1’000 filters.

The main idea for the detector is the following: if we apply the filter on a sound file we get a temporal response curve for each filter. From the

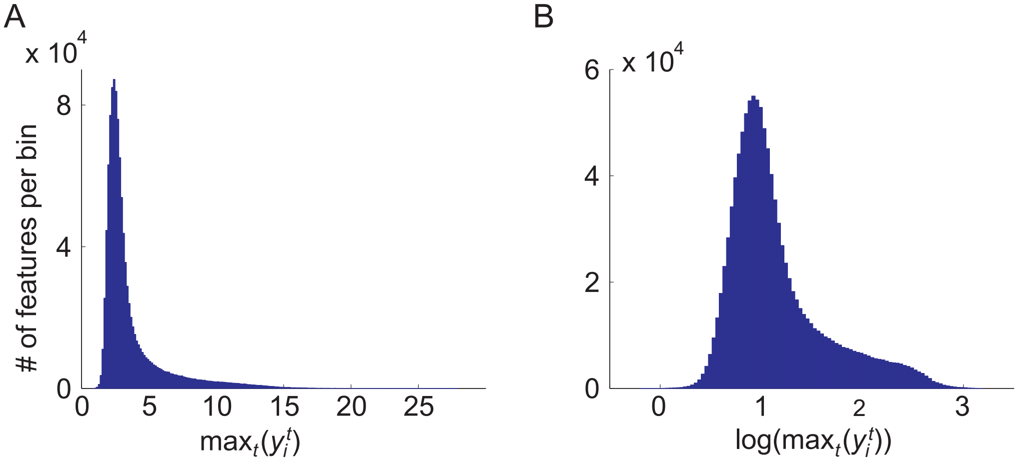


Fig. 6.2: Distribution of features. (A) The histogram of features looks like a unimodal log-normal distribution. (B) When plotting the log-feature histogram we see that the distribution is actually bimodal, consisting in the sum of two log-normal distribution. However, the two distributions are to close to attribute single features to one or the other. $N = 400$.

output histogram (Figure 5.7) we know, that this output curve will mostly be moving around zero, very much like low-pass filtered noise. But at very few points in time, the response will deviate strongly from zero into the positive range. My interpretation of such a curve is that most of the time the sound is just noise in the subspace of the feature encoded by the filter. But at the time of the peaks of the curve, the feature is present in the sound. The amplitude of the peak gives a measure of how strong the feature is. So ideally, if we just take the maximum of a curve, we should either get a small value which is due to the noise, i.e. the feature is not present in the file, or a big value due to at least one presence of the feature (see a schematic drawing in Figure 6.1). If we look at the distribution of the maxima we see, as predicted, a bimodal distribution. But especially with subsong from young birds the distribution of the features that are present is so broad that it is highly overlapping with the distribution of the noise peaks (Figure 6.2). So we are not able to tell with high accuracy, whether or not a single feature is present in a sound file. However, the features are not independent, and the cumulative knowledge makes it easy to separate subsong, as it has a typical feature vector.

After the unsupervised step using our nonsymmmetric sparse coding algorithm, we will have to choose a random set of subsong files and noise files

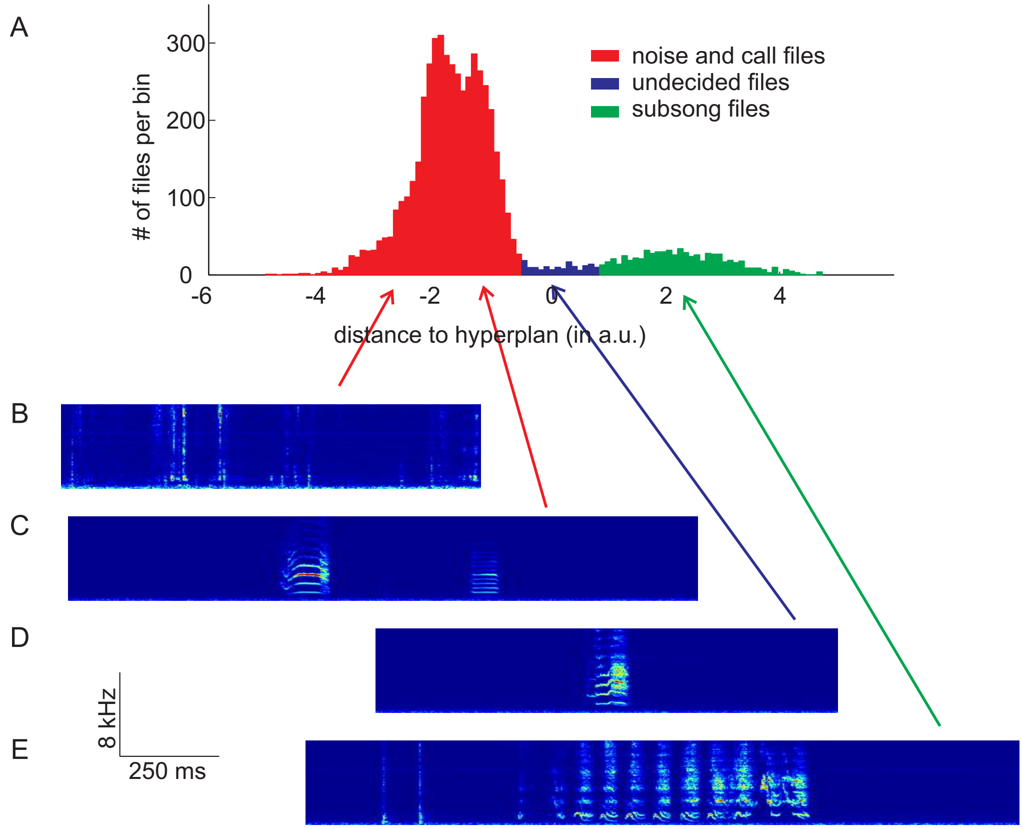


Fig. 6.3: Classification of 5534 files from a single day. The model was trained on the recorded files of the day previous to the classification. (A) Distribution of distance of the files to the hyperplane defined by the SVM classifier. The distribution is clearly bimodal with one mode on each side of the hyperplane. After fitting a Gaussian mixture with two Gaussians to the this distribution we can classify the files with a certain probability to one or the other Gaussian. The 4719 red colored files belong with a probability of 99% and above to the non-subsong distribution, while the 638 green colored files belong to the subsong distribution with a probability of 99% and above. In blue 177 files have no classification above 99%. Roughly one third of the files contain single syllables which are mostly subsong-like long calls. These calls/syllables are hard to classify also for humans as there is a gradual transition. Half of the non-classified files are subsong files classified with probabilities below, but mostly close to 99%. The rest are noise and call files with classification probability below 99%. The red distribution shows two peaks which is due to noise and call files both being classified simply as non-subsong. Examples of classified files are given by (B) noise, (C) calls, (D) long call, and (E) subsong. $N = 400$.

and tag them. For the presented data we choose a total of 50 samples (less than the number of dimensions), roughly 25 each. With the tags and the feature vectors we trained a support vector machine (SVM) with a linear kernel, such that the space of feature vectors was split into a subsong and a non-subsong space. An example of the classification is given in Figure 6.3.

Alternatively this algorithm can also be used from real-time subsong detection. Instead of using whole files we use sliding windows somewhere in the range of the length of a subsong (0.2 - 1 s). For each window a feature vector is calculated and then classified. The exact on- and offset of the subsong can not be determined by this method, however it is simple to estimate the onset of a sound by determining the onset of acoustic power. For the training of such a classifier it has to be trained on such windows completely with or without subsong.

6.2. Tracking Song Development

An open question is the tracking and quantification of how a bird learns his song. Tchernichovski et al. (2001) defined a set of 4 heuristic features that summarized songs: Wiener Entropy, spectral continuity, pitch, and frequency modulation. In Deregnaucourt et al. (2005) they not only had a look at the mean values, but also at the variances of these measures. They thereby found that the variance of Wiener Entropy was raising during training, but reset after sleep to a lower level than the previous evening. These findings were reconfirmed later by Shank and Margoliash (2009).

However, the question arises whether there are more and possibliy non-heuristic features that could describe song learning on a more detailed level and at the same time reproduce the finding from Deregnaucourt et al. (2005) and Shank and Margoliash (2009). As in the last section we therefore trained filters using our nonsymmetric sparse coding algorithm. We once trained them on the first day of training (day 39) and once on the last day (day 102) and calculated feature vectors (see section 6.1) for song files over the whole training phase.

In Figures 6.4C-H we present a number of representative features for the day-102 training set. Filter C is coding for certain downsweeps. As we see, the bird is optimizing his song over the whole period of training in

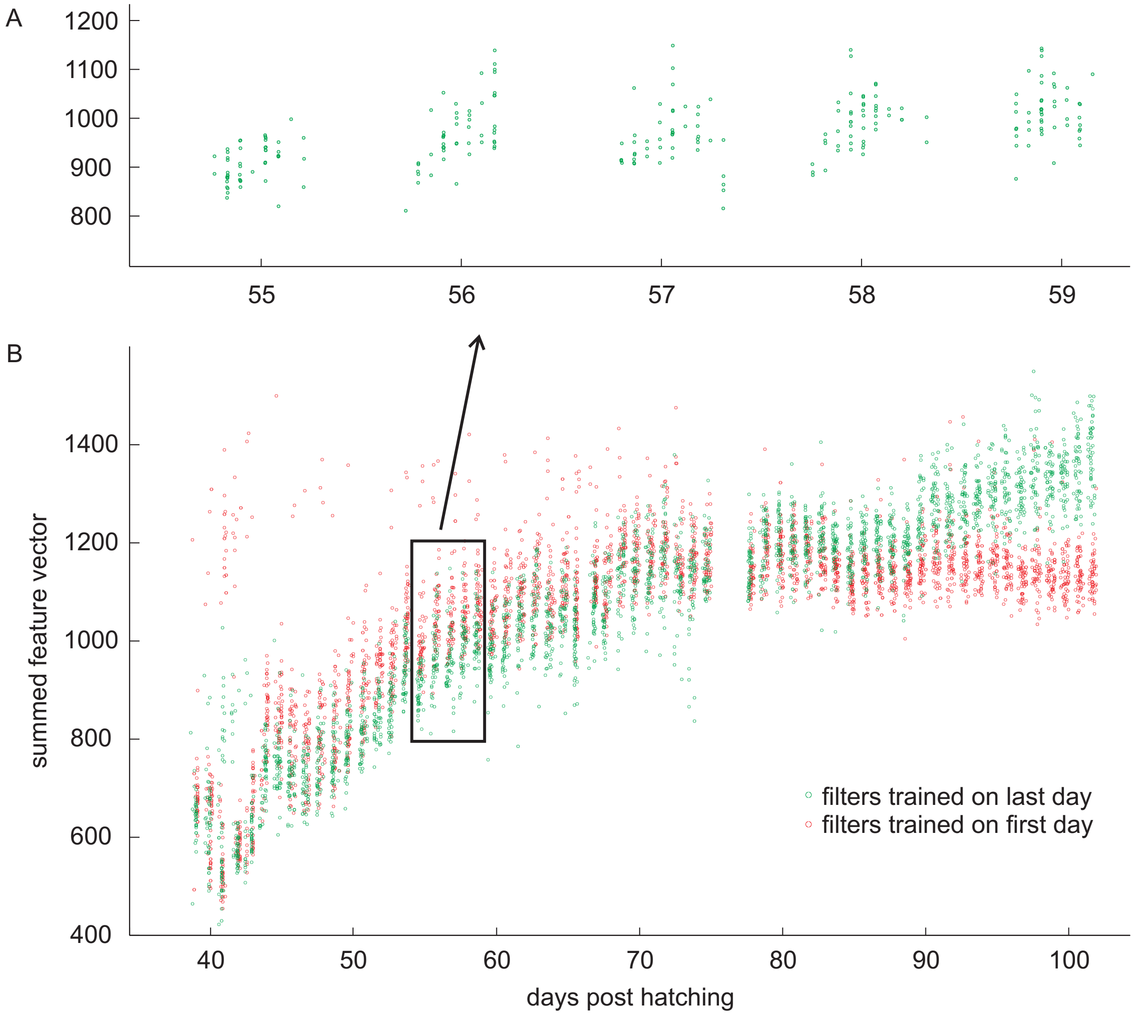
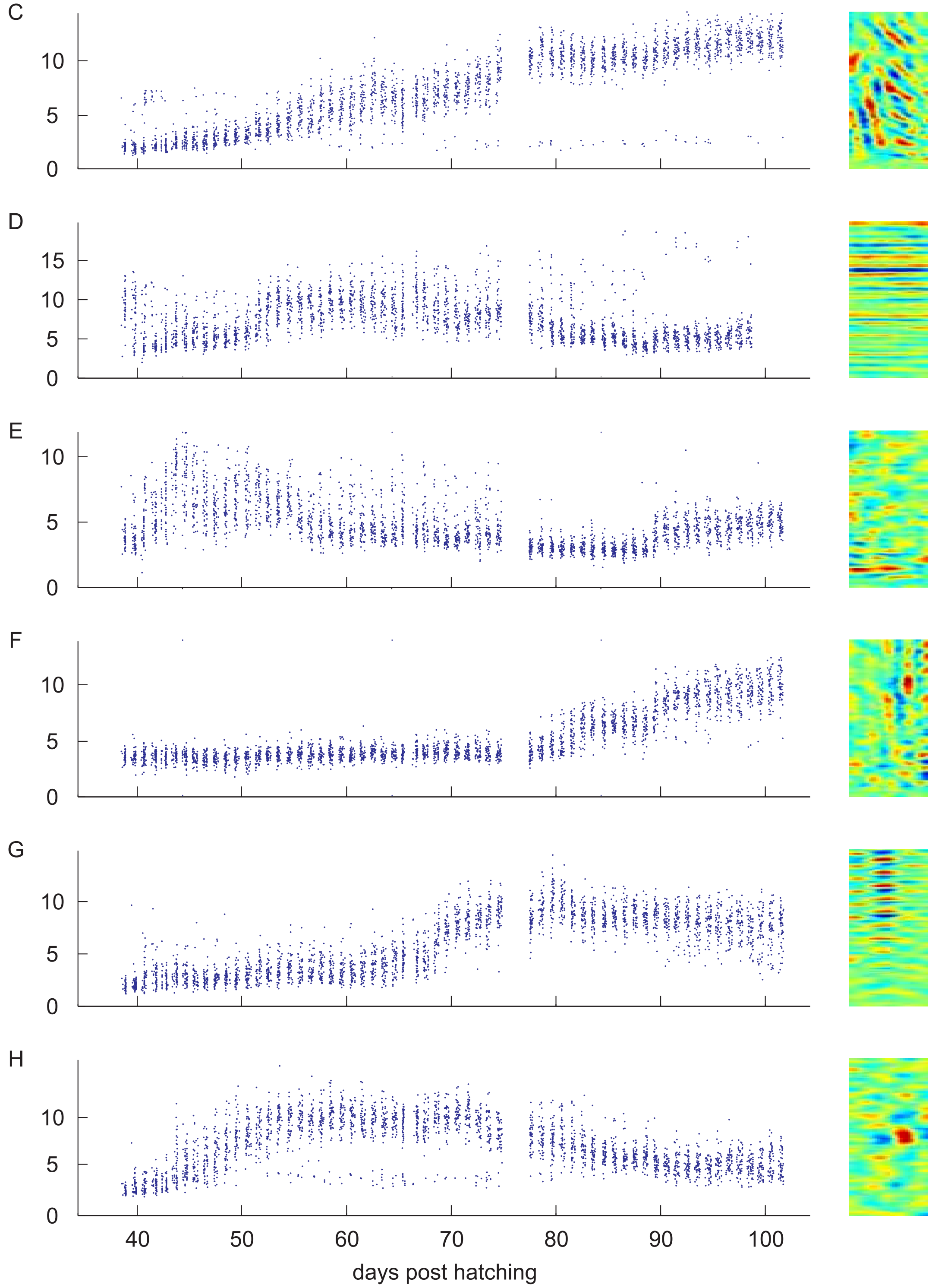


Fig. 6.4: Tracking song development. (A) Magnification of the black box in (B). (B) Example of a bird trained from day 39 to day 102. Each dot represents the summed feature vector of one handclassified song file. The green dots are gained by 200 filters trained on the recordings of day 102, red dots by 200 filters trained on the recordings of day 39. A clear positive trend can be seen for both features sets. However, for the red dots this trend ends around day 75. As one can see from (A) the positive trend can also be seen on an intraday level and a reset on the next morning. (C), (D), (E), (F), (G) and (H) are representative examples of the development of single features (trained on day 102). On the left the development is shown for the filters given on the right. Recordings were provided by Georg Keller.

order to optimally drive this filter. Filter D in contrast is coding for a frequency band somewhere around 6 kHz and is highly suppressed by a



Continued Fig. 6.4: See page 103 for caption.

neighboring frequency band. The bird drives it mainly by his plastic song around day 60, but will no longer in later stages of learning. Filter E codes for a lower frequency band slightly above 1 kHz. As we see, this filter is driven by subsong, but as soon as the bird sings plastic songs, it will no longer be driven. However, there is a slight increase form day 90 on. Filter F is coding for the onset of a harmonic stack, preceded by some high frequency sound. This filter is absolutely not used until day 80. But from then on the bird changes his song continuously in order to drive this filter. Filter G codes for the upper part of a harmonic stack. The filter is not driven until day 68. Within 3 days the bird changes its song towards this filter and leaves it for the rest of training. Filter H is coding for a short sound in a medium frequency band. During subsong phase the bird is optimizing for this filter, but during plastic song phase it gets slowly neglected.

However, are we able to find the development reported in the above mentioned studies? The answer lies in the summed feature vectors. In Figure 6.4B the summed feature vector for the two filter sets are shown. Green dots represent the sum for one song file filtered by the day-102 training set, red dots by the day-39 training set. For both sets a clear raise of the sum can be seen. However for the red dots it saturates somewhere around day 70, while for the green dots it continues until the end of training. If we zoom into single days (Figure 6.4A) we see a clear positive in-day trend: values derived from evening song files have higher values than from morning song files. Additionally we can see a clear reset in the morning, values are lower than on the previous evening. This is in absolute accordance with Deregnaucourt et al. (2005) and Shank and Margoliash (2009) which proves this method to be valuable alternative to heuristic methods.

For the green dots, trained on the data from day 102 this result might not be overly surprising. We can assume, that the bird will have a path from a something random/basic towards his final song which is represented by the subspace spanned by it. Our method analysis this song and finds the dimensions where the complete song can be represented by minimal 1-norm (see section 4.3), leading to many low values and very rarely very high values. As we defined the features as maxima over time, we only take those high values. As the filters are trained on the last day, these filters are optimized for this day and will produce maximal features. Intuitively we now also expect these files to be the ones that produce the maximal features, which in this case proves to be true.

However, if we have a look at the red dots this intuition fails. Even though the corresponding filters were trained on the files from day 39, the feature vector sum continues to raise until around day 70. What does this mean? It means that the basic subspace of zebra finch song is already laid out by innate calls and some randomly produced subsongs. To come back to the findings of (Marler, 1997) and Feher et al. (2009) mentioned in section 1.2.1.1: it might be, that not a proto-song template is innate, but that the song reflects much more the sensory optimum within the boundaries of what the physics of a zebra finch allows to produce. Without a broad auditory experience, previous to the crystallization of the song, the sensory subspace might be not well refined or even degenerate. A song would then be the optimum in such a subspace which would resemble a prematurely crystallized song. The red dots saturating around day 70. If a bird did not hear anything but himself, any further development of his song past day 70 would not be an enhancement in his sensory space. The song would therefore remain unchanged for the rest of his life, unless his sensory subspace would be updated by further experience while the bird is still able to learn.

6.3. Smart Noise Suppression by Selective Neuron Exclusion

The following section is reproduced from the publication Blättler and Hahnloser (2011).

Some of the recorded songs from our colony were contaminated with electrical noise elicited by an old CRT computer monitor. This noise went unnoticed, as it was very low and only noticeable in syllable gaps. However, when such files were part of the training set of our algorithm we found filters encoding this noise, as illustrated in Figure 6.5. When we omit these five neurons for song reconstructions, we are able to effectively suppress the monitor noise. To demonstrate the effectiveness of this smart noise suppression, we iteratively estimated the sound waveform associated with the reconstructed song (Griffin and Lim, 1984). The original and the noise-suppressed waveforms are available in this pdf by clicking on the

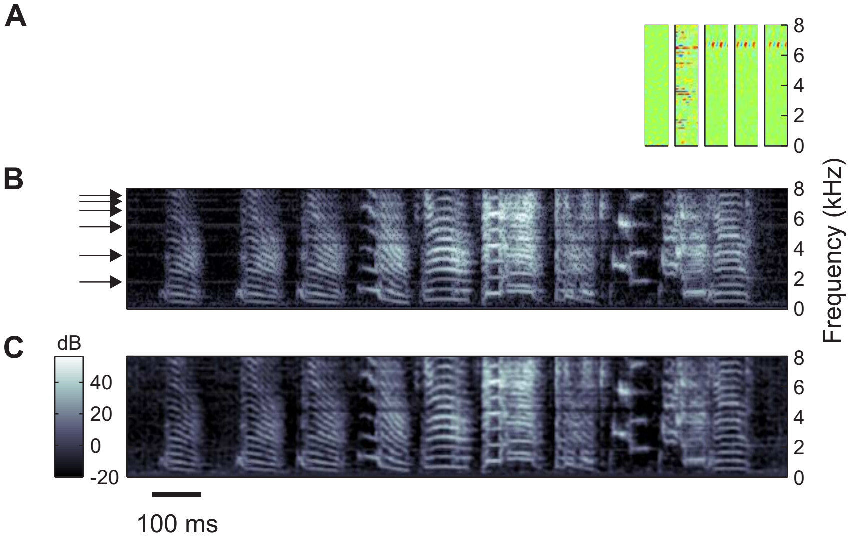


Fig. 6.5: Smart suppression of electrical noise affecting the recordings. (A) The STRFs of five neurons that encoded monitor noise. (B) 🔊 Original BOS spectrogram. The noise is manifest as gray horizontal bands (black arrows) during syllable gaps. (C) 🔊 Reconstruction of the BOS (Equation 4.47) from elicited responses in the network. The thresholds of all neurons were set to $\theta = -\infty$, with exception of the five neurons in A in which the thresholds were set to $\theta = +\infty$. The monitor noise has vanished in the reconstructions, without affecting the birdsong signal. $N = 160$.

corresponding legend of Figure 6.5.

In this example we filtered out noise by adding knowledge about the existence and form of noise and the form of corresponding filter to the system. However, one could also image an immanent system where certain filters are ignored based on the statistics of their output, in this example based on the regular sinusoidal output (see Figure 5.3C, neuron 106).

6.4. Underdetermined Blind Source Separation of Zebra Finch Songs

Blind source separation (BSS, sometimes also for blind signal separation) is a problem we often encounter in our everyday life. It is popularly known as cocktail party problem (CPP), a term coined by Cherry (1953). The objective of BSS is to isolate one signal source from an unknown mixture

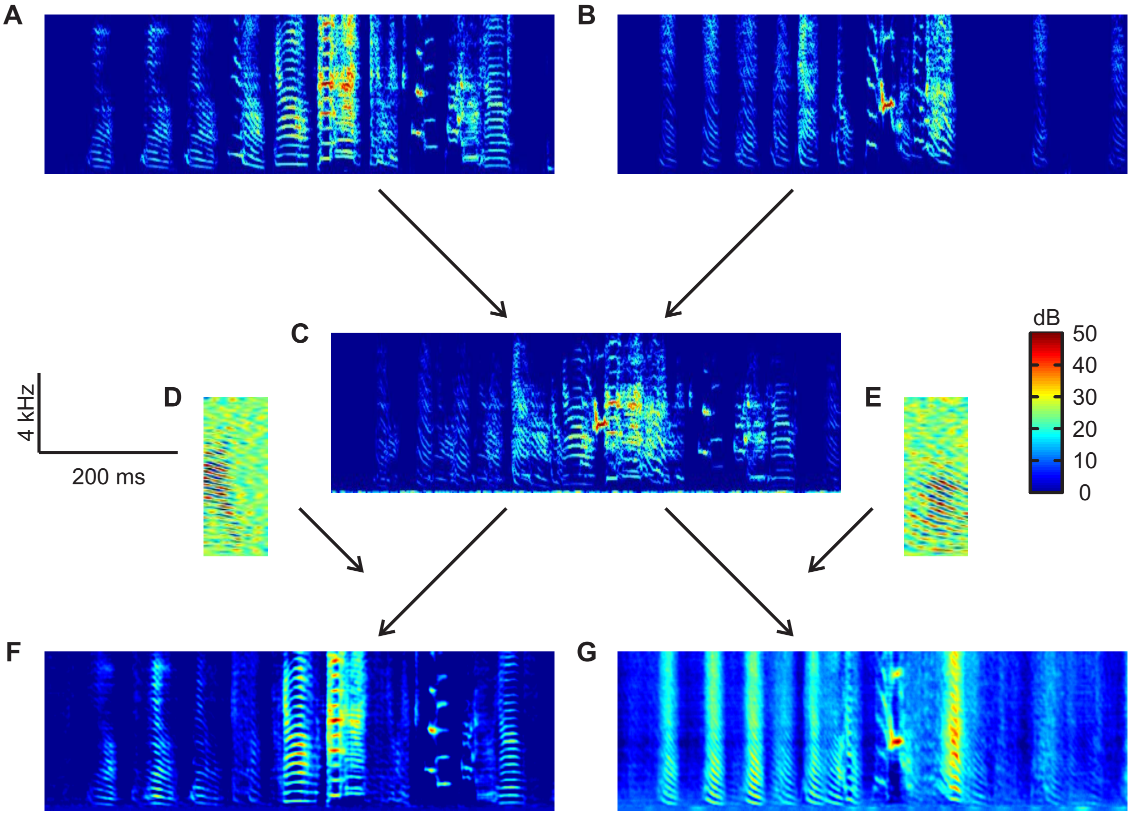


Fig. 6.6: Underdetermined blind source separation. (A) 🔈 Song spectrogram of the bird who provided the majority of songs during training (BOS). (B) 🔈 Song spectrogram of a bird how provided few songs during training (CON). (C) 🔈 Spectrogram of the BOS and CON played simultaneously. (D) Filter most selective for BOS. (E) Filter most selective for the CON . (F) 🔈 Spectrogram of the separated BOS. (G) 🔈 Spectrogram of the separated CON. $N = 400$.

of several signals. Humans are know to be very efficient in performing this task given a mixture of different voices (Cherry, 1953; Arons, 1992), therefore CPP. Several different cues, both monaural and binaural, are considered in order to extract a specific source (Bregman, 1994).

Several algorithms have been proposed to perform BSS in cases where we have the same number of sources as mixtures, or less sources than mixtures (overdetermined) (Bell and Sejnowski, 1995; Hyvärinen and Oja, 2000). However, in real life the number of sources highly outnumbers the mixtures (two in case of our ears). These underdetermined cases pose a so much harder problem. Few algorithms have been proposed to tackle them (Araki et al., 2004; Schmidt et al., 2007).

The downside of most of this algorithms is their iterative character, making them computationally costly and inapt for real time application. Our nonsymmetric sparse coding algorithm in contrast works purely feedforward once trained. It takes the stimuli and projects them onto a higher dimensional space, in which the dimensions are more or less independent, similar to ICA algorithms solving the overdetermined CPP. However, because mixtures in the 1-dimensional audio space will generally not be projected onto this higher dimensional space to be linearly separable, we are prone to nonlinearities. So here I reuse the old assumption: features are either present or not (see Figures 5.7 and 6.2B), anything below a certain threshold is likely random noise. Then the only remaining question is to which source should a certain feature be attributed to. A possible solution to this problem is the selectivity (see section 2.1), but one could imagine different criteria for feature allocation.

We then get the following algorithm:

1. Calculate filters on sound files of the different sources.
2. Decide for a threshold θ.
3. Calculate the selectivity of each filter for each source vs. the other sources.
4. Assign filters with a high selectivity to the sources.
5. Apply the filters assigned to one source on mixtures, threshold the output, and invert everything back to audio domain.

An example of such an underdetermined BSS is given in Figure 6.6. The simulation was one used for sensory modeling (see chapter 5), so the training set was dominated by songs of one bird (BOS) while two other birds provided the remaining songs (CON). The selectivity d'_i of the BOS vs. CON for one bird was determined for each filter at a threshold $\theta = 2$. Filters with a selectivity $d'_i > 0.5$ where labeled as BOS-filters and the ones with selectivity $d'_i < -0.5$ as CON-filters

$$S_B = \left\{ i | d'_i > 0.5 \right\} \tag{6.1}$$

$$S_C = \left\{ i | d'_i < -0.5 \right\} . \tag{6.2}$$

Examples of BOS- and CON-filters are given in Figures 6.6D+E. One BOS (Figure 6.6A) and one CON (Figure 6.6B) were mixed in the auditory domain and then the spectrogram of the mixture was calculated (Figure 6.6C). I calculated the suprathreshold synaptic current Y_+ (equation 4.15) and inverted them back to spectrogram domain using only the labeled filters (see equation 4.45):

$$X^t_{rec\ BOS} = P^{-1} \sum_{i \in S_B} J_i \left(y^t_{i,+} + y^t_{i,E-} \right) \tag{6.3}$$

and equally for CON. The results can be seen for BOS in Figure 6.6F and for CON in Figure 6.6G. For both songs the algorithm was able to retrieve the correct temporal and spectral patterns. Slight distortion can be found in the power amplitude. Interestingly, when listening to the reconstructed waveforms (Griffin and Lim, 1984) the distortions seem to be less for the CON than for the BOS. However, it is only by perception, not by numbers.

The fact that such an algorithm is able to perform the task might be surprising. The mixture has been performed in sound domain. Statistically, the mixture ads up linearly in the power-spectrogram $E\left(\left|\mathcal{F}(\omega)(x+y)\right|^2\right) = \left|\mathcal{F}(\omega)(x)\right|^2 + \left|\mathcal{F}(\omega)(y)\right|^2$, but for sure not in log-power-spectrogram. But our algorithm is splitting up the spectrogram, except for the thresholding, purely linearly. The reason for still being able to separate the two signal lies therein that the filters by themselves build up a model of the songs. The non-linear addition of the two signals and the linear decomposition lead to noise in the encoding, which is statistically distributed over all filter outputs. By thresholding most of the filter outputs are suppressed and only the noise on the suprathreshold outputs goes into the reconstruction. And that the suprathreshold output produces good reconstructions for the data lying in the training subspace has been shown in Figure 5.10.

The question that remains is whether the selectivity is a good way to split up the signals. For sure it is not optimal. When looking at the filters (examples are given in Figure 5.2), it is obvious that certain filters such as the onset filters are roughly equally activated by both, BOS and CON, as both stimuli have clear onsets of sound power (as do most natural sounds). These filters have a selectivity close to zero and are not used for any reconstructions. Optimally, they would be used for both reconstructions, each at its time, depending on which filters are coactivated with certain latencies.

The here presented algorithm for underdetermined BSS is not perfect, nor is it intended to be perfect. The goal was to give a prove of principles. Even in this highly simple version, the algorithm was able to separate the songs. It is not absolutely 'blind', as it has been trained on unmixed signals and, when applied, it knew which sources were present in the mixture. A completely blind algorithm would need to be trained on unlabeled and probably mixed data and would need to identify the sources present in the mixture himself[1]. Further research in this field of UBSS is needed in order to develop working algorithms being close to blind.

The algorithm is most similar to Asari et al. (2006), even though they present a slightly different problem including the head related transfer function. The big difference is that even after training their method requires an iterative optimization making it computationally costly and unsuited for real-time applications.

---

[1] A possibility for identification of sources may be an extension of the algorithm presented in section 6.1

Chapter 7

Discussion

> The moment we want to believe something, we suddenly see all the arguments for it, and become blind to the arguments against it.
>
> George Bernard Shaw

The following chapter is partly reproduced from the publication Blättler and Hahnloser (2011).

By today brain research has mainly been a descriptive science, as any science when it was young. However, at the current stage brain research has grown to breed hypotheses and models as well as therefrom derived principles and theories. The many hypotheses that are proposed and will be proposed, they all have to stand the test of time and data.

7.1. A New Model?

The above presented model offers an understanding of auditory response based on an efficient coding hypothesis. The model falls into the broad category of ICA and sparse coding algorithms which try to linearly transform given (sensory) inputs into independent outputs (e.g., synaptic currents). In most ICA algorithms, independence of outputs is enforced by a symmetric cost such as kurtosis or entropy (Hyvarinen and Oja, 1997; Bell and Sejnowski, 1995; Hyvarinen, 1999). Because of this symmetry, most ICA algorithms fail to account for the asymmetry imposed by the spike threshold. By contrast, the nonsymmetric sparse coding algorithm explicitly includes a rectification nonlinearity which truncates as little information as possible because we minimize an approximate error of reconstructed spectrograms. Among the ICA algorithms that produce asymmetrically distributed outputs with a heavy tail, it is most closely related to non-negative ICA (Plumbley, 2003), although in applications non-negative ICA suffers from the problem that positive synaptic currents are completely unconstrained. A non-negative sparse-coding (NNSC) algorithm with a similar cost function has been described (Hoyer, 2002), but imposes some tighter restrictions on the mixing matrix J (the inverse of W) and has the undesirable property that outputs y_i^t cannot be computed in a single forward pass but require an iterative optimization procedure. Last but not least, our algorithm is different from nonnegative matrix factorization (NMF) (Lee and Seung, 1999), because NMF does not allow for STRFs with inhibitory subfields.

7.1.1. Song Selective Neurons in the Auditory Forebrain

In the model, the prominence of BOS in the training set and the shape of our cost function conjunctively forced neurons to display minimal suprathreshold synaptic currents to BOS (on average). These minimal responses explain why neurons preferred CON over BOS at moderately low firing thresholds, very much like Field-L neurons do. At higher thresholds, neurons responded to specific BOS features more than to other features. In the high-threshold regime, model neurons were BOS selective, and during BOS presentation they fired sparsely and were hyperpolarized on average, all very much like HVC mirror neurons that project to Area X ($\mathrm{HVC_X}$ neurons): $\mathrm{HVC_X}$ neurons are hyperpolarized by playback of the BOS and produce a high frequency burst in response to a very specific song feature (Mooney, 2000; Prather et al., 2008). However, at high thresholds, our model network did no longer function as a universal encoder of the auditory environment. Although BOS reconstructions worsened very gracefully with increasing threshold, the neural representation of non-preferred stimuli degraded rapidly.[1] This observation recommends the high-threshold regime only for specialized auditory areas such as HVC

---

[1] see Figure 5.10

and the low-threshold regime for lower auditory areas such as Field L that respond to a large variety of sounds.

Can selectivity for the TUT also be seen as a corollary of an efficient coding hypothesis? HVC neurons in juveniles tend to prefer TUT over most other stimuli including the BOS (Nick and Konishi, 2005), whereas in adults this selectivity reverses such that HVC neurons tend to prefer BOS over TUT (Margoliash, 1986; Nick and Konishi, 2005). Taken together, these data could be reproduced by our high-threshold model of HVC if TUT originally were the more prominent stimulus than BOS, but then BOS takes over as the most prominent stimulus.

The model may explain response properties in many higher auditory brain areas of songbirds including CM: medial CM responses are shaped by auditory memories and cells typically respond more to familiar than to unfamiliar songs Gentner and Margoliash (2003), in analogy to BOS preference seen in HVC. Interestingly, selective CM cells have lower spontaneous firing rates than non-selective cells, in agreement with the high-threshold regime of our model.

To ensure the dominance of BOS in the model it is the most frequent stimulus in the training set. One could argue that a zebra finch does not hear BOS more often than CON of another bird in the same cage. However, the frequency is just the most simple way to ensure dominance. It could also be induced by attentional factors, i.e. a weighting in cost function depending on the attentional state of the bird (singing, listening, passive, sleeping, ...).

7.1.2. STRFs and Their Relation to Spike Responses

Greene et al. (2009) applied a popular sparse coding algorithm to compute optimal linear kernels on large numbers of birdsong spectrograms. They evaluated model output in terms of receptive field shapes and found that a stronger sparseness prior during training led to stronger resemblance of model STRFs with STRFs in Field L.[2] The STRFs in the similar model presented in this thesis also qualitatively resembled STRFs in Field L.

---

[2] For HVC projection neurons no STRFs have ever been estimated, for reasons explained above.

Without density prior, model STRFs were denser in spectral and temporal modulations than Field L STRFs. However, I showed there is no principled discrepancy because by using a suitable density prior, we were able to modulate the STRF density almost arbitrarily (Figure 5.2B), implying that the model is amenable to fitting a large variety of experimental STRFs. Most importantly, this work suggests that neural firing may constitute a better model read out than receptive fields, because neural firing takes nonlinearities into account (such as the firing threshold), whereas receptive fields are linear and often poor descriptions of spike data. For example, I found that as a function of the firing threshold, the model, despite its fixed underlying STRFs, was able to reproduce qualitatively different responses as seen in Field L and HVC. Hence, a simple STRF may be far from ideal as a characterization of neural firing, because it may be associated with a diverse range of response behaviors.

7.1.3. Function of Selectivity Reversal

The model can reproduce the selectivity reversal seen in $\mathrm{HVC_X}$ neurons Mooney (2000). Based on the model I predict this to be a widespread phenomenon. I predict that sparsely firing neurons, when they are depolarized by constant current injections to fire densely, will display reduced or negative selectivity for their normally preferred stimulus. Similarly, I predict that densely firing neurons should lose or also reverse their selectivity while being hyperpolarized by constant current injection. These predictions apply to the mean and median selectivities in a large population (not to each individual cell) and should be relatively simple to verify using intracellular recordings. Note that these two predictions are surprising for neurons with monotonic frequency-current (F-I) curves f as in our model. I can speculate about the function of such selectivity reversal, if indeed widespread. If it were to be found in other animals and brain areas and were under volitional control, it could be used to attentionally screen the sensory environment for highly familiar stimuli (high threshold case), or to tune in on all kinds of stimuli with preference for unfamiliar ones (low threshold case).

Regarding shifts in excitatory/inhibitory balance, the model predicts that the effect on response selectivity depends on how balance shifts affect firing rates. For example, if increased inhibition leads to decreased firing rates, the model predicts increased response selectivity (for the BOS or an equivalent stimulus). In general, manipulations of excitation or inhibition within a network can lead to highly non-trivial reactions, e.g. disruption of local inhibition onto a cell can lead to lower baseline firing rate and to significant changes in firing patterns. For example, in Rosen and Mooney (2003), decreased G-protein coupled inhibition led to decreased baseline firing, which is counterintuitive and may be caused by nonlinear priming effects. The model is not able to explain such behavior, as it would need to include a more complex neuronal model including synaptic feedback. Nevertheless, because our theory applies in the direction of firing rate changes, the model predicts increased selectivity for the BOS when removal of inhibition decreases firing rates, which is what has been observed Rosen and Mooney (2003).

7.1.4. Applicability to Other Sensory Modalities

The findings may have relevance for the encoding of sensory modalities other than audition, including olfaction. In mammals, strong odorant-selective responses arise immediately downstream of primary sensory inputs (Davison and Katz, 2007). Though the neural mechanisms of this selectivity remain to be studied, in insects the mechanisms giving rise to sparse odor representations have been well characterized. The sparse odor representation in Kenyon cells arises from synchronized excitatory inputs mediated by densely firing projection neurons in the antennal lobe and by nonspecific inhibitory inputs from lateral horn interneurons that in essence set a high firing threshold to Kenyon cells (Laurent, 2002; Perez-Orive et al., 2002). The Kenyon cell's supralinear summation of EPSPs (Perez-Orive et al., 2004) represents a simple biophysical mechanism for achieving a long tail in the distribution of positive synaptic currents, a key element of our model. And, the control of firing threshold in Kenyon cells by global inhibitory input is well suited to endow these cells in principle with the ability to change response selectivity, for example if required by external circumstances.

7.1.5. HVC and Song Learning

As a model of HVC responses, the findings suggests that vocal-auditory mirrored activity in HVC has a sensory origin (activity in $\mathrm{HVC_X}$ neurons is mirrored in that auditory-evoked and singing-related responses in these cells are nearly identical (Prather et al., 2008)). In particular, my interpretation is that initially, HVC responses are shaped by the TUT; thereafter, HVC responses and their selectivity are further shaped by auditory feedback elicited by the BOS (Nick and Konishi, 2005), which during early sensorimotor song development is generated by a motor pathway that excludes HVC (Aronov et al., 2008). During this developmental phase, a network forms among HVC neurons and ultimately produces adult song. A sensory origin of the HVC network would imply that motor responses in HVC neurons learn to mirror sensory responses, not vice versa (HVC neurons learn to use auditory-feedback-elicited responses as future motor outputs, rather than their learning to map auditory feedback onto the HVC neurons that were involved in generating the feedback). In other words, when mirror neurons fire during motor behavior, they do so mainly because they have developed selectivity to the stimulus preceding their firing. More specifically, mirrored activity in HVC neurons could derive from essentially one assumption: that the local HVC network tries to maximize the drive of cells at the moments at which these fire, initially driven by sensory input. Accordingly, HVC synapses would allow for cells to drive each other at time lags at which their preferred TUT/BOS auditory features occur. Such specific function could arise for example by virtue of some spike-time dependent synaptic plasticity mechanisms (Bi and Poo, 1998; Jun and Jin, 2007; Fiete et al., 2010; D'Souza et al., 2010).

In conclusion, the architecture we have described shows that efficient coding constraints can explain the diversity of response specificity in higher sensory areas. Sparse/selective and dense/antiselective responses are at opposite extremes of the same efficient coding principle. It is possible that this link between response specificity and firing sparseness holds true also in other neural systems such as the neocortex. And, by extrapolation, our work shows that efficient coding constraints may guide the formation of sensory pathways all the way up to premotor areas, by which our work can shorten the gap between our understanding of pure sensory and pure motor codes.

7.1.6. The Algorithm from an Engineering Viewpoint

The algorithm offers a powerful method for bioacoustic signal analysis. The diversity of STRFs in the model is well matched with the behavioral richness of birdsong. Stereotyped syllables can be readily detected because they are represented essentially by a single STRF, whereas more variable syllables such as harmonic stacks may be associated with multiple STRFs as in Figure 5.4. Hence, the number of STRFs allocated to a particular syllable or sub-syllable may reflect its variability. The level of song analysis (detailed vs coarse) can be controlled by the number of neurons in the network. One can imagine uses of the algorithm for detecting particular song variants (such as high-pitched versions of harmonic stacks), or for identifying similar notes within different syllables, etc.

From an engineering perspective, one computational benefit of the sparseness transformation in the model is smart noise reduction. Using a high firing threshold during reconstruction, sounds to which cells have not been exposed during training can be effectively filtered out. More interestingly, by omitting certain neurons during reconstruction (e.g., by setting their firing thresholds to infinity), undesirable signals encountered in the training set can be conveniently suppressed. For example, by excluding the five neurons that encoded high-frequency noise (e.g. Neuron 106 in Figure 5.3) BOS could be efficiently cleaned from that noise. Of course the brain may make use of such smart noise reduction without ever explicitly having to reconstruct the original input; for example, feature-based attentional inputs may selectively suppress the firing in some neurons to constrain downstream processing to only relevant sensory features. Although such selective suppression has not been found yet in songbirds, birds may possess attentional selection mechanisms because they can detect subtle acoustic features and adapt their songs when negatively reinforced (Tumer and Brainard, 2007).

An additional application of the algorithm is the qualitative analysis of the development of behavioral data over time. If development is dedicated towards the deployment of distinct features or patterns the algorithm is able to identify such features long before their perfection. Such feature development may be mainly unsupervised and guided e.g. by aesthetics, such as in the bird song or in children's drawings, but also supervised or reward-driven, such as in hunting strategies. The algorithm allows to detect the emergence, the development, and the disuse of features.

Finally the algorithm could be used to recognize and separate known stimuli. Similar to noise suppression one can make use of the algorithm building models of the stimuli. The problem lies only therein to determine which of the features belongs to which stimulus or to which group of stimuli. Once these belongings are determined recognition and separation can be done very fast in a single forward pass with no need for iterative real-time optimization, in contrast to algorithms such as NMF (Lee and Seung, 1999), or the NNSC algorithm proposed by Hoyer (2002).

7.2. Conclusion

In this thesis I presented a short summary of the songbird's auditory system and explained a set of measures to statistically asses neural activity in such systems. I introduced a new non-negative sparse coding algorithm, applied it to model computation performed in the songbird's auditory cortex and compared the characteristics of the model to data from songbird studies. In a last part I indicated a set of possible application of the algorithm as a computational tool.

The model above is not a singular one as others have been presented before for the songbird's auditory system (Greene et al., 2009) or sensory modalities in a variety of animals (Bossomaier and Snyder, 1986; Hancock et al., 1992; Bell and Sejnowski, 1997; Olshausen and Field, 1996, 1997; Hyvarinen and Hoyer, 2001; Smith and Lewicki, 2006). However the model makes a prediction regarding selectivity reversal which to my knowledge has not been predicted by any other model yet. The beauty of the model further lies in its generality - it explains behavior in cortical areas with different functions and functionalities - and in its simplicity - only the firing threshold θ has to be varied to produce the different encodings found in the different cortical areas.

The engineering applications presented are no sophisticated tools but should rather serve as inspiration and open the window to a vast field of tasks that this algorithm and sparse coding in general might help to tackle.

Appendix A

List of Variables and Functions

Rule of thumb: Capital letter variables are matrices or vectors, minor letter variables are scalars or single elements of matrices or vectors. Functions are followed by arguments in parentheses. Actually, there are a few exceptions.

Variable	**Description**	**Type**
$\mathbf{I}$	unity matrix	matrix
$\mathbb{1}$	vector of ones	vector
Δ	chirp factor	scalar
Δt	temporal spacing of the spectrogram	scalar
ΔW	update of W	matrix
δf	spectral resolution	scalar
δt	temporal resolution	scalar
η	(Gaussian white) noise	scalar
η^t	(Gaussian white) noise	vector
η_i^t	noise on the synaptic current of neuron i at time t	scalar
Γ	weight for Tichonov-regularization	matrix
Λ, E	eigenvalue and eigenvector matrices of the stimulus autocorrelation matrix	matrix

Variable	**Description**	**Type**
μ	weight for Tichonov-regularization	scalar
$\Sigma(.), \Sigma(.\|.)$	covariance matrix, conditional covariance matrix	matrix
$\sigma(.)$	signum function	scalar function
$\sigma^2_.$	variance of	scalar
${\sigma_i^{t\|u}}^2$	variance of the i-th signal estimation at time t give the response R^{t+u}	scalar
τ^n	step size in the optimization algorithm	scalar
τ	temporal size of the receptive fields	scalar
θ	firing threshold	scalar
ξ	total transformation	matrix
ξ^{-1}	pseudoinverse of ξ, given by $\xi^{-1} = E \cdot \Lambda^{1/2} \cdot W^{-1}$	matrix
ξ_i	total transformation onto neuron i, STRF	vector
$\xi_i(.,.)$	transformation function onto neuron i	scalar function
ξ_i^u	transformation onto neuron i with time delay of u	vector
$(\xi^{-1})_i^u$	inverse transformation to signal i with delay u	vector
A	mixing matrix	matrix
A^n	gradient (to B) of the cost function $F(.)$ at the n-th optimization step	matrix
B^n	parametrization of the sparseness transformation matrix W at the n-th optimization step	matrix
b_{mn}, a_{mn}	mn-th element of the matrix B, A	scalar
C	covariance matrix	matrix
c	trade-off between sparseness and reconstruction error	scalar

Variable	**Description**	**Type**
c_s	weighting factor of RF density	scalar
C_{SR}	crosscovariance between stimulus and response	matrix
C_{SS}	stimulus covariance	matrix
$\mathrm{covar}(.,.)$	covariance	scalar function
d'	selectivity	scalar
$\mathrm{diag}(.)$	diagonal of a matrix	vector
$E(.)$	expected value	same type as argument
$E(.\|.)$	conditional expected value	same type as first argument
$F(.)$	cost function to be minimized	scalar function
$\mathcal{F}(\omega)(f)$	Fourier transform of function f at frequency ω	scalar function
$f(.)$	elementwise cost function	scalar function
$f(.)$	time dependent frequency of a chirplet	scalar function
f_c	center frequency of a chirplet	scalar
f_n	center frequency of the n-th frequency band	scalar
$g(.)$	elementwise nonlinear function	same type as argument
h	receptive field	matrix
J	left inverse of sparseness transformation W	matrix
K	vector of constant offset	vector
k	extent of smoothing function $U(.)$	scalar
M_n	normalization	scalar
N	number of neurons	scalar
N_0	dimensionality of the cochlear input	scalar
N_2	number of neurons in the second layer	scalar
n_a	fraction of time a source is active, i.e. nonzero	scalar

Variable	**Description**	**Type**
P	PCA transformation	matrix
$p(.)$	prior probability	scalar function
$p(.\|.)$	conditional probability	scalar function
R	firing rates	matrix
$R(.)$	sparseness enforcing function	scalar function
R^t	firing rates at time t	vector
RS_A	response strength of stimulus A	scalar
$\tilde{r}^t$	predicted response at time t	scalar
r_i^t	firing rates of neuron i at time t	scalar
$\bar{r}_A$	mean firing rate in response to stimulus A	scalar
S	(unknown) causes of stimulus X	matrix
s	number of frequency bands per octave	scalar
s	excerpt of sound for frequency analysis	vector
S_A	set of filters encoding a stimulus A	set
s_b	batch size	scalar
$U(.)$	elementwise smoothing function	same type as argument
$V(.)$	derivative of smoothing function $U(.)$	same type as argument
W	sparseness transformation	matrix
W	unmixing matrix	matrix
$W(.)$	chirplet transformation matrix	matrix
w	windowing function	scalar function
X	log-power spectrogram of a stimulus / matrix of any stimuli	matrix
X_P	PCA-transformed stimuli	matrix
$x(.,.)$	spectrogram function	scalar function
$X^{t:u}$	spectrogram from time t until time u	vector
X^t	spectrogram at time t / single stimulus	vector

Variable	**Description**	**Type**
$\hat{X}_p^t$	reconstruction of the PCA-transform X_p^t	vector
X_{rec}^t	reconstruction of the spectrogram $X^{t-\tau:t}$ from $\hat{X}_p^t$	vector
x_i^t	spectrogram amplitude at time t and frequency i	scalar
X_p^t	PCA-transform of the spectrogram $X^{t-\tau:t}$ / PCA-transformed stimulus	vector
Y	synaptic currents	matrix
y_0	subthreshold minimum	scalar
Y^t	synaptic currents at time t	vector
Y_{E-}^t	estimation of subthreshold synaptic currents at time t, zero, where $y_i^t \geq \theta$	vector
Y_+^t	suprathreshold synaptic currents at time t, zero, where $y_i^t < \theta$	vector
y_i^t	synaptic current of neuron i at time t	scalar
z_A	z-score of stimulus A	scalar
$(.)_{ii}^{uu}$	entry in the (correlation) matrix for signal i at time u on the diagonal	scalar
$.^T$	transpose of	matrix function

Appendix B

List of Abbreviations

Abbreviation	**Full Name**
Av	Nucleus Avalanche
BOLD	Blood oxygenation level-dependent
BOS	Bird's own song
BSS	Blind Source (or Signal) Separation
CF	Characteristic frequency: stimulus frequency to which a neuron responds best
CLM	Caudal lateral mesopallium
CM	Caudal mesopallium
CMM	Caudal medial mesopallium
CON	Conspecific song: Song by another bird of the same species
CPP	Cocktail party problem
DLM	Medial nucleus of the dorsolateral thalamus
eMTF	Ensemble modulation transfere function
fMRI	Functional magnetic resonance imaging
FT	Fourier transform
HVC	Letter-based proper name
HVC shelf	Shelf of HVC
ICA	Independent component analysis
ILD	Interaural level difference

Abbreviation	**Full Name**
ITD	Interaural time difference
L	Field L
L*	Subfield L* (where * is 1, 2a, 2b, or 3)
LL	Lateral lemniscal nuclei
LMAN	Lateral magnocellular nucleus of the anterior nidopallium
MID	Maximally informative dimensions
MLd	Nucleus mesencephalicus lateralis pars dorsalis
MMAN	Medial magnocellular nucleus of the anterior nidopallium
MTF	Modulation transfer function
NA	Nucleus angularis
NCM	Caudal medial nidopallium
NIf	Interfacial nucleus of the nidopallium
NL	Nucleus laminaris
NM	Nucleus magnocellularis
NMF	Nonnegative matrix factorization
NNSC	Nonnegative sparse coding
Ov	Nucleus ovoidalis
Ov core	Core of nucleus ovoidalis
Ov shell	Shell of nucleus ovoidalis
Ovm	Nucleus ovoidalis medialis
ParaHVC	Letter-based proper name
PCA	Principle component analysis
RA	Robust nucleus of the archopallium
RA cup	Cup of the robust nucleus of the archopallium
REV	Bird's own song play in reverse
RLC-circuit	Proper name
RF	Receptive Field
SO	Superior olive
STFT	Short-time Fourier transfrom
STRF	Spectral temporal receptive fields (sometimes called spectro-temporal)

Abbreviation	Full Name
SVM	Support vector machine
UVA	Nucleus uvaeformis
VP	Ventral pallidum
VTA	Ventral tegmental area

Bibliography

Aertsen, A. M. and Johannesma, P. I. (1981). A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. *Biol Cybern*, 42(2):145–156.

Aiello, L. C. and Wheeler, P. (1995). The expensive-tissue hypothesis: The brain and the digestive system in human and primate evolution. *Current Anthropology*, 36(2):199.

Airey, D. C., Buchanan, K. L., Szekely, T., Catchpole, C. K., and De-Voogd, T. J. (2000). Song, sexual selection, and a song control nucleus (hvc) in the brains of european sedge warblers. *Journal of Neurobiology*, 44(1):1–6.

Akesson, T., Lanerolle, N. D., and Cheng, M.-F. (1987). Ascending vocalization pathways in the female ring dove: Projections of the nucleus intercollicularis. *Experimental Neurology*, 95(1):34 – 43.

Akutagawa, E. and Konishi, M. (2005). Connections of thalamic modulatory centers to the vocal control system of the zebra finch. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39):14086–14091.

Akutagawa, E. and Konishi, M. (2010). New brain pathways found in the vocal control system of a songbird. *The Journal of Comparative Neurology*, 518(15):3086–3100.

Alvarez-Buylla, A. and Kirn, J. R. (1997). Birth, migration, incorporation, and death of vocal control neurons in adult songbirds. *Journal of Neurobiology*, 33(5):585–601.

Amari, S.-i. (1997). Neural learning in structured parameter spaces - natural riemannian gradient. In *In Advances in Neural Information Processing Systems*, pages 127–133. MIT Press.

Amin, N., Doupe, A., and Theunissen, F. E. (2007). Development of selectivity for natural sounds in the songbird auditory forebrain. *J. Neurophysiol.*, 97(5):3517–3531.

Amin, N., Gill, P. R., and Theunissen, F. E. (2010). The role of the zebra finch auditory thalamus in generating complex representations for natural sounds. *J Neurophysiol*, page jn.00128.2010.

Amin, N., Grace, J. A., and Theunissen, F. E. (2004). Neural response to birds own song and tutor song in the zebra finch field l and caudal mesopallium. *J. Comp. Physiol. [A]*, 190(6):469–489.

Andalman, A. S. and Fee, M. S. (2009). A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc. Natl. Acad. Sci. U.S.A.*, 106:12518–12523.

Anderson, I. (2000). The secret language of birds. [Audio CD]. Fuel 2000.

Araki, S., Makino, S., Sawada, H., and Mukai, R. (2004). Underdetermined blind separation of convolutive mixtures of speech with directivity pattern based mask and ica. In Puntonet, C. and Prieto, A., editors, *Independent Component Analysis and Blind Signal Separation*, volume 3195 of *Lecture Notes in Computer Science*, pages 898–905. Springer Berlin Heidelberg.

Arnold, A. P. (1975). The effects of castration on song development in zebra finches (poephila guttata). *Journal of Experimental Zoology*, 191(2):261–277.

Aronov, D., Andalman, A. S., and Fee, M. S. (2008). A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science*, 320:630–634.

Arons, B. (1992). A review of the cocktail party effect. *Journal of the American Voice I/O Society*, 12:35–50.

Asari, H., Pearlmutter, B. A., and Zador, A. M. (2006). Sparse Representations for the Cocktail Party Problem. *J. Neurosci.*, 26(28):7477–7490.

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network Computation in Neural Systems*, 3(2):213–251.

Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1:371–394.

Bauer, E. E., Coleman, M. J., Roberts, T. F., Roy, A., Prather, J. F., and Mooney, R. (2008). A synaptic basis for auditory-vocal integration in the songbird. *J. Neurosci.*, 28:1509–1522.

Bell, A. J. and Sejnowski, T. J. (1995). Blind separation and blind deconvolution: an information-theoretic approach. *Proc. Internat. Conf. Acoust. Speech Signal Process., Detroit*, 5:3415–3418.

Bell, A. J. and Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision research*, 37(23):3327–3338.

Berwick, R. C., Okanoya, K., Beckers, G. J., and Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences*, 15(3):113 – 121.

Bi, G. and Poo, M. (1998). Synaptic modification in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.*, 18:10464–10472.

Bigalke-Kunz, B., Rübsamen, R., and Dörrscheidt, G. (1987). Tonotopic organization and functional characterization of the auditory thalamus in a songbird, the european starling. *Journal of Comparative Physiology A*, 161:255–265.

Blakers, M., Davies, S., and Reilly, P. (1984). *The Atlas of Australian Birds*. Melbourne University Press, Melbourne.

Blättler, F. and Hahnloser, R. H. R. (2011). An efficient coding hypothesis links sparsity and selectivity of neural responses. *PLoS ONE*, 6(10):e25506.

Blättler, F., Kollmorgen, S., Herbst, J., and Hahnloser, R. (2011). Hidden markov models in the neurosciences. In Dymarski, P., editor, *Hidden Markov Models, Theory and Applications*, pages 169–186. InTech.

Böhner, J. (1990). Early acquisition of song in the zebra finch, taeniopygia guttata. *Animal Behaviour*, 39(2):369 – 374.

Bonke, B., Bonke, D., and Scheich, H. (1979). Connectivity of the auditory forebrain nuclei in the guinea fowl (numida meleagris). *Cell and Tissue Research*, 200:101–121.

Bossomaier, T. and Snyder, A. W. (1986). Why spatial frequency processing in the visual cortex? *Vision Research*, 26(8):1307 – 1309.

Bottjer, S. W., Halsema, K. A., Brown, S. A., and Miesner, E. A. (1989). Axonal connections of a forebrain nucleus involved with vocal learning in zebra finches. *The Journal of Comparative Neurology*, 279(2):312–326.

Brainard, M. S. and Doupe, A. J. (2000). Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature*, 404(6779):762–766.

Brauth, S. E., Liang, W., Tang, Y., Galdzicka, E., and Hall, W. S. (2007). Rapid contact call-driven induction of nr2a and nr2b nmda subunit mrnas in the auditory thalamus of the budgerigar (melopsittacus undulatus). *Neurobiology of Learning and Memory*, 88(1):33 – 39.

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. A Bradford Book.

Brenowitz, E., Nalls, B., Wingfield, J., and Kroodsma, D. (1991). Seasonal changes in avian song nuclei without seasonal changes in song repertoire. *The Journal of Neuroscience*, 11(5):1367–1374.

Bressloff, P. C. (1995). Dynamics of a compartmental model integrate-and-fire neuron with somatic potential reset. *Physica D: Nonlinear Phenomena*, 80(4):399 – 412.

Brownell, W., Bader, C., Bertrand, D., and de Ribaupierre, Y. (1985). Evoked mechanical responses of isolated cochlear outer hair cells. *Science*, 227(4683):194–196.

Bultan, A. (1999). A four-parameter atomic decomposition of chirplets. *Signal Processing, IEEE Transactions on*, 47(3):731 –745.

Cade, T. J., Tobin, C. A., and Gold, A. (1965). Water economy and metabolism of two estrildine finches. *Physiological Zoology*, 38(1):9–33.

Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L., and Woolley, S. M. N. (2011). A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS ONE*, 6(1):e16104.

Cardin, J. A., Raksin, J. N., and Schmidt, M. F. (2005). Sensorimotor nucleus NIf is necessary for auditory processing but not vocal motor output in the avian song system. *J. Neurophysiol.*, 93:2157–2166.

Cardin, J. A. and Schmidt, M. F. (2003). Song system auditory responses are stable and highly tuned during sedation, rapidly modulated and unselective during wakefulness, and suppressed by arousal. *J. Neurophysiol.*, 90:2884–2899.

Cardin, J. A. and Schmidt, M. F. (2004a). Auditory responses in multiple sensorimotor song system nuclei are co-modulated by behavioral state. *Journal of Neurophysiology*, 91(5):2148–2163.

Cardin, J. A. and Schmidt, M. F. (2004b). Noradrenergic inputs mediate state dependence of auditory responses in the avian song system. *J. Neurosci.*, 24:7745–7753.

Cardoso, J.-F. (1997). Infomax and maximum likelihood for source separation. *IEEE Letters on Signal Processing*, 4(4):112–114.

Cardoso, J.-F. (2000). *Unsupervised adaptive filters*, volume 1, chapter Entropic contrasts for source separation: geometry and stability, pages 139–190. John Wiley & sons, Simon Haykin editor. First presented at the NIPS*96 workshop on ‘Blind Signal Processing’ organized by A. Cichocki.

Cariani, P. A. (2001). Temporal coding of sensory information in the brain. *Acoustical Science and Technology*, 22(2):77–84.

Carr, C. E., Fujita, I., and Konishi, M. (1989). Distribution of GABAergic neurons and terminals in the auditory system of the barn owl. *J. Comp. Neurol.*, 286:190–207.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25(5):975–979.

Chew, S. J., Mello, C., Nottebohm, F., Jarvis, E., and Vicario, D. S. (1995). Decrements in auditory responses to a repeated conspecific song are long-lasting and require two periods of protein synthesis in the songbird forebrain. *Proc. Natl. Acad. Sci. U.S.A.*, 92:3406–3410.

Cohen, Y. E., Theunissen, F., Russ, B. E., and Gill, P. (2007). Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *J. Neurophysiol.*, 97(2):1470–1484.

Coleman, M. J. and Mooney, R. (2004). Synaptic transformations underlying highly selective auditory representations of learned birdsong. *J. Neurosci.*, 24:7251–7265.

Coleman, M. J., Roy, A., Wild, J. M., and Mooney, R. (2007). Thalamic Gating of Auditory Responses in Telencephalic Song Control Nuclei. *J. Neurosci.*, 27(37):10024–10036.

Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314.

Correia, M., Eden, A., Westlund, K., and Coulter, J. (1982). Organization of ascending auditory pathways in the pigeon (columbia livia) as determined by autoradiographic methods. *Brain Research*, 234(2):205 – 212.

Dave, A. S., Yu, A. C., and Margoliash, D. (1998). Behavioral state modulation of auditory activity in a vocal motor system. *Science*, 282(5397):2250–2254.

David, S. V., Vinje, W. E., and Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of v1 neurons. *J. Neurosci.*, 24(31):6991–7006.

Davison, I. G. and Katz, L. C. (2007). Sparse and selective odor coding by mitral/tufted neurons in the main olfactory bulb. *J. Neurosci.*, 27:2091–2101.

Delfosse, N. and Loubaton, P. (1995). Adaptive blind separation of independent sources: a deflation approach. *Signal Process.*, 45(1):59–83.

Delgado, K. K., Murray, J. F., Rao, B. D., Engan, K., Lee, T. W., and Sejnowski, T. J. (2003). Dictionary learning algorithms for sparse representation. *Neural Comput.*, 15(2):349–396.

Deregnaucourt, S., Mitra, P. P., Feher, O., Pytte, C., and Tchernichovski, O. (2005). How sleep affects the developmental learning of bird song. *Nature*, 433(7027):710–716.

Devoogd, T. J., Krebs, J. R., Healy, S. D., and Purvis, A. (1993a). Relations between song repertoire size and the volume of brain nuclei related to song: Comparative evolutionary analyses amongst oscine birds. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 254(1340):75–82.

Devoogd, T. J., Krebs, J. R., Healy, S. D., and Purvis, A. (1993b). Relations between song repertoire size and the volume of brain nuclei related to song: Comparative evolutionary analyses amongst oscine birds. *Proceedings: Biological Sciences*, 254(1340):pp. 75–82.

Doupe, A. J. (1997). Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J. Neurosci.*, 17:1147–1167.

Doupe, A. J. and Konishi, M. (1991). Song-selective auditory circuits in the vocal control system of the zebra finch. *Proceedings of the National Academy of Sciences*, 88(24):11339–11343.

D'Souza, P., Liu, S. C., and Hahnloser, R. H. (2010). Perceptron learning rule derived from spike-frequency adaptation and spike-time-dependent plasticity. *Proc. Natl. Acad. Sci. U.S.A.*, 107:4722–4727.

Durgin, F. H. (1995). Texture density adaptation and the perceived numerosity and distribution of texture,. *Journal of Experimental Psychology: Human Perception and Performance*, 21(1):149 – 169.

Eagleman, D. M. (2001). Visual illusions and neurobiology. *Nat. Rev. Neurosci.*, 2:920–926.

Eales, L. A. (1985). Song learning in zebra finches: some effects of song model availability on what is learnt and when. *Animal Behaviour*, 33(4):1293 – 1300.

Endler, J. A. and Basolo, A. L. (1998). Sensory ecology, receiver biases and sexual selection. *Trends in Ecology & Evolution*, 13(10):415 – 420.

Farries, M. A. (2004). The avian song system in comparative perspective. *Annals of the New York Academy of Sciences*, 1016(1):61–76.

Feenders, G., Liedvogel, M., Rivas, M., Zapka, M., Horita, H., Hara, E., Wada, K., Mouritsen, H., and Jarvis, E. D. (2008). Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS ONE*, 3(3):e1768.

Feher, O., Wang, H., Saar, S., Mitra, P. P., and Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. *Nature*, 459(7246):564–568.

Fiete, I. R., Senn, W., Wang, C. Z., and Hahnloser, R. H. (2010). Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron*, 65:563–576.

Fischer, F. P. (1994). General pattern and morphological specializations of the avian cochlea. *Scanning Microsc.*, 8:351–363.

Fiser, J., Berkes, P., Orbn, G., and Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, 14(3):119 – 130.

Fortune, E. S. and Margoliash, D. (1992). Cytoarchitectonic organization and morphology of cells of the field l complex in male zebra finches (taenopygia guttata). *The Journal of Comparative Neurology*, 325(3):388–404.

Fortune, E. S. and Margoliash, D. (1995). Parallel pathways and convergence onto HVc and adjacent neostriatum of adult zebra finches (Taeniopygia guttata). *J. Comp. Neurol.*, 360(3):413–441.

Foster, E. F. and Bottjer, S. W. (1998). Axonal connections of the high vocal center and surrounding cortical regions in juvenile and adult male zebra finches. *The Journal of Comparative Neurology*, 397(1):118–138.

Foster, E. F., Mehta, R. P., and Bottjer, S. W. (1997). Axonal connections of the medial magnocellular nucleus of the anterior neostriatum in zebra finches. *The Journal of Comparative Neurology*, 382(3):364–381.

Gahr, M. and Metzdorf, R. (1999). The sexually dimorphic expression of androgen receptors in the song nucleus hyperstriatalis ventrale pars caudale of the zebra finch develops independently of gonadal steroids. *The Journal of Neuroscience*, 19(7):2628–2636.

Gale, S. D. and Perkel, D. J. (2010). A basal ganglia pathway drives selective auditory responses in songbird dopaminergic neurons via disinhibition. *The Journal of Neuroscience*, 30(3):1027–1037.

Gale, S. D., Person, A. L., and Perkel, D. J. (2008). A novel basal ganglia pathway forms a loop linking a vocal learning circuit with its dopaminergic input. *The Journal of Comparative Neurology*, 508(5):824–839.

Gehr, D. D., Capsius, B., Grabner, P., Gahr, M., and Leppelsack, H. J. (1999). Functional organisation of the field-L-complex of adult male zebra finches. *Neuroreport*, 10(2):375–380.

Gentner, T. Q. (2004). Neural systems for individual song recognition in adult birds. *Ann. N. Y. Acad. Sci.*, 1016:282–302.

Gentner, T. Q. and Margoliash, D. (2003). Neuronal populations and single cells representing learned auditory objects. *Nature*, 424:669–674.

Gill, P., Zhang, J., Woolley, S., Fremouw, T., and Theunissen, F. (2006). Sound representation methods for spectro-temporal receptive field estimation. *Journal of Computational Neuroscience*.

Gleich, O. and Manley, G. A. (1988). Quantitative morphological analysis of the sensory epithelium of the starling and pigeon basilar papilla. *Hearing Research*, 34(1):69 – 85.

Gleich, O., Ryals, B., and Dooling, R. (1998). The number of auditory nerve fibers in normal canaries and in belgian waterslager canaries. *Abstr. 21st Midwinter Meeting ARO, St. Petersburg Beach, FL*, page 198.

Gobes, S. M. and Bolhuis, J. J. (2007). Birdsong memory: a neural dissociation between song recognition and production. *Curr. Biol.*, 17:789–793.

Goelet, P., Castellucci, V. F., Schacher, S., and Kandel, E. R. (1986). The long and the short of long-term memory–a molecular framework. *Nature*, 322(6078):419–422.

Goldberg, J. H. and Fee, M. S. (2011). Vocal babbling in songbirds requires the basal ganglia-recipient motor thalamus but not the basal ganglia. *Journal of Neurophysiology*, 105(6):2729–2739.

Grace, J. A., Amin, N., Singh, N. C., and Theunissen, F. E. (2003). Selectivity for Conspecific Song in the Zebra Finch Auditory Forebrain. *J. Neurophysiol.*, 89(1):472–487.

Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. John Wiley & Sons Ltd, New York.

Greene, G., Barrett, D. G., Sen, K., and Houghton, C. (2009). Sparse coding of birdsong and receptive field structure in songbirds. *Network: Computation in Neural Systems*, 20(3):162–177.

Griffin, D. and Lim, J. (1984). Signal estimation from modified short-time fourier transform. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 32(2):236–243.

Hahnloser, R. H. and Kotowicz, A. (2010). Auditory representations and memory in birdsong learning. *Current Opinion in Neurobiology*, 20(3):332 – 339. ¡ce:title¿Sensory systems¡/ce:title¿.

Hahnloser, R. H., Kozhevnikov, A. A., and Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, 419(6902):65–70.

Hahnloser, R. H. R., Wang, C. Z.-H., Nager, A., and Naie, K. (2008). Spikes and bursts in two types of thalamic projection neurons differentially shape sleep patterns and auditory responses in a songbird. *The Journal of Neuroscience*, 28(19):5040–5052.

Hancock, P., Bradley, R., and Smith, L. (1992). The principal components of natural images. *Network 3*, pages 61–70.

Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology*, 121:400–415.

Hauber, M. E., Woolley, S. M., Cassey, P., and Theunissen, F. E. (2013). Experience dependence of neural responses to different classes of male songs in the primary auditory forebrain of female songbirds. *Behavioural Brain Research*, 243(0):184 – 190.

Heinrich Walter, Elisabeth Harnickell, D. M.-D. (1975). *Climate-diagram maps of the individual continents and the ecological climatic regions of the earth*. Springer-Verlag, Berlin.

Herculano-Houzel, S. (2010). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, 4(0):12.

Hessler, N. A. and Doupe, A. J. (1999). Singing-related neural activity in a dorsal forebrainbasal ganglia circuit of adult zebra finches. *The Journal of Neuroscience*, 19(23):10461–10481.

Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (Lond.)*, 117(4):500–544.

Hosino, T. and Okanoya, K. (2000). Lesion of a higher-order song nucleus disrupts phrase level complexity in Bengalese finches. *Neuroreport*, 11(10):2091–2095.

Hoyer, P. O. (2002). Non-negative sparse coding. *Neural Networks for Signal Processing*, XII:557–565.

Hromdka, T., Deweese, M. R., and Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol.*, 6(1):e16.

Hsu, A., Woolley, S. M. N., Fremouw, T. E., and Theunissen, F. E. (2004). Modulation Power and Phase Spectrum of Natural Sounds Enhance Neural Encoding Performed by Single Auditory Neurons. *J. Neurosci.*, 24(41):9201–9211.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol. (Lond.)*, 148:574–591.

Hyson, R. L. (2005). The analysis of interaural time differences in the chick brain stem. *Physiology & Behavior*, 86(3):297 – 305. Florida State University Special Issue - Florida State University Special Issue.

Hyvärinen, A. (1998). New approximations of differential entropy for independent component analysis and projection pursuit. In *Advances in Neural Information Processing Systems 10*, pages 273 – 279. MIT Press.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural. Netw.*, 10:626–634.

Hyvärinen, A. (1999). Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128.

Hyvarinen, A. and Hoyer, P. O. (2001). A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Res.*, 41:2413–2423.

Hyvärinen, A. and Inki, M. (2002). Estimating overcomplete independent component bases for image windows. *J. Math. Imaging Vis.*, 17(2):139–152.

Hyvarinen, A. and Oja, E. (1997). A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Comp.*, 9(7):1483–1492.

Hyvärinen, A. and Oja, E. (1998). Independent component analysis by general nonlinear hebbian-like learning rules. *Signal Process.*, 64(3):301–313.

Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Netw*, 13:411–430.

Immelmann, K. (1959). Experimentelle Untersuchungen über die biologische Bedeutung artspezifischer Merkmale beim Zebrafinken (Taeniopygia castanotis Gould). *Zoologische Jahrbücher, Abt. f. Systematik*, 86:437–592.

Immelmann, K. (1962). Beitrge zu einer vergleichenden biologie australischer prachtfinken (spermestidae). *Zool Jb Syst*, 1(90):1–196.

Immelmann, K. (1965). Prägungserscheinungen in der gesangsentwicklung junger zebrafinken. *Naturwissenschaften*, 52:169–170.

James, W. (1890). *The Principles of Psychology.* New York: Henry Holt.

Janata, P. and Margoliash, D. (1999). Gradual Emergence of Song Selectivity in Sensorimotor Structures of the Male Zebra Finch Song System. *J. Neurosci.*, 19(12):5108–5118.

Janik, V. and Slater, P. (1997). Vocal learning in mammals. volume 26 of *Advances in the Study of Behavior*, pages 59 – 99. Academic Press.

Jarvis, E. D., Gunturkun, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., Medina, L., Paxinos, G., Perkel, D. J., Shimizu, T., Striedter, G., Wild, J. M., Ball, G. F., Dugas-Ford, J., Durand, S. E., Hough, G. E., Husband, S., Kubikova, L., Lee, D. W., Mello, C. V., Powers, A., Siang, C., Smulders, T. V., Wada, K., White, S. A., Yamamoto, K., Yu, J., Reiner, A., and Butler, A. B. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nat. Rev. Neurosci.*, 6(2):151–159.

Jarvis, E. D. and Nottebohm, F. (1997). Motor-driven gene?expression. *Proceedings of the National Academy of Sciences*, 94(8):4097–4102.

Jarvis, E. D., Scharff, C., Grossman, M. R., Ramos, J. A., and Nottebohm, F. (1998). For whom the bird sings: Context-dependent gene expression. *Neuron*, 21(4):775 – 788.

Jeanne, J. M., Thompson, J. V., Sharpee, T. O., and Gentner, T. Q. (2011). Emergence of learned categorical representations within an auditory forebrain circuit. *The Journal of Neuroscience*, 31(7):2595–2606.

Jonas, J. B., Schneider, U., and Naumann, G. O. H. (1992). Count and density of human retinal photoreceptors. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 230(6):505–510.

Judd, D. B. and Wyszecki, G. (1975). *Color in business, science, and industry.* Wiley, New York :, 3d ed. edition.

Jun, J. K. and Jin, D. Z. (2007). Development of neural circuitry for precise temporal sequences through spontaneous activity, axon remodeling, and synaptic plasticity. *PLoS ONE*, 2:e723.

Jutten, C. and Herault, J. (1991). Blind separation of sources, part 1: an adaptive algorithm based on neuromimetic architecture. *Signal Process.*, 24(1):1–10.

Kao, M. H., Doupe, A. J., and Brainard, M. S. (2005). Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature*, 433(7026):638–643.

Karten, H. J. (1967). The organization of the ascending auditory pathway in the pigeon (columba livia) i. diencephalic projections of the inferior colliculus (nucleus mesencephali lateralis, pars dorsalis). *Brain Research*, 6(3):409 – 427.

Karten, H. J. (1968). The ascending auditory pathway in the pigeon (columba livia) ii. telencephalic projections of the nucleus ovoidalis thalami. *Brain Research*, 11(1):134 – 153.

Keller, G. B. and Hahnloser, R. H. (2009). Neural processing of auditory feedback during vocal practice in a songbird. *Nature*, 457:187–190.

Kennedy, C. and Sokoloff, L. (1957). An adaptation of the nitrous oxide method to the study of the cerebral circulation in children; normal values for cerebral blood flow and cerebral metabolic rate in childhood. *J. Clin. Invest.*, 36:1130–1137.

Konishi, M. (1970). Comparative neurophysiological studies of hearing and vocalizations in songbirds. *J. Comp. Physiol. A*, 66:257–272.

Köppl, C. (1997a). Number and axon calibres of cochlear afferents in the barn owl. *Auditory Neurosci*, 3:313–334.

Köppl, C. (1997b). Phase Locking to High Frequencies in the Auditory Nerve and Cochlear Nucleus Magnocellularis of the Barn Owl, Tyto alba. *J. Neurosci.*, 17(9):3312–3321.

Köppl, C., Manley, G. A., and Konishi, M. (2000a). Auditory processing in birds. *Current Opinion in Neurobiology*, 10(4):474 – 481.

Köppl, C., Wegscheider, A., Gleich, O., and Manley, G. A. (2000b). A quantitative study of cochlear afferent axons in birds. *Hearing Research*, 139(1-2):123 – 143.

Kozhevnikov, A. A. and Fee, M. S. (2007). Singing-related activity of identified hvc neurons in the zebra finch. *Journal of Neurophysiology*, 97(6):4271–4283.

Krützfeldt, N. O., Logerot, P., Kubke, M. F., and Wild, J. M. (2010a). Connections of the auditory brainstem in a songbird, taeniopygia guttata. i. projections of nucleus angularis and nucleus laminaris to the auditory torus. *The Journal of Comparative Neurology*, 518(11):2109–2134.

Krützfeldt, N. O., Logerot, P., Kubke, M. F., and Wild, J. M. (2010b). Connections of the auditory brainstem in a songbird, taeniopygia guttata. ii. projections of nucleus angularis and nucleus laminaris to the superior olive and lateral lemniscal nuclei. *The Journal of Comparative Neurology*, 518(11):2135–2148.

Laheld, B. and Cardoso, J.-F. (1994). Adaptive source separation with uniform performance. In *Proc. EUSIPCO*, pages 183–186, Edinburgh.

Langner, G., Bonke, D., and Scheich, H. (1981). Neuronal discrimination of natural and synthetic vowels in field l of trained mynah birds. *Experimental Brain Research*, 43:11–24.

Laurent, G. (2002). Olfactory network dynamics and the coding of multidimensional signals. *Nat. Rev. Neurosci.*, 3:884–895.

Leblois, A. and Perkel, D. J. (2012). Striatal dopamine modulates song spectral but not temporal features through d1 receptors. *European Journal of Neuroscience*, 35(11):1771–1781.

Lee, D. D. and Seung, S. H. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.

Lee, P. and Schmidt-Nielsen, K. (1971). Respiratory and cutaneous evaporation in the zebra finch: effect on water balance. *Am. J. Physiol.*, 220:1598–1605.

lehlov, L., Voldrich, L., and Janisch, R. (1987). Correlative study of sensory cell density and cochlear length in humans. *Hearing Research*, 28(2-3):149 – 151.

Leonardo, A. and Konishi, M. (1999). Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature*, 399(6735):466–470.

Leppelsack, H. J. and Vogt, M. (1976). Responses of auditory neurons in the forebrain of a songbird to stimulation with species-specific sounds. *J. Comp. Physiol. [A]*, 107('3):263–274.

Levitin, D. J. and Rogers, S. E. (2005). Absolute pitch: perception, coding, and controversies. *Trends in Cognitive Sciences*, 9(1):26 – 33.

Lewandowski, B. C. and Schmidt, M. (2011). Short bouts of vocalization induce long-lasting fast gamma oscillations in a sensorimotor nucleus. *The Journal of Neuroscience*, 31(39):13936–13948.

Lewicki, M. S. and Arthur, B. J. (1996). Hierarchical Organization of Auditory Temporal Context Sensitivity. *J. Neurosci.*, 16(21):6987–6998.

Lewicki, M. S. and Sejnowski, T. J. (2000). Learning overcomplete representations. *Neural Computation*, 12(2):337–365.

Lipkind, D., Nottebohm, F., Rado, R., and Barnea, A. (2002). Social change affects the survival of new neurons in the forebrain of adult songbirds. *Behavioural Brain Research*, 133(1):31 – 43.

London, S. E. and Clayton, D. F. (2008). Functional identification of sensory mechanisms required for developmental song learning. *Nat. Neurosci.*, 11:579–586.

Long, M. A. and Fee, M. S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature*, 456(7219):189–194.

Lovell, P. V., Clayton, D. F., Replogle, K. L., and Mello, C. V. (2008). Birdsong "transcriptomics": neurochemical specializations of the oscine song system. *PLoS ONE*, 3(10):e3440.

Lyon, R. (1982). A computational model of filtering, detection, and compression in the cochlea. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82.*, volume 7, pages 1282 – 1285.

MacDougall-Shackleton, S. A., Hulse, S. H., and Ball, G. F. (1998). Neural correlates of singing behavior in male zebra finches (taeniopygia guttata). *Journal of Neurobiology*, 36(3):421–430.

Machens, C. K., Wehr, M. S., and Zador, A. M. (2004). Linearity of cortical receptive fields measured with natural sounds. *J. Neurosci.*, 24(5):1089–1100.

Manley, G. and Gleich, O. (1992). Evolution and specialization of function in the avian auditory periphery. In Webster, D. B., Popper, A. N., and Fay, R. R., editors, *The Evolutionary Biology of Hearing*, pages 561–580. Springer New York.

Manley, G. A. (1981). A review of the auditory physiology of the reptiles. *Progr Sens Physiol*, 2:49–134.

Manley, G. A., Gleich, O., Leppelsack, H. J., and Oeckinghaus, H. (1985). Activity patterns of cochlear ganglion neurones in the starling. *J. Comp. Physiol. A*, 157:161–181.

Mann, N. and Slater, P. (1995). Song tutor choice by zebra finches in aviaries. *Animal Behaviour*, 49(3):811 – 820.

Mann, S. and Haykin, S. (1991). The chirplet transform: A generalization of Gabor's logon transform. *Vision Interface '91*, pages 205–212. ISSN 0843-803X.

Mann, S. and Haykin, S. (1995). The chirplet transform: physical considerations. *Signal Processing, IEEE Transactions on*, 43(11):2745 –2761.

Margoliash, D. (1986). Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J. Neurosci.*, 6(6):1643–1661.

Marler, P. (1997). Three models of song learning: evidence from behavior. *J. Neurobiol.*, 33(5):501–516.

Martin Wild, J., Karten, H. J., and Frost, B. J. (1993). Connections of the auditory forebrain in the pigeon (columba livia). *The Journal of Comparative Neurology*, 337(1):32–62.

McCasland, J. (1987). Neuronal control of bird song production. *The Journal of Neuroscience*, 7(1):23–39.

Meliza, C. D., Chi, Z., and Margoliash, D. (2010). Representations of conspecific song by starling secondary forebrain auditory neurons: Toward a hierarchical framework. *Journal of Neurophysiology*, 103(3):1195–1208.

Mello, C., Nottebohm, F., and Clayton, D. (1995). Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene's response to that song in zebra finch telencephalon. *The Journal of Neuroscience*, 15(10):6919–6925.

Mello, C. V. and Clayton, D. F. (1994). Song-induced ZENK gene expression in auditory pathways of songbird brain and its relation to the song control system. *J. Neurosci.*, 14(11 Pt 1):6652–6666.

Mello, C. V., Vates, E., Okuhata, S., and Nottebohm, F. (1998). Descending auditory pathways in the adult male zebra finch (taeniopygia guttata). *The Journal of Comparative Neurology*, 395(2):137–160.

Mello, C. V., Vicario, D. S., and Clayton, D. F. (1992). Song presentation induces gene expression in the songbird forebrain. *Proc. Natl. Acad. Sci. U.S.A.*, 89(15):6818–6822.

Mink, J. W., Blumenschine, R. J., and Adams, D. B. (1981). Ratio of central nervous system to body metabolism in vertebrates: its constancy and functional basis. *Am J Physiol Regul Integr Comp Physiol*, 241(3):R203–212.

Moiseff, A. (1989). Bi-coordinate sound localization by the barn owl. *J. Comp. Physiol. A*, 164(5):637–644.

Mooney, R. (2000). Different subthreshold mechanisms underlie song selectivity in identified HVc neurons of the zebra finch. *J. Neurosci.*, 20:5420–5436.

Mooney, R. and Prather, J. F. (2005). The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. *J. Neurosci.*, 25:1952–1964.

Mulcahy, N. J. and Call, J. (2006). Apes Save Tools for Future Use. *Science*, 312(5776):1038–1040.

Nagel, K. and Doupe, A. (2008). Organizing principles of spectro-temporal encoding in the avian primary auditory area field L. *Neuron*, 58:938–955.

Naie, K. and Hahnloser, R. H. (2011). Regulation of learned vocal behavior by an auditory motor cortical nucleus in juvenile zebra finches. *J. Neurophysiol.*, 106(1):291–300.

Neuweiler, G. (1984). Foraging, echolocation and audition in bats. *Naturwissenschaften*, 71:446–455. 10.1007/BF00455897.

Nick, T. and Konishi, M. (2005). Neural auditory selectivity develops in parallel with song. *J. Neurobiol.*, 62:469–481.

Nordeen, E. and Nordeen, K. (1990). Neurogenesis and sensitive periods in avian song learning. *Trends in Neurosciences*, 13(1):31 – 36.

Nottebohm, F. (1981). A brain for all seasons: cyclical anatomical changes in song control nuclei of the canary brain. *Science*, 214(4527):1368–1370.

Nottebohm, F. and Arnold, A. P. (1976). Sexual dimorphism in vocal control areas of the songbird brain. *Science*, 194(4261):211–213.

Nottebohm, F., Kasparian, S., and Pandazis, C. (1981). Brain space for a learned task. *Brain Research*, 213(1):99 – 109.

Nottebohm, F., Kelley, D. B., and Paton, J. A. (1982). Connections of vocal control nuclei in the canary telencephalon. *J. Comp. Neurol.*, 207:344–357.

Nottebohm, F., Nottebohm, M. E., and Crane, L. (1986). Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei. *Behavioral and Neural Biology*, 46(3):445 – 471.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.

Olshausen, B. A. and Field, D. J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by v1? *Vision research*, 37(23):3311–3325.

Parks, T. N. and Rubel, E. W. (1975). Organization and development of brain stem auditory nuclei of the chicken: organization of projections from n. magnocellularis to n. laminaris. *J. Comp. Neurol.*, 164:435–448.

Perez-Orive, J., Bazhenov, M., and Laurent, G. (2004). Intrinsic and circuit properties favor coincidence detection for decoding oscillatory input. *J. Neurosci.*, 24:6037–6047.

Perez-Orive, J., Mazor, O., Turner, G. C., Cassenaer, S., Wilson, R. I., and Laurent, G. (2002). Oscillations and sparsening of odor representations in the mushroom body. *Science*, 297:359–365.

Pham, D. T., Garrat, P., and Jutten, C. (1992). Separation of a mixture of independent sources through a maximum likelihood approach. In *Proc. EUSIPCO*, pages 771–774.

Plumbley, M. (2003). Algorithms for non-negative independent component analysis. *IEEE Trans. Neural Netw.*, 14(3):534–543.

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., and Watwood, S. (2005). Animal behaviour: elephants are capable of vocal learning. *Nature*, 434:455–456.

Prather, J. F., Peters, S., Nowicki, S., and Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature*, 451:305–310.

Raksin, J. N., Glaze, C. M., Smith, S., and Schmidt, M. F. (2012). Linear and nonlinear auditory response properties of interneurons in a high-order avian vocal motor nucleus during wakefulness. *Journal of Neurophysiology*, 107(8):2185–2201.

Rao, Y. and Principe, J. (2002). Robust on-line principal component analysis based on a fixed-point approach. In *Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on*, volume 1, pages I–981 – I–984 vol.1.

Reiner, A., Perkel, D. J., Bruce, L. L., Butler, A. B., Csillag, A., Kuenzel, W., Medina, L., Paxinos, G., Shimizu, T., Striedter, G., Wild, M., Ball, G. F., Durand, S., Gunturkun, O., Lee, D. W., Mello, C. V., Powers, A., White, S. A., Hough, G., Kubikova, L., Smulders, T. V., Wada, K., Dugas-Ford, J., Husband, S., Yamamoto, K., Yu, J., Siang, C., Jarvis, E. D., and Guturkun, O. (2004a). Revised nomenclature for avian telencephalon and some related brainstem nuclei. *J. Comp. Neurol.*, 473(3):377–414.

Reiner, A., Perkel, D. J., Bruce, L. L., Butler, A. B., Csillag, A., Kuenzel, W., Medina, L., Paxinos, G., Shimizu, T., Striedter, G., Wild, M., Ball, G. F., Durand, S., Guturkun, O., Lee, D. W., Mello, C. V., Powers, A., White, S. A., Hough, G., Kubikova, L., Smulders, T. V., Wada, K., Dugas-Ford, J., Husband, S., Yamamoto, K., Yu, J., Siang, C., and Jarvis, E. D. (2004b). The Avian Brain Nomenclature Forum: Terminology for a New Century in Comparative Neuroanatomy. *J. Comp. Neurol.*, 473:E1–E6.

Reiner, A., Perkel, D. J., Mello, C. V., and Jarvis, E. D. (2004c). Songbirds and the revised avian brain nomenclature. *Ann. N. Y. Acad. Sci.*, 1016:77–108.

Reinke, H. and Wild, J. (1998). Identification and connections of inspiratory premotor neurons in songbirds and budgerigar. *The Journal of Comparative Neurology*, 391(2):147–163.

Roberts, T. F., Klein, M. E., Kubke, M. F., Wild, J. M., and Mooney, R. (2008). Telencephalic neurons monosynaptically link brainstem and forebrain premotor networks necessary for song. *The Journal of Neuroscience*, 28(13):3479–3489.

Rose, M. (1914). Über die cytoarchitektonische gliederung des vorderhirns der vgel. *J. Psychol. Neurol.*, 2:278–352.

Rosen, M. J. and Mooney, R. (2003). Inhibitory and excitatory mechanisms underlying auditory responses to learned vocalizations in the songbird nucleus hvc. *Neuron*, 39(1):177 – 194.

Rosen, M. J. and Mooney, R. (2006). Synaptic interactions underlying song-selectivity in the avian nucleus hvc revealed by dual intracellular recordings. *Journal of Neurophysiology*, 95(2):1158–1175.

Ryan, M. J. (1990). Sexual selection, sensory systems and sensory exploitation. *Oxford Surveys in Evolutionary Biology*, 7:157–195.

Sachs, M. B. and Sinnott, J. M. (1978). Responses to tones of single cells in nucleus magnocellularis and nucleus angularis of the redwing blackbird (Agelaius phoeniceus). *J. Comp. Physiol. A*, 126:347–361.

Sakata, J. T. and Brainard, M. S. (2008). Online contributions of auditory feedback to neural activity in avian song control circuitry. *The Journal of Neuroscience*, 28(44):11378–11390.

Scharff, C. and Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *The Journal of Neuroscience*, 11(9):2896–2913.

Schmidt, M., Larsen, J., and Hsiao, F.-T. (2007). Wind noise reduction using non-negative sparse coding. In *Machine Learning for Signal Processing, 2007 IEEE Workshop on*, pages 431–436.

Schmidt, M. F. and Konishi, M. (1998). Gating of auditory responses in the vocal control system of awake songbirds. *Nat. Neurosci.*, 1:513–518.

Schmidt, R. and Altner, H. (1978). *Fundamentals of sensory physiology.* "Springer study edition.". Springer-Verlag.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80(1):1–27.

Schwartzkopff, J. and Winter, P. (1960). Zur anatomie der vogel-cochlea unter naturlichen bedingungen. *Biologisches Zentralblatt*, 79:607–625.

Sen, K., Theunissen, F. E., and Doupe, A. J. (2001). Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.*, 86(3):1445–1458.

Shaevitz, S. S. and Theunissen, F. E. (2007). Functional connectivity between auditory areas field l and clm and song system nucleus hvc in anesthetized zebra finches. *Journal of Neurophysiology*, 98(5):2747–2764.

Shank, S. S. and Margoliash, D. (2009). Sleep and sensorimotor integration during early vocal learning in a songbird. *Nature*, 458(7234):73–77.

Sharpee, T. and Bialek, W. (2007). Neural decision boundaries for maximal information transmission. *PLoS ONE*, 2:e646.

Simpson, H. and Vicario, D. (1990). Brain pathways for learned and unlearned vocalizations differ in zebra finches. *J. Neurosci.*, 10(5):1541–1556.

Singh, N. C. and Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *J. Acoust. Soc. Am.*, 114(6 Pt 1):3394–3411.

Slater, P., Jones, A., and Ten Cate, C. (1992). Can lack of experience delay the end of the sensitive phase for song learning? *Netherlands Journal of Zoology*, 43(1-2):80–90.

Smith, C. C. and Reichman, O. J. (1984). The evolution of food caching by birds and mammals. *Annual Review of Ecology and Systematics*, 15(1):329–351.

Smith, E. C. and Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, 439:978–982.

Sossinka, R. (1972). Langfristiges durstvermögen wilder und domestizierter zebrafinken(taeniopygia guttata castanotis gould). *Journal of Ornithology*, 113:418–426.

Sossinka, R. and Böhner, J. (1980). Song types in the zebra finch poephila guttata castanotis1. *Zeitschrift fr Tierpsychologie*, 53(2):123–132.

Stripling, R., Volman, S. F., and Clayton, D. F. (1997). Response modulation in the zebra finch neostriatum: Relationship to nuclear gene regulation. *The Journal of Neuroscience*, 17(10):3883–3893.

Takahashi, T. and Keller, C. (1992). Commissural connections mediate inhibition for the computation of interaural level difference in the barn owl. *Journal of Comparative Physiology A*, 170:161–169.

Takahashi, T. T. and Konishi, M. (1988). Projections of the cochlear nuclei and nucleus laminaris to the inferior colliculus of the barn owl. *J. Comp. Neurol.*, 274:190–211.

Tchernichovski, O., Mitra, P. P., Lints, T., and Nottebohm, F. (2001). Dynamics of the vocal imitation process: How a zebra finch learns its song. *Science*, 291(5513):2564–2569.

Ten Cate, C. (1982). Behavioural differences between zebrafinch and bengalese finch (foster) parents raising zebrafinch offspring. *Behaviour*, 81(2-4).

Theunissen, F. E., Amin, N., Shaevitz, S. S., Woolley, S. M. N., Fremouw, T., and Hauber, M. E. (2004). Song Selectivity in the Song System and in the Auditory Forebrain. *Ann. N.Y. Acad. Sci.*, 1016(1):222–245.

Theunissen, F. E., David, S. V., Singh, N. C., Hsu, A., Vinje, W. E., and Gallant, J. L. (2001). Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network*, 12:289–316.

Theunissen, F. E. and Doupe, A. J. (1998). Temporal and spectral sensitivity of complex auditory neurons in the nucleus hvc of male zebra finches. *The Journal of Neuroscience*, 18(10):3786–3802.

Theunissen, F. E., Sen, K., and Doupe, A. J. (2000). Spectral-Temporal Receptive Fields of Nonlinear Auditory Neurons Obtained Using Natural Sounds. *J. Neurosci.*, 20(6):2315–2331.

Thompson, J. V. and Gentner, T. Q. (2010). Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. *Journal of Neurophysiology*, 103(4):1785–1797.

Tramontin, A. D. and Brenowitz, E. A. (1999). A field study of seasonal neuronal incorporation into the song control system of a songbird that lacks adult song learning. *Journal of Neurobiology*, 40(3):316–326.

Tumer, E. and Brainard, M. (2007). Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. *Nature*, 450:1240–1244.

van der Helden, J., Boksem, M. A. S., and Blom, J. H. G. (2010). The importance of failure: Feedback-related negativity predicts motor learning efficiency. *Cerebral Cortex*, 20(7):1596–1603.

Vates, G., Broome, B., Mello, C., and Nottebohm, F. (1996). Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *J. Comp. Neurol.*, 366:613–642.

Vates, G. E. and Nottebohm, F. (1995). Feedback circuitry within a song-learning pathway. *Proceedings of the National Academy of Sciences*, 92(11):5139–5143.

Vates, G. E., Vicario, D. S., and Nottebohm, F. (1997). Reafferent thalamo-cortical loops in the song system of oscine songbirds. *The Journal of Comparative Neurology*, 380(2):275–290.

Veit, L., Aronov, D., and Fee, M. S. (2011). Learning to breathe and sing: development of respiratory-vocal coordination in young songbirds. *Journal of Neurophysiology*, 106(4):1747–1765.

Vellema, M., van der Linden, A., and Gahr, M. (2010). Area-specific migration and recruitment of new neurons in the adult songbird brain. *The Journal of Comparative Neurology*, 518(9):1442–1459.

Vicario, D. S. and Yohay, K. H. (1993). Song-selective auditory input to a forebrain vocal control nucleus in the zebra finch. *Journal of Neurobiology*, 24(4):488–505.

Voss, H. U., Tabelow, K., Polzehl, J., Tchernichovski, O., Maul, K. K., Salgado-Commissariat, D., Ballon, D., and Helekar, S. A. (2007). Functional mri of the zebra finch brain during song stimulation suggests a lateralized response topography. *Proceedings of the National Academy of Sciences*, 104(25):10667–10672.

Wang, C. Z. H., Herbst, J. A., Keller, G. B., and Hahnloser, R. H. R. (2008). Rapid interhemispheric switching during vocal production in a songbird. *PLoS Biol*, 6(10):e250.

Wang, J., Sokabe, M., and Sakaguchi, H. (2001). Functional connections between the HVC and the shelf of the zebra finch revealed by real-time optical imaging technique. *Neuroreport*, 12(2):215–221.

Whitfield, I. C. (1967). *The Auditory Pathway.* Williams and Wilkins Company, Baltimore.

Wild, J. M. (1994). Visual and somatosensory inputs to the avian song system via nucleus uvaeformis (uva) and a comparison with the projections of a similar thalamic nucleus in a nonsongbird, columbia livia. *The Journal of Comparative Neurology*, 349(4):512–535.

Wild, J. M., Krützfeldt, N. O., and Kubke, M. F. (2010). Connections of the auditory brainstem in a songbird, taeniopygia guttata. iii. projections of the superior olive and lateral lemniscal nuclei. *The Journal of Comparative Neurology*, 518(11):2149–2167.

Williams, H., Kilander, K., and Sotanski, M. L. (1993). Untutored song, reproductive success and song learning. *Animal Behaviour*, 45(4):695 – 705.

Williams, S. M., Nast, A., and Coleman, M. J. (2012). Characterization of synaptically connected nuclei in a potential sensorimotor feedback pathway in the zebra finch song system. *PLoS ONE*, 7(2):e32178.

Woolley, S., Fremouw, T., Hsu, A., and Theunissen, F. (2005). Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat. Neurosci.*, 8:1371–1379.

Woolley, S. M., Hauber, M. E., and Theunissen, F. E. (2010a). Developmental experience alters information coding in auditory midbrain and forebrain neurons. *Developmental Neurobiology*, 70(4):235–252.

Woolley, S. M., Hauber, M. E., and Theunissen, F. E. (2010b). Developmental experience alters information coding in auditory midbrain and forebrain neurons. *Dev Neurobiol*, 70:235–252.

Woolley, S. M. N. (2012). Early experience shapes vocal neural coding and perception in songbirds. *Developmental Psychobiology*, 54(6):612–631.

Woolley, S. M. N. and Casseday, J. H. (2004). Response properties of single neurons in the zebra finch auditory midbrain: Response patterns, frequency coding, intensity coding, and spike latencies. *Journal of Neurophysiology*, 91(1):136–151.

Woolley, S. M. N. and Casseday, J. H. (2005). Processing of modulated sounds in the zebra finch auditory midbrain: Responses to noise, frequency sweeps, and sinusoidal amplitude modulations. *Journal of Neurophysiology*, 94(2):1143–1157.

Woolley, S. M. N., Gill, P. R., Fremouw, T., and Theunissen, F. E. (2009). Functional Groups in the Avian Auditory System. *J. Neurosci.*, 29(9):2780–2793.

Woolley, S. M. N., Gill, P. R., and Theunissen, F. E. (2006). Stimulus-Dependent Auditory Tuning Results in Synchronous Population Coding of Vocalizations in the Songbird Midbrain. *J. Neurosci.*, 26(9):2499–2512.

Yanagihara, S. and Hessler, N. A. (2006). Modulation of singing-related activity in the songbird ventral tegmental area by social context. *European Journal of Neuroscience*, 24(12):3619–3627.

Yang, L., Monsivais, P., and Rubel, E. W. (1999). The superior olivary nucleus and its influence on nucleus laminaris: A source of inhibitory feedback for coincidence detection in the avian auditory brainstem. *The Journal of Neuroscience*, 19(6):2313–2325.

Yates, G. K., Manley, G. A., and Koppl, C. (2000). Rate-intensity functions in the emu auditory nerve. *The Journal of the Acoustical Society of America*, 107(4):2143–2154.

Young, S. and Rubel, E. (1983). Frequency-specific projections of individual neurons in chick brainstem auditory nuclei. *J. Neurosci.*, 3(7):1373–1378.

Zann, R. (1985). Ontogeny of the zebra finch distance call: I. effects of cross-fostering to bengalese finches. *Zeitschrift für Tierpsychologie*, 68:1–23.

Zann, R. (1990). Song and call learning in wild zebra finches in south-east australia. *Animal Behaviour*, 40(5):811 – 828.

Zann, R. (1993). Structure, sequence and evolution of song elements in wild australian zebra finches. *Auk*, 110:702–715.

Zann, R. A. (1996). *The Zebra Finch: A Synthesis of Field and Laboratory Studies (Oxford Ornithology Series).* Oxford University Press, USA.

Curriculum Vitae

Personal Data

Name	Florian Blättler
Date of birth	June 21, 1978
Nationalities	Swiss
Civil status	Single

Education

2006 -	Ph.D.–Student in Physics, ETH Zurich Dissertation written at the Institute for Neuroinformatics under the supervision of Prof. Dr. R. Hahnloser.
1998 - 2005	Studies in Physics, ETH Zurich Diploma thesis written at the ETH Zurich under the supervision of Prof. Dr. R. Hahnloser and Prof. Dr. R. Douglas.
1993 - 1998	Gymnasium, Kreuzlingen

Work Experience

2006 -	Research and Teaching Assistant, Institute of Neuroninformatics and Department of Physics, ETH Zurich
2000 - 2011	Substitution Teacher in Physics, Mathematics and Informatics at the Gymnasia of Wattwil, Olten, Luzern, and Zurich Birch
1999 - 2003	Member of the board and from 2001 on President of the Students Organization of Mathematics and Physics Students at the ETH (VMP)

Publications

Blättler, F. and Hahnloser, R. H. R. (2011). An efficient coding hypothesis links sparsity and selectivity of neural responses. *PLoS ONE*, 6(10):e25506

Blättler, F., Kollmorgen, S., Herbst, J., and Hahnloser, R. (2011). Hidden markov models in the neurosciences. In Dymarski, P., editor, *Hidden Markov Models, Theory and Applications*, pages 169–186. InTech