Doctoral Thesis

Auditory feedback processing in the songbird
how the zebra finch listens to itself sing

Author(s):
Keller, Georg Bernhard

Publication Date:
2009

Permanent Link:
https://doi.org/10.3929/ethz-a-005843597

Rights / License:
In Copyright - Non-Commercial Use Permitted
Auditory feedback processing in the songbird
- How the zebra finch listens to itself sing -

A B H A N D L U N G
zur Erlangung des Titels

DOKTOR DER WISSENSCHAFTEN
der

ETH Zürich
Vorgelegt von

GEORG BERNHARD KELLER
Dipl. Phys. ETH
geboren am 31.05.1980
von
Amriswil TG

Angenommen auf Antrag von

Richard H.R. Hahnloser
Rodney J. Douglas

2009
"It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."

Arthur Conan Doyle
1 INTRODUCTION
1.1 Vocal learning
1.2 Nyi-nyi – the zebra finch
  1.2.1 Vocalizations
  1.2.2 Function of song
  1.2.3 Song development
1.3 The song system
1.4 The template theory of song learning

2 THE AUDITORY SYSTEM OF THE SONGBIRD
2.1 Performance – can the bird hear everything it sings?
2.2 Anatomy
  2.2.1 Note on the usage of the terms ‘sensory area’ and ‘motor area’
  2.2.2 The auditory network
  2.2.3 Auditory input to the song system
  2.2.4 Feedback connections from the song system to auditory areas
2.3 Function
  2.3.1 Electrophysiology
  2.3.2 Immediate early gene expression

3 AUDITORY FEEDBACK PROCESSING
3.1 Song-evoked activity is similar to BOS-playback-evoked activity
3.2 Song-perturbation responses cannot be predicted from passive playback responses
3.3 Sound amplitude influences response amplitude but not response timing
3.4 Field L exhibits song locked responses in the deafened bird

4 NEURAL CORRELATES OF SONG MEMORIES
4.1 Playback-related NCM responses habituate to repeated stimulus presentation
4.2 Singing-related NCM activity cannot be predicted from passive playback-related responses
4.3 NCM lesion during song development impair imitation accuracy

5 SENSORIMOTOR INTEGRATION
5.1 NIf activity of singing birds
  5.1.1 NIf spike patterns specifically precede song syllables
  5.1.2 NIf spike activity is not suited to conveying auditory feedback to HVC
  5.1.3 No significant differences exist between NIf activity in juveniles and adults
  5.1.4 Discussion

6 BEHAVIORAL EXPERIMENTS
6.1 Come again…? - or, how many repetitions form a template
6.2 Model over-abundance reduces imitation accuracy
6.3 Females prefer stereotyped song
Acknowledgements

For as early as I can remember my only answer to the question of what I would want to be when I grow up was “physicist” – generally to the bewilderment of the one asking. All throughout my early education, I was lured and inspired by the idea of one day understanding nature’s laws. In due course I later took to the study of theoretical physics and was convinced that I would spend the greater part of my life working on one of the dazzling problems of the field. Close to finishing my studies, my curiosity was drawn to a lecture on theoretical neuroscience held by Klaus Hepp, in the consequence of which, not least due to Klaus, most of my interests were diverted to this – to me – new and unexplored field of science. What started out as a few small projects, ended in my decision to start a PhD in the field of neuroscience. I am particularly grateful to Klaus for this small but significant “push” in the right direction.

I started my PhD with Richard Hahnloser in the spring of 2005 still clinging to the idea that I would make better use of my time concerning myself with the theoretical problems of neuroscience than to bother with the more cumbersome and intense task of gathering data. Were it not for Richard’s guidance and his great intuition as both an experimenter and a theoretician, this work would not have been possible, and I would have likely never discovered the virtue of experimenting.

I was also fortunate to be under the experienced co-supervision of Rodney Douglas, and sometimes the watchful eye of Kevan Martin, both of whom have been more than inspirational.

Not least, my deepest appreciation goes to many friends and colleagues for invaluable support and discussion.
in memory of my father
***
who never doubted that I would follow in his footsteps.
Kurzfassung

Das Imitieren einer Vokalisierung bedingt sowohl die Fähigkeit ein Gedächtnis des zu imitierenden auditorischen Stimulus zu bilden, als auch einen Mechanismus um die eigene Vokalisation zu überwachen um eventuelle Fehler darin korrigieren zu können. Bis heute ist es jedoch noch unklar wie diese Aufgaben vom Gehirn gelöst werden.

Im ersten Teil dieser Arbeit untersuchten wir Mechanismen zur Überwachung der eigenen Vokalisation, und konnten zeigen, dass Neuronen in auditorischen Gehirnarealen eines jungen Singvogels, der im Begriff ist seinen Gesang zu erlernen, wirklich dem eigenen Singen „zuhören“, d.h. stereotype Aktivitätsmuster zeigen, während der Vogel singt - dies jedoch nicht so wie man es sich bis anhin vorgestellt hatte. Nämlich finden wir gewisse Neuronen, die nur so tun als ob sie zuhören würden, während andere wiederum nur darauf warten, dass etwas Unerwartetes geschieht. Wir begannen damit aufzudecken wie die verschiedenen Neuronen auditorische Signale während des Singens verarbeiten, indem wir dem Vogel, während er singt, ein Störgeräusch vorspielen, so dass das was der Vogel tatsächlich hört von dem abweicht was er zu hören erwartet. Der erste typ Neuron, der nur zuzuhören scheint, verhält sich ganz gewöhnlich auditorisch solange der Vogel nicht singt, zeigt sogar stereotype Aktivitätsmuster während er singt, lässt sich aber nicht von diesem Aktivitätsmuster abbringen durch die vorgespielten Störgeräusche. Der zweite typ Neuron, der Unerwartetes detektiert, ist kaum auditorisch solange der Vogel nicht singt und reagiert fast ausschliesslich auf die während des Gesanges vorgespielten Störgeräusche. Beide diese Signale sind grundlegende Voraussetzung für das Imitieren einer Vokalisierung, da man zuerst wissen muss, was man zu hören erwartet, unbeeinflusst vom tatsächlich gehörten, um dann, Unterschiede zwischen Erwartung und tatsächlich Gehörtem zu detektieren.


Zusammenfassend, konnten wir die ersten direkten Anzeichen der Funktionsweisen sowohl der Mechanismen der Fehlererkennung als auch derjenigen der Gedächtnisbildung aufdecken.
Abstract

Vocal imitation requires both the ability to form a memory of the auditory stimulus to be imitated and a mechanism to monitor auditory feedback of self generated vocalizations in order to correct imitation errors. Only little is known about how the brain solves these tasks.

In a first part of this work we investigated mechanisms of auditory feedback monitoring and could show that neurons in an auditory brain area of a juvenile songbird that is in the process of song learning, actually “listen” to the auditory feedback during singing – not, however, as one would have predicted. We find a subset of neurons that only pretend to listen during song, while others seem to be only waiting for something unexpected to happen. We began to understand how these different neurons process auditory feedback signals, by playing a loud perturbing stimulus to the bird while he is singing, such that what it actually hears deviates from what it expects to hear. The first type of neuron that only pretends to listen, responds in a purely auditory manner as long as the bird does not sing and even shows stereotyped activity patterns while the bird sings. This type of neuron is, however, almost entirely unresponsive to even loud perturbing stimuli. The second type of neuron, that detects unexpected sounds, shows no auditory responses and reacts almost exclusively to perturbation stimuli presented during song. Both of these signals are a fundamental prerequisite for the imitation of a vocalization, as one first needs to know what to expect to hear, independent of what one actually hears, to then detect the differences between what is actually heard an what one would predict to hear.

In a second part of this work, we investigated possible representations of auditory memories of the tutor song that the juvenile songbird aspires to imitate. By recording the activity patterns of neurons in a secondary auditory area, we could show that, in a subset of neurons, responses during singing more closely resembled the responses elicited by playback of the tutor song, than those elicited by playback of the bird’s own song. This indicates that instead of responding to the actual auditory feedback during singing, these neurons replay an activity pattern similar to the tutor song response while the bird sings. Such a replay of the tutor-song-related activity during singing would naturally correspond to a sensory memory of the tutor song, and to our knowledge this is the first direct evidence of a sensory long-term memory.

In sum, we believe that we were able to unveil the first direct correlates of the mechanisms underlying both error detection and memory storage, necessary for learning in the juvenile songbird.
Foreword

The results presented here come from work conducted during the second part of my PhD. Most of the data shown stands on solid ground and is either published or on the way to being published; other data however is based on small or even single samples and is preliminary to the point of being speculative. We have consistently indicated preliminary data as such in the text. The nature of the data presented in the different chapters is outlined in the summary below. Note that only methods specific to a certain experiment are described in the results chapters (3 to 6), all other methods used multiple times throughout this work are described in chapter 8.

The initial two chapters are intended as a general introduction, the first of which should give the reader novel to the field of songbird research an overview of the basic concepts of vocal learning, the template theory of song learning, and the song system. It also introduces the animal model studied - the zebra finch. The second chapter gives a brief overview of the current knowledge on performance, anatomy, and function of the avian auditory system.

The third chapter is a slightly extended version of our work on auditory feedback processing published elsewhere [Keller and Hahnloser 2009]. In addition to the data presented in the paper, there is a brief discussion on the effect of sound amplitude on response latency, and a short summary of the field L recordings in the deafened bird.

The fourth chapter summarizes preliminary data recorded in the caudal medial nidopallium (NCM) in the singing bird. It discusses correlations between song-related activity and bird’s own song or tutor song playback-evoked activity.

The fifth chapter summarizes our work recordings of neurons in the nucleus interface of the nidopallium (NIf) in the singing bird. Together with inactivation and lesions studies conducted by Katja Naie, this work is in preparation for publication [Naie et al. 2009].

The sixth chapter is a collection of behavioral data we have stumbled upon along the way, but that has had significant influence on our thinking of the memory mechanisms underlying song learning.

The seventh and eighth chapters contain a brief discussion of the results and their relation to previous models of song learning, in addition to a description of the methods used to gather the data presented. In particular we have attempted to include a complete guide to building a miniature motorized-microdrive according to designs developed for and used in the experiments discussed above.
1 INTRODUCTION

We see, hear, smell, taste and feel - everything we experience of the world around us, we experience through our senses. Physical stimuli impinge on our sensory organs where they are transduced to neural signals. These signals are then processed and transformed by a hierarchy of brain areas and eventually give rise to perceptions. In other words, the sensory system transforms a peripheral representation of the stimulus to a more complex high-level representation that can be processed by associative areas of the brain. In the visual system this idea of feed-forward processing has yielded – amongst many other things – the very fruitful concept of the feature detector [Barlow 1953] from which our notion of receptive field stems. The same idea was later used by Hubel and Wiesel [Hubel and Wiesel 1959] in explaining the origin of simple and complex cell responses found in the primary visual cortex. Even today, the models most successful in explaining size-, translation- and even rotation-invariance of the responses found in higher order visual neurons are based on feed-forward sensory-processing hierarchies [Riesenhuber and Poggio 1999].

However, there is accumulating evidence for activity in sensory areas of active, behaving animals that cannot be explained by a unimodal feed-forward processing hierarchy. Activity of visual neurons, for example, is modulated by eye movements [Leopold and Logothetis 1998; Martinez-Conde et al. 2000]. Similarly, a large part of the responses in marmoset auditory cortex are suppressed during vocalization [Eliades and Wang 2003; Eliades and Wang 2005], while a subset of neurons is selective for unexpected auditory perturbations exclusively during vocalization [Eliades and Wang 2008]. All of these effects fail to be explicable in the absence of input from motor- or planning-related brain areas. The detection of unexpected feedback during motor output for example, requires the comparison of actual sensory feedback with predicted sensory feedback.¹

It is not altogether surprising to find such effects, as perception has long been known to be directly influenced by motor output. The sensory system can compensate for self-initiated movements, as demonstrated in the ingenious experiments of von Helmholtz [von Helmholtz 1867]. He found that gently pushing on one’s own eye with a finger will create the illusion of a moving world, whereas during a normal movement of the eye, no such illusion occurs. Von Helmholtz’s explanation of this phenomenon was that during active movement of the eye a copy of the motor command controlling the eye movement is combined with sensory feedback to generate a stable percept of the visual scene. Illusory movements would arise only during passive movements of the eye, in the absence of a motor command. These ideas were later experimentally validated and rephrased by Sperry, von Holst and Mittelstaedt as what is known today as the efference copy theory [Sperry 1950; von Holst and Mittelstaedt 1950].

Moreover, based on the fact that only a small fraction of the synapses in cortex come from its main sensory input, the thalamus [Garey and Powell 1971; Ahmed et al. 1994], one would already predict that a substantial component of the activity observed in even early sensory areas is feedback driven. The question then simply is: does this feedback originate solely in higher-order sensory areas of the same sensory modality, or is there significant contribution of cross-modal or motor related feedback? The experiments on awake, behaving animals support the idea of significant motor related feedback.¹

¹ Even within the visual processing hierarchy itself, it has even been speculated, that feedback from higher order visual areas functions a predictor of the activity of lower order visual areas [Rao and Ballard 1999].
feedback in early sensory areas by demonstrating that there is an active component of sensory processing that goes undetected in passive stimulation experiments.

The influence of such motor-related feedback on sensory processing is best studied in a system in which a well-defined interplay between motor output and sensory feedback is crucial for successful functioning. This is most elegantly fulfilled in the case of imitation learning. To imitate successfully, sensory feedback from motor output has to be compared to a memory of the gesture to be imitated, to then in turn improve motor output. This requires both an indirect (via sensory feedback) and a direct (via neural connections) transfer of information between the sensory system and the motor system.

The first neurophysiological evidence for a direct transfer of information from the motor to the sensory system came from the experiments of Bell on the electro-sensory system of electric fish [Bell 1981]. He could show that an efference copy actively inhibited electro-sensory neurons during the discharge of the electric organ. Similar evidence of an efference copy signal was later found in the visual system of the macaque monkey [Leopold and Logothetis 1998; Martinez-Conde et al. 2000], the auditory system of the cricket [Poulet and Hedwig 2002], and in the auditory system of the marmoset [Eliades and Wang 2003].

Evidence for the converse, the sensory activation of motor related areas, came from the work of Williams and Nottebohm [Williams and Nottebohm 1985], who could show that in anesthetized songbirds motor neurons of the nucleus of the twelfth cranial nerve (nXIIIts) are responsive to auditory stimulation. It was postulated that activity elicited by a sensory stimulus in motor areas would match the motor-related activity during the production of the same stimulus, similar to the motor theory of speech perception in humans [Liberman et al. 1967]. It was however not until the discovery of mirror neurons [di Pellegrino et al. 1992] that this speculation found direct neurophysiological confirmation. Mirror neurons live in a premotor grasp control area (they were originally discovered in F5, but are also found in other brain areas) of monkey cortex and get their name from the fact that they respond similarly during the monkey’s grasping movements, both with and without visual feedback, and when the monkey observes another monkey or the experimenter perform the same hand gesture. Similar types of neurons were later found in premotor areas of the songbird pallium [Dave and Margoliash 2000; Prather et al. 2008].

These findings were of great significance because they provided an experimental basis for our intuitive concept of imitation learning. Even the most primitive model of imitation learning requires the ability to detect errors, that is, the ability to detect differences between actual sensory feedback and intended sensory feedback. This can be achieved, for example, by subtracting predicted sensory feedback from actual sensory feedback, the result of which is an error signal. What is still lacking is a neurophysiological correlate of such a sensory prediction signal in a learning system. The inherent difficulty in detecting such signals is that they are not easily distinguishable from purely sensory signals.

To disentangle signals coding for actual sensory feedback from signals coding for predictions of sensory feedback, we will use an idea similar to the one so successfully used to demonstrate the existence of central motor control, more than 50 ago. The question then was: what component of a motor pattern is independently generated by the motor system and how much is driven by sensory feedback? This was investigated by removing sensory feedback and searching for motor activity patterns that persisted in the absence of sensory feedback. The first neurophysiological evidence for
such motor activity came from Donald Wilson’s recordings in motor neurons of the locust’s flight system [Wilson 1961]. Instead of asking about the influence that sensory feedback exerts on the motor system, the question now is: how does the motor system influence sensory processing? To address this question, instead of simply removing sensory feedback, we will investigate the effects of altered sensory feedback during imitation learning.

### 1.1 Vocal learning

Vocal learning refers to the ability to learn to imitate sounds using a vocal organ (in mammals the larynx, at the top of the trachea, and in birds, the syrinx at the bottom of the trachea). Vocal learning requires, but is distinct from, the ability to recognize sounds, referred to as auditory learning. A dog, for example, can learn to associate meaning to an acoustic command and is thus capable of auditory learning, but it cannot learn to imitate the sound of the command. The capability for vocal learning has only been observed in 3 of the 23 major families of birds: parrots, hummingbirds and songbirds, and in 3 families of mammals: cetaceans, bats and humans (note however that recent evidence has suggested that also elephants [Poole et al. 2005] and seals [Sanvito et al. 2007] are capable of vocal learning). In evolutionary terms this implies that vocal learning has either independently evolved three times, or has been independently lost four times in the avian family tree [Jarvis 2004] and independently evolved 3 times, or independently lost 7 in the mammalian family tree. The different species of vocal learners possess the ability to varying degrees. This ranges from single motif songbirds like the zebra finch, which learn to imitate one brief sequence of sounds, to humans, the most versatile among vocal learners, which are capable of one-shot imitation of a large range of sound sequences.

Speech learning in humans should not be confounded with vocal learning. Although speech critically relies on the ability of vocal learning, vocal learning does not automatically imply vocal communication. Zebra finches for example use a set of twelve different non-learned calls for most of their vocal communication, whereas learned song appears to have much less communicative value.²

### 1.2 Nyi-nyi³ – the zebra finch

The zebra finch belongs to the group of estrildine finches and is the most numerous and widespread passerine finch endemic to Australia [Zann 1996]. The species is divided into two different subspecies, the Australian zebra finch (Taeniopygia guttata castanotis) and the Lesser Sundas Zebra finch (Taeniopygia guttata guttata) native to the Lesser Sundas archipelago of eastern Indonesia.⁴

---

² This is indicated by the fact that a male zebra finch will sing almost independent of social setting and even in complete isolation (see “Function of song” for a discussion of the behavioral significance of song), where in contrast most calls are only used in very specific social context.

³ Nyi-nyi is one of the over ten different native Australian names for zebra finch.

⁴ Most authors, when describing experimental subjects, will not distinguish between the two and will only specify the species (Taeniopygia guttata), omitting the subspecies identity of the animals used. However it is safe to assume that almost exclusively descendents of Taeniopygia guttata castanotis are used in laboratory experiments. There are a number of differences between the two subspecies: lesser Sundas zebra finch males have smaller breast bars (the thick black stripe across the chest), and nearly completely lack throat stripes, their song is of higher pitch and on average faster. Visually, lesser Sundas females are nearly indistinguishable from their Australian counterpart but also have calls of higher pitch.
Zebra finches live in colonies of up to a few hundred birds. The exact size of the population varies with seasonal changes in water and food abundance. Its daily food requirements are estimated to be roughly 3 g of dry seed and 3 ml of water. Nevertheless, the zebra finch is well acclimatized to an arid environment. In the laboratory they have been found to be able to survive without water (on dry seeds alone) almost indefinitely [Cade et al. 1965].

A number of properties – aside from its rather unique vocalizations – have made the zebra finch so apt for scientific study. Predominantly due to predation, the zebra finch in the wild has a life expectancy of 51 days [Zann 1996]5. These harsh environmental conditions appear to have put large evolutionary pressure on the zebra finch for rapid reproduction and early sexual maturity. The average age, independent of sex, at first breeding is two to three months. The average interval between clutch initiations is 52 (± 16) days and the average clutch size is 5 (±1) juveniles. More importantly, the zebra finch also breeds readily in captivity. Easy breeding and the rapid reproduction and maturation have made the zebra finch an ideal laboratory animal.

### 1.2.1 Vocalizations

Zebra finches are highly vocal and extremely lively, especially when in larger groups. Both sexes produce a wide variety of innate calls, but only the male zebra finch sings. A wild zebra finch, independent of sex, has a repertoire of twelve different calls. The calls most frequently uttered are tet⁶ calls, stacks and long distance calls (see Figure 2), the latter of which is the only learned call, and is only learned in males [Simpson and Vicario 1990]. They use wsst calls when attacking rivals, thuk calls to warn the young to flee, and utter distress calls when in extreme pain. Kackle, ark, and whine calls are used as communication calls during nesting, and copulation calls are used for sexual appeasement during mating. Young zebra finches will predominantly utter begging calls to induce feeding and long-tonal calls when separated from parents. The long-tonal calls later develop into the distress calls of the adult bird. In a laboratory setting, one typically observes only tet calls, stacks, and long calls in adults, and begging calls in juveniles.

---

⁵ In the laboratory the zebra finch’s maximum lifespan is more than seven years.
⁶ Call names in italic are onomatopoetic.
Figure 2: The most frequently uttered zebra finch calls. Shown, from left to right, are stacks, tet calls, and the long calls of three different males (M), and of three different females (F). Note that the long call of the male bird in general is slightly more complex and usually is composed of two spectrally distinct parts (notes). It is the only learned call. Lesions of brain areas responsible for singing in males will, aside from their effect on song, also result in a loss of structure in the long call, rendering it female-like.

The song of the male zebra finch is composed of a sequence of typically 2-8 repetitions of a highly stereotyped motif. Song motifs are roughly one second in duration and are made up of a sequence of 2-8 syllables each 50-250 ms in duration (see Figure 3). Syllables are separated by silent intervals, roughly 50 ms long, referred to as the inter syllable intervals (ISI). A good singer will sing up to 2000 motifs per day.

Intuitively, syllables are the basic elements of song. Evidence for this comes from experiments measuring air sac pressure [Franz and Goller 2002; Cooper and Goller 2006] and electromyographic activity of respiratory muscles [Wild et al. 1998; Goller and Cooper 2004] during singing that show that the syllable rhythm correlates well with transitions between expiration and inspiration, and from perturbation experiments using flashes of light [Cynx 1990; Franz and Goller 2002] and electrical stimulation [Vu et al. 1994; Wang et al. 2008] that show that perturbation induced stopping occurs at discreet time points within the motif, usually at syllable transitions. The matter is slightly complicated by the fact that even though sound is typically generated during exhalation, sometimes the bird will also produce sounds during inhalation, in which case there is either no silent interval between syllables or it is too short to be reliably detected. In practice, as one generally does not have access to respiratory patterns or perturbation-induced stopping times, syllables and ISI’s are identified as peaks and troughs in song sound amplitude. Based on spectral properties syllables can be further subdivided into notes (see [Du and Troyer 2006] for a discussion). Notes are segments of syllables with constant spectral features (e.g. the second syllable of the motif shown in Figure 3)

Note that electrical stimulation during singing at low current amplitudes induces song stoppings at syllable transitions and increases the frequency of rare syllable transitions (similar results are obtained with acoustic stimulation [Sakata and Brainard 2006]). At high current amplitudes electrical stimulation will reliably induce immediate song stopping. It is only at intermediate current amplitudes that electrical stimulation reveals an intricate pattern of inter-hemispheric dominance alternation that no longer seems to correlate with syllable transitions as reported in [Wang, et al. 2008].
would be subdivided into three notes and the third syllable into two notes). Note however, that both syllables and notes are not well defined concepts and different authors use the terms differently.

1.2.2 Function of song

The male zebra finch sings two different variants of song depending on social context: directed and undirected song. Both sound almost identical to the untrained observer. They differ however in a number of aspects. Compared to undirected song, directed song is faster (on the order of 5%) and, on average, directed song bouts contain more motifs and introductory notes [Cooper and Goller 2006; Woolley and Doupe 2008]. Directed song is also less variable both in terms of motif duration [Cooper and Goller 2006] and spectral composition [Olveczky et al. 2005; Kao and Brainard 2006] (see Figure 4 for a sample comparison).

Figure 4: Undirected vs. directed song. Two song bouts of two motifs each, sung as the bird was alone (undirected) and immediately after the introduction of a female to the cage (directed). Note that there are slight differences in spectral structure between the first and the second motif of the undirected song bout (compare the spectral structure as highlighted by red and green circles), and much less so in the directed song bout.

Directed song is used by the male zebra finch as a courtship display directed towards a female, and is used in vocal fighting directed towards a male bird (personal observation). Note however that unlike in other songbird species, in zebra finches song appears to have no territorial function [Immelmann 1962; Immelmann 1965]. The function of undirected song is less clear and it was even proposed to have no function at all and to be mere sign of a tranquil mood [Immelmann 1968]. Yet, the male zebra finch will sing 500-1000 song bouts of undirected song per day, and by doing so not only takes the risk of being discovered by predators but also expends a significant amount of energy. Thus we are led to believe, that undirected singing is more than probable to have an important function. It is conceivable, for example, that a certain minimal rate of song production is necessary for song maintenance (see section 1.4 for a discussion). Another possible function of undirected song that has been suggested is that it serves kin recognition. Mated females will prefer the song of their mate over unfamiliar song [Woolley and Doupe 2008]. In addition the relatively high level of undirected singing activity has been suggested to increase the chance of extra-pair copulations by attracting other females.

---

8 Territorial behavior in general is found to be confined to the breeding nest only.
9 ZENK (an immediate early gene) expression is found to be increased in song control nuclei to an even greater extent during undirected singing than during directed singing [Jarvis, et al. 1998].
1.2.3 Song development

Song development can be divided into two periods. In an initial sensory period that starts at roughly 18-20 days post hatch (dph), the juvenile bird does not yet attempt to sing, and only vocalizes calls. During this period the bird will form a memory of the song of a tutor (typically the bird’s father). In the second, sensorimotor, period the bird starts to sing and gradually matches his song to that of his tutor. The first phase of the sensorimotor period is referred to as subsong phase and starts at approximately 30 dph. Subsong sounds soft and squeaky, and roughly corresponds to babbling in humans. It lacks both rhythm and spectral structure. Syllables are not stereotyped and sequences appear to be almost random. Between 55 dph and 65 dph there is relatively rapid transition from subsong to the first complete motifs within just a few days. This transition marks the start of the plastic song phase that is characterized by a comparatively slow refinement of the initially still variable and imprecise motif. That this process is by no means a linear was elegantly demonstrated by Tchernichovski et al. [Tchernichovski et al. 2001], who showed that learning trajectories are not simply determined by the differences between the bird’s own song and the tutor song. Syllable pitch, for example, can follow a period doubling trajectory. If the initial imitation attempt has a pitch higher than that of the tutor song, instead of reducing pitch, the bird will sometimes increase it until it has reached twice the value of tutor song to then half it. Variability gradually decreases with song development and birds typically achieve highly accurate copies of the tutor song by 80-90 dph. At this point song is said to be crystallized and changes only little for the rest of the bird’s life.

Figure 5: Time line of song development. See text for details.

---

10 Individual syllables that already strongly resemble syllables of the tutor song can be observed already earlier.
11 Evidence of such a period doubling trajectory can be seen in Figure 7. Between 73 dph and 77 dph the pitch of syllable 6 (red arrow) gradually increases until on day 78 it halves. Note however, that unlike previously reported [Tchernichovski, et al. 2001], the period doubling trajectory as observed here is not monotonous. Period doubles and halves in rapid succession between different motifs (see e.g. the rendition marked by the red arrow that is flanked by renditions with a doubled period).
Figure 6: Song development I. Songs recorded in the time period from 39 dph to 63 dph in a bird separated from his father at 10 dph, kept without tutor until 35 dph, and then tutored in a perch-hopping triggered-playback tutoring paradigm. All recordings are from the first two hours after lights on. The bottom two spectrograms show one rendition of the song used for tape tutoring and a song motif of the bird’s father.
Figure 7: Song development II. Continuation of Figure 6. Songs recorded in the time period from 64 dph to 90 dph. Recordings on 75 and 76 dph are missing due to a computer crash. Note that this bird’s motif (red box) at 90 dph differs from tutor motif in that syllables 2 and 3 are reversed in order and that an additional syllable (7) is added to the end of the motif. Red arrow: see footnote 11.
1.3 The song system\textsuperscript{12}

In 1976 Nottebohm et al. reported the discovery of the first non-human forebrain vocal control area\textsuperscript{13}, in the canary [Nottebohm et al. 1976]. Using targeted brain lesions, they were able to show that a region they termed HVC (originally named hyperstriatum ventrale caudale, later renamed high vocal center now used as a proper name) was necessary for singing. Moreover they could show that canaries exhibit a left\textsuperscript{14} hemisphere dominance of song control. Song deteriorated stronger after lesions of left HVC, as compared to lesions of right HVC. Using tract tracing, they could further show that HVC projects to the robust nucleus of the arcopallium (RA) in the pallium and to Area X in the striatum. Soon thereafter the afferents of HVC, the nucleus interface of the nidopallium (NIf), the nucleus uvaeformis (Uva), and the magnocellular nucleus of anterior nidopallium (MAN), were described [Nottebohm et al. 1982]. With these discoveries there came an explosion of studies on pallial song control areas that have led to our current understanding of the connectivity of what became known as the song system (see Figure 8).

The song system is the network of premotor (see section 2.2.1 for a discussion of the definition of ‘premotor’) brain areas controlling song. It is traditionally divided into the anterior forebrain pathway (AFP), composed of Area X, the medial nucleus of the dorsolateral thalamus (DLM) and lateral and medial parts of MAN (lMAN and mMAN respectively), and the posterior motor pathway consisting of Uva, NIf, HVC and RA. In female zebra finches HVC, RA and lMAN are underdeveloped and area X seems to be entirely absent [Nottebohm and Arnold 1976; Gurney and Konishi 1980; Gurney 1981; Nordean and Nordean 1988]. This difference was in part responsible for the identification of the song control nuclei [Mark Konishi during a talk].

\textbf{Figure 8: The song system}. The network of song control areas is divided into the posterior motor pathway (blue) and the anterior forebrain pathway (AFP, green). See text or section 9.1.3 for explanation of abbreviations.

\textsuperscript{12} Note on terminology: In 2002 the avian brain nomenclature forum held at Duke University revised most of the terminology describing the areas of avian forebrain. Throughout this work we will adhere to this revised nomenclature. A detailed list of changes of nomenclature can be found at \url{www.avianbrain.org}.

\textsuperscript{13} More than 100 years earlier Paul Broca discovered an area of human cortex (corresponding to Brodmann areas 44 and 45) essential for speech production.

\textsuperscript{14} Note that in zebra finches song production appears to be right dominant [Williams, et al. 1992].
1.4 The template theory of song learning

One of the major discussions in the early 20th century in neuroscience was that of central versus peripheral motor control; that is, the question of whether a motor program, once initiated, was reliant on sensory feedback from the periphery in order to be continued to completion, or whether activity of the central motor control areas was sufficient for motor output with sensory feedback merely serving a modulatory role. The obvious method of testing if the main contribution to motor control was indeed central in origin was to eliminate sensory feedback. It was argued, that if coordinated motor patterns would remain even in the absence of sensory feedback, then this was good evidence in favor of central motor control.

Upon this background Mark Konishi took to studying motor control in songbirds. In his initial experiments he deafened songbirds at different stages of song development [Konishi 1963; Konishi 1964; Konishi 1965] and discovered that a bird deafened after tutoring in the early sensorimotor phase fails to imitate the tutor and develops a highly aberrant song. He also knew from his own experiments and from those of Peter Marler that the bird would imitate tutor song even if only tutored during the sensory phase of song learning and not during the sensorimotor phase. Based on these two observations (the fact that birds are capable of storing a long-term memory of tutor song and that they need intact auditory feedback to match their own vocalizations to the tutor’s song) Konishi formulated the template theory of song learning. It postulates that in a first phase the juvenile bird, by listening to its tutor sing, acquires an auditory template of the tutor song. In a second phase the juvenile bird begins to sing itself, giving rise to auditory feedback that in turn is compared to the template of tutor song. This comparison generates a mismatch signal that is subsequently used to adapt the motor program, thus improving the imitation (see Figure 9).

Figure 9: The template theory of song learning. a: first phase, template formation. b: second phase, the template is compared to auditory feedback of own vocalizations, giving rise to an error signal which in turn adaptively changes motor output improving the imitation.

A substantial amount of research has been performed both on trying to find neurons responsive to auditory feedback while the bird is singing [Leonardo 2004; Kozhevnikov and Fee 2007; Prather et al. 2008; Sakata and Brainard 2008], and on trying to anatomically localize the neural correlates of the song template [Chew et al. 1996; Bolhuis et al. 2000; Adret 2004; Phan et al. 2006; London and

---

15 Konishi performed his experiments in juncos, white crowned sparrows, robins and black-headed grosbeaks; only later were similar effects of deafening during song development observed in zebra finches [Price 1979].

16 It could later be shown that the songs of birds deafened during early song development still contained a remarkable amount of species specific song structure [Marler and Sherman 1983].
However, most results so far have been rather elusive and have raised more questions than they have answered.

Auditory feedback sensitivity has only very recently been found in HVC interneurons [Sakata and Brainard 2008] of Bengalese finches, but not in HVC-RA projection neurons [Kozhevnikov and Fee 2007] or in IMAN neurons [Leonardo 2004] of zebra finches. If feedback sensitivity is really absent from the premotor pathway, this would imply that learning does not happen online, but would rather have to be performed offline. This hypothesis is complicated by the fact, that imitation using offline error correction would imply the existence of additional forms of memory. Either the feedback signal or the result of the comparison would have to be stored in memory to be later used, offline, to generate a correction signal directed to premotor neurons. Efforts to anatomically pinpoint the neural correlates of the song template have been equally elusive and are undermined by the difficulty of separating purely auditory signals from memory related signals.
2 THE AUDITORY SYSTEM OF THE SONGBIRD

2.1 Performance – can the bird hear everything it sings?

The basic principles of sound transduction in the ear are similar in both birds and mammals. Sound impinges upon the tympanic membrane; the resulting oscillations are transferred to the fluids of the cochlea via bone "levers" where the pressure oscillations of the inner ear fluid induce movement in the hair cells of the cochlea. Even though the basic mechanisms are similar in birds and mammals, there are a few marked differences. Most apparent is the fact that the avian ear has no external pinnae, which in mammals form the first stage of sound amplification. Birds also lack a second mechanism of sound amplification found in the mammalian ear; instead of the intricately formed chain of the three ossicles, the avian ear contains only one bone connecting the tympanic membrane to the oval window, the columella (see Figure 59). Additionally, mammalian hair cells of the auditory epithelium are arranged such that they form a regular pattern of one row of inner hair cells and three rows of outer hair cells. In the avian basilar papilla no such arrangement is found [Gleich et al. 1994]. And more interestingly, birds are capable of regenerating lost hair cells [Ryals and Rubel 1988].

When compared to the human cochlea, the cochlea of a zebra finch is considerably smaller. The basilar papilla of the zebra finch measures approximately 1.6 mm in length and contains 3600 hair cells [Gleich et al. 1994]. In contrast, the human homolog of the basilar papilla, the organ of Corti, measures on average is 31.5 mm in length [Hardy 1938] and contains on the order of 30000 hair cells at birth. The differences in geometry and the number of hair cells both suggest that the hearing acuity and/or bandwidth are considerably smaller in birds than in humans. Theoretical arguments further imply that the reduced size of the cochlear duct leads to impaired low-frequency hearing, and has less effect on high frequency hearing.

![Figure 10: Frequency range of perception and production](image)

**Figure 10: Frequency range of perception and production.** a: A typical human audiogram and a behavioral audibility curve in the zebra finch. Data is reproduced from [Ryals et al. 1999]. b: Average power spectrum of zebra finch song in arbitrary units linear in sound amplitude. Solid black line is the average of the songs of 54 male zebra finches raised in our breeding colony. Dotted lines indicate average ± standard deviation. The peak close to zero kHz is an artifact of noise on many of the recordings. Note that the best frequency range of hearing is well matched to the range of high spectral power of song.

Hearing threshold as a function of frequency is referred to as an audiogram or an audibility curve. In humans determining hearing thresholds is a trivial matter, while in birds one must resort to more
intricate methods. Two such methods are the auditory brainstem response (ABR) and behavioral testing. In measuring the ABR, similar to EEG measurements, superficial electrodes are used to record an auditory-stimulus evoked potential. In behavioral testing, the bird is trained to report the presence of an auditory stimulus for a food or water reward. An audibility curve measured in such a way in the zebra finch is shown in Figure 10a alongside a typical human audiogram. Note that hearing thresholds are higher in zebra finches than in humans at all frequencies. Moreover, the zebra finch audibility curve shows a much narrower region of high sensitivity in the frequency range of 2 to 6 kHz. Outside of this region the zebra finch is nearly deaf (a sound amplitude level of 60 dB corresponds to the sound amplitude level of normal conversation). At birth, hearing thresholds in birds are typically higher and only reach adult levels by 20-25 days post hatch (as demonstrated in canaries and budgerigars [Brittan-Powell and Dooling 2004]), coincident with the onset of the sensory phase of song learning.

During vocalizations an additional mechanism is trigger that acts to attenuate oscillations of the columella, thus attenuating the amplitude of the oscillations of the hair cells. This mechanism is known as the middle ear acoustic reflex [Borg et al. 1982; Counter and Borg 1982]. Oscillations of the columella are dampened by the columellar muscle, both anticipatorily during self-initiated vocalizations and in response to very loud sounds. The achieved attenuation can be as large as 20 dB [Counter and Borg 1979].

2.2 Anatomy

2.2.1 Note on the usage of the terms ‘sensory area’ and ‘motor area’

The classification of areas as sensory or as premotor is often influenced more by the chronology of discoveries of the different characteristics of a particular area, rather than by analytic description. The resulting classifications in turn fundamentally influence the way we think about the system we are analyzing. For example, the classification auditory (or sensory) has a strong connotation of passive processing. For questions regarding song learning this has led to a research focus on the song system, and to a certain negligence of the possible active aspects of processing occurring in ‘auditory’ areas. We will use the terms in this work relatively consistent with previous usage, however to avoid using mere intuitive notions of ‘auditory’ and ‘premotor’ we shall use definitions as outlined below.

In general, a functional classification can be based on the following measurements:

i) Electrophysiological responses

ii) Immediate lesion (or inactivation) induced effects

iii) Anatomical connectivity

Gene expression patterns seem to be unsuited for functional classification, as the underlying mechanisms are not well understood. In zebra finches, for example, ZENK expression after auditory stimulation is up-regulated in most secondary auditory areas yet not in a primary auditory area [Mello and Clayton 1994]. Also unsuitable for classification are long term lesion effects, as it is more
Premotor areas are characterized by the following criteria:

1. A lead of electrophysiological activity on sound amplitude during singing, i.e. electrophysiological activity is predictive of motor output,
2. Immediate detrimental lesion effects on song, during at least one stage of song development,
3. Existence of a projection pathway of axonal connections to sub-pallial motor areas (such as nXIIIts).

Auditory areas are characterized by the following criteria:

1. A correlation between auditory stimulation and electrophysiological activity at positive time lags, i.e. the auditory stimulus is predictive of the electrophysiological activity,
2. A lack of immediate detrimental lesion effects on song, during all stages of song development,
3. Existence of a projection pathway of axonal connections from the cochlea.

Note that this classification is not easily extended to sub-pallial brain areas due to the fact that many brainstem structures are vital for survival (e.g. one cannot assess the effects of lesions of respiratory areas on song because damage to these areas is usually lethal). Also note, that the list of characteristics is not exclusive (premotor areas of the song system have been shown to respond to auditory stimulation [Williams and Nottebohm 1985]), however point ii) ensures that the two categories are mutually exclusive. The category of auditory areas is thus – as in most previous work – held artificially large and includes areas that would correspond to associative areas of mammalian cortex.

2.2.2 The auditory network

In the early days of avian brain anatomy, research was confined to cataloging the different brain areas based on cyto-architectonic structure, similar to the efforts of Korbinian Brodmann in cataloging the areas of cortex. Unlike in cortex, however, most of the nomenclature has not survived the years, as more recent interest has been directed mainly at the song motor control system that is found in the pallial nuclei and not the anatomically less well-defined areas surrounding the nuclei. One of the remnants of this time is the name of the avian auditory cortex analogue Field L, introduced by Rose in 1914 who used letters of the alphabet to label the different areas (or fields) in his map of the avian brain [Rose 1914]18.

It was not until fifty years later that the first auditory pathways to Field L were discovered. After the ascending connections from the cochlea to the mesencephalic lateralis dorsalis (MLd) had been

---

17 In cases of severe epilepsy in humans, for example, entire cortical hemispheres have been removed in early childhood without inducing severe long-term behavioral deficits.
18 Note that today, based on cytoarchitectonic organization, field L is subdivided into different sub-fields: L1, L2a, L2b, L3 and L. See [Fortune and Margoliash 1992] for details.
described and the analogy between MLd and the mammalian inferior colliculus was established [Boord 1968], Harvey Karten described the ascending projections of MLd to the thalamus [Karten 1967] and from there to the pallium [Karten 1968]. This laid the basis for an extensive series of studies analyzing the connectivity of the avian auditory network depicted schematically in Figure 3, and listed, with references to the work describing each projection, in Table 1. Data used to build the network is pooled across avian species for the hind- and midbrain connectivity, as these structures are thought to be homologous in all avian species. Most of the early work on the subpallial auditory network was conducted in the pigeon and the owl. Most of the data on the thalamic and the pallial connectivity comes from work on songbirds.

Figure 11: Connectivity diagram of the auditory system of the songbird and its interface to the posterior motor pathway. See Table 1 for details of individual connections and “Appendix” for abbreviations. The premotor nature of paraHVC, as suggested by connectivity [Foster and Bottjer 1998], is still unconfirmed. Other brain areas involved in song production that are not included in the diagram are the AFP and sub-pallial premotor areas (such as respiratory areas). Areas are arranged in a hierarchy based on distance from cochlea in synapses. This distance is the theoretically shortest distance based on the known inter-area connectivity. In the case of CSt the synapse count suggested by location on the diagram is incorrect, the correct distance is 5. Only connections that contribute auditory input to the right forebrain are shown, pathways that would require two crossings of the midline are not shown.
<table>
<thead>
<tr>
<th>FROM</th>
<th>TO [REFERENCES]</th>
<th>COMMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cochlea</td>
<td>--- &gt; NA [Correia et al. 1982]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; NM [Correia et al. 1982]</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>--- &gt; SON [Correia et al. 1982; Yang et al. 1999]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; LL nuclei [Takahashi and Konishi 1988a]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; MLD [Conlee and Parks 1986]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-x&gt; SON [Correia et al. 1982; Conlee and Parks 1986]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-x&gt; LL nuclei [Takahashi and Konishi 1988a]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-x&gt; MLD [Takahashi and Konishi 1988a]</td>
<td></td>
</tr>
<tr>
<td>NM</td>
<td>--- &gt; NL [Boord 1968; Parks and Rubel 1975; Takahashi and Konishi 1988b]</td>
<td>both projections are tonotopic [Parks and Rubel 1975] and the sole efferent connections of NM [Takahashi and Konishi 1988a]</td>
</tr>
<tr>
<td></td>
<td>-x&gt; NL [Boord 1968; Parks and Rubel 1975; Takahashi and Konishi 1988b]</td>
<td></td>
</tr>
<tr>
<td>LL nuclei</td>
<td>--- &gt; MLD [Conlee and Parks 1986]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; Uva [Coleman et al. 2007]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-x&gt; MLD [Conlee and Parks 1986]</td>
<td></td>
</tr>
<tr>
<td>SON</td>
<td>--- &gt; NA [Correia et al. 1982; Yang et al. 1999]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; NM [Yang et al. 1999]</td>
<td>provides GABAergic inhibitory feedback to laminaris neurons</td>
</tr>
<tr>
<td></td>
<td>--- &gt; NL [Yang et al. 1999]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; MLD [Correia et al. 1982]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-x&gt; MLD [Correia et al. 1982]</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>--- &gt; SON [Yang et al. 1999]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-x&gt; LL nuclei [Conlee and Parks 1986]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-x&gt; MLD [Boord 1968; Takahashi and Konishi 1988a]</td>
<td></td>
</tr>
<tr>
<td>MLD</td>
<td>--- &gt; Ov [Karten 1967; Akesson et al. 1987]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; Ov shell [Durand et al. 1992]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-x&gt; Ov [Karten 1967; Akesson et al. 1987]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-x&gt; MLD [Karten 1967; Akesson et al. 1987]</td>
<td></td>
</tr>
<tr>
<td>Ov core</td>
<td>--- &gt; L2a [Karten 1967; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td>Ovm</td>
<td>--- &gt; L2b [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td>Ov shell</td>
<td>--- &gt; L1 [Karten 1967; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; L3 [Karten 1967; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; NCM [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; CMM [Durand et al. 1992; Wild et al. 1993; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; CLM [Durand et al. 1992; Wild et al. 1993; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td>L2b</td>
<td>--- &gt; L2a [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; CST [Kelley and Nottebohm 1979; Vates et al. 1996; Cohen et al. 1998]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; CLM [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td>CST</td>
<td>--- &gt; NCM [Mello and Clayton 1994]</td>
<td></td>
</tr>
<tr>
<td>L2a</td>
<td>--- &gt; L2b [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; L3 [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; L1 [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; NCM [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--- &gt; CLM [Fortune and Margoliash 1995; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td>L3</td>
<td>---→ L2a [Fortune and Margoliash 1995]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ L1 [Fortune and Margoliash 1995]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ NCM [Fortune and Margoliash 1995]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ CLM [Fortune and Margoliash 1995]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ RA cup [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ HVC shelf [Fortune and Margoliash 1995; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ HVC [Saini and Leppelsack 1981; Fortune and Margoliash 1995], see however [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td>---→ L2a [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ L3 [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ CLM [Fortune and Margoliash 1995; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ RA cup [Fortune and Margoliash 1995; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ HVC shelf [Fortune and Margoliash 1995; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ HVC [Fortune and Margoliash 1995], see however [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td>NCM</td>
<td>---→ CMM [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CSt [Mello and Clayton 1994]</td>
<td></td>
</tr>
<tr>
<td>CMM</td>
<td>---→ NCM [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ CLM [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td>CLM</td>
<td>---→ L2b [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ L2a [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ L3 [Fortune and Margoliash 1995; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ L1 [Fortune and Margoliash 1995; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ CMM [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ RA cup [Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ HVC shelf [Fortune and Margoliash 1995; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ NIf [Vates et al. 1996; Bauer et al. 2008]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ HVC [Fortune and Margoliash 1995; Bauer et al. 2008]</td>
<td></td>
</tr>
<tr>
<td>RA cup</td>
<td>---→ LL nuclei [Wild et al. 1993; Mello et al. 1998]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ MLD [Wild et al. 1993; Mello et al. 1998]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ Ov [Wild et al. 1993; Mello et al. 1998]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ RA [Mello et al. 1998]</td>
<td></td>
</tr>
<tr>
<td>HVC shelf</td>
<td>---→ RA cup [Mello et al. 1998]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ HVC [Saini and Leppelsack 1981; Vates et al. 1996]</td>
<td></td>
</tr>
<tr>
<td>RA</td>
<td>---→ HVC [Roberts et al. 2008]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ nXlIts [Nottebohm et al. 1976]</td>
<td></td>
</tr>
<tr>
<td>HVC</td>
<td>---→ CLM [Nottebohm et al. 1982; Fortune and Margoliash 1995]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---→ RA [Nottebohm et al. 1976]</td>
<td></td>
</tr>
<tr>
<td>Nif</td>
<td>---→ HVC [Nottebohm et al. 1982; Fortune and Margoliash 1995]</td>
<td></td>
</tr>
<tr>
<td>Uva</td>
<td>---→ HVC [Nottebohm et al. 1982; Akutagawa and Konishi 2005]</td>
<td></td>
</tr>
</tbody>
</table>

However, without retrograde confirmation.

RA has no known projections to the auditory system.

CLM is also referred to as field avalanche [Nottebohm et al. 1982].
### Table 1: List of the known projections in the avian auditory system.

Completeness is bounded only by shortcomings of the author’s knowledge. “---->” indicates ipsilateral projection. “---x->” indicates contralateral projection. Areas are arranged as in Figure 11, bottom to top, left to right. Projection targets in italic indicate projections whose existence has been called into question. See section 9.1.3 for explanation of abbreviations.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th>---</th>
<th>---</th>
<th>---</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>paraHVC</td>
<td>---</td>
<td>CST [Foster and Bottjer 1998]</td>
<td>---</td>
<td>NCM [Foster and Bottjer 1998]</td>
</tr>
</tbody>
</table>

2.2.3 Auditory input to the song system

Of particular interest are connections that link auditory and premotor areas, as such connections are necessary for sensorimotor integration and could mediate error signals necessary for song learning.

Vates et al. have shown that the NIf receives a direct projection from the auditory area caudolateral mesopallium (CLM) [Vates et al. 1996]. This combined with the strong projection from NIf to HVC led to the suspicion that NIf is the major source of auditory input to HVC. This idea awaited conformation until the experiments of Cardin et al. who showed that auditory activity in HVC almost entirely vanishes after NIf inactivation [Cardin and Schmidt 2004a; Cardin et al. 2005]. Bauer et al. subsequently demonstrated that there is a separate auditory input to HVC directly from CLM, and that only inactivation of NIf and CLM completely abolishes auditory responses in HVC [Bauer et al. 2008].

Additionally, there is some evidence for a direct projection from L1 and L3 to HVC that comes from retrograde labeling found in field L after tracer injections in HVC and anterograde labeling found in HVC after tracer injections in field L [Fortune and Margoliash 1995; Vates et al. 1996]. However, retrogradely labeled cell bodies found in L1 and L3 could have resulted from spread of tracer to the HVC shelf, and similarly, anterograde labeling found in HVC could have resulted from spread of tracer to NIf (see [Vates et al. 1996] for a discussion). Nevertheless, L1 and L3 both have a strong projection to the HVC shelf [Vates et al. 1996], and the strongest evidence for a projection from the shelf to HVC proper comes from tracer injections into ventral HVC that show that neurons along the ventral border of HVC have spiny dendrites that protrude far into the shelf [Vates et al. 1996]. Thus, it is possible that the axons of L1 and L3 cells synapse directly onto HVC neuron dendrites in the shelf.

2.2.4 Feedback connections from the song system to auditory areas

Currently we have only little knowledge of the relevant connections that feed premotor activity back to auditory areas. There are two possible candidate pathways by which this could occur, evidence for both pathways, however, is relatively sparse.

19 Note that the authors claimed that auditory activity completely vanished after NIf inactivation.
Nottebohm et al. first reported the existence of a connection from HVC to a restricted area within anterior CLM, they termed field avalanche [Nottebohm et al. 1982]. Fortune and Margoliash later confirmed this observation, showing that tracer injection into anterior CLM retrogradely labels neurons in HVC [Fortune and Margoliash 1995]. Vates et al. however find no such projection. They speculate that the reason for this might be the fact that they were looking at posterior as opposed to anterior CLM [Vates et al. 1996].

The second possible feedback pathway discovered by Foster and Bottjer runs from HVC via paraHVC to the caudal medial nidopallium (NCM) [Foster and Bottjer 1998]. While there is convincing evidence for a projection from HVC to paraHVC, the projection from paraHVC to NCM stands on less solid ground. Due to the small thickness of paraHVC, injections are hard to target exclusively at paraHVC, to such an extent that in all reported injections there was spread of the tracer to dorsal NCM directly underlying paraHVC [Foster and Bottjer 1998].

2.3 Function
2.3.1 Electrophysiology
There have been two major approaches to understanding auditory processing in the songbird. The first is a top down approach, motivated by the idea of finding a basis for the bird’s own song (BOS) selectivity observed in HVC. The second is a bottom up approach driven by the attempt to describe the input to the auditory pallium coming from the auditory hind- and midbrain in a receptive field framework.

2.3.1.1 Origin of BOS selectivity
With the discovery of auditory neurons in HVC [Katz and Gurney 1981] there arose the hope of finding neural correlates of the song template. It was reasoned that if HVC was imprinted with an auditory memory of tutor song during the sensitive phase of song learning, it should respond selectively to auditory stimulation with a playback of BOS or of tutor song. This argument is constructed as follows. Stimulating the juvenile bird with tutor song would lead to a certain auditory activity pattern in HVC. By being repeated thousands of times, this auditory activation would lead to a strengthening of synapses between any given neuron and those that fire in close succession (based on a mechanism such as spike-timing dependent plasticity (STDP)). As a result, this process would lead to a synfire-chain-like network within HVC (electrophysiological recordings of the activity of HVC neurons during singing appear to strengthen this hypothesis [Hahnloser et al. 2002; Prather et al. 2008]). Depending on whether the differences between BOS and the tutor song stem from poor memorization or from poor reproduction, the network should thus be particularly well activated by auditory stimulation with BOS or tutor song playback. The phenomenon was coined song selectivity, and was indeed found in a subset of HVC neurons [Margoliash 1983; Margoliash and Konishi 1985].

Because the bird itself produces thousands of imperfect versions of the motif in the early plastic song phase that would ‘overwrite’ the original tutor template, the hypothesis that HVC encodes an auditory template of tutor song further requires that either HVC ceases to be activated by auditory stimulation at the transition from subsong to plastic song, or that auditory responses are suppressed while the bird is singing. Support for the latter came from the discovery, that auditory responses in HVC are attenuated in the first few seconds after singing [McCasland and Konishi 1981], demonstrating that HVC does indeed function in two separate processing modes.
However, later experiments showed that tutor song selectivity is not yet present at the end of the sensory phase and only develops during the sensorimotor phase of song learning. Neurons in HVC and the AFP show are not selective for tutor song or for BOS after tutoring but are BOS selective during plastic song [Volman 1993; Doupe 1997; Nick and Konishi 2005a], indicated that song selectivity is not an artifact of template acquisition but most likely an artifact of song learning. This idea was supported by the finding that most neurons in the AFP of 60 day old birds respond stronger to BOS playback than to tutor song playback [Solis and Doupe 1997]. Moreover, in the same experiments Solis and Doupe find neurons in the AFP that are both BOS and tutor song selective (i.e. show responses that are significantly higher during BOS and tutor song playback, as compared to responses during playback of conspecific song). The existence of neurons that exhibit such a dual selectivity at a stage in development where song selectivity has just emerged, would suggest that both BOS and tutor song selectivity in the AFP develop in parallel during the early plastic song phase. However BOS and tutor song are typically acoustically similar in the plastic song phase and thus this dual selectivity was not uniquely interpretable. To circumvent this problem Solis and Doupe performed surgical transections of the tracheosyringeal portion of the hypoglossal nerve (nXIIts) in juvenile birds before song learning [Solis and Doupe 1999; Solis and Doupe 2000]. This procedure severely distorts the bird’s song, resulting in large dissimilarities between BOS and tutor song. In spite of this dissimilarity, dual selectivity persisted in a subset of AFP neurons. This demonstrated that song selectivity cannot simply be explained by an auditory memory of tutor song, but rather is the result of a dynamical process that involves both the tutor song memory and the current BOS, and which emerges only well after tutoring.

Further evidence against the idea that HVC stores a template of tutor song comes from the more recent discovery that auditory responses in HVC and NIf in adult zebra finches depend strongly on the bird’s behavioral state [Cardin and Schmidt 2003; Cardin and Schmidt 2004a]. Auditory responses are strongest under sedation and absent in the aroused state. It could further be shown that this modulation is mediated by noradrenergic inputs [Cardin and Schmidt 2004b], thus suggesting that auditory information reaches the song system only in states of low attention. Clearly not what one would have predicted from the hypothesis that HVC stores an auditory song template.

Having discovered a strong selectivity for BOS in HVC, the natural question to ask was how this selectivity arises. Is selectivity a result of the network activity within HVC itself, or does HVC simply form the top of an auditory processing hierarchy? It was shown, for example, that auditory responses in NIf are song selective, yet less so than in HVC [Lewicki and Arthur 1996; Janata and Margoliash 1999]. In field L no evidence for any consistent BOS or tutor song selectivity was found [Margoliash 1986; Lewicki and Arthur 1996; Janata and Margoliash 1999]. These findings seemed to confirm the idea of an auditory song selectivity hierarchy that ascends from field L to HVC via NIf and sparked interest in primary and secondary auditory areas of the pallium.

Electrophysiological recordings in anesthetized animals have shown that many neurons in field L and in the adjacent areas CM and NCM are auditory. Amin et al., for example, report to find 352 of 647 (54%) recording sites in field L and CM that are responsive (as judged by firing rates significantly

\[\text{See however [Nick and Konishi 2005b]. They could show that tutor song responses in HVC depend on behavioral state. In the late subsong/early plastic song phase responses to tutor song are stronger in the awake than in the sleeping state, and show a preference for the tutor song during wakefulness.} \]

\[\text{Indeed, by inducing song decrystallization (by vocal nerve section) it could be shown that IMAN neurons can become selective for decrystallized BOS [Roy and Mooney 2007].} \]
different from baseline, t-test, p<0.05, all recordings in the range 1.2-1.6 mm from midline) to auditory stimuli [Amin et al. 2004] (see Figure 12 for sample responses of an L2b neuron).

While no consistent BOS or tutor song selectivity has been found in any of the regions of the auditory forebrain, there is evidence for selective processing of familiar stimuli, such as BOS and tutor song, in secondary auditory areas CM and NCM [Chew et al. 1995; Gentner and Margoliash 2003; Phan et al. 2006] (see also section 2.3.2 for further discussion). In NCM the rate of response adaptation seems to be an indicator of the novelty of a stimulus. Phan et al. claim that NCM responses adapt faster when presented with a novel song as compared to stimulation with tutor song [Phan et al. 2006]. Gentner and Margoliash show that CLM responds more strongly to song stimuli with a behavioral significance (stimuli used in a song discrimination task) as opposed to novel songs [Gentner and Margoliash 2003]. All of these results indicate that auditory processing is indeed modulated by experience. The resulting effects in field L, however, appear to be too weak to be measured simply by averaging responses over the entire duration of a song stimulus, as is done when determining song selectivity using z-scores or d’-scores.

Assuming that HVC is not the primary site of storage of an auditory template of tutor song, and considering that neurons of the auditory forebrain are capable of highly adaptive responses, we can speculate that the auditory forebrain stores a sensory template of tutor song that is mapped onto HVC during the plastic phase of song learning. It is only then that this mapping process gives rise to the observed song selectivity in HVC and the AFP.

2.3.1.2 Auditory receptive fields

In an attempt to understand the pallial processing of auditory information, there have been extensive efforts to characterize the function of the processing hierarchy that provides the input to the auditory pallium. Auditory signals are passed from the cochlea to the cochlear nuclei (nucleus magnocellularis (NM) and nucleus angularis (NA)), where they are passed on to a number of hindbrain structures (lateral lemniscal nuclei (LL), the superior olivary nucleus (SON) and the nucleus laminaris (NL)). This intricate network of hindbrain structures projects to the auditory midbrain (Mesencephalicus lateralis dorsalis (MLd)), which in turn projects to the thalamic nucleus ovoidalis (Ov). Ovoidalis provides the main auditory input to the pallium and projects to field L2a and L2b (see Figure 11 for a schematic representation of the entire auditory network). Most of the projections of

---

22 This naturally begs the question what the other 46 % are doing?
23 Note that because the example data shown appears to be contradictory (see figure 2 in [Phan, et al. 2006], responses to tutor song adapt faster than responses to a novel song between repetitions 5 and 20) we are left uncertain, as to how strong this effect is.
the pathway leading up to field L have been found to be tonotopic, either by anatomical connectivity [Parks and Rubel 1975; Takahashi and Konishi 1988b; Wild et al. 1993] or by systematic response mapping [Zaretsky and Konishi 1976; Müller and Leppelsack 1985; Bigalke-Kunz et al. 1987; Gehr et al. 1999; Gehr et al. 2000; Woolley and Casseday 2004; Terleph et al. 2006].

One way to characterize the response properties of auditory neurons beyond merely measuring preferred frequency is to use spectro-temporal receptive fields (STRF). Similar to the concept of visual receptive fields, STRFs quantify the linear component of the stimulus-response characteristic of a given auditory neuron. STRFs can be constructed from spike-triggered average spectrogram of the auditory stimulus (reverse correlation). It can be shown however, that this method only produces reasonable results, if the set of auditory stimuli used spans the entire stimulus space (such as e.g. white noise). To measure STRFs based on data that does not fulfill this criterion, Theunissen et. al. have proposed the use of a method that normalizes the spike-triggered average by the autocorrelation function of the test stimulus [Theunissen et al. 2000]24. With this they had a potent tool at hand that allowed them to compare the response properties of auditory neurons to different stimulus categories. They could show that in field L, STRFs calculated from noise stimuli are different from the STRFs calculated from natural stimuli, and moreover, that the STRFs measured from natural stimuli generalized better to other stimulus categories than the noise STRFs25. The results also showed however, that a large part of the responses in field L are highly non-linear and cannot be captured by a linear STRF.

To understand what part of the responses observed in field L was intrinsically generated and what part was input driven, it was then necessary to characterize auditory responses in subpallial auditory areas. Woolley and Casseday found that MLd neurons are well suited for encoding a wide range of complex sounds with a high degree of temporal accuracy rather than selectively responding to only some sounds [Woolley and Casseday 2005]. Additionally, it was shown that, similar to field L neurons, tuning properties of MLd neurons also changed with a change in stimulus category [Woolley et al. 2006], suggesting that already MLd is in different processing regimes depending on the auditory stimulus category. It was also shown that response properties of both MLd and field L neurons are tuned to maximize information content of natural stimuli by enhancing acoustic differences between different natural sounds [Hsu et al. 2004; Woolley et al. 2005]. Both of these results would suggest that a large part of the passive processing characteristics of field L is inherited from an extensive subpallial processing network.

### 2.3.2 Immediate early gene expression

Immediate early genes (IEG) are genes whose expression is rapidly co-modulated with cellular stimuli such as neuronal activity and growth. They are activated before any new proteins are synthesized and lay the ground for the genomic response. An IEG that has become prominent in songbird research is ZENK (an acronym for the avian orthologue of the mammalian genes *zif-268, egr-1, ngfi-a* and *krox-24*). ZENK expression in the auditory forebrain is increased in response to auditory stimuli.

---

24 The intuition behind this can be illustrated by considering a set of test stimuli containing a constant 50 Hz noise. Independent of spike times, the spike-triggered average will always contain a constant 50 Hz band. Normalizing by the autocorrelation will reduce the amplitude of this 50 Hz band, because it is constant, to the noise level.

25 The performance of an STRF is measured by convolving it with the stimulus to yield a predicted firing rate and comparing this prediction to the measured firing rate.
and in song system nuclei in response to singing. Because ZENK has been implicated in processes of synaptic plasticity involved in learning [Rose 1991; Kaczmarek 1993], the hope arose that one might be able to identify the site of template storage using IEG expression levels in response to auditory stimuli. As a result, early studies focused on measuring ZENK expression induced by auditory stimulation. Claudio Mello et al. [Mello et al. 1992] found a strong ZENK induction in the auditory forebrain in response to conspecific song but less so in response to heterospecific song and none at all in response to artificial tone stimuli. Later this finding was refined when Mello and Clayton showed that in response to conspecific song, ZENK is induced in the entire auditory forebrain (L1,L3,NCM,CM,HVC shelf, RA cup) with the exception of L2 [Mello and Clayton 1994], which demonstrated that ZENK expression is not simply linearly correlated with neural activity, as L2 exhibits the highest firing rates of the auditory forebrain. Mello and Clayton further confirmed this by showing that the administration of metrazol (a GABA antagonist) during auditory stimulation leads to a non-uniform increase in ZENK expression, with only a small increase in the auditory forebrain [Mello and Clayton 1995].

Despite intensive searching, no auditory-induced ZENK expression was found in the nuclei of the song system [Mello and Clayton 1994; Jarvis et al. 1995; Mello and Clayton 1995], even though song system nuclei can show auditory evoked electrophysiological responses. Eric Jarvis and Fernando Nottebohm were able to demonstrate that ZENK expression is up-regulated in song system nuclei during singing in a motor-driven fashion [Jarvis and Nottebohm 1997]. They compared ZENK expression patterns in normal singing birds, deafened singing birds, and birds listening to playbacks of their own songs. They find an almost 20-fold ZENK gene-induction increase in HVC in both normal singing and deafened singing birds. Thus the increase in ZENK expression in song control nuclei during singing is not caused merely by the bird hearing itself sing. They also find no singing-induced ZENK expression in the HVC shelf and in NCM in deafened birds. This implies that ZENK expression in song system nuclei is motor driven, while expression in auditory areas is sensory driven during singing. Surprisingly there is no area that in which ZENK expression is both sensory and motor driven.
3 AUDITORY FEEDBACK PROCESSING

Songbirds are capable of vocal learning and communication [Marler and Tamura 1964; Thorpe 1975] and are ideally suited to study neural mechanisms of complex sensory and motor processing. Vocal communication in a noisy bird colony and vocal learning of a specific song template both require the ability to monitor auditory feedback [Konishi 1965; Brainard and Doupe 2000a] to distinguish self-generated vocalizations from external sounds and to identify mismatches between the developing song and a memorized template acquired from a tutor [Immelman 1969]. However, neurons that respond to auditory feedback from vocal output have not been found in song-control areas despite intensive searching [Leonardo 2004; Kozhevnikov and Fee 2007; Prather et al. 2008]. We investigate feedback processing outside the traditional song system, in single auditory forebrain neurons of juvenile zebra finches that were in a late developmental stage of song learning. Overall, we found similarity of spike responses during singing and during playback of the bird’s own song, with song responses commonly leading by a few milliseconds. However, brief time-locked acoustic perturbations of auditory feedback revealed complex sensitivity that could not be predicted from passive playback responses. Some neurons that responded to playback perturbations did not respond to song perturbations, reminiscent of sensory-motor mirror neurons [Rizzolatti and Craighero 2004; Prather et al. 2008]. By contrast, some neurons were highly feedback sensitive in that they responded vigorously to song perturbations, but not to unperturbed songs or perturbed playback. These findings suggest that a computational function of forebrain auditory areas may be to detect errors between actual feedback and mirrored feedback deriving from an internal model of the bird’s own song, or that of its tutor. Such feedback-sensitive spikes could constitute the key signals that trigger adaptive motor responses to song disruptions [Leonardo and Konishi 1999; Tumer and Brainard 2007] or reinforce exploratory motor gestures for vocal learning [Fiete et al. 2007b].

3.1 Song-evoked activity is similar to BOS-playback-evoked activity

Field L and the caudolateral mesopallium (CLM) are interconnected brain areas not part of the traditional song-control system and are analogous to auditory cortex in mammals in that they receive the main stream of auditory input from the thalamus, as well as feedback from motor-related areas [Karten 1968; Nottebohm et al. 1982; Vates et al. 1996; Foster and Bottjer 1998; Zeng et al. 2004]. Neurons in field L and CLM of awake and anaesthetized animals respond robustly to a large variety of auditory stimuli such as white noise, the birds’ own song (BOS), and conspecific songs [Janata and Margoliash 1999; Cardin and Schmidt 2003; Amin et al. 2004; Bauer et al. 2008]. These features make field L and CLM potential substrates for the integration of self-generated and external sounds and for monitoring singing-related auditory feedback. To explore this hypothesis, we made extracellular recordings from CLM and field L neurons in juvenile male zebra finches using chronically implanted miniature motorized microdrives (Suppl. Fig. 1). Our strategy was first to probe singing-related firing in these neurons for evidence of motor-specific processing beyond passive auditory responses elicited by playback of the BOS, and then to investigate the feedback sensitivity of singing-related spikes by delivering brief acoustic stimuli during singing. In total we recorded from 92 cells during song and playback of the BOS at physiological sound levels at which birds hear each other’s songs, in 67 of these we also performed perturbation experiments.

26 Most of the data in this chapter has appeared elsewhere [Keller and Hahnloser 2009].
Figure 13: Active responses were similar to passive responses, but more stereotyped and slightly anticipatory. a: Activity during song (green rasters) leads on activity during BOS playback (red rasters) in this example neuron. Top: spectrogram of an example song motif. Bottom: average firing rate curves. b: Scatter plot of firing rates (Z scores) in individual cells (black circles: cells stimulated with one version of the BOS; red crosses: many versions). c: The mean coherency function between singing and playback-related spike trains peaked 6.8 ms (dotted line) after singing-related spikes. d: The spike rasters in this neuron are anticipatory to song onset (blue line) and delayed to playback onset. Significant deviations from baseline firing are marked by green and red horizontal bars, onset times by asterisks. Top: Mean sound amplitude (solid) ± standard deviations (dashed). e: Cumulative distribution of response onset times. f: Median firing stereotypy (black bars) during song (n=92) is similar to stereotypy during playback of a single BOS (n=24) but higher than stereotypy during playback of many versions of the BOS (n=68). Second and third quartiles are shown by coloured boxes.

Singing and playback-related firing was similar in most cells (see Figure 13a), despite the large differences in sound amplitudes in vocal and playback conditions and despite the variable direction of the playback source relative to the bird’s head. In most cells, average firing rates increased from baseline in either vocal or playback condition (see Figure 13b, vocal: increase in 52/92 cells and decrease in 9/92 cells; playback: increase in 50/92 cells and decrease in 2/92 cells, significant changes from baseline were detected by Z scores >0.75 or <-0.75, see methods section 8.4 for details). Note that in both the auditory cortex of marmosets and the auditory ganglion of crickets there is a dominant suppression of activity during vocalization. The fact that we rarely observed
suppression during vocalization could not be explained by lower background firing rates in our study (25 Hz versus 11 Hz reported in marmosets).

Average firing rates in vocal and playback conditions were highly correlated (correlation coefficient 0.77, p<10\(^{-10}\), n=92; see Figure 14 and Figure 15). Firing rates were also well matched on a finer time scale. In 69/92 cells, the spike-coherency function [Kimpo et al. 2003; Amin et al. 2004; Hahnloser et al. 2006] averaged over all pairings of song and playback trials displayed a significant peak (>2 jacknife standard deviations above zero) within ±20 ms. Interestingly, the peak of the mean coherency function (averaged over 92 cells) was significant and occurred 6.8 ms after singing-related spikes (see Figure 13c), indicating that overall, singing-related activity slightly preceded playback-evoked responses (only 0.5 ms of this lag can be explained by the closer proximity of the sound source to the bird’s ears during singing). This anticipatory behaviour of singing-related activity suggests that in addition to auditory inputs, cells in field L and CLM also received inputs from a vocal-related, non-auditory source. Consistent with this view, in roughly one fifth of the cells we observed firing increases before onset of the first introductory note of a song bout (see Figure 13d-e and Figure 16), clearly demonstrating a source of non-auditory drive in these cells during singing. Hence, some neurons seemed to integrate auditory with non-auditory signals, the latter of which may have reflected information about song-motor activity, for example, as part of a motor estimate of auditory feedback.

Motor-specific processing was also evident by analysis of firing stereotypy which was higher during singing than during playback of different versions of the BOS (Wilcoxon rank-sum (WRS) test, p<10\(^{-8}\), see Figure 13f). This stereotypy difference could not be attributed to intrinsic differences between song and playback stimuli (since the latter were copies of the former), or to differences in average firing rates (see Figure 15). By contrast, the firing stereotypy during singing did not differ significantly from the stereotypy during playback of one unique BOS stimulus (WRS test, p=0.42). Thus, singing-related firing stereotypy was higher than predicted by passive responses, but was commensurate with intrinsic synaptic noise, suggesting that auditory responses may be partly subsumed during singing by a motor-specific source of stereotyped synaptic input.

### 3.2 Song-perturbation responses cannot be predicted from passive playback responses

Brief acoustic stimuli delivered during singing provide an effective means for operant conditioning of song features [Tumer and Brainard 2007]. Such stimuli are thus ideally suited for probing auditory feedback sensitivity. In 50% of song motifs (randomly selected) we presented a brief perturbing stimulus through a second loudspeaker that was time-locked to a given syllable (see methods section 8.1, Figure 17a). In agreement with previous reports in adult birds [Leonardo 2004], we found that feedback perturbations did not induce immediate spectral or temporal changes in vocal output (all analyses were restricted to song motifs with conserved syllable sequences, see Figure 18, Figure 19, Figure 20 and methods section 8.1).

We quantified the propensity of cells to respond to perturbations either in the vocal or the playback condition in terms of a response bias \(b\) confined to values between \(b=−1\) (no perturbation response during song) and \(b=1\) (no perturbation response during playback). Similarly, we quantified the sensitivity of responses to perturbed feedback in terms of a selectivity \(s\) that was normalized to \(s=0\) if the firing did not change during feedback perturbations and to \(s=1\) if the firing doubled during
perturbations (see methods section 8.4). Overall, 66/67 cells significantly responded to feedback or playback perturbations (54/67 to feedback, and 53/67 to playback, see methods section 8.4).

**Figure 14: Similarity of song and BOS playback evoked activity.** Spike-raster stack plots of cells recorded in two birds (left and right) showing similarity of vocal-related activity (red) and BOS-playback-related activity (blue). Spike raster plots in different neurons (22 on the left, 7 on the right) are drawn on alternating white and grey backgrounds. The spike raster plots from four putative CLM neurons are indicated by black vertical bars.
Figure 15: Average singing and BOS playback related firing rates. Scatter plot of average firing rates during singing versus BOS playback in 92 cells. Red crosses indicate cells stimulated by a single version of the BOS and black circles indicate cells stimulated by multiple versions. The red dashed line is a linear regression through the red crosses and the black dashed line is a linear regression through the black circles. These regression curves almost coincide with the diagonal, illustrating similarity of firing rates in song and playback conditions. Singing- and playback-related firing rates were not statistically different from each other, Wilcoxon signed-rank test, P=0.10.

Figure 16: Anticipatory activity. Shown is a spike raster plot of a neuron with anticipatory firing to the first introductory note of a song bout in the vocal condition (green) and delayed firing in the playback condition (red). The vertical blue line marks the onset of the first introductory note. Shown above the spectrogram is the mean sound amplitude (solid black line) ± standard deviation (dashed black lines). Green and blue horizontal bars mark time intervals in which firing rates are significantly different from baseline (30 ms moving window analysis, see methods section 8.4). The asterisks mark onsets of perturbation responses.
Many cells responded robustly to perturbations in both conditions (see Figure 17b). However, almost 20% of cells had a strong playback bias and small selectivity for distorted feedback ($b < -0.5$ and $|s| < 0.5$, $n=12$); these cells responded to playback perturbation, but did not respond to even high-intensity song perturbation (Figure 17c). About 10% of cells showed very high selectivity for distorted feedback ($s > 3$, $n=8$), many of which tended to be quite unresponsive to perturbed playback (see Figure 17d,e; and Figure 22a). Feedback perturbations predominantly induced firing increases ($s > 0$), though we observed firing suppression in a few cells (see Figure 21a, Figure 22b,c). Averaged over the time intervals in which responses to perturbed song were significant (30-ms moving window analysis), the firing rates in 29/67 neurons more than doubled (such as cell in Figure 22b) but in only 9/67 neurons they less than halved (such as in the cell shown in Figure 22c). In 11/67 neurons we observed a combination of firing increases and decreases, suggesting that underlying these responses were mixtures of synaptic inhibition and excitation. In both conditions, sound amplitudes during perturbation responses were significantly lower than average (Student’s t-test, $p < 10^{-10}$), in agreement with a monotonic relation between response selectivity $s$ and perturbation amplitude (see Figure 23).

Figure 17: Example perturbation responses. a: Bottom: Spectrogram of a song that was twice perturbed by a long-call stimulus (red shading). Song motifs are delimited by red boxes. Top: Extracellular voltage trace of a simultaneously recorded neuron (inset: spike burst). b: Neuron with increased firing during perturbation of song (top) and of BOS playback (bottom). Perturbing stimuli are indicated by red shaded areas: lower sound amplitudes by lighter shading. Average firing-rate curves are shown for cases without perturbation (blue line) and with perturbation (solid red line, dashed and dotted red lines are averages over trials with high and low perturbation amplitudes, respectively). Also shown are average spontaneous firing rate (dashed horizontal lines) and the times of significant perturbation responses (black horizontal bars). $b=0.0$, $s=0.42$. c: A neuron that responds to perturbations only in the playback, but not the vocal condition. $b=-0.95$, $s=0.0$. d,e: Two feedback-sensitive neurons in the same bird with selective (d: $b=0.38$, $s=1.38$) and highly selective (e: $b=0.38$, $s=7.11$) responses for song perturbations (e: same cell as in a).
Figure 18: Song perturbations. Spectrograms of 10 renditions of unperturbed song motifs (top) and perturbed song motifs (bottom). The perturbation stimulus is a long call. Perturbations did not affect the timing of song syllables produced thereafter.
Figure 19: Perturbations did not induce immediate changes in vocal output. Shown are sample song motifs from three birds (from top to bottom row: perturbed motifs original, perturbed motifs masked, unperturbed motifs masked, and unperturbed motifs original). Cross correlations of spectrograms were calculated in the regions marked by red shading (syllable gaps were excluded). Cross correlations between perturbed masked vs. unperturbed masked groups were not significantly different from cross correlations in the unperturbed masked group (unpaired t-test in three birds, P=0.88, P=0.12, P=0.29). Note that we excluded inter-syllable intervals from the analysis because perturbation stimuli could not be well masked in syllable gaps (trade off between threshold level and mask size).
Figure 20: Control for perturbation induced song amplitude increase. a: Perturbations did not lead to increases in sound-pressure levels (SPLs, or sound amplitudes) of underlying song (as in a Lombard-like effect). We delivered white-noise stimuli with zero spectral power in the frequency band from 4-6 kHz (indicated by red rectangles). Top: perturbed song; bottom: unperturbed song. SPLs of perturbed and unperturbed songs were not different from each other in a frequency window from 4.5-5.5 kHz (in which the broadband noise stimulus had zero power (n=2 birds, Wilcoxon rank sum test, P=0.8 and P=0.24). b: A potential Lombard-like effect would have on- and offset latencies. It would predict an increased sound amplitude level post perturbation. However, sound amplitude returns to unperturbed values within 5 ms post offset of the perturbation. Most of this delay can be explained by the standard deviation of detection time (roughly 3 ms).

To examine dependences of the response bias $b$ and selectivity $s$ on perturbation (sound) amplitudes, we occasionally presented the perturbing stimulus at sound pressure levels ±6 dB from normal levels (n=40 cells, Figure 17b and Figure 23). Excluding either the high or low-intensity trials led to only small changes in $b$ and $s$ (median absolute change in $b$: 0.16; median absolute change in $s$: 0.16).

We analyzed the timing of perturbation responses by their onset latencies and by testing for correlations with underlying song amplitudes. A correlation analysis revealed excesses of perturbation responses during syllable gaps and soft sounds: Across the entire neuronal population, song amplitudes averaged in 30 ms windows in which perturbation responses were significant, were smaller than amplitudes averaged in windows in which perturbations were delivered (Student’s t-test, $P<10^{-10}$). Because this observation applied to both song and playback conditions, we infer that the relative amplitude of perturbation stimuli modulated the strength of perturbation responses. Though highly significant, the modulation by perturbation amplitudes was not strong enough to predict onsets of perturbation responses from underlying song amplitudes. Namely, onset latencies of perturbation responses in both active and passive conditions were distributed over a wide range, starting just a few milliseconds after perturbation onset (Figure 21b). A wide range of onset latencies were also observed in individual birds (Figure 21b), demonstrating that response onsets were not a mere function of the relative amplitude of perturbation stimuli, but that other factors played a role as well such as ongoing neural activity. Hence, perturbation responses across the population of cells were not only diverse in terms of their sensitivity and state dependence, but also with respect to their timing, thus revealing a cell-specific temporal modulation of perturbation sensitivity.
In 6/8 cells with selectivity $s > 3$, peak firing rates during perturbed song were >3 times higher than peak firing rates during unperturbed song (in the remaining 2/8 cells peak firing rates were >40% higher). Thus, responses to perturbed feedback could largely exceed all of unperturbed singing-related activity, suggesting that high selectivity for distorted auditory feedback derives from a precisely timed and strongly coherent synaptic drive. We also explored whether high selectivity was associated with suppression of neural activity during self-initiated vocalizations, because such suppression is a common gain-control mechanism found in the auditory brain areas of animals as diverse as crickets [Poulet and Hedwig 2006], bats [Marler and Tamura 1964], and marmosets [Eliades and Wang 2003]. The baseline firing in highly selective cells ($s > 3$) was lower than for all other cells (5.7 Hz versus 19.5 Hz, $p=0.028$, Figure 24a), nevertheless, highly selective cells were relatively more suppressed during singing (WRS test of equal median Z scores, $p=0.005$, Figure 24b).
Figure 22: Spike raster plots of three additional example neurons. (Top: song spectrograms, Bottom: firing rate curves). a: A highly perturbation-selective neuron that is inhibited during unperturbed song, but vigorously bursts during most perturbed motifs with a burst of spikes. The neuron spikes irregularly in response to playback, and does not respond to perturbed playback. Response bias $b=0.60$, response selectivity $s = 9.69$, $z$-score song = -0.57, $z$-score playback = -0.04. b: Highly perturbation-selective neuron that is inhibited during song (top) but responds to playback (bottom), $s=6.2$. The firing behaviour of this cell qualitatively agrees with $e=|s-c|^2$ (where $s$ is the sensory signal and $c$ is the corollary discharge, see Figure 47): it fires during playback ($c=0$), it is suppressed during song ($s=c$), but fires during perturbations (where $s \neq c$). Response bias $b=-0.24$, $z$-score song = -1.36, $z$-score playback = 0.36. c: Neuron that is excited by song and BOS playback (baseline firing almost zero) and responds to perturbations with a firing decrease. Response selectivity $s=-0.54$, response bias $b=0.04$, $z$-score song = 3.21, $z$-score playback = 2.80.

In conclusion, field L and CLM responses equally emphasize the importance of self-generated and external auditory inputs, as evidenced by similarity of average firing rates in active and passive conditions (Figure 13b and Figure 15), by the uniformly distributed bias index (Figure 21a), and by similarity of onset latency curves (Figure 21b). Hence, the auditory forebrain seems to form an invariant representation of actively and passively perceived songs for integrating and comparing auditory feedback with the songs of other birds. Dissimilarities between active and passive sound processing were evident in terms of a motor-specific drive. Consequently, some neurons showed singing-related activity that resembled playback-evoked activity but was insensitive to perturbed feedback. Such behaviour is reminiscent of auditory-vocal mirroring reported in HVC [Prather et al. 2008] neurons and could arise from corollary discharges elicited by an efference copy of motor commands. On the other hand, neurons that were largely quiescent during singing, except when the auditory signal was perturbed, are reminiscent of some neurons in primate auditory cortex that strongly respond to frequency-shifted auditory feedback [Eliades and Wang 2008]. Vocal-mirror spikes in playback-biased cells could contribute to the generation of high perturbation selective responses, provided that such spikes are able to counterbalance precisely the excitatory drive elicited by sensory feedback. We find some evidence for such counterbalancing in perturbation-selective neurons in terms of their relatively strong firing suppression during song. However, suppression was rare overall, unlike in monkey auditory cortex [Eliades and Wang 2003] and in an auditory ganglion of crickets [Poulet and Hedwig 2006] (note that in crickets, responses are not perturbation selective despite this suppression).
Relatively few cells specialized into highly selective perturbation detectors, yet their mere existence suggests that auditory feedback is analyzed in the auditory forebrain with reference to an internal model (see Figure 47). For example, vocal mirror responses could represent predicted auditory feedback, which helps the bird to generate a stable perception of its song in the midst of a noisy colony. Accordingly, highly feedback-sensitive responses would reflect prediction errors of auditory feedback; such errors could signal song disruptions or simplify vocal learning according to some forward-model theories [Jordan and Rumelhart 1992b]. Alternatively, given that the birds in our study were in the process of learning a tutor song, vocal mirroring could constitute online replay of the tutor memory, as evidenced by similar firing stereotypies during song and during playback of a single song template (Figure 13f). Similarity of vocal and playback-related firing could thus be a reflection of a good match between the actual song and the memorized auditory template, the latter of which may feed into CLM and field L via the caudomedial nidopallium (NCM) [Phan et al. 2006; London and Clayton 2008]. According to such a template-replay interpretation, responses in highly perturbation-selective neurons would represent performance errors signalling the dissimilarity between the perturbed song and the tutor memory, a property with obvious benefits for song learning [Jordan and Rumelhart 1992b; Fiete et al. 2007b]. The ability of birds to correlate perturbations with subtle motor variability [Tumer and Brainard 2007] suggests a functional connection between perturbation-selective neurons and premotor neurons, though a direct link between perturbation selectivity and song learning remains to be seen.

By design, acoustic perturbations transiently increased the variability of auditory feedback in a brief song-locked time window. Hypothetically, if auditory feedback is analyzed with respect to an internal model, then the drive provided by such a model to field L and CLM neurons could exhibit increased variability during perturbations (commensurate with the variability of auditory feedback). To test for this possibility, we analyzed the firing stereotypy in perturbation windows and compared it to stereotypy in equally-sized windows prior to perturbations. Perturbations of feedback did not evoke increases in spike variability: there was no difference between the coefficients of variation (CVs) of spike counts in these two windows (Wilcoxon rank sum test, P=0.23). Similarly, during unperturbed singing there was no difference in spike-count CVs measured in these two windows (P=0.30, of course no perturbation was delivered in the ‘perturbation window’ in these trials). Thus, for either vocal condition, the externally-imposed variability of auditory feedback did not produce a coincident
increase in firing variability. Hence, putative internal models display little trial-to-trial variability in spite of highly variable feedback.

Figure 24: Perturbation responsive neurons are usually quiescent. a: A scatter plot of response selectivity \( s \) versus baseline firing rate. The median firing rate in highly perturbation-selective neurons (\( s > 3 \), 5.7 Hz) was significantly lower than the baseline firing rate of neurons with \( s < 3 \) (19.5 Hz, Wilcoxon rank sum, \( P = 0.028 \)). b: A scatter plot of response selectivity \( s \) versus z-score in 67 neurons (black circles) indicates an increased tendency of suppressed firing during song in highly perturbation-selective neurons: The median z-score of neurons with \( s > 3 \) was -0.48, which is significantly smaller than the median z-score of neurons with \( s < 3 \), which was 1.1 (\( P = 0.005 \), Wilcoxon rank sum test).

3.3 Sound amplitude influences response amplitude but not response timing

The possible effect of sound amplitude and the potential confounds it could introduce especially when comparing singing-related to playback-related responses, merit separate discussion. There are several inherent differences between auditory processing of self-initiated vocalizations and processing of external stimuli. These differences mainly pertain to bone conduction and the middle ear reflex, both influencing auditory signals before they are processed by the cochlea. During singing, bone-conducted inputs have relatively high power at low frequencies < 1 kHz [Fukushima and Margoliash 2007]. The middle ear reflex, elicited during vocalization, acts to attenuate both bone and air-conducted inputs. The effects of bone-conduction and the middle-ear reflex cannot be removed when comparing song with playback responses. Also non-removable is that playback through a fixed loudspeaker to a freely moving bird cannot replicate the head-related transfer function (unless the loudspeaker is positioned right at the syrinx). Hence, there exist several caveats in such comparative experiments that are hard to overcome.

The perturbation experiments are less affected by these caveats. Because bone conduction is equal during both normal and perturbed feedback, its effects should be removed during the perturbed-baseline subtraction used in the analysis of perturbation responses. Hence, we infer that our perturbation analyses were unaffected by bone conduction and middle-ear reflexes.

Due to these inherent differences between active and passive auditory processing there is not one clear choice for the BOS playback amplitude. Even matching the sound amplitude of playback to the sound amplitude of song at the bird’s ear does not guarantee identical stimulus amplitudes due to bone conduction and middle-ear reflex. In our view the most reasonable choice is playing back BOS stimuli at physiological sound levels at the bird’s ear, rather than at levels of self-generated songs.
We estimate that sound amplitudes in vocal and playback conditions differed up to 30 dB (point-source assumption, 6 dB per distance doubling, the loudspeaker was roughly 20 cm away from the ear and the syrinx roughly 1 cm). By choosing physiological sound-playback levels, we could compare sound processing in the only two naturally occurring conditions (hearing oneself and hearing others). Given that the bird matches his own vocalizations to the ones of his tutor he hears at similar sound amplitude levels, we further speculate, that increasing playback amplitudes at the bird’s ear to those of self-generated sounds will leave most results unchanged. This idea is further supported by the following findings: First, many field L neurons are remarkably invariant to changes in sound amplitudes of even more than 30 dB [Billimoria et al. 2008; Nagel and Doupe 2008]. Second, even though playback and vocal conditions may have been sub-optimally matched, mean firing rates are not significantly different in vocal and playback conditions in most field L cells (see Figure 15), indicating that field L is in similar operating modes in both conditions.

**Figure 25: Effect of sound amplitude on response latency.** In this neuron singing-related responses had a lead of 5.5 ms on playback-related responses, thus slightly less than the average lead of 6.8 ms (same neuron as shown in Figure 13a). a: Top (green raster): Response to playback of a long call stimulus at low sound amplitude. Middle (red raster): Response to playback of the same stimulus at high sound amplitude (12 dB louder). Bottom: Overlay of the two firing rate curves. The dashed blue line is the high sound amplitude response shifted by 6.8 ms to the left. b: Close up of the raster plots (colors as in a). Blue raster plot again is a shifted version of the high sound amplitude response. Vertical grey lines are positions of peak response. Trials of high and low amplitude playbacks are aligned by peak cross correlation of sound stimuli. Alignment by threshold crossing would shift the high amplitude spikes (red) to the right (to even later times) by roughly 2.5 ms.

Another confound that needs to be ruled out, is a possible influence of sound amplitude on response timing. Given that the responses in a subset of field-L neurons varies significantly with changes in sound amplitude [Billimoria et al. 2008], it is conceivable that an increase in response amplitude is accompanied by a shift of peak response to earlier times. Especially when comparing response latencies between song and playback condition (e.g. as in a cross correlation analysis) we need to
exclude the possibility that the observed response lag during playback is simply due to the increased sound amplitude during song. In many cases such an effect can be ruled out to be the dominant contribution to the observed response lag, simply due to the fact that during song in roughly 20% of the neurons response onsets precede the onset of sound (see Figure 13e). However, to further investigate possible sound amplitude induced response shifts, we analyzed responses to playbacks of the perturbation stimulus at high (+ 6 dB) and low (- 6 dB) playback amplitudes (see Figure 25). We found that the 12 dB amplitude difference did not lead to any significant shift in the response (see Figure 25b), even when there was a considerable change in response amplitude. The responses of the cell shown in Figure 25 vary significantly with playback amplitude. Peak response at high playback amplitudes (+6 dB) is 192 Hz, peak response at low playback amplitude (- 6 dB) is 163 Hz. This response change corresponds to a slope of 2.4 Hz / dB, a value comparable to the response change of 3.0 Hz / dB of the example amplitude sensitive neuron shown in figure 1 of [Billimoria et al. 2008]. Thus even though this cell is sound-amplitude sensitive, time of peak response does not vary with changes in playback amplitude, and the observed anticipatory shift during song of 5.5 ms in this cell is not explicable by the higher sound amplitude of song.

3.4 Field L exhibits song locked responses in the deafened bird

The field L recordings in the deaf bird were intended as a proof of existence of pre-motor signals in field L. However, because there is a direct projection from the apical part of the hyperpallium (HA) to field L [Wild and Williams 1999; Wild and Williams 2000], we cannot rule out the possibility that activity in field L in the deaf bird is somatosensory-feedback-induced activity. Further complicating the matter is that the fiber tracts connecting Uva to HVC and NIf to HVC run through field L. Thus activity could also arise from axons of passage. In addition there are a number of problems inherent to the method of recording in a deaf bird that we could not solve to our satisfaction. When removing the cochlea by means of pulling it from the cochlear duct (see methods section 8.3 for details), one is bound to damage the lagena – a vestibular organ. Damage to the lagena leads to a mild form of opisthotonus, in which the bird assumes a posture in which his head is tilted backwards, typically such that the peak points upward at an angle of 45 - 90 degrees. In this position, the microdrive will continually be in contact with feathers, leading to large movement artifacts on the recordings, most likely induced by small electrostatic discharges. One could attempt (we did not) to implant the microdrive tilted forward. One would however, need to be very careful not to let the implant hinder the bird in eating and drinking. Repeated blows to the microdrive from eating- or drinking-related head movements will usually lead to a loss of the implant. Another possible solution would be to deafen a large number of birds, as opisthotonus occurs in most, but not all bilaterally deafened animals, and subsequently record in animals which retained a normal posture. This solution however does not adhere to the principle of the 3Rs [Russel et al. 1992] and is thus inapplicable.

Cochlear removal and subsequent electrophysiological recording in Field L with a microdrive was only performed successfully in one bird in a pilot study. Due to the above mentioned caveats and problems, we did not follow up on this initial experiment. For reasons of completeness we chose to present the data here, however in light of the possible confounds (somatosensory input, and fibers of passage) one should interpret the data with care.

27 The value of 3.0 Hz / dB was estimated from the data shown in the figure.
28 As was the case in “V1 recordings” – prior to Hubel and Wiesel – “evidencing” the center surround LGN-like receptive fields of V1 neurons [Hubel 1959], [Hubel and Wiesel 1959].
Figure 26: Song locked field L activity in a deaf bird. a: Anticipatory firing to the first introductory note of a song bout. Top: Spectrogram of one introductory note. Middle: The top raster plot (green) is song related activity. The bottom raster plot (blue) is a set of randomly chosen baseline intervals (intervals of no sound at least one second before or after any sound on the microphone). Bottom: Firing rate curves for song related activity and baseline activity. Average baseline activity is indicated by the dotted black line. Black horizontal lines indicate significant difference between baseline activity and song related activity. 

b: Same as in a, neural recordings are now aligned to the first syllable in the motif.

We were able to record singing-related activity from 14 neurons. Seven of these 14 neurons showed song-locked activity during singing (see Figure 26 and Figure 27). Song locked activity is defined as stereotyped activity that significantly deviates from baseline activity during, immediately before or immediately after song. Stereotypy is assessed by a human observer and no efforts were undertaken to rigorously quantify this, as stereotypy of firing across trials was either clearly present (as in the neuron shown in Figure 26) or clearly absent. Three of the seven neurons that had song locked activity patterns exhibited a motif offset response only (see Figure 27). These neurons were almost inactive for the duration of the song motif but fired at high rates after motif offset, reminiscent of rebound from inhibition. Post deafening this bird only sang song bouts containing one single motif,29 thus unfortunately we cannot distinguish between end of motif and end of song bout related activity.

In light of these results we can rule out one of the two possible confounds. Evidence against the possibility of the observed activity being that of fibers of passage comes from the fact that the activity does not resemble any known singing-related activity pattern of projection neurons30. We cannot however rule out the possibility that song-locked field L activity is somatosensory in nature. A

29 This has been a tendency in most of our deafening experiments. The fact that deaf birds displayed a propensity to terminate the song bout after one motif raises the suspicion that the bird does indeed rely on auditory feedback for certain weak syllable transitions. A transition from syllable A to syllable B is said to be strong if A is always followed by B; it is considered weak if A is also frequently followed by other syllables or if the motif frequently terminates after syllable A. If the motif has no repeated syllables, the transition from the last to the first syllable of the motif is thus the weakest transition.

30 Uva (Dmitriy Aronov, personal communication), NIf (see section “NIf activity of singing birds”), HVC [Hahnloser, et al. 2002] and RA [Leonardo and Fee 2005a] projection neurons all show highly stereotyped activity patterns, unlike the ones observed in field L of the deafened bird.
possible experiment to test this would be to bilaterally deafferentiate the syrinx in addition to deafening (as has been done before [Bottjer and Arnold 1984]). However, we refrained from attempting this experiment due not only to its difficulty (both procedures, deafening and syringeal deafferentiation, have high failure rates), but also due to the fact that the severity of the experiment does not, we are of opinion, justify the expected knowledge gain as activity anticipatory to onset of singing already is strong evidence of premotor activity in field L.

Figure 27: Offset response of a field L neuron in a deaf bird. Conventions as in Figure 26. This neuron is barely active during song but shows a significant motif (or bout) - offset response.
4 Neural correlates of song memories

Almost fifty years have passed since Mark Konishi’s groundbreaking discovery that songbirds are capable of forming a highly accurate long-term memory of the tutor song (also referred to as song template) [Konishi 1963]. Numerous efforts have made to identify this song memory, to this day however, only little is known about its location and its form. Part of the reason for this is, that neural correlates of memory are exceedingly hard to identify due to possible confounds with sensory activity. Detrimental effects of lesions of a given brain area on imitation accuracy for example, cannot be ruled out to have been caused by sensory deficits alone. Similarly, as it cannot be excluded that a memory of a sensory stimulus stored as a neural activity pattern is excitable by sensory stimulation, memory-related activity cannot easily be distinguished from purely sensory driven activity. In spite of these difficulties, there has been recent evidence that secondary auditory areas NCM and CLM may play key roles in the process of template formation [Gentner and Margoliash 2003; London and Clayton 2008]. Here we present further evidence that NCM might indeed be a storage site for the song memory. We recorded electrophysiological activity in NCM during singing and while the bird was listening to playbacks of his own song and of his tutor song. We find that in a subset of neurons, activity during singing is more similar to the activity during tutor song playback than to the activity during BOS playback. If confirmed, this could be the first direct electrophysiological correlate of the song template.

4.1 Playback-related NCM responses habituate to repeated stimulus presentation

NCM is a relatively large region encompassing most of the caudomedial pallium without well defined cytoarchitectonic boundaries [Mello and Clayton 1994]. A subset of NCM neurons has been classified as auditory based on electrophysiological responses [Bonke et al. 1979; Müller and Leppelsack 1985; Stripling et al. 1997; Amin et al. 2004]31. NCM also shows a strong increase in ZENK expression in response to auditory stimulation [Mello and Clayton 1994]. This is consistent with the fact that NCM receives input from auditory areas L3 [Fortune and Margoliash 1995], the shelf region of ovoidalis [Vates et al. 1996], and the caudal striatum (CSt) [Mello and Clayton 1994]. In addition, NCM is strongly interconnected with the caudomedial mesopallium [Vates et al. 1996], that has been suggested to be involved in higher order auditory processing and in auditory memory formation [Gentner and Margoliash 2003]. It was also shown that bilateral lesions in NCM have no effects on song structure [Nottebohm et al. 1976; Gobes and Bolhuis 2007]. Based on these findings NCM has been classified as an auditory area32.

There is increasing evidence that the NCM participates in the formation or the retrieval of auditory memories. A number of studies now have shown that NCM activity habituates to successive presentation of identical auditory stimuli [Chew et al. 1995; Stripling et al. 1997; Stripling et al. 2001;  

31 Müller and Leppelsack report to find 231 of 303 (76 %) NCM neurons that respond to auditory stimuli [Müller and Leppelsack 1985]. Note however that their definition of NCM includes L1, L3 and likely also L and L2b. Amin et al. report to find even less responsive units (54 %) in roughly similar areas [Amin, et al. 2004], recording however primarily in field L not in NCM. Nevertheless, a subset of their recording sites probably lies in NCM (see figure 1 of their paper).

32 Note that NCM fulfills all the criteria of an auditory area, as defined in “Note on the usage of the terms ‘sensory area’ and ‘motor area’”.
Phan et al. 2006) \(^{33}\) (see Figure 28). The observed habituation is extremely rapid (the largest response attenuation occurs between the first and second playback of the stimulus [Stripling et al. 1997]) and surprisingly strong (exemplary single unit data reduces spiking activity by a factor of 2 [Stripling et al. 1997], from RMS data we would predict an underlying change in spiking activity of a factor of 5, see Figure 28). Moreover, the rate of habituation depends significantly on the type of stimulus, and in juveniles is largest for conspecific song [Stripling et al. 2001]. It was even reported, that the rate of habituation correlates with the imitation accuracy of tutor song [Phan et al. 2006] (see Figure 28). Furthermore, if indeed involved in the storage of an auditory memory of tutor song one would expect changes in NCM activity during the sensory period of song learning. Comparing the activity in NCM in juveniles at 20 dph and 30 dph, Stripling et al. could show that even though electrophysiological responses did not change, ZENK activation is significantly modulated with age [Stripling et al. 2001]. Whereas ZENK induction in juveniles 20 dph is persistently high and shows no further increase in response to auditory stimulation, it is overall reduced and becomes inducible by auditory stimulation in juveniles at 30 dph.

**Figure 28: NCM response habituates with stimulus repetition.**

\(^{a}\) Figure is reproduced from [Phan et al. 2006]. (Top) Multi unit NCM response to song playback during the first and during the 25\(^{th}\) trial. (Bottom) RMS of neural response to a novel stimulus decreases with stimulus repetition by roughly a factor of 2 from the first to the 25\(^{th}\) trial. \(^{b}\) (Top) Simulated multi unit activity. The mean firing rate of a population of neurons (with spike amplitude inversely proportional to cell density) is increased from 0.2 Hz baseline to 10 Hz (top trace) and to 2 Hz (bottom trace) in the time window 1-2 seconds (simulated stimulus). (Bottom) Fold reduction of RMS as a function of percent activity relative to 10 Hz mean firing rate. The noise level of the simulation was chosen, such that the difference in RMS between 0.2 Hz and 10 Hz mean firing rate was a factor of 3 (this number was estimated from the figure shown in \(^{a}\)). A reduction of RMS by a factor of two would thus indicate a reduction of neural activity (number of spikes) by a factor of five to ten.

\(^{33}\) Note that all of this data comes from either multi unit recordings, or highly activated single units. Thus it is well possible that such adaptation is only observed in a subset of neurons that exhibit high firing rates. All studies cited also preselect their recording sites based on auditory responses.
In a template storage and recall-related brain area, one would expect to find gender differences similar to those found in the premotor areas of the song system [Nottebohm and Arnold 1976]34. And indeed, NCM is the only auditory forebrain region in which any gender differences have been described. ZENK expression levels in juvenile birds (30 dph) but not in adult birds, in response to conspecific song is higher in males than in females [Bailey and Wade 2003]. In addition, even though there are no apparent cytoarchitectonic gender differences [Saini and Leppelsack 1981], it was found that cells immunolabeled for the calcium-binding protein calbindin, primarily localized to caudal NCM are almost twice as numerous in males as in females [Pinaud et al. 2006]. Calbindin-positive cells constitute a subset of GABAergic neurons in NCM. It has been estimated that GABAergic neurons constitute roughly half of the neurons in NCM. GABAergic signal transmission has been shown to significantly contribute to the temporal pattern of song-evoked responses in NCM without significantly influencing frequency tuning properties [Pinaud et al. 2008]. Thus it is plausible that this difference in calbindin-positive cell density is the correlate of a gender-specific auditory processing capability directly related to song learning.

To investigate the role of NCM during singing and its possible involvement in song learning we recorded single unit activity35 in freely behaving juvenile (60-90 dph) zebra finches that were in the plastic phase of song learning. We compared this activity to the BOS and tutor song playback evoked activity. And similar to the experiments in field L (see chapter ) we perturbed the bird during singing with an auditory stimulus and compared these perturbation responses to those elicited by perturbing song playbacks.

4.2 Singing-related NCM activity cannot be predicted from passive playback-related responses

We find that responses in NCM are highly heterogeneous both in terms of baseline firing rates and playback and song responses. Some neurons are unresponsive in all conditions tested, others are strongly activated by both auditory stimuli and singing, confirming previous reports of heterogeneous playback responses [Bonke et al. 1979; Stripling et al. 1997]. Cells of high and low activity were not intermingled randomly, as was frequently observed in field L (unpublished observation), but seemed to be segregated into entire regions of high and low activity. This is reflective of the fact that ZENK expression in NCM in response to auditory stimulation is highly heterogeneous [Mello and Clayton 1994] and of the idea that NCM is composed of neurochemically-distinct sub-domains [Pinaud et al. 2006].

A substantial fraction of NCM neurons is activated during both singing and BOS playback. Activity in the two conditions was on average similar both in terms of Z scores and cross-correlation, with a slight lead during singing (see Figure 29 and Figure 35c). Consistent with previous reports [Chew et al. 1995; Chew et al. 1996; Stripling et al. 1997; Phan et al. 2006], we find that in a subset of NCM neurons, BOS-playback responses habituate rapidly with successive stimulus presentation (see Figure 30a). We also find that this habituation is rapidly and completely reset by intermittent intervals of

34 Even though females can learn to recognize songs, one would expect to find gender differences in auditory areas specifically related to song learning, as male birds need to recall this memory in a highly specific fashion during singing.

35 All data were recorded in the right hemisphere. This was motivated by the fact that motor control of song is right side dominant [Williams, et al. 1992] and that there are no reported differences between left and right NCM activity [Chew, et al. 1996]
singing (see Figure 30b). Moreover, singing-related activity is modulated with a habituation-opposing trend to increased activity with successive song bouts. Both of these effects are functions of song bout repetition number and not of time as judged by overall responses (see Figure 30c). On a finer time scale, both habituation-related decreases of activity during BOS playback and modulation-related increases of activity during singing occur between song bouts and not between motifs of the same song bout (see Figure 31).

To investigate auditory feedback processing in NCM, we played perturbation stimuli through a loudspeaker while the bird was singing and while it was listening to playbacks of its own song. Perturbation stimuli were time locked to a particular song syllable. We found that some NCM neurons were unresponsive to auditory perturbations during singing. On average, NCM neurons seemed less responsive to auditory perturbation during singing than neurons in field L. Thus, this suggests that a subset of NCM neurons encodes a correlate of an efference copy or a template replay.

We also found that NCM responses can be predictive of motor output. This is based on one recording site at which we were able to record stable multiunit activity for the duration of a few four hours. This multiunit showed significant and long lasting perturbation responses during singing (response differences were still significant more than 50 ms after the end of the perturbation stimulus, see Figure 32a). In a small number of motifs (16) the bird spontaneously stopped singing after the third syllable. Neural activity during these spontaneous stopping trials was significantly lower than the activity during completed motifs prior to the change in song (see Figure 32b). In addition, in trials in which the perturbation stimulus induced song stoppings (40) after the third syllable of the motif, neural activity was significantly higher than during perturbed completed motifs (see Figure 32c). This indicates that activity is at least in part driven by an efference copy of motor command or a template replay. It is also interesting to note that deviations from the activity pattern during unperturbed trials towards higher or lower firing rates appear to be indicative of song stopping (see Figure 32d).

The strongest evidence of a correlate of template replay comes from a subset of neurons whose activity during singing is more similar to the activity during tutor song playback than to the activity during BOS playback (see Figure 33, Figure 34 and Figure 35). Moreover, average motor lead of 2.1 ms on tutor song playback responses was smaller than the average lead of 9.6 ms on BOS playback responses (Figure 35c). These effects are trivially explained, if one assumes, that the activity observed during singing is driven in part by the replay of a sensory template of the tutor song.

On a more anecdotal note, we found one neuron that selectively responded to perturbations of BOS playback (see Figure 36). The neuron showed no significant activation during song, perturbed song, BOS playback or the PS stimulus presented in isolation (data not shown), implying that the response is highly non-linear. One simple model that would explain such a response is that of a deviation from expectation. Assuming that BOS (or tutor song) playback triggers a replay of the auditory memory of tutor song, responses selective for deviations from tutor memory could be generated in a similar manner to the mismatch selective responses observed in field L (as described in “Song-perturbation responses”). Such responses could in principle be a neural correlate of the phenomenon of mismatch negativity (MMN) [Näätänen et al. 1978]. MMN is a characteristic response in an EEG signal that occurs after a rare and unpredictable change in a series of repetitive sensory stimuli. It is also observed in subjects listening to a sentence read aloud in response to mispronounced words or grammatical errors. In this sense, such a response would be a correlate of surprise.
Figure 29: NCM activity during song and BOS playback are similar. (Top) Spectrogram of song motif. (Middle) Raster plot of song (blue) and BOS playback (red) related multi-unit activity. Trials are arranged chronologically from bottom to top. (Bottom) Average firing rate curves for song (blue) and BOS playback (red) trials.
Figure 30: Counteracting NCM response adaptations during song and BOS playback. a: (Top) Sound spectrogram of BOS. (Bottom) Traces of multiunit (most likely 3 neurons with S/N > 4) activity recorded in NCM during BOS playback at the beginning and the end of a playback session. b: Average MU activity (number of S/N > 4 spikes, averaged over the 600 ms duration of the motif) as a function of song (blue) and BOS playback (red) motif repetitions. Vertical dashed lines mark times when the bird sang during BOS playback. Red and blue circles indicate average baseline firing rate. c: As in b but now as a function of time. Same data as in Figure 29.

Figure 31: Activity adaptations occur between, not within song bouts. (Top) Schmatic of representation of two song bouts with the motif labels as used below. (Bottom) Plots of the difference of activity between the second to last and the last motif of song bout A (b-a), the difference of activity between the last motif of song bout A and the first motif of song bout B (c-b) and the difference of activity between the first and the second motif of the song bout B (d-c) for all song bout transitions in BOS playback (left) and song (right) trials. p-values (Student’s t-test) for deviations from mean 0 are indicated above the bar plots.
Figure 32: NCM activity can be predictive of song changes. a: Multi unit recording in NCM, voltage signal is thresholded at four times the signal standard deviation, every event that crosses threshold is considered a spike. (Top) Example spectrogram and raster plot from song trials without perturbation, (Middle) song trials with perturbation and (Bottom) average firing rates of unperturbed (black) and perturbed song trials (red). Note that for clarity only a subset of the data is shown in the raster plots. Significant difference is indicated by the horizontal red bars. Gray shading marks region where song is not stereotyped. Numbers indicate the number of song trials in the unperturbed (220) and the perturbed (277) condition. b: Conventions as in a, but now for unperturbed (black) and unperturbed stopped (blue) trials. Light blue shading marks regions of significant difference between the average spectrograms of the two conditions. Note changes in firing rate occur prior to changes in song. c: Conventions as in a, but now for perturbed (red) and perturbed stopped trials (green). Light green shading indicates stopping times. Song was either truncated in the middle of the second syllable (at time 125 ms) or after the second syllable (at time 225 ms). Note that changes in firing rate occur prior to song stopping. d: Average firing rates during unperturbed song trials (black), unperturbed stopping trials (blue) and perturbed stopping trials (green). Note that deviations from normal firing pattern are predictive of changes in song.
Figure 33: Similarity of singing and tutor song playback related activity, example neuron I. a: (Top) spectrogram of song, (Middle) raster plot of single unit NCM response, (Bottom) average firing rate curve (blue line). b: as in a, but for BOS playback responses. c: as in a, but for tutor playback responses. d: overlay of the firing rate curves in all three conditions. Note that tutor song playback-related activity slightly leads BOS playback-related activity.

Figure 34: Similarity of singing and tutor song playback related activity, example neuron II. Conventions as in Figure 33. This neuron shows almost no response during BOS playback, but a very strong response during the last syllable of tutor song playback.
Figure 35: Similarity of song-related and tutor song playback-related responses. a: Median response to tutor song playback (TUT) is higher than the median response to BOS playback (Signed rank test, p=0.008). The red data point corresponds to the neuron shown in Figure 33, and the green data point to the neuron shown in Figure 34. b: In 11 of 16 neurons the peak cross covariance between song-related and tutor song playback-related activity is higher than the cross covariance between song-related and BOS playback-related activity. Peak was cross covariance was determined as the maximum cross covariance for time lags in the window $-30$ ms to $+30$ ms. c: Average cross-covariance between song-related and BOS playback-related spikes (red), and song-related and tutor song playback-related spikes (black). The average lead of song spikes on tutor playback spikes is 2.1 ms, the average lead of song spikes on BOS playback spikes is 9.6 ms.

Figure 36: Surprise neuron. Conventions as in Figure 22. a: Song (top) and BOS playback (bottom) responses of a neuron that selectively responds to perturbations of BOS playback. b: Even though the neuron selectively responds to perturbations of BOS playback, it shows almost no response to the playback of the PS in isolation.
4.3 NCM lesion during song development impair imitation accuracy

To assess the influence of NCM during song development, we lesioned a relatively large area of the auditory pallium, encompassing parts of paraHVC and parts of NCM. In adults such lesions have been shown to have no effect on song production [Nottebohm et al. 1976; Gobes and Bolhuis 2007]. In juveniles we find however, that such lesions can have drastic effects on adult song and imitation accuracy.

Bilateral electrolytic lesions of the medial auditory pallium of two juvenile zebra finches (36 dph) were made based on stereotactic coordinates (see Table 2 for a list of lesion coordinates). Prior to the experiments birds were housed normally in a clutch with a tutor. After lesioning birds were transferred to isolation chambers for song recordings. A third control bird without lesion from the same batch was transferred to an isolation chamber at the same time. Aside from infrequent intermissions, songs were recorded continuously up to an age of 100 dph. Animals were then all returned to a social cage, and housed with other male birds but not the tutor 36.

All analysis was performed by visual inspection of the song spectrograms and no attempt was made to quantify the effects. During the post-lesion subsong phase or the early plastic song phase no difference between lesioned and control animals was observed. Overall song structure, song variability and singing frequency was unchanged in lesioned animals. In the late plastic song phase (75-90 dph) however differences between songs of the two experimental groups began to emerge. The songs of the lesioned animals remained variable and showed little tendency to become more stereotyped. Variability was especially apparent in syllables that contained harmonic stacks (see Figure 37). Syllable sequence stereotypy also showed a tendency to increase less rapidly in lesioned animals. The songs of both lesioned animals had still not crystallized at the end of the continuous recording phase (100 dph). All birds were subsequently recorded again at 160 dph. At this time all songs appeared to have crystallized. The tutor song imitations of the lesioned birds however were less accurate than in controls (see Figure 37, as a reference, the control bird achieved an almost perfect imitation of the tutor song).

Note that the data presented here is preliminary in nature due to the fact that the lesion experiment was only performed in two birds. We did perform a second similar experiment also on three siblings with the same tutor. Lesions also encompassed parts of paraHVC and NCM. The extent of the lesions however was significantly smaller (2 lesion sites instead of 13, at similar current and duration settings) and lesions were made later in development (50 dph instead of 36 dph). Nevertheless, similar effects, albeit much less strong, were observed in these birds (data not shown).

Based on these experiments we can conclude that NCM and paraHVC play a crucial role during song development. The observed deficits in song learning could be caused either by a loss of tutor song memory or simply by an impairment of auditory processing system. Unfortunately both the absence of a suitable tutor during development [Eales 1985; Eales 1987; Morrison and Nottebohm 1993] and impaired auditory feedback [Funabiki and Konishi 2003] can lead to delayed crystallization. The experiments of Gobes and Bolhuis however, that have shown that bilateral NCM lesion in adult males significantly reduce tutor song preference, without affecting the birds’ preference for female calls, would argue for the idea that the observed song deficits are caused by a loss of tutor memory and not merely by damage to the auditory processing system [Gobes and Bolhuis 2007].

36 For all practical purposes the control animal could have functioned as a tutor for the lesioned animals.
Figure 37: Song development after NCM lesion at 36 dph. (Top) Song development of a bird that received an extensive bilateral electrolytic lesion to NCM/pHVC (coordinates as listed in Table 2). (Middle) Control bird. (Bottom) NCM/pHVC lesion bird, as in (Top). All three birds are brothers from the same batch and had the same tutor. Note the apparent fuzziness of the 160 dph spectrogram is artificially caused by a lower sampling rate of the sound recording.

Table 2: Coordinates of electrolytic lesions (20 seconds @ 20 uA) in the birds shown in Figure 37 (Top and bottom).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>0,50</td>
<td>0,50</td>
<td>0,50</td>
<td>1,00</td>
<td>1,00</td>
<td>1,50</td>
<td>1,50</td>
<td>1,00</td>
<td>1,00</td>
<td>1,50</td>
<td>1,50</td>
<td>1,50</td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>0,50</td>
<td>0,50</td>
<td>0,50</td>
<td>0,50</td>
<td>0,50</td>
<td>0,50</td>
<td>0,50</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td></td>
</tr>
<tr>
<td>DV</td>
<td>0,50</td>
<td>1,00</td>
<td>1,50</td>
<td>0,50</td>
<td>1,00</td>
<td>1,50</td>
<td>0,50</td>
<td>1,00</td>
<td>1,50</td>
<td>0,50</td>
<td>1,00</td>
<td>0,50</td>
<td>1,00</td>
</tr>
</tbody>
</table>
5 Sensorimotor Integration

Sensorimotor learning can be described by two interacting systems, the motor system generating movement, and the sensory system processing input and feedback. In order for sensory feedback to be able to exert influence on motor output, as is required if motor output is to follow a specific pattern, there needs to be a site of interaction between the two systems. This interaction is generally referred to as sensorimotor integration. The identification of such a site of sensorimotor integration in the songbird song system is a long standing and unsolved problem. By definition, a candidate brain area, or a candidate set of brain areas, would need to display two general characteristics:

i) motor-related activity during singing

ii) auditory-perturbation-induced changes in singing-related activity patterns

The former need be true by definition, the latter to signal mismatch (or error) between actual and desired feedback. There are three possible frameworks of sensorimotor integration that all entail slightly different anatomical constraints.

In an online learning model in which feedback mismatch leads to immediate activity-induced (e.g. via spike-timing-dependent plasticity) changes of synaptic strength in the motor system, there must be a subset of neurons encoding both motor- and mismatch-related signals. That is, on a population level we would expect to find at least one brain area containing evidence for both types of signals.

A second possibility would be a neuromodulator (e.g. dopamine) induced change of synaptic strength. In this scenario auditory input would drive neuromodulatory neurons that in turn exhibit mismatch sensitivity. Thus this scenario would require no direct connections between auditory and premotor brain areas. Furthermore, although not strictly necessary it would be of significant advantage if neuromodulatory neurons were mismatch responsive during song only; initial evidence seems to contradict this. Zebra finch Area X projecting VTA neurons do show auditory responses to unexpected perturbation stimuli, but do so both during and outside of song [Liora Las, personal communication].

A third possibility would be an offline learning model in which a trace of a feedback or a mismatch signal is stored in memory and is later used to change the motor system. Here motor- and mismatch-related signals need not be present in the same neurons. Thus it is conceivable that the two signals are divided across two different brain areas, one of these areas we would classify as premotor, the other as auditory. Note however, that this interpretation is somewhat problematic because it requires the existence not only of a highly variable and accurate memory system storing time and type of mismatch, but also of a mechanism of transfer of this mismatch information to the motor system.

It has generally been assumed that an online learning model is the most likely scenario. HVC displays premotor-related activity [McCasland 1987; Yu and Margoliash 1996; Hahnloser et al. 2002] and its BOS-selective playback responses in anesthetized and sleeping animals [Margoliash 1983; Margoliash and Konishi 1985; Cardin and Schmidt 2003]. As such it is believed to be an ideal candidate for a site

37 Most of the data presented in this chapter is intended for publication in [Naie, et al. 2009].
of sensorimotor integration. Recent evidence supports this hypothesis. Prather and Mooney were able to show that in swamp sparrows, song-related HVC activity in area X projecting neurons is mirrored during BOS playback [Prather et al. 2008]. Shortly thereafter Sakata and Brainard found feedback-perturbation induced changes in song-related HVC activity in Bengalese finches [Sakata and Brainard 2008]. This came as a surprise, given that previous studies did not find any evidence for such perturbation induced changes in zebra finches [Leonardo 2004; Kozhevnikov and Fee 2007] and swamp sparrows [Prather et al. 2008].

It has been shown that NiF and CLM are necessary for auditory responses in HVC [Cardin and Schmidt 2004a; Cardin et al. 2005; Bauer et al. 2008]. Thus it is likely, although not necessary, that perturbation responses in HVC are mediated via at least one of these areas. We were able to show that CLM shows perturbation selective responses during song (see section 3.2). However, we were not able to determine where these neurons projected.

**Figure 38: Loss of song rhythm during NiF inactivation in juveniles.** a: NiF inactivation broadened the distribution of syllable duration: In the same bird, syllable-duration distribution of song produced during the subsong phase (49 days, blue line), during the plastic song phase (80 days, black line), and during NiF inactivation (80 days, red line). The red boxes enclose renditions of the emerging stereotyped song motif. The sound amplitude is indicated by the blue line. The rhythm function just below (black line) was constructed by thresholding (red line) and mapping the sound amplitude to values of zero and one. c: Rhythm spectrum of song at 49 dph and at 80 dph before and after NiF inactivation. The rhythm spectrum is generated by Fourier transforming the rhythm function. Stereotyped syllable structure emerges as vertical lines in the rhythm spectrum (see [Saar and Mitra 2008] for details). NiF inactivation (black arrow) flattened the rhythm spectrum towards the rhythm spectrum observed during the subsong phase (49 dph). Figure is reproduced from [Naie et al. 2009]

Multiunit recordings in adult zebra finches have shown that NiF activity precedes the onset of introductory notes and calls [McCasland 1987]. Aside from this, only little is known about song-related activity in NiF and its contribution to song learning and song production. There is evidence from lesion and inactivation studies that NiF might not be involved in song learning or production at all. It was shown for example, that in the adult zebra finch NiF lesions do not impair song production
66

It was also shown, that bilateral excitotoxic NIf lesions during early development (45 - 50 dph) do not impair song learning [Gardner and Fee 2007]. In contrast we were able to show recently that transient NIf inactivation during the plastic song phase leads to an almost complete loss of song structure (see Figure 38). NIf has the additional advantage that HVC-projecting neurons are easily identifiable by antidromic stimulation in HVC. Thus, to further investigate the exact nature of NIf's premotor role and to test if NIf is perturbation responsive, we recorded single- and multiunit data in freely behaving juvenile zebra finches (n= 3).

5.1 NIf activity of singing birds

NIf was localized and identified using antidromic stimulation in HVC as previously described [Hahnloser et al. 2002]. Neurons were classified as either NIf interneurons (NIfI) or NIf HVC-projection neurons (NIfHVC) by near threshold stimulation response latency and variability as described in [Hahnloser et al. 2006]. We recorded the activity of 13 NIfHVC neurons (8 of the 13 neurons were identified by antidromic spike response, the remaining 5 by firing pattern, see [Hahnloser and Fee 2007]), we also recorded from 3 NIf during song. In many recordings with identifiable single-unit activity we also simultaneously recorded multiunit activity (see Figure 39a). Single-unit activity was analyzed in terms of spike times and multiunit activity was analyzed by thresholding the root-mean square (RMS) of the voltage trace (see methods section 8.4 for details).

We found that NIf neurons were spontaneously active in the non-singing bird and most neurons increased their firing during singing (14 of 16, see Figure 39a and Figure 40c). In agreement with McCasland’s findings [McCasland 1987], we find that the activity of most NIf neurons deviates significantly from baseline prior to the first introductory note of a song bout (11 of 16, see Figure 40a). Both single and multiunit activity displayed rhythmic patterns that were quite stereotyped over the time course of song motifs (see Figure 39b,c). The firing pattern of single NIfHVC neurons was time locked to song syllables and consisted to a large part of high frequency (100 – 600 Hz) bursts and to a lesser part of noisier time windows of low firing rate activity. By visual inspection, the stereotypy and precision of NIfHVC activity appeared to be considerably lower than that of RA and HVC projection neurons in adult birds. We analyzed the time jitter of 10 isolated bursts (1 to 3 different bursts in 6 different NIfHVC neurons) and found a RMS burst onset jitter of 2.92 ± 0.76 ms std. This average jitter was larger than that reported in RA neurons (1.04 ms) [Chi and Margoliash 2001] and in HVC projection neurons (0.66 ms) [Hahnloser and Fee 2007]. Note that to estimate spike jitter, we discounted for significant variability of tempo between different renditions of the motif (see Figure 39c) by alignment of spike trains with the nearest syllable onset.

5.1.1 NIf spike patterns specifically precede song syllables

We investigated whether NIf activity could be classified as pre-vocal (leading on motor output) or as post-vocal (such as when activity reflects auditory feedback). To this end, we performed a fine-structured analysis of song patterns. One bird (motif and data shown in Figure 39c) sang a motif that contained a repeated syllable (labeled C). On both occasions, NIf activity was nearly identical before, but not after each syllable C, suggesting that NIf spikes correlate more strongly with following, rather than preceding, sounds. Furthermore, this bird sang a syllable (labeled B2) that strongly resembled

38 Note however, that in Bengalese finches bilateral NIf lesion leads to a significant increase in syllable sequence stereotypy [Hosino and Okanoya 2000]. And NIf lesions in adult zebra finches lead to long-term deficits in song [Coleman, et al. 2008].

39 Interestingly, NIf lesions does however lead to highly aberrant long distance calls [Gardner and Fee 2007].
syllable B. NIf activity was highly similar prior to both syllable onsets, suggesting that NIf activity is not only premotor, but also indicative of the spectral composition of the subsequent syllable. From the two neurons shown in Figure 39b we were able to record activity during over 150 renditions of the quasi-crystallized song motif. In roughly 4% of the renditions, the bird truncated the song motif by dropping the last syllable. Both neurons fired at least one spike within the 50 ms following the second last syllable in all renditions of the complete motif (n=164, Figure 39b top; n=165, Figure 39b bottom; in both cases only a subset of all complete renditions is shown). However, in the truncated motifs, both neurons remained virtually silent in the same 50 ms window (top neuron: no spike in all six truncated renditions, bottom neuron: one spike in one of the nine truncated renditions). Hence, NIf activity following the second last syllable was predictive of the last syllable and can thus be classified as premotor.

To estimate the time lag at which vocal output followed NIf_{HVC} spikes, we calculated cross-correlation functions between song rhythm functions and NIf_{HVC} spike trains. We found that cross-correlation functions exhibited significant peaks at time lags in the range -50 to -40 ms (p<0.01, n=13 NIf_{HVC} cells in 3 birds), indicating that NIf_{HVC} spikes on average preceded song syllables by roughly 45 ms. We obtained similar estimated time lags between NIf activity and song syllables when computing cross-correlation functions between song rhythm functions and the RMS of the voltage trace of NIf multiunit signals: these cross-correlation function also peaked in the range -50 to -40 ms relative to syllables (p<0.01, n=6 multiunit sites, Figure 39b,d).

NIf activity appeared to not only be anticipatory to vocal output, but also to be predictive of spectral content. To demonstrate this, we combined the sequentially recorded spike trains of neurons recorded in the same bird into pseudo-population activity that would have resulted had we recorded the neurons simultaneously. Then we extracted the 50 ms windows of this pseudo-population activity prior to syllable onsets and trained a linear classifier to predict the subsequent syllable. We evaluated the performance of the classifier on test data and found correct prediction of the subsequent syllable in over 90% of test cases. Because all song motifs consisted of more than 4 syllables, this fraction of correct prediction was much higher than the chance level (25% or less with 4 or more syllables). Thus, HVC neurons may in principle read out the identity of the upcoming song element from just a few NIf_{HVC} spikes.

5.1.2 NIf spike activity is not suited to conveying auditory feedback to HVC

In the anaesthetised zebra finch, NIf is a major source of auditory input to HVC [Cardin et al. 2005; Bauer et al. 2008]. To test if NIf could convey auditory feedback to HVC during singing, we perturbed the bird with the playback of an auditory stimulus time-locked to the song motif (see Figure 39b, as described in [Keller and Hahnloser 2009]). NIf neurons did not respond to even high amplitude perturbations during song, making it unlikely that NIf conveys auditory input to HVC during singing. Given that NIf input to HVC appears to be essential for plastic song, and that HVC neurons are not sensitive to perturbations during this phase [Kozhevnikov and Fee 2007], it is not surprising to find no auditory sensitivity in NIf projection neurons during singing.
Figure 39: NIf premotor activity in juveniles. a: Top, sound waveform. Bottom, raw extracellular trace in NIf, the shaded orange area indicates the time intervals during which the root-mean square (RMS) voltage exceeds a threshold of one standard deviation. Insets: close ups of spike bursts and multiunit signals. b: Activity of two NIfHVC neurons. The spectral derivatives of one sample song motif are shown on top. The black arrow indicates the alignment point for the raster plot below. Top: Spike raster plot of the activity of one NIf neuron during 63 unperturbed and 38 perturbed (perturbations are indicated by red shading in the raster plot, mean perturbation on- and offset are indicated by vertical dashed red lines) song motifs. The mean firing rate curves for unperturbed (blue) and perturbed (red) trials are shown below. Note that there is no significant difference between the two firing rate curves. Bottom: Spike raster plot of the activity of another NIf neuron during 65 complete motifs and 7 renditions of the motif where the bird truncated the motif by dropping the last syllable (below the black line). Spiking probability for this cell within the 50 ms window before the last syllable (delimited by the vertical black dashed lines) is 1 (165/165 - not all data is shown in the raster plot) for complete motifs and 0.11 (1/9) for truncated motifs. Underlying light gray shading depicts the binarized RMS voltage (Multi unit activity). The mean firing rate curves for complete (blue) and truncated (red) trials are shown below. Horizontal black bar identifies times of significant difference between the two. Due to technical difficulties, we were unable to antidromically identify the two neurons displayed here, but classified them as NIfHVC neurons based on firing pattern (for spike train statistics of two different NIf neuron classes see [Hahnloser and Fee 2007]). c: Activity of 6 NIf neurons in a
second bird. The top four neurons are antidromically identified NIf\textsubscript{HVC} neurons, the lower two are identified NIf neurons. The rhythm function (light green shading) illustrates the variability in song tempo across renditions. Note that syllables B and C are repeated and that activity patterns prior to both renditions each are highly similar. d: Cross correlations between spike trains and rhythm function (thick lines) and multiunit signal and rhythm function (thin lines) of the two neurons shown in b.

![Figure 40: NIf premotor and auditory responses](image)

a: Cumulative response onset probability distribution (as in Figure 13e) aligned to onset of first syllable of song bout, during song trials (black line) and BOS playback trials (red line). b: Scatter plot of Z score of song trials versus Z score of BOS playback trials. Gray shading indicates areas of Z scores that are not significantly different from zero (Z score < |0.75|). Open circles are putative HVC projection neurons and filled circles are putative interneurons. c: Scatter plot of Z score versus the average inter trial spike train cross correlation normalized by the baseline cross correlation (C score) of BOS playback trials (red circles) and song trials (black circles). Note that song-related activity patterns of all but one neuron are highly stereotyped, whereas BOS-playback-related responses, even in responsive neurons with z-score>0.75, are typically of low stereotypy. Open and filled circles as in b.

5.1.3 **No significant differences exist between NIf activity in juveniles and adults**

To test whether NIf activity in adults differed strongly from that in juveniles, we recorded from NIf neurons in two singing adult birds. In the adult, we found that NIf neurons similarly fired high frequency bursts during singing as we had observed in juveniles. Adult burst onsets were slightly more precise than in juveniles (RMS burst onset jitter 1.92 ± 0.52 ms std, n=8 bursts, compared to 2.92 ± 0.76 ms std, n=10 bursts in juveniles), though the jitter difference did not reach significance (p=0.06). Adult NIf neurons also fired anticipatory to vocal output: cross-correlation functions between sound amplitude and NIf spike trains exhibited peaks in the interval 20 to 50 ms after NIf spikes (n=3 neurons). We conclude that adult and juvenile NIf neurons fired in similar premotor manners, and that overall, firing patterns in NIf projection neurons appeared to be more variable than those of HVC and RA projection neurons.

5.1.4 **Discussion**

We found spiking activity in NIf neurons to be unsuitable for conveying auditory feedback to HVC. Rather, our electrophysiological recordings have revealed a motor code predictive of vocal output in NIf. Previous studies in adults have demonstrated that motor-related multiunit activity in NIf leads introductory notes and distance calls by several milliseconds [McCasland 1987; Cardin and Schmidt 2004a]. Our single-unit recordings have shown that in both juveniles and adults, NIf neurons fired distinct spike patterns before each syllable and some sub-syllables. Spiking activity in only few NIf projection neurons already contained enough information to predict the subsequent song element, thereby allowing HVC in principle to read out the identity of the upcoming song element.
Our correlation analysis revealed that activity in NIf projection neurons leads vocal output by roughly 40 to 50 ms. Similar estimates were previously obtained for motor-related activity in RA and HVC [McCasland and Konishi 1981; Schmidt 2003; Kozhevnikov and Fee 2007]. However, unlike in the present study, extensive recordings from HVC and RA neurons have provided only weak support of anticipatory spiking in RA and HVC neurons [Hahnloser et al. 2002; Leonardo and Fee 2005b; Kozhevnikov and Fee 2007]. By contrast, the NIf projection neurons in our study fired robustly in 20 to 50 ms anticipation of song syllables, which to our knowledge is the most striking premotor correlate of song to date.

Syllable trigger signals are believed to originate in coupled respiratory-vocal networks in the brainstem [Wild et al. 1998; Ashmore et al. 2008]. From there, trigger signals are relayed to HVC via Uva\textsuperscript{40}. Our findings suggest that the direct pathway from Uva to HVC provides the primary trigger signals to HVC and that the indirect pathway via NIf provides secondary trigger signals that may interfere constructively or destructively with the primary triggers. If we assume that HVC relies on several afferent trigger signals but provides the remaining premotor commands for song syllables, then the lower onset precision of NIf\textsubscript{HVC} bursts compared to bursts in HVC and RA projection neurons can explain the increased behavioral elasticity of gap durations compared to syllable durations [Glaze and Troyer 2006]. The behavioral importance of precise control of phrase-level complexity in birdsong may have facilitated the evolution of such parallel neural pathways controlling sequence variability.

\textsuperscript{40} Multiunit Uva activity shows bursts that correlate with on- and offsets of syllables and some notes (Dmitriy Aronov, personal communication).
6 BEHAVIORAL EXPERIMENTS

This chapter contains a brief summary of two behavioral experiments we have conducted, the results of which pertain directly to our understanding of the limits of a zebra finches memory storage capability. The data from these experiments have allowed us to substantially refine our speculations in the discussion of the matter.

6.1 Come again…? - or, how many repetitions form a template

A question that naturally arises in the context of song imitation learning is how often does the juvenile bird need to hear the tutor motif to be able to form a template and to then accurately imitate the tutor song? Different studies have addressed this question and have shown that a bird can be successfully tutored with relatively few repetitions of the tutor motif. Petrinovich, for example, finds two birds that were able to copy tutor song with a mere 252 repetitions of the motif, he also finds that none of the birds tutored with less than 120 repetitions were able to imitated tutor song [Petrinovich 1985]. Tchernichovski et al. report to have successfully tutored birds with 10 playbacks of the tutor motif per day starting at 30 dph [Tchernichovski et al. 1999]. This would imply that roughly 250 motifs are sufficient, given that birds usually show the first recognizable imitation of the tutor motif at around 55 dph. Overall this suggests that on the order of 200-300 repetitions of the motif distributed over the course of 25 days (30-55 dph) is enough auditory input, for the juvenile bird to form a template of, and later faithfully imitate, tutor song.

![Figure 41: Model overabundance impedes learning.](image)

As it turns out, there appears to be not only a lower but also an upper limit on the ideal number of tutor motif repetitions. In a first series of experiments Tchernichovski and Nottebohm [Tchernichovski and Nottebohm 1998] found that the accuracy with which birds copied their tutor showed a negative correlation with the number of male offspring in the clutch. In other words, the more male siblings a bird has the worse his imitation of tutor song will be (see Figure 41a). They find
consistently that the first bird to produce a recognizable copy of the tutor song, independent of the order of hatching, achieves the highest imitation accuracy.

They explain this effect by making the bold assumption that an overabundance of the model leads to a decrease in imitation accuracy. To confirm this hypothesis they conduct a second series of experiments [Tchernichovski et al. 1999] in which they tape-tutor birds, using an experimental setup where birds can trigger tutor song playback by pecking a key. Different birds are assigned different maximum number of motif playbacks per day. They find that the more often the bird has access to the tutor song playback the less accurate his imitation is (see Figure 41b). At first thought this stands in stark contrast with their initial results, which suggested that a bird raised in isolation with a live tutor will achieve near perfect imitation, considering that a live tutor will sing a few thousand motifs per day.

6.2 Model over-abundance reduces imitation accuracy

In the wild the juvenile bird will leave the nest between 20 to 30 dph and will stop feeding with the family by 40 to 50 dph [Zann 1996], and is thus typically well isolated from tutor and sibling song during the plastic song phase.

![Figure 42: Father stops singing as juvenile enters plastic song phase.](image)

(Top) Spectrograms of the juvenile’s song at 50, 52 and 54 dph. Note that between 52 and 54 dph the motif emerges. (Middle) An automated algorithm was used to record song-like sound files (2 – 10 seconds in duration) of a juvenile from 40 to 60 dph, kept in a sound attenuation chamber with his father (tutor). A set of 50 randomly chosen files per day was retained for analysis. Shown are the fractions of song files containing juvenile (light gray bars) and the tutor (dark gray bars) song as a function of age of the juvenile. Note that concurrent with the juvenile’s first recognizable copy of the tutor song (52 to 54 dph) the father significantly reduced its own singing activity. Colored boxes correspond to the spectrograms shown in the top panel. (Bottom) Sample spectrograms of the song of the juvenile at 90 dph and the tutor song. Highlighted in red boxes are the motifs. Note that the bird achieves a highly accurate copy of the relatively complex tutor song motif. He omits syllable 7 of the tutor motif, and adds a syllable after syllable 4 not present in the original.
We now find evidence that imitation accuracy correlates with the singing activity of the father as the juvenile is in the plastic phase of song learning. We conducted a total of three experiments in which a juvenile bird was raised with his father at the time of transition from subsong to plastic song. For the duration of the experiment, birds were housed in a sound-attenuating recording chamber.

In two of the three cases the father significantly reduced its singing activity concurrent with the juvenile’s first recognizable copies of the tutor song (see Figure 42). In one of these two cases we removed the juvenile after the father had decreased its singing activity and found that the father resumed normal singing activity levels within one day after removal of the juvenile. Both juveniles produced good imitation of the tutor song.

In one of the three cases the tutor did not reduce its singing activity at any time during the entire duration of the experiment (45 to 100 dph). This juvenile developed a highly aberrant song that showed little similarity to the tutor song (see Figure 43). More interestingly most syllables were still highly plastic and the song motif was not crystallized even at the age of 100 dph. This phenomenon was only observed once and will need to be confirmed in further experiments; if confirmed however, it would leave us with two possible interpretations. One could either argue that the tutor did not reduce its singing activity in the plastic song phase directly resulting in a poor imitation via an effect of model over-abundance, or one could argue that the juvenile never produced a recognizable copy of the tutor song, and thus never induced the father to reduce its singing activity. With the data at hand we, cannot distinguish between these two possibilities.

All these experiments only provide preliminary results and need to be repeated. However, the implications of the results – if confirmed – are intriguing. We know from the experiments of Tchernichovski et al. [Tchernichovski et al. 1999] that model over-abundance if present during the entire developmental window (30 to 90 dph) has detrimental effects on song imitation. Our results now suggest that this detrimental effect of over-stimulation is specific to the plastic phase of song learning.

In principle this is predicted by the fact that on average imitation accuracy decreases with the number of male siblings [Tchernichovski and Nottebohm 1998]. Additional stimulation with the
siblings’ versions of the tutor song, however, only occurs once the first bird enters the plastic song phase. This, combined with the fact that the father does not show reduced singing activity levels while the juvenile is in the subsong phase of song learning [Zann 1996], would already lead to the conclusion, that over-stimulation with tutor song, only has detrimental effects on song imitation if it occurs during the plastic phase of song learning.

6.3 Females prefer stereotyped song

Female zebra finches have been shown to prefer an unfamiliar male’s directed song over undirected song [Woolley and Doupe 2008]. Directed song differs from undirected song in a number of parameters. Directed song bouts contain more motifs, more introductory notes; directed song motifs are sung slightly faster, and are more stereotyped. It is generally assumed that preference for directed song is caused by the higher stereotypy of directed song. When correlating song preference with all these parameters, Woolley and Doupe find that only syllable stereotypy significantly correlates with preference [Woolley and Doupe 2008]. If females can indeed distinguish the two song types based on stereotypy, this would imply that females have an extremely rapid and precise auditory memory system. Detecting similarity between two subsequent stimuli requires holding the first stimulus in memory to then compare it to the second stimulus. Complicating the task here is the fact that even the differences between renditions of undirected motifs are comparably small – to the extent, that they are nearly undetectable by a human observer. To confirm the hypothesis that females actually use stereotypy to accomplish this task we tested female preference to a set of natural and artificial stimuli:

i) Undirected song (U) – a set of 20 undirected song bouts.

ii) Directed song (D) – a set of 20 directed song bouts.

iii) Pseudo-directed song (PD) – every motif in a set of 20 undirected song bouts is replaced with a copy of one randomly chosen motif (see Figure 45b).

iv) Introductory notes directed, motifs undirected (iDmU) – a set of 20 directed and 20 undirected song bouts were each split into two segments, one containing the introductory notes the other containing the motifs. 20 artificial song bouts were formed by combining introductory note segments of directed song with motif segments of undirected song.

v) Introductory notes undirected, motifs directed (iUmD) – Same as above, but here the 20 artificial song bouts were formed by combining introductory note segments of undirected song with motif segments of directed song.

6.3.1 Experimental design

Females in groups of 5-15 birds were isolated from males at least three days prior to the experiments. For the duration of the experiments, birds were kept alone or in groups of five in a custom built experimental setup (see Figure 44) divided into three chambers of 40 x 40 x 40 cm arranged in a row. Chambers were connected by holes in the wall separating them and were freely accessible to the birds throughout the experiment. Each chamber contained one perch connected to a switch that would be triggered by the weight of at least one bird sitting on the perch. Trigger signals were sampled at 10 Hz and were stored on a computer. Adjacent to the two outside
There were two independently controllable speakers used to broadcast the test stimuli. Food and water was available *ad libitum* in the central chamber.

![Figure 44: Preference test setup. Three chambers with one perch each (1-3) connected by openings in the wall between the chambers. Speakers were located adjacent to chambers 1 and 3 (4 and 5, speakers not shown).](image)

Birds were allowed to adapt to the novel environment for 4 - 24 hours before testing. Test sessions consisted of a 30 minute pre-testing baseline period, playback session 1 (30 minutes), a 30 minute inter stimulus period, playback session 2 (30 minutes) and a 30 minute post-testing period. Tests were commenced independently of where the birds were in the setup. In playback session 1 stimulus A was broadcast through speaker 1 and stimulus B through speaker 2, in playback session 2 stimuli were reversed (stimulus A broadcast through speaker 2, stimulus B through speaker 1). All data analyzed were pooled over both playback sessions to eliminate possible effects of chamber preference. To exclude a possible delayed effect of initial location of the bird the first six minutes (20%) of every playback session were excluded from further analysis.

During playback sessions stimuli were broadcast alternating between stimulus A and stimulus B in 30 second intervals (corresponding to 4-7 song bout stimuli played in sequence). For each stimulus type song bouts were drawn randomly from a set of 20 song bouts. The test stimuli used were undirected song (U), directed song (D), pseudo-directed song (PD, see Figure 45b) and directed song motifs with a preceding series of introductory notes form an undirected song bout (iUmD, introductory notes U and motifs D) and vice versa (iDmU).

There are a number of inherent problems associated with this experimental paradigm:

- Females rapidly lose interest in the stimulus playbacks. Strong behavioral responses with significant preference to one stimulus are only observed in the first few experiments and wane quickly with the successive number of experiments. We found that we could do no more than one or two experiments with the same bird, and only so if experiments were spaced by at least 24 hours. Even returning birds to the home cage for a week did not revive interest in the playback stimuli upon renewed testing. To reduce the impact of this effect we later performed the experiments with groups of five females. The idea was to create a competitive environment, thus potentially increasing interest in the song stimuli. We did indeed find that in groups the experiments could be repeated more often and we were able to get consistent results from up to four successive experiments with the same group of birds.
Unlike in previous experiments [Woolley and Doupe 2008] that employed video surveillance we had no continuous information on the birds whereabouts. With our setup birds were only detected when sitting on one of the three perches. Note however, that most birds will only leave the perches for feeding and drinking.

6.3.2 Results

In all experiments females displayed a significant interest in the playback stimuli. Throughout the playback sessions, birds frequently flew back and forth between the two perches adjacent to the two loudspeakers in rapid succession and hopped on perches flipping back and forth between facing left and facing right. To measure the behavioral response, we counted the number of perch on- and offset triggers (number of “perch hops”) during both playback sessions and compared this to the average number of perch hops during the inter stimulus period (see Figure 45a,b). Activity during the playback sessions was significantly higher than during the inter-stimulus period (signed rank test p = 0.027, Student’s t-test p = 0.031).

In agreement with the results of Woolley and Doupe [Woolley and Doupe 2008], we found that females displayed a preference for directed over undirected song. They spent significantly more time in the chamber adjacent to the speaker broadcasting directed song (65 % ± 9 %, mean ± std), than in the one adjacent to the speaker broadcasting undirected song (28 % ± 8 %). Woolley and Doupe find that naïve females spent on average 57 % ± 21 % in the directed song chamber and 8 % ± 3 % in the undirected song chamber [Woolley and Doupe 2008]. This verifies that our experimental paradigm is capable of measuring preference effects in spite of the lack of a continuous video surveillance as was used in the previous experiments [Woolley and Doupe 2008].

One of the parameters that distinguish directed from undirected song, is the number of introductory notes. Directed song bouts are on average preceded by more introductory notes than undirected song bouts [Sossinka and Böhner 1980; Cooper and Goller 2006; Kao and Brainard 2006]. To test the influence of the number of introductory notes on preference we used sets of directed and undirected song bouts in which we swapped the introductory note sequences (named iUmD and iDmU, see description above). The experiment was only successfully performed once (see Figure 45c, blue circles), the results would however suggest that the number of introductory notes has little if any significant effect on song preference.

Woolley and Doupe [Woolley and Doupe 2008] have suggested that female preference predominantly correlates with differences in stereotypy. To investigate if the preference for directed song is caused by an increased stereotypy, we tested females on playbacks of pseudo-directed song bouts and undirected song bouts. In 9 of the 10 experiments conducted, birds showed a preference for pseudo-directed song (see Figure 45d). The average time spent in the pseudo-directed song chamber was 42 % ± 23 % as compared to 24 % ± 13 % spent in the undirected song chamber.

41 Note that actual values are not reported in the paper, and that values presented here are estimates based on the data shown in Figure 3b in [Woolley and Doupe 2008].
Figure 45: Females can distinguish directed from undirected song based on stereotypy. 

**a:** Sample activity plot of the average number of perch hops per minute as a function of time. Gray shading indicates stimulus playback sessions. Note that the group of 5 birds is almost exclusively active during playback sessions. **b:** Stimulus playback significantly increases activity. Hops per minute during the inter stimulus period, and during the stimulus periods. Thick horizontal lines are medians, boxes the inter-quartile range (difference is significant, signed rank test $p = 0.027$, Student’s t-test $p = 0.031$). Data in green corresponds to the example shown in a. **c:** Schematic of the structure of undirected (U) and a pseudo directed (PD) song bout. All motifs of an undirected song bout are replaced with one version of the motif to artificially increase stereotypy without changing structure of the song bout, average tempo, average amplitude or number and timing of introductory notes. **d:** Average time during which there is at least one bird on the perch in the chamber adjacent to the speaker broadcasting undirected (U) and directed (D) song. Thick horizontal lines are medians, boxes the inter-quartile range (signed rank test $p = 0.125$, note however, that this is the minimum p value possible with $n = 4$, Student’s t-test $p = 0.008$). Data shown in blue is an experiment where introductory note sequences of U trials and D trials were exchanged (iDmU and iUmD). Data come from two experiments with a group of 5 birds each and two experiments with a single bird each. **e:** Average time during which there is at least one bird on the perch in the chamber adjacent to the speaker broadcasting undirected (U) and pseudo directed (PD) song. Data come from 9 experiments in 3 groups of 5 birds each (3 experiments per group, gray circles) and one experiment in an isolated bird (red circles). In all but one case the time spent in the PD chamber is longer than the time spent in the U chamber; on average the median time spent in the PD chamber is significantly longer (signed rank test $p = 0.023$, Student’s t-test $p = 0.057$).

### 6.3.3 Discussion

Our results demonstrate that females can distinguish undirected from pseudo-directed song and show a preference for the latter. Distinguishing these two sets of stimuli requires the ability to measure stereotypy with relatively high precision; the differences between two undirected song motifs are comparably small (see Figure 4). Measuring stereotypy is achieved by determining the similarity of two stimuli presented in sequence. In the case of song stimuli, this requires storing an
auditory memory of the first stimulus presented and then comparing this memory to the second stimulus presented. From the minimum difference that is still distinguishable we are in general able to infer a lower bound on the accuracy of the combined process of memory storage and comparison. The less accurate the memory stored the bigger the differences that will go unnoticed. To illustrate this point we can use the “find the 10 differences” game (commonly found in entertainment sections of newspapers) in which one has to identify differences between two images shown side by side. To perform this task with the images shown sequentially (with intermediate masking) instead of simultaneously requires almost photographic memory if the differences are small. The concept of “change blindness” for example, has its root in the comparably poor capacity of our visual memory.

Thus from the fact that female birds are able to hear the differences between different undirected motifs we can conclude that representations of auditory stimuli in memory are of higher resolution than the variability observed in undirected song. This is ever more surprising if we consider the speed at which this memory formation must occur. Song stimuli are alternately presented in intervals of 30 seconds (this corresponds to roughly 15 repetitions of the motif) and assessment of stereotypy will have to be made within that time. In the absence of an efficient attentional gating mechanism that would enable the bird to completely suppress auditory responses to the unattended stimulus, these auditory response are likely to severely interfere with the memory of the attended stimulus by masking it. Thus comparisons of motifs across different stimulus sessions are severely complicated by the intermittent presentation of another, very similar stimulus.

Our results also show, as previously suggested [Woolley and Doupe 2008], that stereotypy is one of the main determinants of female song preference. However, it is most likely not the only contributing parameter. Preference for pseudo-directed song is less strong than the preference for directed song (PD preference just fails to be significantly smaller than D preference, Student’s t-test p

Figure 46: Time course of song preference during playback sessions. Difference of time spent in PD chamber minus time spent in undirected chamber, normalized by the total time spent in both chambers. Data is averaged over three minutes sampled at one minute resolution.
= 0.075, Wilcoxon’s ranksum test p = 0.076), even though pseudo-directed song is more stereotyped than directed song\textsuperscript{42}.

To further tease apart the contributions of the different parameters that distinguish directed and undirected song, one would need to use an entire assortment of different types of artificial song. To test the contribution of all parameters apart from stereotypy for example, one could use pseudo-undirected song constructed by replacing all motifs of directed song bouts with randomly chosen and sped-up undirected versions of the motif. Or one could test the influence of song tempo by testing preference for sped-up undirected song versus undirected song at normal speed. Experiments should always be designed to reduce the difference between directed and undirected song step by step, to note with which parameter change preference is reduced to chance. Once a major contributing parameter has been identified in this way, preference should be tested on a set of artificial directed and undirected song bouts in which this parameter has been swapped between the two (e.g. as in iUmD and iDmU). In this way one could prove or disprove that a certain parameter is sufficient and necessary for female song preference.

\textsuperscript{42} Note that the question of whether stereotypy is the dominant factor in female song preference is most advantageously addressed by testing PD song versus D song directly.
7 Discussion

There are two big gaps in our understanding of song learning. The first is that of error detection and correction. We know that the songbird gradually matches his own vocalizations to those of his tutor and in order to achieve this task, critically relies on auditory feedback. This suggests that the bird compares feedback from his own vocalizations to a memory of the tutor song and uses the result of this comparison to improve the imitation. To this day, however, we have only little knowledge about how the differences between actual and desired sensory feedback are detected, and how the resulting error signals influence the motor system.

The second gap is that of memory storage. We know that the songbird is able form a long-term memory of tutor song in the early sensory period of song learning. Birds separated from their tutor before the onset of the sensorimotor period will also imitate faithfully without any further presentations of the tutor song. We know however alarmingly little about where and in what form, such a memory is stored.

Answers to either of these problems will significantly increase our understanding of also the other, as both are intricately tied. We would expect to find error- and template-related signals in close proximity, as the computation of the error signal is inherently based on a template signal. Mark Konishi has formulated this idea as follows: “In control system terms an acquired template can be conceived as a memory subsystem which also evaluates auditory feedback information, […].”[Konishi 1965]

7.1 Error correction

To selectively induce changes in the motor system contingent on errors, we can infer that either the activity of a single subset of neurons or the coincident activity of multiple subsets of neurons with direct anatomical connections to the motor system must selectively code for errors. To ease discussion we will refer to such a subset or subsets of neurons as the ‘error detection system’.

From deafening experiments we know that such an error detection system reliant on auditory feedback must be active during song development. In adult zebra finches it was, however, initially thought to be no longer active, as initial evidence had suggested that in age-limited learners (such as zebra finches, Bengalese finches or song sparrows) deafening in adulthood no longer has detrimental effects on song, i.e. auditory feedback is no longer necessary for song maintenance once song is fully developed [Konishi 1965; Nottebohm 1968; Price 1979; Bottjer and Arnold 1984]. Later experiments however showed that even in adult age-limited learners maintenance of song stability requires auditory feedback [Nordeen and Nordeen 1992; Woolley and Rubel 1997; Lombardino and Nottebohm 2000]. Zebra finches deafened between 80 and 170 dph show strong song deterioration within 2-3 weeks. Birds deafened at the age 170 to 360 dph show signs of deterioration within one year from deafening. Birds deafened after the age of one year only showed significant effects two

43 This could for example be the case for X-projecting VTA neurons that have been speculated to convey dopaminergic error signals. These neurons do respond to auditory perturbations during song, however they also respond to the same perturbation stimuli when presented in isolation [Liora Las, personal communication]. Such signals can only selectively induce changes in the motor system contingent on errors if they are coupled with a singing related signal.
years post deafening. It is interesting to note that even the small residual auditory feedback that remains after targeted cochlear-hair-cell lesions [Dooling et al. 1997], or continuous white-noise exposure, can almost completely prevent song deterioration [Marler et al. 1973]. In addition to this, it has been shown that distorted auditory feedback can lead to decrystallization of adult song [Leonardo and Konishi 1999; Leonardo 2004; Kozhevnikov and Fee 2007], and can even be used in operant conditioning of pitch shifts [Tumer and Brainard 2007]. More importantly, song recovers from decrystallization once normal auditory feedback is restored [Funabiki and Konishi 2003; Zevin et al. 2004].

Thus it seems that the neural processes of error detection and motor adaptation underlying song development do not simply stop functioning with the crystallization of song but remain active for an extensive period of time. We are even led to believe that the observed gradual reduction of song plasticity with age is the result of a reduction of neural plasticity in the motor system (as it is observed in the connectivity of RA [Herrmann and Arnold 1991; Kittelberger and Mooney 1999]) not a reduction in sensitivity of the error detection system.

Two different systems have been proposed to underlie error signaling: the anterior forebrain pathway and the auditory pathway. We will briefly review the evidence that has pointed to the AFP as potential candidate for error detection, and then the more recent evidence including our own work that has suggests a direct involvement of the auditory pathway.

### 7.1.1 The anterior forebrain pathway is permissive for vocal plasticity

The architecture of the AFP has long intrigued and inspired theoreticians and experimentalists alike. The two projection pathways from HVC to RA, a direct and an indirect pathway via the AFP, go well with ideas like supervised learning – a framework in which the AFP would provide a teaching signal for the HVC to RA connections. Early support for these ideas came from experiments that evidenced a role for the AFP specific to song learning. INMAN lesions during song development result in poor imitations and early crystallization [Bottjer et al. 1984; Scharff and Nottebohm 1991], but have little effect on song maintenance in adults [Bottjer et al. 1984]. Basham et al. later devised an experiment in which they tutored birds in sessions every other day [Basham et al. 1996]. With this they could show that birds that received an infusion of an NMDA-receptor antagonist in INMAN prior to each tutoring session achieved significantly less accurate imitations of the tutor song than normal controls, or birds that received similar infusions on non-tutoring days. Both of these experiments demonstrated that INMAN activity is crucial for developmental plasticity during song learning. When Brainard and Doupe could then show that INMAN lesion also eliminate deafening-induced song deterioration in adults [Brainard and Doupe 2000b], it became clear that INMAN is part of an active corrective mechanism. It can be argued, that in the absence of auditory feedback an error detection mechanism would produce only aberrant error signals. A lesion of the error detection mechanism would then retain the status quo. These findings led to the speculation that the AFP computes an error signal that is then sent to RA via INMAN [Doupe 1997; Brainard and Doupe 2000a; Rosen and Mooney 2000].

Two possible mechanisms for error detection have been suggested. In the first, auditory feedback is directly compared to the tutor template [Brainard and Doupe 2000b]. This could occur via BOS and tutor song selective neurons in INMAN [Doupe and Konishi 1991]. In this model tutor selective

---

44 NMDA receptors are thought to underlie certain types of learning (e.g. NMDA-receptor dependent LTP).
neurons, acting like filters, would only respond if song-related auditory feedback is a good match of the template. One potential problem of this is the delay between premotor activity and sensory feedback, which has been estimated to be on the order of 50 ms. A second model that circumvents this problem is the efference copy model [Troyer and Doupe 2000b; Troyer and Doupe 2000a]. Here a prediction of auditory feedback, not the feedback itself, is compared to the template. Both of these models assume a mechanism capable of online (i.e. during singing) detection and correction of errors. Another way to circumvent the problem of the feedback delay, is to assume that only error detection occurs online and that correction happens offline, e.g. during sleep [Dave and Margoliash 2000; Margoliash 2001; Derégnaucourt et al. 2005; Shank and Margoliash 2008].

Another possible explanation of these findings is that activity of the AFP is only permissive for plasticity in the HVC to RA connectivity, but does not itself carry an error signal [Brainard and Doupe 2000b]. This idea is supported by the fact that IMAN lesions in juveniles rapidly and profoundly alter RA circuitry [Kittelberger and Mooney 1999], suggesting that IMAN-lesion-induced deficits in song development are not caused by a resulting lack of error signals but rather by an induced premature change in neuronal connectivity in the motor system. Leonardo later directly put the AFP error-correction model to the test, by recording from IMAN-HVC projection neurons during singing [Leonardo 2004]. He could show that these neurons are not responsive to loud broad-band noise perturbations during singing. These findings are clearly at odds with an AFP error-correction model in which auditory feedback is compared to the template to give rise to an error signal; they are however, compatible with a model in which an efference copy is used to compute error. Activity in the AFP would thus be an efference copy of motor command and carry no sensory feedback signal. More recently Prather et al. have speculated that already HVC X-projecting neurons carry an efference copy signal [Prather et al. 2008].

Further evidence against the idea that the AFP computes an error signal comes from AFP lesions during HVC-lesion-induced song destabilizations [Thompson et al. 2007]. ‘Micro’ lesions in HVC can cause transient song destabilizations from which the bird typically recovers within a matter of one week. This recovery is dependent on auditory feedback – deafened birds do not recover their song after such HVC ‘micro’ lesions. A further lesion to IMAN can lead to immediate recovery of song45. This, combined with the assumption that error signals can only drive slow adaptation of the motor program 46, would suggest that the AFP is not the source of an error signal to RA.

Another very promising interpretation of the role of the AFP that would explain why the AFP is permissive for plasticity, yet is not the source of an error signal, is that of the experimenter in an actor, experimenter and critic framework [Doya and Sejnowski 1995; Troyer and Doupe 2000b; Troyer and Doupe 2000a; Fiete et al. 2007a]. The actor corresponds to the motor system – the critic is still unknown. Given that IMAN has been shown to be able to drive transient song perturbations [Kao et al. 2005; Olveczky et al. 2005] and that IMAN lesions reverse HVC ‘micro’-lesion-induced destabilization of song [Thompson et al. 2007], it is not implausible that the AFP should introduce exploratory variability in the motor system.

45 Birds that receive IMAN lesions at the same time as HVC ‘micro’ lesions show almost no signs of song degradation.
46 This assumption is based on the fact that neither deafening, nor auditory perturbation, nor ovoidalis stimulation have immediate effects on motor output. If error signals could drive immediate changes, then the observed recovery of song after IMAN lesion is exactly what one would expect if IMAN calculated an error signal based on an impaired efference copy signal or an impaired auditory signal (one could expect these signals to be impaired as a result of the HVC ‘micro’ lesion).
7.1.2 The auditory pathway can compute an error signal

In an auditory framework calculating error is a simple matter. Subtract from the actual auditory feedback the predicted or desired auditory feedback – the resulting difference is the error. Depending on whether predicted or desired auditory feedback is used for this calculation one speaks of prediction error or performance error. The prediction error is the difference between actual sensory feedback and a prediction of sensory feedback based on motor commands. Such a prediction can be described by a coordinate transformation between sensory and motor coordinates and is referred to as a forward model [Jordan and Rumelhart 1992a]. Prediction errors are inherent to most goal directed behaviors. The performance error, on the other hand, is the difference between actual sensory feedback and a memory of desired sensory feedback. This memory would need to be replayed in synchrony with the sensory feedback (see section 7.2 for a discussion). The ability to detect performance errors is essential for imitation learning in general.

In our analysis of perturbation responses during singing in field L and CLM neurons we have found neural correlates of actual sensory feedback (neurons that respond equally to perturbation during singing and during BOS playback) and neural correlates of predicted or desired sensory feedback (neurons that are unresponsive to perturbation stimuli during singing but not during BOS playback) side by side in field L and CLM. This together with the existence of perturbation selective neurons in the same brain regions is strong evidence for the idea that the auditory processing stream is capable of calculating an error signal. Note that in principle the observed perturbation-selective responses could be a correlate of a disruption detector and not an actual error signal. Evidence that these responses are indeed a correlate of an error signal comes from the neuron shown in Figure 17e. On close inspection of the unperturbed trials it becomes evident, that this neuron also shows a small response in the absence of a perturbation stimulus. Given that the birds are in the plastic phase of song learning one would expect to see “mistakes” in unperturbed trials. Thus we would predict a small naturally occurring error signal during unperturbed trials to be simply amplified by perturbation. And this is exactly what we observe in the responses of this neuron. Furthermore, assuming that these neurons function as disruption detectors, one would have to assume that spikes during unperturbed motifs correspond to errors in disruption detection.

Another very strong indication for the idea that primary and secondary auditory areas of the pallium are directly involved in error detection comes from the work of Lei and Mooney ([Lei et al. 2007] and personal communication). They were able to show that electrical stimulation of Ovoidalis during singing, targeted at one syllable only, leads to rapid deterioration of this syllable within three to four days. This is significantly faster than the deterioration observed in deaf birds, where in age matched animals (Lei and Mooney use birds roughly 90 dph) song deterioration effects are visible after two to three weeks [Lombardino and Nottebohm 2000]. Thus, a stimulation-induced increase in input to field L leads to more rapid deterioration than a deafening-induced attenuation of input.

The data we have however does not allow for a distinction between performance and prediction error. This would require interfering with normal vocal output either by unilateral lesions of HVC, syringeal denervation or operant conditioning of song features. Perturbations of this altered song

47 Evidence for the idea that song learning makes use of forward-model-like principles comes from the work of Liu and Nottebohm, who show that the chipping sparrow develops set of precursor songs, of which it then selects and further develops the one closest to the tutor song to finally achieve an accurate imitation [Liu and Nottebohm 2007].

would not lead to selective performance errors, as unperturbed song itself is far off template. Any potential remaining perturbation selectivity could then be classified as prediction error.

Assuming that error $e$ is calculated as the absolute difference of sensory signal $s$ and corollary discharge signal $c$, $e = |s - c|$, we would expect that passive auditory stimulation would lead to aberrant error signals, as the sensory signal is not cancelled out by corresponding corollary discharge. This can clearly not be the case as birds typically grow up in a very noisy environment, where the time spent singing is significantly smaller than the time exposed to auditory stimuli, such that aberrant error signals would outweigh song-related error signals. Thus birds must have a filter or gating mechanism that selectively reduces the influence of auditory stimuli when the bird is not singing\textsuperscript{49}. One possibility would be to simply suppress all auditory input to such an error detection system except during singing, similar to gating mechanisms that are suggested to suppress auditory activity in NIf and HVC in the awake zebra finch [Nick and Konishi 2001; Coleman et al. 2007]. These gating mechanisms are known to be rapidly driven by behavioral state. Auditory responses in NIf and HVC are suppressed in the aroused state, but not in a sedated or sleeping state [Cardin and Schmidt 2004a]. Although in field L no behavioral state dependent modulation is observed [Schmidt and Konishi 1998; Cardin and Schmidt 2003], evidence for such a gating mechanism comes from our finding that a subset of field L and CLM neurons is unresponsive to auditory perturbations during song [Keller and Hahnloser 2009] and from the preliminary evidence that BOS playback response adaptation is counteracted during singing (see Figure 30). Both findings suggest that auditory processing during passive listening and singing is significantly different.

A second possibility would be a filter mechanism that only passes sensory stimuli to the error detection system that are already a sufficiently close match to the BOS or the tutor song\textsuperscript{50}. However such a mechanism would require direct access to a memory of BOS and tutor song. In principle the dual selectivity for BOS and tutor song observed in HVC [Volman 1993] and the AFP [Solis and Doupe 2000] could be an artifact of such a filter mechanism. With the data at hand we cannot distinguish between these two possibilities, it is however, very likely that the bird uses a combination of the two strategies\textsuperscript{51}.

Previous results have suggested [Tchernichovski and Nottebohm 1998; Tchernichovski et al. 1999] and our behavioral data now appears to confirm that auditory stimulation with tutor song\textsuperscript{52} during the plastic song phase has detrimental effects on imitation accuracy. This can readily be explained if

\textsuperscript{49} In principle one could imagine that error is calculated as $e = [c - s]$ in which case auditory stimulation alone would not result in an error. The problem here is that any sensory signal that exceeds the corollary signal will not lead to an error.

\textsuperscript{50} A similar effect is exemplified by the fact, that when we introduce distortions in the auditory feedback of our own voice, there is an upper limit of distortions after which one no longer has the perception of hearing one’s own voice.

\textsuperscript{51} A possible experiment to test the effects of auditory stimulation during the plastic song phase directly would be continuous stimulation with song stimuli. Playback of BOS, tutor song and conspecific song could be used to disentangle the different interference effects. The hypothesis based on our data would be that BOS and tutor song playback would lead to detrimental effects on song development, whereas playback of conspecific song would be without effect. In addition, stimulation during the night or during the day could be used to investigate possible behavioral state-dependent gating mechanisms.

\textsuperscript{52} Songs similar to tutor song, such as BOS or the songs of siblings seem to have the same effect. This is based on the fact that the accuracy of imitation correlates negatively with the number of males in a clutch [Tchernichovski and Nottebohm 1998].
one assumes that the aforementioned gating or filtering mechanism is not perfect. In addition, it would suggest that the mechanism involves filtering, as stimulation with other auditory stimuli (i.e. those not closely related to tutor song) has no such effects.

Given that these effects only occur during the plastic song phase and not during the subsong phase, an intriguing possibility would be that the filtering is specific to the BOS motif (and not to tutor song or to subsong - due to its large variability subsong itself would be difficult to selectively filter). In this model the filter would develop in parallel with the emergence of the BOS motif (and not already with tutor-song template formation). This hypothesis is supported by the fact that auditory stimulation during the subsong phase has no detrimental effects, and, assuming that BOS-selective responses in HVC are indeed an artifact of such a filter, that these BOS-selective responses in HVC emerge only during the plastic song phase [Volman 1993; Doupe 1997].

The existence of such a filter would also explain why HVC activity is not responsive to broadband-noise perturbations during singing [Kozhevnikov and Fee 2007], but does respond to delayed auditory feedback [Sakata and Brainard 2008].

7.2 The template replay hypothesis

7.2.1 Song memories are stored in sensory coordinates

We will distinguish two different forms of memory, sensory and motor. ‘Sensory memory’ shall refer to a memory stored in sensory coordinates, that is, in the coordinate system used to represent sensory stimuli in a sensory area (see section 2.2.1 for definitions). The equivalent shall be true for ‘motor memory’. In general this bipartite distinction is a gross oversimplification and is only introduced to ease discussion.

An everyday example that well illustrates this distinction of sensory and motor forms of memories is computer passwords; in particular those, that do not merely consist of words, such that they will need to be memorized symbol by symbol. At first, one will recite the password while typing it, later it happens every so often, that after using it for months or even years one is still able to type the password, but can no longer recite it. Thus one can no longer reconstruct a sensory memory and only has access to a motor memory of the password. This loss of a sensory memory is most likely abetted by the fact that typing of passwords is typically hidden, such that one is deprived of the sensory stimulus when using it.

---

53 A second possible explanation for this effect would be interference with the template, see “The template replay hypothesis” for a discussion.
54 We can conclude this from the fact that birds typically grow up in noisy environments where conspecific song is abundant.
55 Tchernichovski and Nottebohm report that in a clutch with multiple males the juvenile that produces the first recognizable copy of the tutor song typically develops a good imitation of the tutor song while the others produce significantly poorer imitations [Tchernichovski and Nottebohm 1998]. As both the father and the male siblings sing throughout the juveniles’ subsong phase, this is a good indication that only BOS-like song stimulation in the plastic song phase has detrimental effects on song learning.
56 Note however, that there are a number of more trivial explanations for this discrepancy. Kozhevnikov and Fee recorded from HVC X-projecting neurons in zebra finches whereas Sakata and Brainard speculate that their recordings conducted in Bengalese finches are dominated by HVC interneuron activity.
57 Note that these definitions are not consistent with previous usage of the term ‘sensory memory’, which in psychophysics has been used to describe after-image-like effects, i.e. a memory that retains a brief (on the order of 1 second) impression of the sensory stimulus after the stimulus itself has ended.
We know that birds have a motor memory of song, because they are able to sing without auditory or proprioceptive feedback [Bottjer and Arnold 1984], and because lesions of sensory areas have no immediate detrimental effects on song production [Nottebohm et al. 1976]58. In other words, we know that song is centrally controlled and central control of motor output requires a motor memory. Conversely, although it has long been suspected that the juvenile bird initially forms a sensory memory of tutor song that is later transferred to a motor memory [Nottebohm 2000; Bolhuis and Eda-Fujiwara 2003], direct evidence for this is relatively sparse, and the suspicion is based on a number of findings that only provide indirect support. Given that the bird can acquire a memory of tutor song before the onset of the subsong phase 30 dph59, and that the connections from HVC to RA are only grown between 30 - 35 dph [Konishi and Akutagawa 1985; Mooney and Rao 1994], we can conclude that the initial memory of tutor song is not stored in the HVC to RA connections or in the motor pathway downstream thereof. It is also unlikely that such a fundamental change in projection architecture should occur in HVC without influencing its functional properties. We know for example that HVC activity significantly changes between the subsong phase and the plastic song phase [Crandall et al. 2007], indicating that also intrinsic HVC connectivity changes fundamentally between 20 (roughly the age of template acquisition) and 60 dph (roughly the age of first recognizable motif). This together with the finding that subsong is driven by IMAN and not HVC [Aronov et al. 2008] makes it highly unlikely that HVC is the storage site of the initial memory of tutor song60.

Thus we can speculate, that the template of song – the initial memory of tutor song – is stored in sensory coordinates (in sensory areas) and is only later during song development transferred to motor coordinates (in premotor areas). The largest part of this transfer most likely occurs prior to, or during the transition from subsong to plastic song. Direct evidence for such a transfer comes from the findings that in swamp sparrows auditory responses in HVC correlate well with perceptual boundaries [Prather et al. 2009], and that activity during BOS playback closely resembles activity during singing in HVC [Prather et al. 2008] and RA [Dave and Margoliash 2000].

Based on behavioral data we can make a number of inferences about acquisition and form of this sensory template. The following observations for example would suggest that the memory of the template is acquired and stored on a syllable-by-syllable basis. First, juvenile birds will sing recognizable copies of individual syllables during subsong before the emergence of a motif, indicating that they practice song on a syllable-by-syllable basis. Second, white-crowned sparrows can be successfully tutored with syllable pairs only [Rose et al. 2004]. Thus the bird learns to produce a

---

58 Note that this argument is circular both because premotor areas HVC and RA were initially identified as pallial brain areas necessary for song production and because sensory areas are defined by precisely this absence of lesion induced deficits (see “Note on the usage of the terms ‘sensory area’ and ‘motor area’”).
59 On careful inspection of the song development shown in Figure 6 and Figure 7 we note that some syllables of the early motif resemble syllables of his father’s song, even though the juvenile was separated from his father at 10 dph. This would indicate that memories can be acquired well before the onset of the subsong phase.
60 One could for example imagine that HVC initially stores a memory of tutor song in sensory coordinates and that learning then modifies the HVC to RA connections thus translating this sensory memory to motor memory. In such a framework however, one would predict that HVC activity is necessary for subsong.
complete motif without ever having heard it in one piece. In addition, Funabiki and Konishi have shown that recovery of song syntax (syllable order) and syllable structure after white noise exposure is limited by two different time scales [Funabiki and Konishi 2003]. Zebra finches exposed to continuous white noise after tutoring in a time window 35 dph to 200 dph, initially produce highly aberrant song after restoration of auditory feedback. Within the following 30 days they then develop recognizable imitations of individual syllables but not of motif syntax. Both syntax and syllable structure is imitated correctly however, if noise exposure is terminated at 80 dph.

We can infer a lower bound on the accuracy of the auditory template from the fact that females can distinguish pseudo-directed (or directed) song from undirected song. To be able to perform this task with the relatively limited data available in one song bout, the variability of the auditory memory needs to be smaller than the variability of undirected song.

We also know that the template is acquired with relatively few repetitions of the motif. Two to three hundred repetitions are sufficient for accurate imitation of tutor song [Petrinovich 1985; Tchernichovski et al. 1999] (and personal observation). The fact that females can relatively rapidly, within a matter of minutes, tell the difference between directed and undirected song [Woolley and Doupe 2008] (Figure 45d) and between pseudo-directed and undirected song (Figure 45e and Figure 46), further suggests that initial memory formation occurs with substantially fewer repetitions of the motif61.

7.2.2 NCM is involved in memory formation and storage

There has been accumulating evidence that the initial sensory memory of the tutor song is stored in higher order auditory areas, the most promising candidate of which at the moment is NCM. This idea dates back to response habituation experiments of Chew et al. [Chew et al. 1995; Chew et al. 1996]. They were able to show that NCM auditory responses habituate with successive presentations of an auditory stimulus. Upon initial presentation of a certain song stimulus, auditory responses are vigorous; they then rapidly decay over the course of 10 to 20 presentations to firing rates on the order of 20% of the initial response (see Figure 28). This habituation is highly stimulus specific, as intermittent presentations of different auditory stimuli do not influence habituation. Later Phan et al. have suggested that the rate of habituation inversely correlates with novelty of the song stimulus [Phan et al. 2006]. The more familiar the bird is with a certain song stimulus, the faster the responses habituate. We now find evidence that singing almost immediately resets BOS playback induced habituation and that the singing-related activity follows an opposite, activity-increasing trend (see Figure 30). Conversely, BOS playback appears to reset this singing-related activity increase. This is true, in spite of the fact that average BOS-playback-related activity is highly similar to average singing related activity (see Figure 29). These data would suggest that activity in NCM is highly non-linear and critically dependent on motor activity, as would be expected from memory-related activity.

Further evidence for memory-related activity comes from work on IEG expression. In response to song playback, there is a significant increase in ZENK expression throughout NCM [Mello et al. 1992], in response to single syllable playbacks however, only certain sub-regions of NCM exhibit significant increases in ZENK expression [Ribeiro et al. 1998]. If indeed a correlate of memory this would also

---

61 This is based on the argument that measuring variability requires comparing the current rendition of the motif with a memory of one previous rendition or an average of multiple previous renditions.
provide further evidence for a syllabic representation of sensory song memories. Additionally, ZENK expression in response to tutor song playback in NCM but not in the song system [Bolhuis et al. 2000], the hippocampus or CLM [Terpstra et al. 2004] correlates with accuracy of tutor song imitation. Zebra finches that developed a more accurate imitation of tutor song exhibited higher levels of ZENK induction upon re-exposure to tutor song.

Upon lesion to the sensory song-memory system, one would expect, impaired song recognition as well as an impaired ability for song imitation. And indeed, NCM lesion in adult zebra finches impair song recognition but not song or call production [Gobes and Bolhuis 2007]. NCM lesions in juvenile zebra finches after tutoring in the early subsong phase lead to poor imitations of tutor song (see Figure 37). Suppression of protein synthesis in NCM and the caudomedial mesopallium (CMM) during tutoring also leads to poor imitations of tutor song [London and Clayton 2008]. A fundamental problem of this lesion or inactivation approach however is, that one cannot exclude beyond any doubt that the observed deficits are simply due to impaired auditory processing.

7.2.3 A sensory memory of the tutor song is replayed during singing

A further conjecture we will need to hazard concerns the functional properties of memories. Assuming that memory formation is constrained to changes in synaptic strengths (including formation and loss of synapses), there are two possible neural implementations of memory:

1. **Pattern matching.** Synaptic connections between sensory neurons and a population of ‘readout’ neurons are adapted such that the readout neurons selectively respond to the presence of one particular pattern in the sensory stimulus. BOS-selective HVC neurons could function as readout neurons in such a framework.

2. **Activity propagation.** Synaptic connections between sensory or motor neurons are adapted such that connections between neurons that fire in sequence in response to a certain stimulus or during a certain motor output are enhanced. The synfire chain architecture proposed to underlie HVC function corresponds to such an activity propagation mechanism.

The main difference between these two functional implementations is that a pattern-matching mechanism cannot be used to store stimuli with durations that exceed synaptic time constants. Given that typical song stimuli exceed this duration and that motor memories by definition are manifest as sequences of motor-activity patterns, such a pattern-matching mechanism on its own is of little practical use in the storage of auditory memories. Thus we will hypothesize that the sensory

---

62 Note however that the data presented does not allow to exclude beyond doubt the possibility that NCM is selectively activated by BOS playback, and that tutor-song playback in birds that have achieved good imitations more potently induces ZENK expression simply because tutor song and BOS are more similar in these birds. The argument presented against this hypothesis is, that it would imply that the total ZENK expression in response to BOS was higher than the expression in response to tutor song; the data however, seems to contradict this, as both stimuli induce roughly equal levels of ZENK expression.

63 London and Clayton argue against this by showing that birds’ ability to perform auditory discrimination tasks is not impaired by this suppression of protein synthesis.

64 Note that this in principle does not necessarily imply that motor memories are also stored as sequences of activity. However, based on the very sparse singing related activity pattern of HVC_{RA} neurons [Hahnloser, et al. 2002], and the fact that electrical stimulation will interrupt motor output with different current thresholds at different points in the song [Wang, et al. 2008], we have good evidence to speculate that motor memories in HVC are stored as synfire-chain-like sequences of activity patterns. We also know that such activity patterns are spontaneously replayed during sleep [Dave and Margoliash 2000], evidencing the fact that activation of one activity pattern can directly drive the subsequent activity pattern.
memory of song is stored as a sequence of activity patterns that by virtue of their connectivity form a synfire-chain-like structure. With other words, activation of one activity pattern directly drives or facilitates activation of the next activity pattern\(^{65}\).

From the perturbation experiments in field L we cannot infer if the corollary signals we encounter are the correlates of an efference copy of motor command or those of a singing related template replay (see Figure 47).

Figure 47: Template replay hypothesis. An efference copy of motor command or a template replay signal is compared to the sensory feedback to detect errors in motor output.

Conceptually the efference copy model and the template replay model are similar. Both require the transfer of a signal from the motor system to the sensory system. In the case of the efference copy model the signal transferred is the efference copy, in the case of the template replay model this is a trigger signal, initiating replay. Both models would predict the same neural activity patterns. The difference between the two simply is that the efference copy model would require a much stronger projection from the motor system to the sensory system. Thus without detailed anatomical information we cannot definitively exclude one or the other. The fact that in some NCM neurons activity during singing is a closer match to the activity elicited by tutor song playback than to the activity elicited by BOS playback would however suggest that a sensory memory of tutor song is replayed while the bird sings\(^{66}\).

Memory replay could be triggered either by premotor signals or by auditory feedback signals. The fact that we find song locked activity in field L of a deafened bird (see section 3.4) suggests that replay is at least in part motor-triggered. The fact however that there is no motor driven ZENK expression in the auditory pallium in deafened birds [Jarvis and Nottebohm 1997] indicates that normal NCM function heavily relies on auditory feedback. Thus it is most likely a combination of premotor and auditory feedback signals that trigger memory replay. The assumption that memory replay is reliant on auditory feedback could explain why deafening-induced [Lombardino and Nottebohm 2000] song deterioration is slower than ovoidalis stimulation [Lei et al. 2007] or

---

\(^{65}\) It could for example be shown that in humans auditory cortical areas can be activated even in the absence of sound and that this could correspond to the phenomenological experience of imagining music [Zatorre and Halpern 2005]

\(^{66}\) The efference copy model and the template replay model are not mutually exclusive. It is well possible that both effects contribute to vocal learning in general. The main advantage of the forward/inverse model architecture is that, once an inverse model is trained, this allows for rapid (i.e. without further learning) imitation of arbitrary auditory stimuli.
feedback-perturbation [Leonardo and Konishi 1999] induced song deterioration, as the error signal is a function of the difference between the memory replay signal and the auditory feedback signal.

Moreover, if we thus assume that NCM activity functions as a memory replay during singing, then the fact that BOS playback significantly reduces singing related NCM activity (see Figure 30) could also explain why over-exposure to tutor song during the plastic song phase has detrimental effects on song learning (see section 6.2)\(^\text{67}\). Similar to a hard disk, that cannot read and write at the same time, auditory stimulation with tutor song would thus interfere with template recall. This is not altogether unexpected, as memory acquisition and memory retrieval are functionally distinct processes that both need to interact with the same neuronal structures.

Brain areas of the avian pallium classified as sensory are vastly larger\(^\text{68}\) than those classified as premotor and are substantially less well understood. A very impressive example hereof is the HVC shelf – RA cup pathway (see Figure 48). Even though it is clear that HVC and its underlying shelf area are functionally connected [Vates et al. 1996] and probable that a similar interaction between RA and its surrounding cup exists, to date no experiments addressing functional properties of these ‘auditory’ areas have been conducted.

We have every reason to believe, that future insight on questions of song memories and of error detection will only be possible with a deeper understanding of the functional properties of the brain areas we have so loosely classified as auditory.

Figure 48: Projection form the HVC shelf to the RA cup. Darkfield photomicrograph of RA after the injection of biocytin in the shelf region of HVC. Picture is taken from [Mello et al. 1998].

\(^{67}\) Interesting to note is that both over-stimulation with a BOS-like song during the plastic song phase and NCM/pHVC lesions appear to have similar effects (compare Figure 37 and Figure 43). The fact that auditory stimulation with BOS significantly reduces singing-related NCM activity could be the underlying cause for this similarity.

\(^{68}\) By a factor of at least 10 in both volume and estimated number of cells!
7.3 Active sensory processing

For the first time we have now been able to detect correlates of error and memory related signaling. It is not altogether surprising to find such signals, as they have long been predicted to exist, the surprising thing however is that we find them in primary and secondary auditory areas of the pallium.

We speculate that these motor related signals are an inherent component of any higher order (i.e. cortical or pallial) sensory processing system. The reason such effects of active sensory processing were not detected earlier most likely is shortcomings in both experimental paradigm and recording technique. Most of our knowledge of sensory processing systems comes from experiments on anesthetized animals that are passively seeing, hearing, smelling or feeling the respective sensory stimuli presented. Assuming that feed-back input from planning and motor related areas has significant influence on the activity of sensory areas, these experiments would only evidence the passive aspects of sensory processing - while active aspects go undetected. David Field and Bruno Olshausen have estimated that in primary visual cortex, the most extensively studied cortical area, only a small fraction of the activity has been characterized (“What is the other 85% of V1 doing?” in [Hemmen and Sejnowski 2005]). Based on the evidence of differences between active and passive processing, we have reason to believe that this gap can only be filled by characterizing the activity of sensory neurons in active, behaving animals.

In addition, in most of this work neural activity in vivo is measured with electrophysiological recording techniques. Such methods are significantly biased to record from neurons with continuously high levels of activity or those that are easily driven with simple stimuli; neurons that respond only infrequently easily go undetected. This bias is not easily overcome with current recording techniques. It is however quite important to keep this in mind when searching for neurons: the most interesting neurons might not be the loudest ones.

The answers to the mysteries of sensory processing will only be unveiled, once we begin to ask how the sensory system functions in interaction with the rest of the brain. We can expect that it will not be without a deep understanding of this active sensory processing that we shall discover the key principles by which brains work.
7.4 Afterword

As I began my research on auditory feedback processing in the songbird, I came across a very inspiring review by Erich Jarvis on ‘brains and birdsong’ in which he writes:

“[T]he search for the auditory template and the site for auditory feedback in the songbird brain has been a major area of research. The search has led to findings that are both interesting and elusive.”

[Marler and Slabberkoorn 2004], p.245

It was this search that was then to become the main theme of also this thesis – and I now in retrospect hope that we were able to amend to it an interesting finding or two.
8 METHODS

8.1 Experimental design

All experiments were carried out in accordance with protocols approved by the Veterinary Office of the Canton of Zurich, Switzerland. In total (including all experiments, failed and successful and practice surgeries) 39 juvenile (age 60-90 days) and 23 adult male zebra finches, and 7 adult female zebra finches were used to acquire the data presented here.

8.1.1 Song experiments

During all electrophysiological recording sessions birds were housed in a sound isolation chamber equipped with a microphone and two loudspeakers (for playback of the BOS and perturbing stimuli) in a triangular arrangement around the centre of the experimental cage (distance 25 cm, see Figure 49).

Experiments were conducted in a sound isolation chamber housing a Plexiglas bird cage (24 x 24 x 20 cm (L x W x H)), two loudspeakers, a microphone and a webcam (not shown in diagram).

Experiments were only performed once the birds reached song activity levels of over 1000 motifs per day. This typically occurred 1-5 days post surgery. To accelerate recovery and increase singing activity birds were housed with a companion bird (usually an adult female bird). Although it is generally assumed that directed song is only produced in adulthood, even birds as young as 60 days post hatch can be induced to sing by removing the companion female for a few minutes and subsequently replacing it. Songs produced in this way look very similar to the directed songs of adults. In all experiments we did not distinguish between directed and undirected song, or the juvenile equivalents thereof. Furthermore, we used the companion female to induce the bird to sing in cases where we had a high signal to noise neural recording but no song trials and were running danger of losing the recording.

8.1.2 Playback experiments

All song files used in playback experiments (BOS and TUT playback) were recorded in the same or similar recording setup as the microdrive recordings were performed in. Sound amplitude level of playback was chosen to match the sound amplitude level of the singing bird on the microphone.
To minimize potential differences between playback stimuli and the birds own song in BOS playback experiments, the song files used for playback were files recorded on the same day or the previous evening. Unless noted otherwise BOS and TUT playback experiments were conducted with a set of 10-20 different song files, each containing one song bout composed of at least two motifs. Songs were undirected unless noted otherwise.

To reduce the effects of a potential response-adaption process, BOS playbacks were interspersed with songs as much as possible (see section 8.2.4 for details). In the case of experiments with both BOS and tutor song playback, the stimuli were randomly alternated. We ignored all data in which birds countersang with song playback.

8.1.3 Perturbation experiments

The acoustic perturbing stimuli used for all experiments were an introductory note, a long call and a broad band noise stimulus (see Figure 50). The stimulus set was chosen based on the following considerations. The set had to be small to enable testing all stimuli for every cell without significant loss of recording time. The white noise stimulus was included to compare results to previous studies [Leonardo 2004; Kozhevnikov and Fee 2007]. The introductory note and long call stimulus were included because broad band noise often fails to activate neurons of higher auditory areas. Note however that the exact choice of the two call stimuli was relatively arbitrary and was simply guided by the idea of using salient stimuli.

![Figure 50: Perturbation stimuli.](image)

Initially, experiments were performed using one single perturbing stimulus sound-amplitude level, set to 6 dB above maximum sound amplitude level of song, as measured on the microphone. Later, experiments were performed using three different, discreet sound-amplitude levels, max sound-amplitude level of song +0 dB, +6 dB and +12 dB, alternated at random. This corresponds to equal, double and quadruple perturbing stimulus sound amplitude, as compared to maximum sound amplitude during song.

For every cell recorded, we first determined the perturbing stimulus that elicited the largest response when presented in isolation and then used this stimulus for the rest of the experiments on this cell. This was done simply to ensure that the selected stimulus was capable of driving the cell, and thus to avoid having experiments where cells showed responses to neither song nor playback perturbations. Quantification of response was done visually “on the fly” and no attempt was made to quantify the different responses objectively. If for a particular cell none of the stimuli triggered a significant response or responses were approximately equal, we used the long call stimulus.
Perturbations were presented at random with a probability of 50%. To minimize possible effects of song deterioration, perturbations were confined to recording sessions only (lasting roughly 0.5-2 h each).

### 8.2 Data acquisition

A zebra finch will only sing if it is allowed to move freely and is comfortable in its environment. To record single unit neural activity in a freely behaving zebra finch we used two different versions of an implantable miniature motorized microdrive based on the design of Fee and Leonardo [Fee and Leonardo 2001].

#### 8.2.1 Three channel microdrive

Initially the experiments (the field L and NIf recordings) were performed with a three channel motorized microdrive, that contained three independently movable electrodes.

![Three channel microdrive](image)

Figure 51: Three channel microdrive.

However, with a weight of 1.5 g and an assembly time of roughly 6 hours the disadvantages of the design outweighed the advantage of being able to record from three neurons simultaneously, such that we went on to design a simpler and lighter single channel microdrive. For brevity we will only describe the mechanical design and details of the assembly procedure of this new version of the microdrive here.

#### 8.2.2 Single channel microdrive

This version of the microdrive was designed with the help of Yannick Riesen who drew the CAD plans and programmed the Labview microdrive controller software during his semester project. The microdrive without tether cable and dental cement (used as fixation for implantation) weighs roughly 600 mg and can be assembled in 2-4 hours. The mechanical components are the body (1), the tip (2), the shuttle (3), the motor (4 – Faulhaber Minimotor), the Omnetics connector (5), and the lateral positioner (see Figure 52). In the following we describe the associated assembly and implantation procedures.

**List of components:**

1. Body, tip and shuttle, custom made (see section 9.2 of the Appendix for blueprints, nr. 1,2 & 3 in Figure 52).
ii. M 0.5 Bergeon screws (to connect tip and body and as lateral positioner) 2.25 mm length. No. 2646A-1L-0.5.

iii. Faulhaber Minimotor micro brushless DC-motor 0206 A 0.5 B 021 47:1 2825. (nr. 4 & 8 in Figure 52).

iv. Faulhaber Minimotor MCBL 07002 motion controller.

v. MicroLumen polyimide tubing 745-IV, inside diameter 0.0745 inch (1.89 mm), used as motor casing (nr. 9 in Figure 52).

vi. Omnetics connector, NPD-14-DD 14 Pin dual row male nano connector (nr. 5 in Figure 52) and the corresponding female connector for assembly and tether cable NSD-14-WD-20.0 (nr. 10 in Figure 52).

vii. A-M Systems polyimide tubing, inside diameter 0.0040 inch, wall 0.0005 inch, catalog # 823600, used as guide tube for the electrode (nr. 4 in Figure 53).

viii. Thomas recording microelectrode, Type EF8025, pulled to specified impedance.

ix. A-M Systems platinum-iridium wire (0.001 inch bare, 0.003 inch coated) part nr. 775000 (nr. 5 in Figure 53).

x. A-M Systems silver wire (0.005 inch bare, 0.007 inch coated) part nr. 786000 (nr. 1 in Figure 53).

xi. Hardman extra fast setting non-sag epoxy, #04008, used for all glue joints.

xii. Heraeus Paladur dental cement, used to fix the microdrive to the skull during implantation.

xiii. Wax oil mixture. 60 % paraffin wax and 40 % mineral oil.

*Figure 52: Single channel microdrive after primary assembly (before electrode is mounted). 1: body, 2: tip, 3: shuttle, 4: motor, 5: omnetics connector, 6: lateral positioner, 7: fixation bar, 8: gear box, 9: polyimide tubing encasing motor, 10: assembly mount.*
Mounting the motor on the body (primary assembly see Figure 52):

1. Connect the body to the tip with two M0.5 (2.25 mm long) screws.

2. Cut the pins of the bottom (top) row of a male omnetics connector to 2 mm (1 mm) length, and glue the connector to the flat surface of the body (as shown in Figure 52). The exact placement of the connector on the body is not crucial and depends on the tether cable used.

3. Cut a piece (~4mm) of microlumen polyimide tubing (diam. 1.89mm, not labeled on package, see “List of components”) and encase the motor and gearbox with it. Motor and gearbox can easily be pulled apart if not fastened. Put a shuttle on the motor axis such that the tip of the axis protrudes from the shuttle by about 0.5mm. Insert the motor into the assembled body such that the end of the axis fits into the central hole of the tip and the polyimide tube encasing the motor just touches the body (take care to place the motor cables at the desired location relative to the omnetics plug) and glue gearbox, body and polyimide coating by placing a drop of non-sag glue in the shuttle guide slit. Be careful not to glue the axis to the motor. Also strengthen the motor cables with a drop of glue where they leave the motor. Once hardened, solder the motor cables to the appropriate pins on the omnetics connector.

4. Test the motor! Whenever using a new motor, or a new motion controller, test the motor with a small piece of paraffin wax placed on the motor casing. Motors will overheat and melt within seconds in case of a malfunction (e.g. in case of a short circuit between any of the motor cables).

5. On the tip insert the lateral positioner screw, and a fixation bar (anything that fits into the 0.5 mm hole, and protrudes on both sides by roughly 1 mm will do) and glue it in place. At this stage the microdrive assembly should look as depicted in Figure 52.

Mounting an electrode on the MD (the following steps assume that you have already completed primary assembly):

1. Cut an approx. 3 cm piece of 0.005” mm diameter coated silver wire (see “List of components”) and uncoat 3 mm on one end. Insert the uncoated end of the silver wire into the hole in the shuttle from below, such that the uncoated part protrudes from the top of the shuttle. To stabilize the wire it to the side both on top and on the bottom of the shuttle and glue it firmly to the shuttle. Make sure that enough of the uncoated end of the wire
remains glue-free to solder the electrode to it and that you do not accidentally glue the shuttle to the body. Lead the loose end of the wire around the shuttle relatively tightly, cut it to length and solder it to the appropriate Omnetics pin.

2. Use a short piece of elastic string to tie the silver wire leading from the shuttle to the Omnetics connector, to the fixation bar (see nr. 2 in Figure 53). Make sure that the elastic string is not completely relaxed even if the shuttle is in the lowest position.

3. Cut a 3 cm piece of 0.001 inch diameter coated platinum-iridium wire (see “List of components”) and uncoat and solder one end to the virtual ground pin on the Omnetics connector. Lead the wire in the shortest possible path towards the tip of the microdrive and glue it in place (this might require holding it in place with a toothpick). At the tip bend and cut it such that the tip of the wire is parallel to the electrode guide channel and extends 0.5-1 mm from the tip. Uncoat the wire on the first 200 um (see nr. 5 in Figure 53). This can be done using a cauterizer, or for the more dexterous microdrive builder, using cutting forceps.

4. Cut a 6 cm piece of 0.005” mm diameter coated silver wire (see “List of components”) and uncoat the wire everywhere except for the first 15 mm. Uncoat another 3mm from the still coated end and solder it to the ground pin on the Omnetics connector. Lead the wire to the end of fixation bar on the opposite side of the silver wire leading to the shuttle and wrap it once around and glue it in place.

5. Position the shuttle as close as possible to the motor end of the axis. Be sure not to block the shuttle by driving it up the axis too far.

6. Cut a ~6.5 mm piece of polyimide tubing (.0040” inside diameter, see “List of components”) to use as a guide tube for the electrode. The guide tube should span the entire length of the electrode guide channel and protrude at the tip by about 500-1000 μm. Cut the desired electrode to length (~12.5 mm – measure and adjust this value for every experiment. Ideally the tip of the electrode should be just flush with the tip of the guide tube when the shuttle is fully retracted. This prevents accidental damage to the electrode tip during assembly.) and uncoat 1 mm by crushing the glass coating with forceps. Now carefully slip the guide tube over back end of the electrode and place it such that it covers the tip.

7. Place the electrode (with the guide tube) in the electrode guide channel such that the back end rests on the shuttle. Make sure the electrode lies flat in the electrode guide channel, glue the electrode in place on the shuttle and solder the uncoated end of the electrode to the silver wire using Cerro Matrix. Fix the guide tube with a drop of non-sag glue in place at the end of the electrode guide channel closer to the shuttle (such that it is fixed at the top and can be moved by the lateral positioned at the bottom).

8. Cover both the lateral positioned screw and the electrode guide channel with wax to prevent them from being fixed in place by the dental acrylic used during implantation.

9. Cut and bend a piece (~ 4 x 2 mm) clear plastic foil (use e.g. a piece an overhead projector transparency) to shape and glue it to the body to cover the electrode to prevent the bird from reaching it with a claw.
10. If planning to implant stimulation electrodes, cut two ~5 cm pieces of copper wire, uncoat and tin-coat both ends. Bend one end of each wire to a hook and solder the others to the appropriate stimulation pins on the Omnetics connector. Glue both wires to the body close to the Omnetics connector to avoid putting strain on the solder joint during implantation.

11. For implantation the electrode should extend by about 200 μm from the tip of the guide tube. (Extend maximally if planning to search for the target location, e.g. NIf.)

**Implantation procedure.** The following description assumes basic knowledge of general surgery and anesthesia protocols. Expect the entire procedure to last 2-4 hours (not including the time required to find the target area, e.g. NIf).

1. Induce anesthesia with 3-5% Isofluorane in nitrous oxide or oxygen. Measure depth of anesthesia by monitoring respiratory frequency. This can be done either visually by observing the tail or neck (the latter of which is more convenient when working under the microscope) feathers, that move up and down synchronously with the respiratory rhythm. Normal respiratory frequency is in the range of 2-3 Hz. Depth of anesthesia should be adjusted such that the respiratory frequency is 1 Hz. Reduce the Isofluorane concentration to 2% as soon as this level is reached. Continuously monitor respiratory frequency throughout the entire surgery and adjust the Isofluorane concentration to keep it at 1 Hz (this might require concentrations as low as .5% or as high as 5%).

2. Measure and adjust the head angle with the head angle measurement tool. This method is favorable over measuring the head angle directly on the skull with a bar, because it alleviates the need to have the incision in the skin (as described in 4) extend all the way to the beak, thus greatly reducing the extent of the inflicted wound.

3. Remove the feathers on the head, and apply Xylocaine gel to the scalp and wait for 10 minutes before proceeding to let it reach full effectiveness.

4. Cut with a scalpel the skin from about 1 cm anterior to about 0.5 cm posterior of the estimated position of lambda. Avoid damaging large blood vessels in the skin (this sometimes requires cutting in S-like-shapes).

5. Drill away the upper bone layer over lambda and in at least 1.5 x 1.5 mm region centered on the planned implantation location.

6. Drill 3-4 sets of two holes spaced by 1-2 mm around the opening in the upper bone layer and insert insect pins.

7. Implant stimulation electrodes if required.

8. Measure and mark the exact location of implantation on the second bone layer. Cut 500 x 500 μm opening in the second bone layer by incising it along the edges with a hypodermic needle and removing the piece of bone with fine forceps.

9. Place the microdrive in a stereotax holder, adjust it to be exactly vertical and move it to the correct AP and ML coordinates. Mark the point of entry on the dura of both electrode and virtual ground wire and retract the microdrive.
10. Use a syringe to fill the electrode guide tube with mineral oil. Holding a drop of oil to the tip of the guide tube suffices to fill it.

11. Use a fresh hypodermic needle to make a small hole in the dura by puncturing it at the point of entry of electrode and virtual ground wire.

12. Lower the microdrive until the guide tube penetrates the surface of the brain by 50-100 µm. Make sure that also the virtual-ground wire is in the brain.

13. Fill the space between microdrive and brain surrounding the guide tube with wax (60% paraffin wax and 40% mineral oil) to ensure that the electrode and guide tube are not cemented in place by the dental cement.

14. Fix the microdrive to the skull with a few drops of dental acrylic, and allow it to dry.

15. Drill a small hole in the outer bone layer and insert the ground wire. If implanting stimulation electrodes connect them to the appropriate copper wires attached to the Omnetics connector on the microdrive.

16. Cover the entire area of the exposed skull (including ground and stimulation wires) with dental acrylic to fix the microdrive firmly in place. Be sure to only contact the tip and not the body of the microdrive with the dental acrylic. If any dental acrylic comes in direct contact with the electrode (especially at the upper end of the guide tube) the electrode will not be moveable.

**Note on electrodes.** For most of the experiments we used Thomas Recording glass coated platinum-tungsten electrodes with impedances in the range of 4-10 MΩhms. We have successfully mounted and used other types of electrodes, such as nichrome wire tetrodes and conventional epoxy-resin-coated metal electrodes. The main advantage of the glass coated electrodes (especially when used with the three channel microdrive where you have no visual control of the electrodes once the microdrive is implanted) is that the electrode will break if it is clogged and the shuttle is moved downwards (typically the electrode will break when the shuttle has moved 500 µm). With an electrode that will bend instead of break it takes a lot of experience and usually much more time to realize that the electrode is no longer moving. The choice of impedance depends on the targeted recording site. In the areas L1, L3, NCM and the dorsal part of CLM 4-6 MΩhm electrodes should produce good single unit isolation, and in the areas NIf, HVC, para HVC and the ventral part of CLM (and possibly dorsal and ventral parts of NCM), 6-10 MΩhm electrodes should produce good single unit isolation.

### 8.2.3 Troubleshooting microdrive recordings

1. **Large movement artifacts and no neural signal (spikes).** Combined with an increase of noise RMS, this is a good indication for an open loop signal, i.e. a broken electrode or a loose connection. This can be tested by moving the electrode into a non-essential brain area (i.e. hippocampus) and passing a small current between electrode and virtual ground. The most common place of a loose contact is the connection between electrode and silver wire on the shuttle. This can be easily fixed by resoldering the connection with Cerro Matrix and gluing the silver wire firmly to the shuttle. If this does not fix the problem, and the gold pins on the omnetics connectors are all clean, then most likely the problem is the electrode itself. Usually a broken electrode implies the end of the experiment. The design of single channel microdrive would in principle allow for online electrode replacement.
2. **Good neural signal (spikes) but large movement artifacts.** This is either caused by a bad connection in the tether cable, or a grounding problem. Grounding problems can be easily detected by testing if movement artifacts are reduced in amplitude when holding – and thus grounding - the animal and manually moving the tether cable. In case of a grounding problem a second ground wire can be implanted. If replacement of the tether cable solves the problem, the most likely cause is corrosion in the tether cable. Check the cable for corrosion (black discoloration) of the copper wires close to the connectors. Corrosion of the cables is greatly accelerated by cleaning them with Deconnex or water but can also occur naturally with use.

3. **Audible sound on the electrode channel when the bird sings (microphonics).** This is caused by a resonating wire or electrode. The most likely cause is the silver wire connecting the electrode to the Omnetics connector. This problem can be solved by putting more tension on the wire by fitting or readjusting the elastic string.

4. **Electrode does not move (clogged electrode).** This is typically caused by either dental acrylic (during the surgery) or blood (throughout the experiment) leaking into the electrode guide tube. This can be avoided by implanting the microdrive such that the guide tube penetrates the surface of the brain by 100-200 μm. If left unmoved for more than a few hours (i.e. overnight) the electrode will usually be stuck. To avoid breaking it when moving it down, always retract the electrode for 300-500 μm before moving downwards again.

5. **The shuttle does not immediately reverse direction when the motor reverses direction (Backlash).** On the microdrive there are two sources of backlash: a) the thread on the shuttle is cut relatively loose to guarantee smooth travel, especially on shuttles that have been used multiple times this can lead to large backlash, b) the gearbox of the Faulhaber motor itself has large backlash. As of now there are no workarounds for this problem. This backlash has significant effects on the recording protocol. Moving the electrode back and forth to find the place of best signal – as can be done in an acute head fixed recording – fails to work. Retracting the electrode on the microdrive almost invariably leads to signal decrease. Further complicating the matter is the fact that, because the shuttle can slightly tilt and thus function as a lever, the electrode will initially continue to move downwards when the shuttle is moved back up and will slightly shift in horizontal location.

6. **The motor overheats and melts.** When using the Faulhaber MCBL 07002 motion controller, this can occur if one of the three motor control cables is shorted to ground. If this happens in the setup, the resulting current can trigger the Hall sensor leading to a spinning commutator. Aside from damaging microdrive and tether cable, THIS WILL SERIOUSLY HARM THE ANIMAL!

7. **The same neurons (or none at all) are encountered on every penetration.** This problem requires a lot of experience to be diagnosed correctly. Most likely cause is a non-functioning lateral positioner. This will occur if during implantation dental acrylic penetrates the wax surrounding the tip of the microdrive and fixes guide tube in place. Be aware that the dental acrylic solvent contains alcohol and thus dissolves the wax to a certain degree.

### 8.2.4 Recording protocol

The recording protocol that has proven to be most successful was as follows. The electrode is advanced in steps of ~10 um until spikes are audible on the sound monitor. Usually at this point
spikes are not or only barely visible on the oscilloscope. Pause for 30 s or more to allow the tissue to relax from compression, this typically already improves signal. The electrode is again advanced in steps of 2-5 um until spikes on oscilloscope become sortable (S/N > 3) or until they start decreasing in amplitude. Attempting to retract the electrode at this point almost invariably leads to a decrease in spike amplitude, most likely due to backlash (as described in the section Troubleshooting microdrive recordings). For perturbation experiments the setting of the perturbation probability is now increased from 0% to 50%. If the bird is already singing, 10-20 song motifs are recorded before playback trials, otherwise if the bird is quiescent, playback trials are initiated immediately. If the bird refuses to sing the female can be removed from the recording chamber for a few minutes to elicit directed song upon replacement. Song and playback trial sets are then alternated (10-20 motifs of song and one set of playback trials) until the neuron is lost. Avoid recording long stretches of just song or just playback trials, because in the event that song and playback related activity differ, one can no longer exclude the possibility that the recordings are from two different neurons if song and playback trials were not intermingled. The time for which stable recording was possible was typically short (on the order of 5-20 minutes, see Figure 54).

![Figure 54: Average stable recording time.](image)

Perturbation probability is again set to 0% to minimize the chance of adaptation effects. Search is continued in this manner until the maximum depth is reached at which point the electrode is slowly fully retracted again. The bird is removed from the recording setup and the lateral positioned is advanced by 1/8 or 1/4 of a turn of the screw. Attempting to record from the same lateral coordinate a second time will, if anything, lead to the same neurons being recorded again, and should be avoided. After being handled, the bird normally will not sing for a few hours. In this way up to two penetrations can be completed per day.

### 8.2.5 Time course of experiment

Birds typically recovered from surgery within 1-2 days, and were back to normal song activity levels (1000-4000 motifs per day) within 3-4 days. Attempts to record before the bird had regained normal song activity levels were usually unsuccessful due to a lack of song trials. (see Figure 55)
Figure 55: Experimental time window. Histogram of the number of recording sites as a function days post surgery. Light gray shading indicates recordings where the experiment could not be completed because the bird did not sing enough or not at all, or recordings where spikes were not sortable. Dark gray shading indicates recording sites that were sortable and for which enough data for further analysis could be collected. Recordings were only attempted once the bird had reached normal song activity levels. Data from a total of 293 recording sites in seven birds are shown.

There was a large variability of baseline song activity levels across birds. “Bad” singers with low song activity levels (as determined from pre-implantation song recordings) were not used for microdrive experiments. Also, to maximize experimental yield, recording schedules were adapted as well as possible to the birds’ individual daily song activity level pattern, that – interestingly – were usually well preserved across days. Examples of such daily patterns are shown in Figure 56 and Figure 57.

Figure 56: Vocal activity timeline I. Vocal activity histograms (measured as number of files recorded per time) of two brothers recorded simultaneously on seven consecutive days. Bold black line is an average. The number of files recorded is not a direct measure of the number of motifs sung, but a good correlate thereof. Recordings were triggered if a certain number of pitch and duration criterions were fulfilled. These criterions were typically only fulfilled by sequences of syllables as in song bouts, or series of consecutive calls. Note that both birds have peaks in activity that seem to be conserved across days (y5o-s3k: 7:00 – 11:00 and 11:30 – 12:30, y6o-s3k: 12:00 – 13:00). Birds were on a 14/10 hours light/dark schedule (gray shading indicates lights off). Exact lights on or off times could not be reconstructed and are estimated based on the following considerations. Birds start singing 20 to 60 minutes after lights on and remain active for a short period after lights off.
8.2.6 Reconstruction of recording sites

Throughout the experiment electrode position was monitored by integration motor commands sent to the motors, using either the indicator on the MP 285 motor controller used for the three channel microdrive or the depth reading of the Labview motor controller used for the single channel microdrive. Unfortunately this method is rather unreliable because the measured depth is a correlate of motor command (i.e. desired position of motor) not actual motor position. Thus it is oblivious to the fact that the motor does not always turn on command, as for example when the shuttle briefly gets stuck and comes loose again, a situation not infrequently occurring. To partially overcome this problem we measured the depth of the shuttle mechanically using specially constructed depth gages both at the beginning and the end of every penetration.

At the end of each experiment, small lesions were made in vicinity of the recording sites. Using 4-10 MOhm electrodes, currents of 15 $\mu$A applied for a duration of 15 s resulted in lesions of roughly 100-200 um in diameter. Immediately thereafter animals were killed with a pentobarbital (Nembutal) overdose, and the brain was removed for histological examination of unstained slices to confirm location of recording sites. In general, size and shape of lesion relative to the electrode tip were too unpredictable and too variable to reconstruct electrode position with an accuracy of more than 50 um.

Overall these considerations lead to an estimate of the recording site reconstruction error of 100 um along the dorso-ventral axis. Because electrode tracks were usually visible in tissue slices, errors in both the anterior-posterior and medial-lateral axis were considerably smaller.

8.2.7 Real-time song recognizer

In order to be able to trigger perturbations time locked to a certain song syllable with a very high temporal accuracy we developed a real time recognizer. This was done in collaboration with Claude Wang who used the same recognizer for his stimulation experiments [Wang et al. 2008]. For brevity we shall only describe the design concepts employed and not the details of the implementation (we
originally used a Matlab Simulink version of the recognizer, Claude later went on to implement a Labview version).

The problem of syllable recognition – or strictly speaking syllable classification – can be reduced to two design choices, the features extracted from the sound data and the classifier used. In our case the latter choice fortunately turned out to have only little influence on the recognition performance. We used a two layer neural net with a hyperbolic tangent nonlinearity in the hidden layer and a single output. However, also support vector machines or even simple perceptrons produced reasonable results. Of greater importance proved to be the choice of features extracted from the sound data. The key idea as presented in the following evolved in a discussion with Daniel Ben Dayan Rubin. The incoming sound data was sampled at 33 kHz and buffered in bins of 256 data points (7.8 ms) overlapping by 128 data points (3.9 ms). Each buffer was Fourier transformed to calculate the total energy and the energy in 19 linearly spaced frequency bands in the range of 0-10 kHz. The final input data vector was then composed of the energies of the current buffer and the energies of a history of 5 buffers, each spaced by one buffer, resulting in a data vector with 120 entries. Thus, loosely formulated, the classification worked on a structure similar to that of a spectrogram of a 46.8 ms piece of sound data.

![Figure 58: Functioning principle of the real time recognizer. Incoming data is buffered, Fourier transformed and vectorized.](image)

In this way we were able to train recognizers for virtually any time point in the motif (except for the first 20-40 ms), with a standard deviation of detection time of less than 4 ms. Detection reliability was typically very high (less than 1% false negatives and false positives) with a well trained recognizer.

### 8.3 Deafening procedure

Note that this surgery poses considerable difficulties and will need a lot of practice to be reliably performed with success\(^69\). We will point to the steps of particular importance or difficulty in the description below. The description assumes basic knowledge of general surgery and anesthesia protocols. Note also that the method described here does not correspond to the one developed by Konishi [Konishi 1964].

1. Anesthetize the bird using gas or preferably injection anesthesia. The latter method will slightly ease the handling of the animal during the surgery (gas anesthesia requires both an

\(^{69}\) A process that can easily require 10-20 practice surgeries (ideally performed on dead animals), and that should only be attempted by a manually skilled experimenter.
intubation tube leading to the animal and a suction tube placed in close proximity). However when inexperienced with the surgery, use a method of anesthesia you are familiar with, i.e. can perform with a high success rate.

2. Wrap the bird in a piece of cloth and fix it with tape to a small board (approx. 10 x 15 cm) that can be tilted and fixed at arbitrary angles (this can be achieved with a clamp on a ball joint swivel arm or with a chunk of plasticine, throughout this procedure you will have to adapt both board and microscope position and angle to keep the working area visible). Initially the board should be at a 70 degree angle relative to horizontal, and the animal should lie on its side with the beak pointing up. If using gas anesthesia take care to properly fix the intubation tube and keep the suction tube in close proximity of the head of the bird (this is done most advantageously by using a board with small holes and placing the suction tube on the opposite side of the board close to the birds head).

3. Remove the feathers surrounding the ear canal. Using forceps pull the skin lining the inside of the ear canal out of the ear canal (the skin is not attached to the underlying tissue in the ear canal) and cut a hole in it using surgical scissors.

4. The tympanic membrane should now be directly visible as a milky semi-transparent “foil” that is spanned taut across the entire breadth of the ear canal. In the center of the tympanic membrane there is a small upward protrusion revealing the location of the columella (the avian equivalent of the mammalian ossicles). Using forceps puncture the tympanic membrane somewhere in the vicinity of the columella and remove it from the ear canal.

5. Mark the exact location of the columella and, using a drill or forceps, clear out all the small bone trabecula in the vicinity of the columella. Exact structure of the bone surrounding the columella varies greatly with different birds. An inexperienced experimenter will easily lose sight of the columella during this step. Note also that there is large blood vessels running through the inner ear cavity, damage to which will almost inevitably lead to failure of the surgery.

6. Change the placement of the bird beneath the microscope such that you can see as much of the oval window as possible (in some birds this is not possible due to the geometry of the inner ear, in which case you will need to do the following steps blindly, see Figure 59). Success of the surgery correlates strongly with how well you are able to get the oval window into view.

7. Grasp the neck of the columella with forceps as close to the oval window as possible and pull gently on the columella to remove the foot of the columella from the oval window. The neck of the columella is fragile and is easily torn if the direction of the pull is not exactly perpendicular to the oval window or pressure exerted on the forceps is too large. If the neck is torn, try again with what is left of it, or if no longer possible, reach directly into the oval window with forceps and remove the rest of the columella.

8. Using a sharpened wire with a hooked tip\textsuperscript{70} remove the cochlea as follows. Advance the hook into the cochlear duct by at most 500um, twist the wire 90 to 180 degrees around its axis to

\textsuperscript{70} This can be produced as follows. Sharpen the tip of a thin steel wire by holding it in the flame of a welding torch. To ease later handling you can stick the other tip of the wire into the end of a small wooden rod, e.g. a
increase the chance of the hook catching on and retract the hook. This step will potentially have to be repeated multiple times until successfully extracting the cochlea. If you have trouble finding the cochlea try varying the angle of penetration. Note that the blood vessels of the inner ear are easily damaged when retracting the hook improvidently.

9. Check that the entire cochlea has been removed (see Figure 59).

10. Close the skin over the ear canal by sowing it shut or using cyanoacrylate to glue it shut. Failure to do so will result in a gradual decline in the bird’s condition post operatively over the course of one to two days, eventually leading to severe vestibular disturbances and death.

11. Repeat the procedure as required on the contralateral ear.

8.4 Data analysis

Extracellular voltage traces were digitized at 33 kHz and recorded for offline spike sorting. Subsequent trials were aligned by sound energy threshold crossing to one of the first syllables in the motif. Unless noted otherwise, firing rates were calculated in 5 ms bins and averaged over multiple renditions or playbacks of the motif. In addition, to reduce the effect of the binning on the visual display firing rates were smoothed with a 10-ms Gaussian kernel for plotting in all figures. This representation of firing rate is comparably robust to misalignment due to variability in motif length but does not well capture the high instantaneous firing rates during bursts. Instantaneous firing rates in bursts of NIfHVC neurons for example ranged up to 600Hz.

Perturbation responses. Differences between average firing rates in unperturbed and perturbed trials were assessed using a WRS test (p=0.05) on the number of spikes in 30-ms time windows. Windows were shifted in 5 ms steps and only sequences of at least two subsequent windows with p<0.05 were considered significantly different.
We defined the response bias as \( b = \frac{|d_f| - |d_p|}{|d_f| + |d_p|} \), where \( d_f \) is the feedback and \( d_p \) the playback perturbation response, each defined as the difference in average firing rates between perturbed and unperturbed conditions in a time window extending from the time of onset of perturbation up to 50 ms post offset. We defined the response selectivity as \( s = d_f / r_s \), where \( r_s \) is the average firing rate during unperturbed song motifs.

Firing stereotypy. The stereotypes of song and playback-related spike patterns were assessed by the average coherency of spike rasters at zero time lag (sound traces were aligned by the amplitude onset of the detected syllable). Differences between song and playback-related firing stereotypies were detected using a WRS test (same BOS: \( p = 0.3 \), different BOS: \( p < 10^{-7} \)).

Z score. To assess changes from background firing, we calculated the Z score of song/playback responses as

\[
Z = \frac{\mu_S - \mu_B}{\sqrt{\sigma_S^2 + \sigma_B^2 - 2 \text{cov}(S,B)}},
\]

where \( \mu_S \) is the average firing rate during song/playback, \( \mu_B \) is the average baseline firing rate assessed in 1-3 s silent intervals interleaved between playbacks and songs, and \( \sigma_S \) and \( \sigma_B \) are the respective standard deviations. Due to a lack of pairing of vocal and playback conditions we set the covariance term \( \text{cov}(S,B) \) to 0. We considered cells with a Z score > 0.75 to be significantly excited, and cells with a Z score < -0.75 to be significantly inhibited (the criterion \(|Z\) score\| > 0.75 corresponds to \( p < 0.05 \) if the number of trials >5, which was the case for all our cells).

RMS voltage trace. Multiunit activity was assessed by calculating the root mean square (RMS) of the voltage trace in a 3.9 ms (corresponding to 128 samples) moving window. This value was then thresholded to yield a binary function. Threshold was chosen for each bird separately, but was identical for all analysis of an individual bird.

Cross correlation. To assess temporal relationship between neural activity and song, we calculated the normalized cross correlation between the rhythm functions \( r(t) \) and spike trains \( \rho(t) \) or RMS voltage traces \( m(t) \):

\[
C_{rp}(t) = \frac{\int_0^T r(\tau) \cdot \rho(t+\tau) d\tau}{\int_0^T \rho(\tau) d\tau} \quad \text{and} \quad C_{rm}(t) = \frac{\int_0^T r(\tau) \cdot m(t+\tau) d\tau}{\int_0^T m(\tau) d\tau}.
\]

These correlation functions measure the amount of neural activity that coincides with syllables under a time shift \( t \). For example, \( C_{rp}(t) \) represents the average rhythm function at a time shift \( t \) before a spike (or after a spike if \( t \) is negative). To measure significance of these cross correlation functions for a particular neuron we used all the data of that neuron, recalculated normalized cross correlation with random time shifts of the spike trains to estimate the probability distribution of correlation values. From this distribution we measured the 95\(^{\text{th}}\) percentile which corresponds to the \( p = 0.05 \) significance level.
9 APPENDIX

9.1 Summary of abbreviations:

9.1.1 General

BOS  Birds own song
CON  Song of conspecific birds
dph  days post hatch
IEG  immediate early gene
PS   Perturbation stimulus
RMS  Root mean squared
SPL  Sound pressure level
STDP Spike timing dependent plasticity
TUT  Tutor song
WRS  Wilcoxon rank sum test
ZENK an acronym for the avian orthologue of the mammalian genes zif-268, egr-1, ngnf-1 and krox-24

9.1.2 Data analysis

b  Response bias (|d_i|−|d_p|)/(|d_i|+|d_p|)
d_c  Difference in average firing rates between perturbed and unperturbed conditions in a time window preceding the perturbation window
d_f  Response induced by perturbation during song (feedback)
d_p  Response induced by perturbation during BOS playback
r_s  average firing rate during unperturbed song motifs
s  Response selectivity d_f/r_s
s_c  Response selectivity that would arise by chance due to small sample sizes d_f/r_s

9.1.3 Avian brain anatomy

All abbreviations and names are in accord with the revised avian nomenclature.

<table>
<thead>
<tr>
<th>Full name</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA  Anterior arcopallium</td>
<td>Arcopallium</td>
</tr>
<tr>
<td>AAC Central nucleus of anterior arcopallium</td>
<td>Arcopallium</td>
</tr>
<tr>
<td>Ac  Nucleus accumbens (ventromedial rostral LPO)</td>
<td>Subpallium (Striatum, pallidum)</td>
</tr>
<tr>
<td>AD  Dorsal arcopallium</td>
<td>Arcopallium</td>
</tr>
<tr>
<td>AFP Anterior forebrain pathway</td>
<td>Pallium, striatum, thalamus</td>
</tr>
<tr>
<td>AI  Intermediate arcopallium</td>
<td>Arcopallium</td>
</tr>
<tr>
<td>AM  Medial arcopallium</td>
<td>Arcopallium</td>
</tr>
<tr>
<td>APH Area parahippocampal</td>
<td>Pallium</td>
</tr>
<tr>
<td>Bas The basorostral pallial nucleus</td>
<td>Nidopallium</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>BSTL</td>
<td>Lateral part of the bed nucleus of the stria terminalis</td>
</tr>
<tr>
<td>BSTM</td>
<td>Medial part of the bed nucleus of the stria terminalis</td>
</tr>
<tr>
<td>CDL</td>
<td>Area corticoidea dorsolateralis</td>
</tr>
<tr>
<td>CLM</td>
<td>Caudolateral mesopallium</td>
</tr>
<tr>
<td>CMM</td>
<td>Caudomedial mesopallium</td>
</tr>
<tr>
<td>CPI</td>
<td>Cortex piriformis</td>
</tr>
<tr>
<td>CSt</td>
<td>Caudal striatum</td>
</tr>
<tr>
<td>DA</td>
<td>Dorsal arcopallial tract</td>
</tr>
<tr>
<td>E</td>
<td>The entopallial nucleus or the entopallium</td>
</tr>
<tr>
<td>FA</td>
<td>Fronto-arcopallial tract</td>
</tr>
<tr>
<td>GP</td>
<td>Globus pallidus</td>
</tr>
<tr>
<td>HA</td>
<td>Apical part of the hyperpallium</td>
</tr>
<tr>
<td>HD</td>
<td>Densocellular part of the hyperpallium</td>
</tr>
<tr>
<td>HI</td>
<td>Intercalated part of the hyperpallium</td>
</tr>
<tr>
<td>HOM</td>
<td>Hypothalamic part of occipitomesencephalic tract</td>
</tr>
<tr>
<td>Hp</td>
<td>Hippocampus</td>
</tr>
<tr>
<td>HVC</td>
<td>HVC</td>
</tr>
<tr>
<td>IHA</td>
<td>Interstitial part of the hyperpallium apicale</td>
</tr>
<tr>
<td>IMN</td>
<td>Intermediate medial mesopallium</td>
</tr>
<tr>
<td>INP</td>
<td>Intrapeduncular nucleus</td>
</tr>
<tr>
<td>L</td>
<td>Field L</td>
</tr>
<tr>
<td>LAD</td>
<td>Dorsal arcopallial lamina</td>
</tr>
<tr>
<td>LaM</td>
<td>Mesopallial lamina</td>
</tr>
<tr>
<td>LL</td>
<td>Lateral lemniscal nuclei</td>
</tr>
<tr>
<td>LMAN</td>
<td>Lateral magnocellular nucleus of anterior nidopallium</td>
</tr>
<tr>
<td>Lst</td>
<td>Lateral striatum</td>
</tr>
<tr>
<td>MD</td>
<td>Dorsal mesopallium</td>
</tr>
<tr>
<td>MFL</td>
<td>Supreme frontal lamina</td>
</tr>
<tr>
<td>MLD</td>
<td>Mesencephalicus lateralis dorsalis</td>
</tr>
<tr>
<td>MMAN</td>
<td>Medial magnocellular nucleus of anterior nidopallium</td>
</tr>
<tr>
<td>MO</td>
<td>Oval nucleus of mesopallium</td>
</tr>
<tr>
<td>MSt</td>
<td>Medial striatum</td>
</tr>
<tr>
<td>MSTm</td>
<td>Magnocellular part of medial striatum</td>
</tr>
<tr>
<td>MV</td>
<td>Ventral mesopallium</td>
</tr>
<tr>
<td>NA</td>
<td>Nucleus angularis</td>
</tr>
<tr>
<td>NAO</td>
<td>Oval nucleus of the anterior nidopallium</td>
</tr>
<tr>
<td>NAOH</td>
<td>Medial oval nucleus of the anterior nidopallium</td>
</tr>
<tr>
<td>NBM</td>
<td>Basal magnocellular cholinergic nucleus</td>
</tr>
<tr>
<td>NC</td>
<td>Caudal nidopallium</td>
</tr>
<tr>
<td>NCL</td>
<td>Caudolateral nidopallium</td>
</tr>
<tr>
<td>NCM</td>
<td>Caudal medial nidopallium</td>
</tr>
<tr>
<td>NDB</td>
<td>Nucleus of the diagonal band</td>
</tr>
<tr>
<td>NF</td>
<td>Frontal nidopallium</td>
</tr>
<tr>
<td>NI</td>
<td>Intermediate nidopallium</td>
</tr>
<tr>
<td>NIF</td>
<td>Nucleus interface of the nidopallium</td>
</tr>
<tr>
<td>NL</td>
<td>Nucleus laminaris</td>
</tr>
<tr>
<td>Code</td>
<td>Term</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>NLC</td>
<td>Central nucleus of the lateral nidopallium</td>
</tr>
<tr>
<td>NM</td>
<td>Nucleus magnocellularis</td>
</tr>
<tr>
<td>OM</td>
<td>Occipito-mesencephalic tract</td>
</tr>
<tr>
<td>OTu</td>
<td>Olfactory tubercle</td>
</tr>
<tr>
<td>Ov</td>
<td>Nucleus ovoidalis</td>
</tr>
<tr>
<td>PoA</td>
<td>Posterior nucleus of the pallial amygdala</td>
</tr>
<tr>
<td>PPT</td>
<td>Pedunculopontine tegmental nucleus</td>
</tr>
<tr>
<td>PSL</td>
<td>Pallial-subpallial lamina</td>
</tr>
<tr>
<td>RA</td>
<td>Robust nucleus of arcopallium</td>
</tr>
<tr>
<td>SFL</td>
<td>Superior frontal lamina</td>
</tr>
<tr>
<td>SL</td>
<td>Lateral septal nucleus</td>
</tr>
<tr>
<td>SM</td>
<td>Medial septal nucleus</td>
</tr>
<tr>
<td>SNC</td>
<td>Substantia nigra, pars compacta</td>
</tr>
<tr>
<td>SNr</td>
<td>Substantia nigra, pars reticulata</td>
</tr>
<tr>
<td>SON</td>
<td>Superior olivary nucleus</td>
</tr>
<tr>
<td>SpA</td>
<td>Subpallial amygdaloid area</td>
</tr>
<tr>
<td>SSp</td>
<td>Supraspinal nucleus</td>
</tr>
<tr>
<td>STN</td>
<td>Subthalamic nucleus</td>
</tr>
<tr>
<td>TnA</td>
<td>Nucleus taeniae of the amygdala</td>
</tr>
<tr>
<td>TPO</td>
<td>Area temporoparieto-occipitalis</td>
</tr>
<tr>
<td>TSM</td>
<td>Septopallio-mesencephalic tract</td>
</tr>
<tr>
<td>TTP</td>
<td>Thalamopallial tract</td>
</tr>
<tr>
<td>Uva</td>
<td>Nucleus uvaeformis</td>
</tr>
<tr>
<td>VP</td>
<td>Ventral pallidum</td>
</tr>
<tr>
<td>VTA</td>
<td>Ventral tegmental area</td>
</tr>
<tr>
<td>X</td>
<td>Area X</td>
</tr>
</tbody>
</table>
9.2 Microdrive technical drawings

The following pages contain blueprints of the microdrive body, tip and shuttle. All measurements are in mm. The material used was a polyamide-imide polymer (trademarked as Torlon, manufactured by Amoco Performance Products). The choice of material was guided by the following requirements:

- High melting point, to prevent damage during soldering (soldering joints are located on the shuttle, and in close proximity to the body on the omnetics connector).

- High solvent resistance. Both shuttle and tip are cleaned in acetone to remove dental cement and glue.
Figure 60: Microdrive technical drawing I/III: Assembly overview.
Figure 61: Microdrive technical drawing II/III: Tip.
Figure 62: Microdrive technical drawing II/III: Body and shuttle.
9.3 Table of Figures

Figure 1: A male zebra finch and his offspring. ................................................................. 14
Figure 2: The most frequently uttered zebra finch calls. .................................................... 15
Figure 3: Zebra finch song. ............................................................................................... 15
Figure 4: Undirected vs. directed song. ............................................................................ 16
Figure 5: Time line of song development. ....................................................................... 17
Figure 6: Song development I. .......................................................................................... 18
Figure 7: Song development II. ......................................................................................... 19
Figure 8: The song system. .............................................................................................. 20
Figure 9: The template theory of song learning. ............................................................... 21
Figure 10: Frequency range of perception and production. ............................................ 23
Figure 11: Connectivity diagram of the auditory system of the songbird and its interface to the posterior motor pathway. ................................................................. 26
Figure 12: Responses of a L2b neuron to a set of different auditory stimuli. .................... 32
Figure 13: Active responses were similar to passive responses, but more stereotyped and slightly anticipatory. ......................................................................................... 36
Figure 14: Similarity of song and BOS playback evoked activity. .................................. 38
Figure 15: Average singing and BOS playback related firing rates. ................................ 39
Figure 16: Anticipatory activity. ......................................................................................... 39
Figure 17: Example perturbation responses. .................................................................... 40
Figure 18: Song perturbations. ......................................................................................... 41
Figure 19: Perturbations did not induce immediate changes in vocal output .................... 42
Figure 20: Control for perturbation induced song amplitude increase. ............................ 43
Figure 21: Summary of perturbation selectivity. ................................................................ 44
Figure 22: Spike raster plots of three additional example neurons. ................................. 45
Figure 23: Response increases with perturbation amplitude. ......................................... 46
Figure 24: Perturbation responsive neurons are usually quiescent. ................................. 47
Figure 25: Effect of sound amplitude on response latency .............................................. 48
Figure 26: Song locked field L activity in a deaf bird. ......................................................... 50
Figure 27: Offset response of a field L neuron in a deaf bird. .......................................... 51
Figure 28: NCM response habituates with stimulus repetition. ........................................ 53
Figure 29: NCM activity during song and BOS playback are similar. ............................. 56
Figure 30: Counteracting NCM response adaptations during song and BOS playback. .. 57
Figure 31: Activity adaptations occur between, not within song bouts. .......................... 57
Figure 32: NCM activity can be predictive of song changes. ......................................... 58
Figure 33: Similarity of singing and tutor song playback related activity, example neuron I. 59
Figure 34: Similarity of singing and tutor song playback related activity, example neuron II. 59
Figure 35: Similarity of song-related and tutor song playback-related responses. ............ 60
Figure 36: Surprise neuron. .............................................................................................. 60
Figure 37: Song development after NCM lesion at 36 dph. ............................................ 63
Figure 38: Loss of song rhythm during NIf inactivation in juveniles. ............................... 65
Figure 39: NIf premotor activity in juveniles. ................................................................. 68
Figure 40: NIf premotor and auditory responses. ............................................................. 69
Figure 41: Model overabundance impedes learning. ....................................................... 71
Figure 42: Father stops singing as juvenile enters plastic song phase. ............................ 72
Figure 43: Tutor song stimulation during the plastic song phase has detrimental effects on song imitation.

Figure 44: Preference test setup.

Figure 45: Females can distinguish directed from undirected song based on stereotypy.

Figure 46: Time course of song preference during playback sessions.

Figure 47: Template replay hypothesis.

Figure 48: Projection from the HVC shelf to the RA cup.

Figure 49: Schematic of experimental setup.

Figure 50: Perturbation stimuli.

Figure 51: Three channel microdrive.

Figure 52: Single channel microdrive after primary assembly (before electrode is mounted).

Figure 53: Assembled microdrive with mounted and extended electrode.

Figure 54: Average stable recording time.

Figure 55: Experimental time window.

Figure 56: Vocal activity timeline I.

Figure 57: Vocal activity timeline II.

Figure 58: Functioning principle of the real time recognizer.

Figure 59: The avian inner ear.

Figure 60: Microdrive technical drawing I/III: Assembly overview.

Figure 61: Microdrive technical drawing II/III: Tip.

Figure 62: Microdrive technical drawing II/III: Body and shuttle.
## 9.4 Index

Note that this index points to passages in the text where the specified concept is defined or discussed; it does not list all the occurrences of the respective terms.

| A | active sensory processing, 91  
actor, experimenter and critic framework, 82  
audibility curve, 23  
auditory area, 25  
auditory feedback, 21, 35, 69, 80  
auditory gating, 84  
auditory memory, 52  
auditory network, 26  
auditory template, 86  
auditory-vocal mirroring, 45 |
|---|---|
| C | calls, 14  
cochlea, 23  
columella, 23  
crystallized song, 17 |
| D | deafening, 80  
directed song, 16, 74 |
| E | efference copy, 11, 12, 82, 89  
error detection system, 80  
error signal, 12, 21, 82, 83, 84 |
| F | female song preference, 74 |
| G | Field L, 25  
gender differences, 54 |
| I | imitation learning, 12  
immediate early gene, 33  
ter syllable interval, 15 |
| M | memory variability, 78  
motor memory, 85 |
| N | note, 15 |
| P | perturbation-response bias, 37  
perturbation-response selectivity, 37  
plastic song, 17  
premtor area, 25  
pseudo-directed song, 74 |
| R | response habituation, 52 |
| S | selectivity, 30  
sensory memory, 85  
song development, 17  
song syllables, 15  
song system, 20  
STRF, 33  
subsong, 17  
surprise neuron, 55 |
| T | template replay, 55, 89  
template theory, 21  
tutor-song over-exposure, 90  
tutor-song over-stimulation, 74 |
| U | undirected song, 16, 74 |
| V | vocal learning, 13 |
| Z | zebra finch, 13  
ZENK, 33 |
“The central problem in the study of neural vocalization control mechanisms is not to locate brain regions which “control” vocalization; it is to determine how the brain as a whole functions during the various vocalizations.”

10 REFERENCES

Akutagawa, E. and M. Konishi (2005). "Connections of thalamic modulatory centers to the vocal
in the zebra finch field L and caudal mesopallium." J Comp Physiol A Neuroethol Sens Neural
Behav Physiol 190(6): 469--489.
Aronov, D., A. S. Andalman and M. S. Fee (2008). "A specialized forebrain circuit for vocal babbling in
the juvenile songbird." Science 320(5876): 630--634.
following auditory stimulation in the juvenile male and female zebra finch." Brain Res Mol
forebrain impairs sensory acquisition in the zebra finch (Poephila guttata)." Neurobiol Learn
453.
characterization of the auditory thalamus in a songbird, the European starling." J Comp
activation in the zebra finch brain is related to the strength of song learning." Proc Natl Acad
Bonke, D., H. Scheich and G. Langner (1979). "Responsiveness of units in the auditory neostriatum of
the guinea fowl (Numida meleagris) to species-specific calls and synthetic stimuli." Journal of
Comparative Physiology A 132: 243-255.
Boord, R. L. (1968). "Ascending projections of the primary cochlear nuclei and nucleus laminaris in the
by strain gauge and volume and impedance change measures." Comp Biochem Physiol A
Comp Physiol 71(4): 619--621.
Bottjer, S. W. and A. P. Arnold (1984). "The role of feedback from the vocal organ. I. Maintenance of
Bottjer, S. W., E. A. Miesner and A. P. Arnold (1984). "Forebrain lesions disrupt development but not


GEORG BERNHARD KELLER

Oerlikonerstrasse 8
8057 Zurich
Switzerland

georg@phys.ethz.ch
Phone: +41 44 635 30 22
Mobile: +41 78 798 42 51

Biographical

Education
Diploma in theoretical physics
ETH Zurich
2000–2005

Thesis work on 'Theoretical model of the head directions system'.

PhD in Neuroscience
Institute of Neuroinformatics
University and ETH Zurich
2005–2009

Supervised by Richard Hahnloser and Rodney Douglas, work on the subject of 'Auditory feedback processing in the songbird'.

Experience
Research project Neo-Cartilage Formation in Microgravity
European Space Agency (ESA)
Space Biology Institute ETHZ
2001–2003

2001 Participation in the 4. ESA Student Parabolic Flight Campaign. This in cooperation with the Space Biology Institute at the ETH Zurich, testing a bioreactor for later use on space missions.
2002 Research project investigating growth of neo cartilage in weightlessness, again in cooperation with the Space Biology Institute. Premature end of the project due to a launch failure of the Russian Soyuz rocket.
2003 Project restart with now successful launch and flight aboard a manned space mission to the International Space Station (ISS) as the first student project worldwide aboard the ISS.

Music Festival
Verein Openair Muri
2002–2004

2002 Head of PR and finances of one the 10 biggest rock music festivals in Switzerland.
2004 Head of the organisation committee of the 17th edition of the festival.

Scientific advisor
2004 - 2005

Personal scientific advisor to a Board Chairman and CEO of a major Swiss technology company.


**ATTENDED CONFERENCES**

- Neuroscience 2006 *Atlanta, USA*
- Swiss Society for Neuroscience Meeting 2006 *Bern, CH*
- Neuroscience 2007 *San Diego, USA*
- Capo Caccio Workshop 2008 *Sardegna, Italy*
- FENS 2008 *Geneva, CH*
- Neuroinformatics 2008 *Stockholm, Sweden*
- Computation in Cortical Circuits 2008 *Ascona, CH*
- Neuroscience 2008 *Washington DC, USA*