

---

# A telemetric approach for characterizing behavioral dynamics and neurophysiology of vocal interactions in Zebra Finches

---



Graduate School of  
Systemic Neurosciences  
LMU Munich

Dissertation  
an der Graduate School of Systemic Neurosciences  
der Ludwig-Maximilians-Universität München

vorgelegt von  
Shouwen Ma  
aus Guangdong, China

München, den 21. Juni 2017



Erstgutachter: Prof. Dr. Manfred Gahr

Zweitgutachter: Dr. Boris Chagnaud

Drittgutachter: PD. Dr. Lars Kunz

Viertgutachter: PD. Dr. Steffen Hage

Tag der mündlichen Prüfung: 29. September 2017

# Contents

<b>Abstract</b>	<b>xi</b>
<b>Zusammenfassung</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 A comparative view of vocalizations . . . . .	4
1.1.1 The call repertoire of zebra finches . . . . .	5
1.1.2 Innate calls in various animal taxa . . . . .	5
1.1.3 Innate semantic calls of zebra finches . . . . .	7
1.1.4 The innate non-semantic calls of zebra finches . . . . .	7
1.2 Plasticity of bird call productions . . . . .	8
1.3 The structure of vocal interactions . . . . .	9
1.4 The models of vocal interactions . . . . .	9
1.4.1 Vocal interactions in the long-time range . . . . .	10
1.4.2 Mathematics of power-law and the empirical data . . . . .	11
1.4.3 Statistic of power-law . . . . .	16
1.4.4 Vocal interaction in the short dynamic range . . . . .	18
1.5 Neurophysiology of call-based vocal communications . . . . .	19
1.5.1 Motor control of bird calls . . . . .	19
1.5.2 Auditory representations of bird calls . . . . .	21
1.5.3 Neurophysiology of vocal interactions . . . . .	23
1.6 Neurophysiology of expectations . . . . .	23
1.7 Sex differences of auditory object formation . . . . .	25
1.8 A telemetric approach for simultaneously measuring vocal and neuronal activity of freely interacting animals . . . . .	26
<b>2 Manuscripts</b>	<b>31</b>
2.1 Power-law scaling of vocal activity in zebra finches . . . . .	31
2.1.1 Abstract . . . . .	31
2.1.2 Introduction . . . . .	32
2.1.3 Results . . . . .	33
2.1.4 Discussion . . . . .	37
2.1.5 Materials and Methods . . . . .	39



2.1.6	Supplementary Information . . . . .	42
2.2	Sensorimotor expectation in call-based vocal communications of zebra finches	57
2.2.1	Abstract . . . . .	57
2.2.2	Introduction . . . . .	57
2.2.3	Results . . . . .	58
2.2.4	Discussion . . . . .	66
2.2.5	Materials and Methods . . . . .	67
2.2.6	Supplementary information . . . . .	70
2.3	Spectrotemporal coding of auditory trains in the auditory forebrains of zebra finches . . . . .	77
2.3.1	Abstract . . . . .	77
2.3.2	Introduction . . . . .	77
2.3.3	Results . . . . .	78
2.3.4	Discussion . . . . .	84
2.3.5	Materials and Methods . . . . .	85
2.3.6	Supplementary Information . . . . .	87
<b>3</b>	<b>Discussion</b>	<b>93</b>
3.1	Power-law scales of information in vocal communication system . . . . .	94
3.1.1	Vocal communication system in birds and humans . . . . .	94
3.1.2	Measurements of information in vocal communications . . . . .	96
3.1.3	Increase in information content by varying inter-call intervals . . . . .	98
3.1.4	Power-law scaling of vocal activity and Barabasi's model for choice of actions . . . . .	100
3.2	Sensorimotor expectation in call-based vocal communications . . . . .	101
3.3	Acoustic representation in the auditory forebrains . . . . .	103
<b>4</b>	<b>Summary</b>	<b>107</b>
<b>A</b>	<b>List of abbreviations and symbols</b>	<b>109</b>
	<b>Acknowledgment</b>	<b>129</b>
	<b>List of Author Contributions</b>	<b>135</b>
	<b>Permissions</b>	<b>137</b>
	<b>Eidesstattliche Versicherung/Affidavit</b>	<b>139</b>

# List of Figures

1.1	Specifications and successions of zebra finches' calls . . . . .	6
1.2	Comparison of the vocal dynamics between zebra finch and mouse . . . . .	8
1.3	Plotting empirical data and fitting to power-law . . . . .	13
1.4	Time series and periodograms of different color noises . . . . .	16
1.5	The song systems of songbirds and non-songbirds . . . . .	20
1.6	The forebrain auditory system and its connectivity with the song system .	22
1.7	A telemetric approach for measuring vocal and neuronal activity of freely interacting animals . . . . .	27
1.8	Simultaneous recordings of vocal and neuronal activity in freely interacting zebra finches . . . . .	28
2.1	Calling activity of zebra finch pairs . . . . .	47
2.2	The self-contained and the reactive callings exhibit power-law dynamics . .	48
2.3	The power-law scaling of the self-contained and of the reactive callings change independently . . . . .	49
2.4	The exponents of different callings changed transiently in response to perturbation experiments . . . . .	50
2.5	Successions of calling interactions between females and males of zebra finch pairs . . . . .	51
2.6	Water-removal influences the distribution of the call rates of both male and female zebra finches . . . . .	52
2.7	Synchronous recording of vocal interactions during sensory perturbation . .	53
2.8	Comparing the tail distributions of different behavioral categories . . . . .	54
2.9	Power-law dynamics of calling intervals is not affected by the daily circadian cycle . . . . .	55
2.10	Isolation changes the exponents of "reactive" callings . . . . .	56
2.11	Vocal interactions between male and female zebra finches . . . . .	59
2.12	Different types of calling associations between male and female zebra finches	62
2.13	Neuronal activity of male HVC during calling interactions . . . . .	63
2.14	Premotor activity of male HVC neurons in relation to female calls are sensory and context dependent during calling interactions . . . . .	64
2.15	Limiting social-sensory cues affect the response times during calling interactions . . . . .	65

2.16	The proportions of “stack” calls . . . . .	71
2.17	Different types of calling associations between of male and female zebra finches	72
2.18	Neuronal activity of male HVC related to different types of calling associations	73
2.19	Premotor activity of male HVC are sensory and context dependent during calling interactions . . . . .	74
2.20	Auditory sensitivity to single stimuli with inserted gaps in male and female auditory forebrain . . . . .	81
2.21	Multi-unit responses to stimulus trains in male and female Field L . . . . .	82
2.22	Pair-wise comparison of habituation rates in relation to different stimulus trains . . . . .	83
2.23	Waveforms and spectrograms of the stimuli . . . . .	89
2.24	Semitone differences between “A” and “B” segments of stimuli . . . . .	90
2.25	The auditory responses to song syllables with different length of inserted gaps in male HVC . . . . .	91
2.26	Inactivation of HVC decreased the habituation rates in response to stimulus trains . . . . .	92
3.1	The relationship between the frequency of syllables and the syllable ranks in zebra finches . . . . .	96
3.2	Song and call sequences of a male zebra finch . . . . .	98
3.3	The information entropy of call and song sequences with respect to different entropic orders . . . . .	99

# List of Tables

2.1	Measured values and estimated parameters of the calling behavior . . . . .	35
2.2	Estimated parameters and statistical tests for different zebra finch pairs during cohabitation before treatments . . . . .	43
2.2	Estimated parameters and statistical tests for different zebra finch pairs during cohabitation before treatments . . . . .	44
2.2	Estimated parameters and statistical tests for different zebra finch pairs during cohabitation before treatments . . . . .	45
2.2	Estimated parameters and statistical tests for different zebra finch pairs during cohabitation before treatments . . . . .	46
2.3	Comparison of neuronal activity of male HVC in relation to their own calls between different types of calling associations . . . . .	75
2.4	Comparison of neuronal activity of male HVC in relation to female calls between different types of calling associations . . . . .	76
A.1	List of abbreviations . . . . .	109
A.1	List of abbreviations . . . . .	110
A.2	List of symbols . . . . .	110



# Abstract

Calls are uttered by many species of fish, birds, mice, whales, non-human primates and by humans. In most cases, their acoustic structure and the context of emission were studied concerning their signaling value. Calls can be categorized as semantic calls with pre-determined meaning and in non-semantic calls lacking a particular meaning. Little is known about the signaling value of non-semantic calls that are both acoustically invariant and produced in large numbers during the course of a day. In particular, since such calls are almost identical from rendition to rendition, the call per se and the sequence of calls cannot contain much information, except an individual signature.

Birds such as the zebra finches (*Taeniopygia guttata*) have a rich repertoire of calls used for vocal communications. One of these calls, the “Stack” call, is non-semantic, acoustically-invariant and their most frequent utterance. The succession of these calls displays, i.e. the inter-call intervals show a profound inter- and intra-individual diversity in the dynamic range, lasting from a fraction of seconds to hours. In this thesis, I investigated the full range of calling dynamics in pair-wise (one male and one female) living zebra finches. In particular, I studied (1) the organizational rules of “Stack” calling in the context of dyadic interactions of male and female, (2) the neurophysiological correlates of the calling interactions, and (3) the auditory processing of single sounds and sound sequences with varying inter-call intervals. I used a telemetric system for simultaneous measurements of all vocal and neuronal activities of male and female zebra finches that freely interacted with each other. The results presented in this thesis show that:

1. The “Stack” calling intervals were characterized as a power-law function, scaling with the exponent ( $\alpha$ ) ranging between  $2 < \alpha < 3$ , which implies that all “Stack” calls in that scale have similar dynamic patterns.
2. The power-law function suggests that information could be encoded by varying the inter-call intervals, which is reflected in the exponent ( $\alpha$ ) of the power-law function.
3. Based on the power-law scaling, I was able to show that the calling activity of zebra finches is physiological- and sensory state-dependent and that reactive calling and self-consistent calling behavior changed independent of each other.

For theoretical reasons, power-law scaling does not cover the whole dynamic range. The distribution of inter-call intervals deviates from the power-law characteristic below a minimum value  $\tau_{min}$ . For those callings, I showed that:

1. Calling interactions are performed alternatively with pauses.
2. Neurophysiological studies of these fast calling interactions showed that male neurons of the sensorimotor song control nucleus HVC (higher vocal center) not only displayed premotor activity related to male's own calls, but also displayed auditory-elicited discharge to incoming female calls and predictive discharge to upcoming female calls that were uttered in bouts of male-female vocal exchange. Thus, HVC is a general vocal communication controlling nucleus rather than a song control nucleus.
3. With separation experiments I showed that the predictive discharge of HVC required sensory cues originating from the mate.

Information coding and decoding is essential to animal communications. To decode the sound signals, animals require a bottom-up and a top-down process [33]. I investigated the top-down process, namely the hierarchical process between sensorimotor (the HVC) and the auditory system (Field L), the primary auditory cortex-like region of birds. I compared the auditory responses of Field L neurons of males and females since females lack a differentiated HVC and thus likely lack a HVC to Field L projection. The results showed that:

1. Male and female Field L responded similarly to single natural sounds.
2. When the natural sounds were played repeatedly as trains with varying inter-sound intervals, the neuronal responses attenuated faster in male Field L than in female Field L.
3. Unfortunately, due to small number of studied animals, I could not (statistically) demonstrate that HVC activity is required for Field L response properties.

# Zusammenfassung

Viele Arten von Fischen, Vögel, Mäusen, Walen, nicht-menschlichen Primaten und Menschen nutzen vokale Kommunikation. In den meisten Fällen wurde deren akustische Struktur und der Zusammenhang der Lautäußerung im Hinblick auf den Signalwert untersucht.. Die Rufe können als semantische Rufe mit vorgegebener Bedeutung und als nicht-semantischen Rufe ohne besondere Bedeutung kategorisiert werden. Über den Signalwert der nicht-semantischen Rufe, die akustisch unveränderlich sind und im Laufe des Tages häufig geäußert werden, ist wenig bekannt. Da diese Rufe von Lautäußerung zu Lautäußerung fast identisch sind, können der Ruf an sich und die Rufabfolge, abgesehen von einer individuellen Signatur, nicht viel Information enthalten.

Vögel, wie z.B. Zebrafinken (*Taeniopygia guttata*) verfügen über ein vielfältiges Repertoire an Rufen für vokale Kommunikation. Einer dieser Rufe, der sogenannte „Stack“ Ruf, zählt zu den nicht-semantischen Rufen, ist akustisch unveränderlich und wird hauptsächlich von domestizierten Zebrafinken geäußert. Die Abfolge dieser Rufe, z.B. die Zwischenruf-Intervalle, zeigt eine tiefgehende inter- und intraindividuelle Diversität im Dynamikumfang, die von Sekundenbruchteilen bis hin zu Stunden andauern können.

In dieser Dissertation untersuchte ich das gesamte Spektrum des Dynamikumfangs bei paarweise lebenden (ein Männchen und ein Weibchen) Zebrafinken. Vor allem untersuchte ich (1) die organisatorischen Regeln der „Stack“ Rufe in einer dyadischen Interaktion zwischen männlichen und weiblichen Zebrafinken, (2) die neurophysiologischen Zusammenhänge der Vogelrufe, und (3) die auditorischen Prozesse einzelner Töne und aufeinanderfolgender Töne mit unterschiedlichen Zwischenruf-Intervallen.

Ich verwendete ein telemetrisches System für gleichzeitige Messungen aller vokalen und neuronalen Aktivitäten bei männlichen und weiblichen Zebrafinken, die frei miteinander kommunizieren. Die Ergebnisse, die ich in dieser Dissertation präsentieren möchte, zeigen, dass:

1. Die „Stack“ Ruf - Intervalle als ‘Power-law’-Funktion charakterisiert wurden, skalierend mit dem Exponenten ( $\alpha$ ) zwischen  $2 < \alpha < 3$ . Das bedeutet, dass alle „Stack“ Rufe in dieser Skalierung ähnliche dynamische Muster haben.



2. Die ‘Power-law’-Funktion nahe legt, dass Informationen durch Variation der Zwischenrufe Intervalle codiert werden könnten, was sich im Exponent ( $\alpha$ ) der ‘Power-law’-Funktion widerspiegelt.
3. Basierend auf der Skalierung der ‘Power-law’-Funktion, die Aktivität der Vogelrufe bei Zebrafinken vom physiologischen und sensorischen Zustand abhängig ist und, dass sich Reaktionsrufe und selbständiges Rufverhalten unabhängig voneinander unterscheiden.

Rein theoretisch deckt die Skalierung der ‘Power-law’-Funktion nicht den gesamten Dynamikumfang ab. Die Verteilung der Zwischenruf-Intervalle weicht von der ‘Power-law’-Funktion unterhalb eines Minimums  $\tau_{min}$ . ab. Für diese Rufe konnte ich aufzeigen, dass:

1. Ruf-Interaktionen finden wahlweise mit Pausen statt.
2. Neurophysiologische Untersuchungen dieser schnell ablaufenden Rufinteraktionen zeigen, dass die Neuronen im männlichen sensomotorischen Kern HVC (higher vocal center) nicht nur prämotorische Aktivität im Zusammenhang mit eigenen Rufe darstellten, sondern auch auditorisch ausgelöste Aktivität auf eingehende weibliche Rufe und prädiktive Aktivität auf aufkommenden weiblichen Rufe, die in wechselseitiger Kommunikation geäußert wurden. Daraus folgt, dass das HVC eher einem allgemeinen Kontrollsystem für die vokale Kommunikation entspricht, als einem Gesangssystem.
3. Anhand von Trennungsexperimenten konnte ich zeigen, dass für die prädiktive Aktivität im HVC vom Partner stammende sensorische Auslöser erforderlich sind.

Informationskodierung und Informationsdekodierung sind für die Tierkommunikation unerlässlich. Um die Tonsignale zu dekodieren, sind die Tiere dem Bottom-up und dem Top-down Prozess [33] unterworfen. Ich untersuchte den Top-down-Prozess, namentlich den hierarchischen Prozess zwischen sensomotorischem System (HVC) und auditorischem System (Field L), dem primären auditorischen Vorderhirn im Vogel. Ich verglich die auditorischen Antworten in Field L-Neuronen bei Männchen und Weibchen weil dem weiblichen Vorderhirn ein differenziertes HVC fehlt und vermutlich eine neuronale Projektion von HVC auf Field L fehlt. Die Ergebnisse zeigen, dass:

1. Das männliche und weibliche Field L reagieren in ähnlicher Weise auf natürliche Laute.
2. Wenn die Töne in Zügen wiederholt gespielt wurden, wurden die neuronalen Antworten im männlichen Field L schneller schwächer, als im weiblichen Field L.
3. Aufgrund der zu geringen Anzahl der zu untersuchenden Tiere, konnte ich leider nicht (statistisch) zeigen, dass die HVC-Aktivität für Reaktionseigenschaften im Field L erforderlich ist.

# Chapter 1

## Introduction

Social animals produce a large number of vocalizations. While learned songs and calls with a predetermined meaning (semantic calls) are widely credited with sound signaling in social communications [135, 162, 54], little is known about those calls that have no predetermined meaning (non-semantic calls), in particular, if they are not contingent upon specific contexts. Zebra finches (*Taeniopygia guttata*) are songbirds of the *Estrildid* family of the *Passeriformes* order. They are highly social animals, which gather and vocalize in flocks. Mates (the male and female forming a reproductive pair) are an inseparable dyad, always keeping in touch and synchronizing their movements with calls [147, 246]. In contrast to a few *Estrildid* species with singing females [68, 70] songs in zebra finches are male specific, whereas male and female zebra finches share a similar call repertoire [54] in which the SStackcalls are the most frequent utterances of domestic zebra finches in flocks [15, 211] that may allow explicit bi-directional communication between zebra finches during rapid vocal exchanges. If mates are physically separated, but allowed visual and auditory contact, they still maintain the pair-bond by the exchange of calls [166].

Semantic calls are generally produced infrequently, whereas non-semantic calls of zebra finches and others are produced in large numbers during the course of a day [246, 15, 54]. In relation, calls such as the “Distance” call of zebra finches [222] and “Alarm” calls of chicken [56] are produced rarely and are semantic, whereas “Stack” calls are non-semantic. However, in rhesus macaques (*Macaca mulatta*), the predetermined “Food” calls are also produced in a nonfood context [93], suggesting that acoustically similar calls occur in diverse socio-ecological contexts. Similar to non-semantic calls of monkeys [93, 161], the “Stack” calls of zebra finches are produced in large amounts and may allow animals to associate their own calls with others. However, this interpretation remains hypothetical due to the methodological difficulties in measuring and quantifying the occurrence of acoustically similar calls in freely behaving animals in a social context. In this thesis, I investigated the organizational rules of non-semantic “Stack” calling of zebra finches. Thus, the term “call” refers to the “Stack” call in zebra finches, if not specified otherwise.

Recent technique has improved the on-bird recording of vocalizations and allows an

accurate assignment of the acoustically similar calls to individual birds in a social context [211]. Traditional time series analysis of behavioral data mainly employs auto-correlations and cross-correlations [80] to analyze self-contained (calls that are uttered after previous own calls without acoustic input from others) and reactive (calls that are uttered after hearing calls of others) behaviors. Their products however only reveal a short range of temporal correlations, e.g., by using lagged cross-correlations, the vocal activity of zebra finch mates were contingent upon the partner activity over a window of 2 seconds [211, 77] and up to 60 seconds for most human dyads [57, 106]. Using on-bird microphones and cross-correlation analysis, Gill et al. found that a small proportion of “Stack” calls is used by sexual partners to maintain vocal contact [77]. The cross-correlation analysis only reveals this very short range of vocal dynamics and concerns what we call reactive calls [211, 77]. It is unclear how the large numbers of “Stack” calls are organized beyond this short dynamic range. In particular, the organization of consecutive (self-contained) calls, as another key component of calling interaction, has not been taken into account.

To investigate the organizational rules of natural behavior in a large dynamic range, the inter-event intervals are widely used to describe the link between dynamic processes and behavioral events based on the time series theory, in which the state transition between events generally depends on the history of mechanical processes [191]. The Poissonian description is well-documented in explaining human dynamics, in which the heavy tailed inter-event intervals of human activity are a consequence of periodic cycles with a cascading random process, namely the Poisson process [129, 128]. The Poissonian model predicts that the time interval between two consecutive events follows an exponential distribution, e.g., the email communication of humans [7] and mobility of humans [200]. It has been discovered that the distribution of time intervals in a specific activity, such as the consecutive sent e-mails by an individual, is heavily tailed distributed (i.e. some e-mails were sent after long waiting times) and is well approximated by a power-law function. In contrast, a non-Poissonian model based on priority queuing mechanisms provides a direct description of human activity by measuring the actual time differences between the consecutive activities [7], but this model is largely influenced by an external factor, namely the circadian cycle that may give rise to heavy-tailed inter-event intervals. Although both models provide an accurate description of heavy-tailed dynamics in human activity, these distinct explanations are controversial. Recently, a multi-modal model was introduced to bring the Poissonian and the priority queuing explanations together. It suggests that the observed heavy-tailed dynamics are tied to the multiple active states of individuals and the interactions between them [185]. Nevertheless, these models, including the Poissonian model, the priority queuing model and the multi-modal models, were based on data only consisting of consecutive events. The questions of whether behavioral dynamics are state-dependent and how the interactive events impact on the observed dynamics remained unaddressed.

In line with previous studies, the first part of this thesis aims at investigating the behavioral process of calling interactions in zebra finches by incorporating the description of behavioral dynamics into a context of dyadic interactions. The goal is to clarify the plasticity

---

of calling behaviors and the biological relevance of “Stack” calls during moment-by-moment interactions in zebra finches over large time-windows using mathematical descriptions such as the power-law. That is, the frequency of interaction between two individuals is approximated by an inverse square power exponent ( $\alpha$ ) of inter-event intervals of interactions [153]. Previous analyze of vocal activities focused on callings that occur within short-time windows (2 seconds) around the focal call of the mate [211, 77]. This analysis showed that only a very small portion of all “Stack” calls are used to maintain vocal contact between sexual partners [77].

Although the short-time range of calling dynamics constitutes a very important part in calling interactions [77], the power-law description can only cover a long-range of dynamics due to theoretical reasons [153, 39]. Thus, the second part of this thesis aims at the organizational rules of short-range call interactions and their neurophysiology, in pair-living zebra finches. As mentioned above, cross-correlation analysis reveals that “Stack” calls are produced in relation to the partner’s “Stack” calls within a window of 2 seconds history [211, 77]. Since zebra finches have rather rapid call exchanges of 250 ms or less, they might have a mechanism to anticipate the upcoming calls of their partners during a calling interaction. Previous studies in humans and monkeys show that the sensorimotor system contributes to the predictive coding in a context-based reaction to experimentally associated objects [100, 114, 131]. However, the predictive coding during a natural social interaction remains unknown. The second goal of this thesis is to search for such predictive mechanisms in the sensorimotor nucleus HVC that is a well-known essential part of the so-called song control system of songbirds [157]. More specifically, are there auditory-vocal correspondent neurons that show anticipation for calls?

The auditory responses in the sensorimotor nucleus HVC may correspond to the vocal performance during song learning and learned vocal communication [119, 42, 172], but the auditory mode of HVC neurons in response to songs was gated off in the awake birds [188]. HVC activity is related to 6 afferent nuclei: nucleus avalanche (Av), Field L, magnocellular nucleus of the anterior nidopallium (mMAN), nucleus interface (NIF), robust nucleus of the arcopallium (RA) and uvaeform nucleus (Uva) [158, 28, 61, 62, 182, 1], among which the primary auditory forebrain Field L has a direct link between the subcortex-like auditory system and sensorimotor system in songbirds [61]. In contrast to the HVC neurons that preferentially respond to the bird’s own songs (BOS) [132], Field L has broader auditory tunes to acoustic stimuli while being more selectivity for natural sounds than synthesized sounds [121, 81, 2]. Furthermore, auditory responses of Field L neurons do not differ between the awake and under anesthesia [188]. As mentioned above, songs and in relation the song control system in zebra finches are male specific, whereas male and female zebra finches share a similar auditory system. Since females can discriminate songs just as males, and Field L is related to spectrotemporal encoding of natural sounds [212, 214, 149], it is suggested that there is a similar structure of Field L in females and males, relating to the function. However, the behavioral studies by using an operant conditioning and a lesion in HVC in various songbird species support the impact of the sensorimotor system on acoustic

discrimination [32, 44, 125, 187]. These results may suggest a sexual dimorphism in the primary auditory forebrain. Female and male zebra finches may have different preferences to contextual sounds in a complex acoustic environment, although both have the auditory preference to natural sounds over mechanic sounds. In other words, the maintenance of auditory sensitivity may differ between female and male zebra finches. To test this hypothesis, a robust paradigm to experimentally examine the auditory maintenance is required.

Subsequently, in Manuscript 2.3 I compared the neuronal activity of a male and female Field L in response to a set of playback trains. The playback trains were designed according to the auditory scene analysis [31]. In a classic paradigm of an auditory scene analysis, a sequence of repeated gallops is presented to the listener. Each gallop consists of two pure tones (“A” and “B”) at different frequencies. If the frequency separation between these two tones is sufficiently large and the playback rate is sufficiently high, a listener is likely to perceive two separated streams (“AAA ...” and “BBB ...”) at the same time instead of one stream (“ABABAB ...”) as displayed paradigm. Such auditory segregation of pure tones has been reported in birds, monkeys and humans [31, 59, 17, 18]. A recent psychoacoustic study in humans reported that auditory experiences, such as listening to trains of repeated words (spoken words with lexical meaning) and non-words (synthetic words without lexical meaning). Lexicon affected the possibility to identify words in playback trains [21]. Therefore, the playback trains of an auditory scene analysis may provide an ideal paradigm to examine the maintenance of auditory sensitivity to contextual sounds in male and female birds.

## 1.1 A comparative view of vocalizations

The need of auditory feedback distinguishes learned vocalizations from innate vocalizations since the latter do not require auditory feedback for development. Human speech and the song of certain bird taxa (parrots, hummingbirds, songbirds) are learned vocalizations, depending strongly on the auditory feedback and vocal practicing during development [46]. In contrast to learned vocalizations, there are surprisingly large numbers of taxa and species that produce only innate sounds, such as squirrel monkeys (*Saimiri sciureus*) [36], macaques [160] and mice [89, 127]. In this section, I compared innate and learned call categories with semantic and non-semantic calls in different animal taxa including insects, fish, birds, mice, meerkats, non-human primates, humans and then zebra finches. In principle, semantics determine whether or not the meanings are linked to previously established referents, and how this linkage can be accomplished [55]. By this principle and using the knowledge and experience acquired from various observations, vocalizations emitted in distinct social contexts from different groups of domestic zebra finches can be classified [54].

### 1.1.1 The call repertoire of zebra finches

The calls of wild zebra finches were originally classified into 12 categories: “Tet”, “Distance”, “Stack”, “Wsst”, “Thuk”, “Distress”, “Kackle”, “Ark”, “Whine”, “Copulation”, “Begging” and “Long Tonal” [246]. Recently, a data driven classification has divided the calls of domestic zebra finches into 10 categories: “Begging”, “Long Tonal”, “Whine”, “Nest”, “Tet”, “Distance”, “Wsst”, “Distress”, “Thuk” and “Tuck” [54], in which, as compared with the nomenclature of Richard Zann, the “Stack” calls were combined with “Tet” calls, the “Tuck” calls were established, the “Nest” calls were used to combine the “Kackle” and the “Ark” calls, and the “Copulation” calls were excluded. Thus, some of these call categories were defined in relation to the social contexts in which these vocalizations were emitted, whereas “Long Tonal” and “Tet” were classified in relation to their acoustic structures. All calls, except the “Distance” calls of males, are innate and all calls show a stereotyped individual acoustic structure.

In this thesis, the “Tet” and the “Stack” are kept as separate categories since they are distinguishable based on their acoustic structures. The sonogram of the “Tet” calls shows a pile of down-sweep shaped harmonics, whereas the sonogram of the “Stack” calls shows a pile of flat shaped harmonics (Figure 1.1 A & C). Neither the “Stack” nor the “Tet” calls are sex-specific (Figure 1.1 A & C). Both call types are uttered over a large dynamic range in that some calls are produced in a very fast succession whereas others are produced after long silent gaps (Figure 1.1 B & D). Further, the successions showed that both could be used in response to others (reactive calling) or following the own previous calls without vocal activity of others (self-consistent calling). Since “Stack” calls occur in diverse circumstances, they are likely produced without a predetermined signaling value. Thus, we classified them as non-semantic calls compared to semantic calls, such as the “Distance” calls and the “Begging” calls, in which the productions occurs in a specific context.

### 1.1.2 Innate calls in various animal taxa

Many arthropods, such as crickets and fruit flies, use sound for localizing and mating rituals [103, 150]. Fish generate low frequency calls to transfer acoustic signals in water [12]. Amphibians and reptiles produce calls to attract mates or frighten off rivals [229, 41]. Like these calls, data from various mammalian species show that their calls, too are innate [160]. Mice vocalizations sound normally even when the mice are raised in isolation or cross-fostered [160]. Most strikingly, mice are able to produce normal vocalizations even when their motor cortex is lesioned [89]. Recent evidence in marmoset monkeys (*Callithrix jacchus*) has shown that vocalizations of infant marmoset monkeys change in acoustic features [208] and vocal interactions [38]. However, this is due to selection based learning [151], and the calls of nonhuman primates have been confirmed to be innate by using an isolation method in all species studied in this way [239, 88, 49].

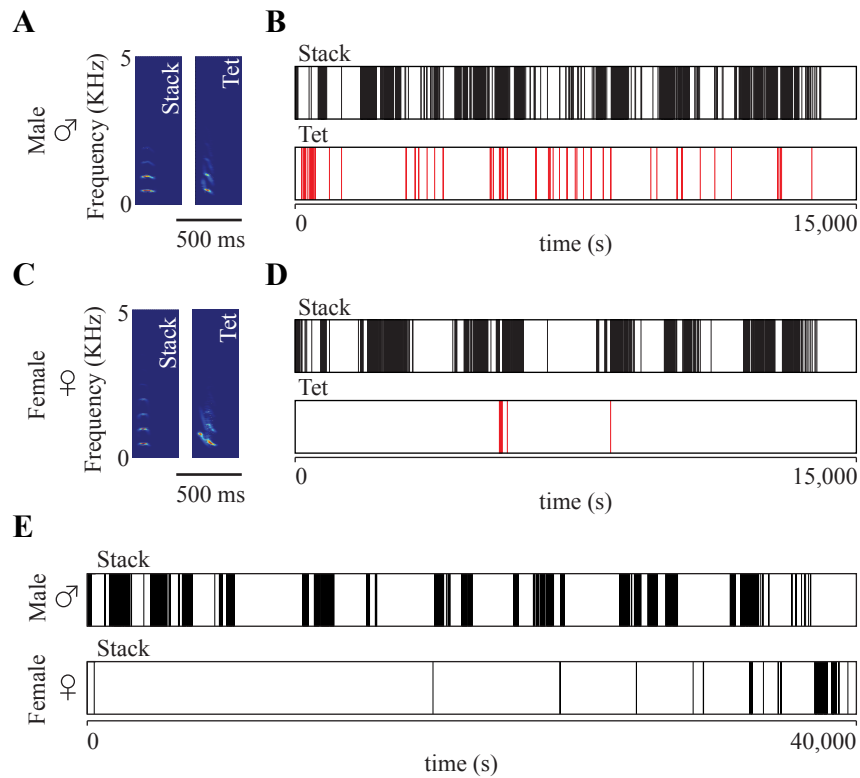


Figure 1.1: Specifications of “Stack” calls and “Tet” calls in a male and a female zebra finch and successions of these two types of calls during calling interactions and isolation of the male and female zebra finches. (A) The sonograms and (B) a succession of the “Stack” calls (black) and the “Tet” calls (red) in a male zebra finch. (C) The sonograms and (D) a succession of the “Stack” calls (black) and the “Tet” calls (red) in a female zebra finch. (E) Successions of the “Stack” calls during isolation in a male and a female zebra finch. The vertical lines in B, D and E correspond to the events of “Stack” (black) and “Tet” (red) calls over time.

Compared with learned bird songs that occur in species of songbirds, parrots and hummingbirds, the acoustic structures of most bird calls are simple and develop independently of experience [135]. One of the known exceptions among songbirds concerns the distance call of zebra finches, which are learned in males but not in females [199]. Budgerigars and parrots of the *Psittaculidae* family have been shown to mimic calls for specific contexts [110]. Among birds, examples of learned calls only come from species that also learn to sing. Since it is, in many cases, difficult to distinguish between songs and calls, learned calls might be seen as simple songs. Nevertheless, as stated above, calls of species of most bird taxa are innate.

In zebra finches, there is a call repertoire firstly described by Richard Zann [246]. and lately refined by Elie et. al. [54]. This list provides a description of each call according

to its acoustic structure and putative meanings. Some calls are associated with unique semantic meanings, whereas other calls, in particular the “Stack” calls, appear to be produced erratically and do not show any semantic meaning. All calls of zebra finches are innate with exception of the males’ “Distance” calls.

### 1.1.3 Innate semantic calls of zebra finches

From a functional perspective, “Begging”, “Distance” and “Alarm” calls provide referential information for conspecifics and heterospecifics to indicate certain contexts [53, 54]. “Begging” calls are almost a sign signal for the parents to take care of their offspring [223]. Loud “Distance” calls are used to contact partners when they are out of sight [222]. “Alarm” calls explicitly refer to danger, such as the presence of predators [56].

### 1.1.4 The innate non-semantic calls of zebra finches

The “Stack” calls together with the “Tet” calls are the most common calls of zebra finches [246]. In the wild, they combine them in a very complex way with different locomotion. “Stack” calls are the most uttered vocalization in the laboratory [246, 15]. A “Stack” call is not sexually dimorphic. Although female “Stack” calls seem to be variable in duration from 0.05 to 0.15 seconds [246], we observed that male “Stack” calls varied in duration from 0.05 to 0.1 seconds, too.

Time series approach of cross-correlation is used to address the possible coordination between two animals. Cross-correlation in the occurrence time of “Stack” calls of male and female zebra finches showed consistent interaction patterns between bonded zebra finch pairs. Although the number of calls differs every day, the possibility of calling and answering between male and female of a pair is consistent after bonding [77]. The correlation of elapsed time indicated that the male “Stack” calls were used to respond more contingently to the mates than to others. Unlike “Alarm” calls, which are produced when animals are presented with predators, “Stack” calls are not an explicit semantic signal. However, the overall daily occurrence of “Stack” calls may provide semantic information. As opposed to the conventional sense of signaling, the successions of zebra finches’ “Stack” calls and mouse’s ultrasonic calls show large variations (Figure 1.2). Neither a single “Stack” call nor a succession of “Stack” calls can easily be determined as signaling. One difficulty in characterizing the usage of the “Stack” calls is due to the lack of predictability.



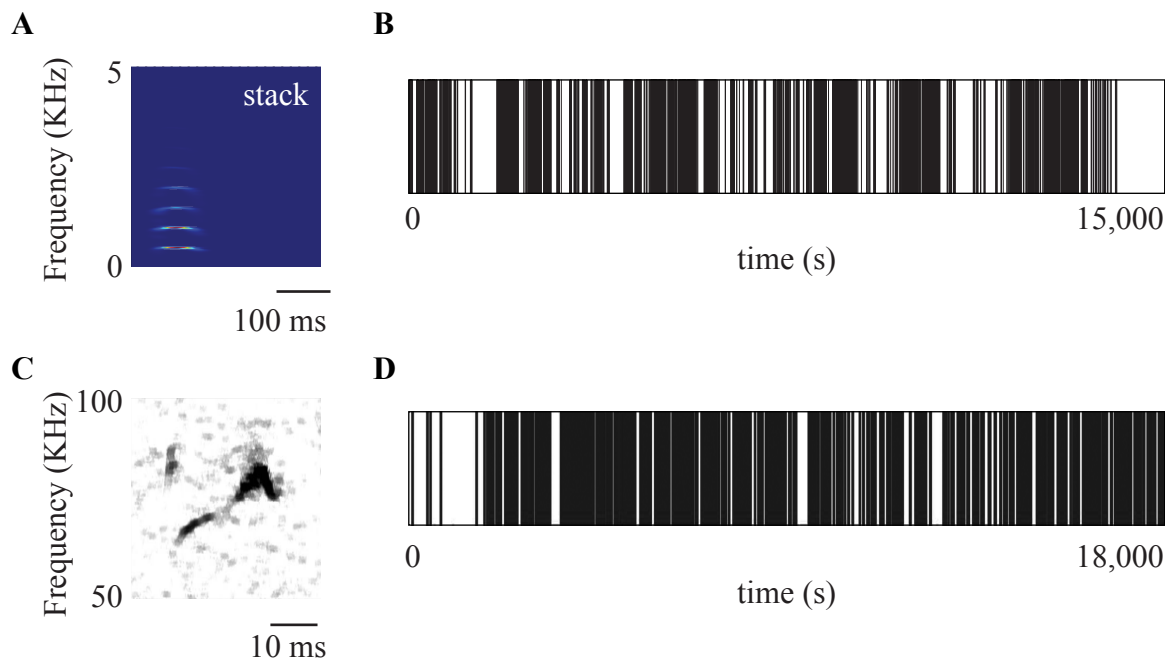


Figure 1.2: Comparison of the vocal dynamics between zebra finch and mouse. (A) A zebra finch’s “Stack” call, (B) a succession of zebra finch “Stack” calls. (C) A mouse’s ultrasonic call, (D) a succession of mouse calls. Figure 1.2 C & D are modified from [152] with CC-BY license 4.0. Black lines indicate the calling events.

## 1.2 Plasticity of bird call productions

Here, I address the plasticity in relation to the usage of the calls, but not in relation to their spectro-temporal structure since most calls are stereotyped innate sounds, as detailed before.

As stated in the introduction, semantic calls can also occur in a nonspecific context. The “Food” calls are semantic calls of monkeys and chicken [136, 93]. These calls are not only given in this context, but are often produced in nonfood contexts [136, 93], indicating that the production of these putative “Food” associated calls is not necessarily “Food” specific, but may have other social consequences or could express other implicit contexts. Further, the “Food” calls in monkeys are produced throughout the day. The analysis of all-day activities showed that the rate of “Food” calls changed but not the structure, suggesting that the call rates convey information about the physiological states of the monkeys [93]. Budgerigars, parrots and chicken changed their call rates adaptively in response to different schedules of reinforcement [117, 78]. Moreover, operant conditioning in budgerigars [130] and humans [116] confirmed that call usage is plastic in association to different contexts. Cumulative evidence suggests that call rates of birds are plastic and that call usage can be adapted to a particular context. The utterance of “Stack” calls in zebra

finches may also be adaptive, i.e. change from moment to moment to cope with changing situations during social interactions. In my thesis, I refer to the term “adaptive” in the sense of behavioral plasticity.

### 1.3 The structure of vocal interactions

There are two major forms of vocal communications: one-way communication and two-way communication. The one-way vocal communication refers to a series of actions that involve the sender, encodes information and sends it to the receiver. In the one-way vocal communication, sensory feedbacks from the receiver can, but do not have to be involved, whereas the two-way communication involves mutual activity. Vocal interaction is a special form of the two-way communication, in which the coordination between the sender and the receiver is rapidly adapted from moment to moment to cope with changing situations. Both, one-way and two-way communications are determined by a set of conditions including the time, the place, the process and the context. In addition, it can be influenced by the circadian rhythm, the environment, the way of vocal performance and the contextual relativeness situated in the vocal communication.

Vocal interactions may be classified in three forms, i.e. alteration, duet and chorus according to their temporal organizations and the number of participating animals [98]. A wide range of birds, for example, magpie larks (*Grallina cyanoleuca*) [87], plain-tailed wrens (*Pheugopedius euophrys*) [236], and white-browed sparrow-weavers (*Plocepasser mahali*) [225] are duet-singing birds. Duet-singing birds produce antiphonal notes for territorial defense in addition to identifying different neighbors. Hall and Magrath discovered that the temporal precision of duet singing is learned during pair bonding in some songbird species [87].

### 1.4 The models of vocal interactions

Developing an approach to quantify the calling activity and determining dynamic patterns of calling interactions, may open the possibility for a comparative study to determine the function of innate calls via model species, such as monkeys, mice and zebra finches. In particular, here I present methods allowing the analysis of calling-based interactions in both, the short-time range (seconds) and the long-time range (entire day).

Insights into behavioral patterns are often gained by using model systems that describe complex behaviors. To describe the dynamics of vocal interactions in humans, a specialized computer system, called the Automatic Vocal Transaction Analyzer (AVTA) [34], was used to sample and detect vocal interactions by which the vocal activity of each individ-

ual was sampled every 250 ms. These time series were transformed as a sequence of four binary numbers: 0, 1, 2 and 3, which represent the four observable dyadic vocal states of the dialogue: 0: partner A and B are both silent; 1: A is vocalizing while B remains silent; 2: B is vocalizing while A remains silent; 3: A and B vocalize simultaneously. This recording approach is an exhaustive classification of everything that can possibly happen in a two-person, on-off vocal stream [106] and has been widely used in many studies of two-person vocal interactions [96, 107, 57, 58, 106]. However, considering the rapid interaction with monosyllabic calls in many animal species, the response time of “Stack” calls (time difference between the offset of initial call and the onset of answering call) between a zebra finch pair can be as short as 250 ms. the 4-Hz sampling rate is much too low for detecting calling interactions in zebra finches. To achieve a better time resolution of vocal states and a sufficient acoustic resolution to differentiate between different vocal types, we incorporated a high sampling rate (22050 Hz) into the dyadic interaction system in zebra finches’ dyads.

Once animals engage in vocal interactions, they have to balance between self-generated action patterns and reactions to cope with changing internal and external environments. By using “Stack” calls, the challenge for the birds during interactions and for us in characterizing self-generated action patterns and reactions during vocal interactions lies in both actions being performed identically in terms of muscle activity and production-related movements. One solution for disassociating self-generated action patterns and reactions is to use a mixture model by dividing a behavioral pattern into self-dynamics and pairwise reaction [11]. With this, we are able to attribute experimental data to the model and demonstrate their statistical properties [11], thereby exploring the changes of specific activity of the model in response to perturbation experiments [11]. By adapting this mixture model for the zebra finch’ dyads, a calling interaction can be divided into self-contained and reactive calls. Then we examined the statistical supports [153, 39, 206] for the attributed model consisting of self-contained and reactive calls, which represent self-dynamics and reactions, respectively. In difference, traditional time series analysis of behavioral data is mainly employing auto-correlations and cross-correlations [80] and the Markovian analysis [84] to analyze interactions between individuals, but their products only reveal a short-time range of temporal correlations.

### 1.4.1 Vocal interactions in the long-time range

Calling interaction in zebra finches is a complex behavior. To explore a complex behavior without much prior knowledge, we first require a quantitative description for this behavior. The “Stack” calls in zebra finches are produced over a large dynamic range. That is: some calls are produced in a very fast succession, whereas some calls are produced after long silent gaps, sometimes many orders of magnitude in time (Figure 1.1). Despite their erratic occurrences, intuitively, people would think that the production of “Stack” calls should not be random. Due to a large dynamic range, power-law distribution can

be an ideal candidate to capture the characteristic of the productions of bird calls. However, power-law distribution is not the only distribution that describes natural phenomena. Some human activity, such as making telephone calls, in which the distribution of the time it takes before next call, is well approximated by exponential functions. A report has shown that the distribution of inter-spike interval in the efferent optic nerve of the horseshoe crab fits well to the lognormal distribution [228]. These right-skewed distributions share a similar characteristic. That is: most events occur with much shorter inter-event intervals than a small number of events with longer inter-event intervals, which results in a tail to the right of the probability density histogram. Debates about the tails of the inter-event intervals is dedicated to the dynamics of natural behaviors [128]. There are critical problems in the tail distributions to describe natural phenomena, in particular it is not clear whether they are power-law, lognormal, exponential or something else [153, 39].

In given situations, the behavioral mean should be fluctuating around a certain point to which animals maintain their behavioral variance within bound. That is essential for the central limit theorem. Not only in the infinite physical system [198], but also in the biological system, the Gaussian distribution has its meaning, so that being away from the mean is not an error. In the biological system, especially in freely behaving animals, the mean and the variance are not Gaussian. In fact, the variances are much larger than the mean. In 1961, Taylor summarized that the population variances of many organisms are proportional to a fractional power of the population mean in nature [210].

In contrast to the studies of physics, where the infinite system helps to predict extreme cases, the studies of the biological system are always limited by finite sampling. Moreover, even in the physical world, new studies in the fields of fractals and chaos characterize objects or events with features with no characteristic scales [198]. The attractive feature of power-law should also be considerable in the biological system. Examples of power-law behaviors have been studied in a wide range of complex phenomena in the biological system, including the allometric scaling in animals [230], network behaviors in human society [9], metabolism [176], neuroscience [204] and the bursty nature of human activity [7]. Nonetheless, if claiming a biological behavior or phenomenon is a power-law function, one must be cautious because the concept of power-law behavior originates from the infinite system [206]. Therefore, a consensus analysis and a systematic statistical test must be done, when using a power-law function. If used with caution, a robust feature of the power-law may help in data interpretation [206].

### 1.4.2 Mathematics of power-law and the empirical data

To determine, whether the data obey a power-law, a framework for quantifying power-law function in empirical data is required. In this section, I will describe some mathematical properties of power-law distribution and compare them with empirical data in calling behaviors. Most of the variables refer to the inter-call intervals. Therefore, I use  $\tau$  to

indicate inter-event interval in the following functions. Mathematically, if variables  $\tau$  obey power-law, the probability density function of  $\tau$  follows:

$$p(\tau) \sim \tau^{-\alpha}$$

, where  $\alpha$  is the exponent. This is an approximation function, meaning that in natural phenomena, not all  $\tau$ , but most of empirical data are applicable for power-law distribution. The limitation is not only because of practical reasons, but also due to mathematical reasons. If considering the probability function follows:  $p(\tau) = C\tau^{-\alpha}$ , where  $C$  is the constant.  $p(\tau)d\tau = C\tau^{-\alpha}d\tau$ . The density function is convergence, only if the value of  $\tau$  is bigger than certain values. For instance, the series of  $\int_{\tau_{min}}^{\infty} \tau^{-\alpha}d\tau$  converges, but the series of  $\int_{\tau}^{\tau_{min}} \tau^{-\alpha}d\tau$  diverges as  $\tau \rightarrow 0$  [153, 39].

To reveal the power-law behavior of the empirical data, it is to plot the histogram on the logarithmic scales of both sides of the probability density equation. The data are normally treated with a certain width of bins. There are several ways to define bin widths, such as linear and logarithmic in order to achieve the best representation of the tail distribution of the data. Using linear bin width, as bins have equal size, the calling behaviors with linear bin often show a high value plateau (Figure 1.3 A), consisting with the data in which some calling events occur after a long waiting time. The exponent will be strongly biased by the high values. Using logarithmic bin width, the data will be sampled with increasing bin widths; in turn, the sampling errors will be decreased. Although the estimated exponents with logarithmic scale may capture the true value, as I choose manually in Figure 1.3 A, the bin width must be selected with caution. However, either using linear or logarithmic bin widths, some information from the data must be discarded because of the sampling by binning.

To capture the data without throwing away any data point, it is useful to transfer the probability density function into a cumulative density function of power-law distribution (Figure 1.3). If the probability density function is  $p(\tau)$  and the cumulative density function is  $P(\tau) = \int_{\tau}^{\infty} p(\tau')d(\tau')$ . For  $\tau \geq \tau_{min}$ , we obtain:  $P(\tau) = (\frac{\tau}{\tau_{min}})^{-\alpha+1}$  [153, 39]. The cumulative distribution also follows a power-law, but displays a different exponent  $\alpha - 1$ . For example, if the exponent of the cumulative distribution is 1.5, the true exponent  $\alpha$  will be  $1.5 + 1 = 2.5$ . All exponents mentioned in this thesis are the “true” exponent  $\alpha$ , except for specific indications. For  $\tau \geq \tau_{min}$ , the tail of the cumulative distribution will still fall onto a straight line. The complementary empirical cumulative distribution may capture the power-law behavior without losing any data points. The plotting procedure follows: 1): Let  $(\tau_1, \dots, \tau_n)$  be the inter-event intervals. 2) The values are sorted in ascending order  $(\tau_1', \dots, \tau_n')$  with a new set of indices:  $(1', \dots, n')$ . 3) The tail distribution  $1 - P(\tau)$  can be seen by plotting  $1 - \frac{indices}{n}$  against sorted inter-event intervals  $(\tau_i)$  on logarithmic scales [153, 39]. 4)  $\tau_{min}$  to  $\tau_{max}$ , where  $\tau_{max}$  is the largest value of the data. Worth noticing that the exponent value  $\alpha = -(b - 1)$ , where  $b$  is slope of the fitting line to the eCDF (Figure 1.3 B, light blue).

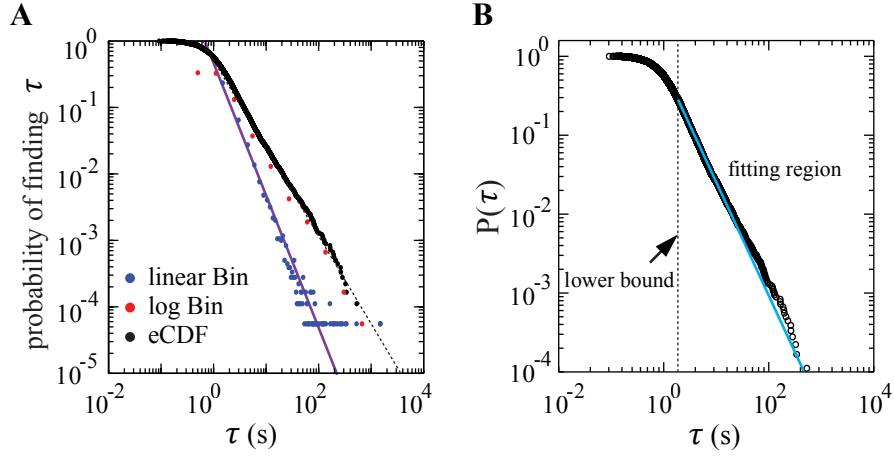


Figure 1.3: Plotting empirical data and fitting to power-law. (A) Comparison of plotting empirical data with a linear bin, a logarithmic bin and plotting data using complementary cumulative distribution (eCDF). Bin width for linear bin is 1.5 seconds. Logarithmic bin has  $k = 11$  bins, thus it creates spanned intervals  $10^{n_1}, \dots, 10^{n_{11}}$ . The values of  $n_i$  are linearly spaced from  $\log_{10}0.5$  to  $\log_{10}\tau_{max}$ , where  $\tau_{max}$  is the largest value of the data. (B) Power-law function makes sense for the values larger than a lower bound.

## Moments

Moments are an important mathematical quantification of a function, which can be used to differentiate power-law distribution from another distribution. The  $n$ th moment of the distribution is defined as:  $\langle \tau^n \rangle \sim \int_{\tau}^{\infty} p(\tau) d\tau$ . The first moment  $\langle \tau \rangle$  is the average inter-event interval. The second moment is the variance. There are higher moments, such as skewness and kurtosis that represent the shapes of a probability distribution. While there is a fixed  $\tau_{min}$ , the maximum inter-event interval could go to  $\infty$ . When integrating the distribution function over the range from  $\tau_{min}$  to  $\tau_{max}$ , the  $n$ th moment of the distribution will be:  $\langle \tau^n \rangle = \frac{\tau_{max}^{-\alpha+n+1} - \tau_{min}^{-\alpha+n+1}}{-\alpha+n+1}$  [8]. By using the limit prediction, the value of  $\langle \tau^n \rangle$  depends on  $-\alpha + n + 1$ . If  $n \leq \alpha - 1$ , the value of  $\langle \tau^n \rangle$  will be finite as  $\tau_{max} \rightarrow \infty$ . If  $n \geq \alpha - 1$ , the value of  $\langle \tau^n \rangle$  will go to infinity as  $\tau_{max} \rightarrow \infty$ . Therefore, all moments diverge if  $n$  is larger than  $\alpha - 1$ .

If the power-law scaling with the exponent  $\alpha$  is in a range between 2 and 3, only the first moment (i.e. the average value) is finite, the second moment (i.e. the variance) and the higher moments go to infinity. Yet strictly speaking, the second moment goes to infinity only if  $\tau_{max} \rightarrow \infty$ . For empirical data, the maximum of inter-event intervals always exists, but the measured variance of the data that follow power-law must be significantly larger than the measured variance of other data, that follow lognormal or exponential distributions.

## Scale

**Scale-free:** Power-law distribution is also called “scale-free” distribution. The name of “scale-free” captures the feature of the distribution that lacks an internal scale. Simply speaking from the perspective of the moments, if the distribution is a bounded distribution that has convergent moments, such as exponential distribution, the randomly chosen value is in the range of  $\langle \tau \rangle \pm \sigma$ . The randomly chosen value from a power-law distribution will be significantly diverse from the average value  $\langle \tau \rangle$ , because the variance can be very large.

Another feature of scale-free distribution is that the distribution is the same, from whatever scale we look at it [153]. That means, the distribution feature is the same if scales of variables (the inter-call interval, in my case) are multiplied by a common factor:  $p(b\tau) = g(b)p(\tau)$ , for any factor  $b$ . Detailed proof of the scale-free property can be found in Newman’s review paper [153]. In this thesis, I simplified the function as being a homogeneous function to better illustrate the scale-free property. Let  $g(b)$  be  $b^{-\alpha}$  that is equivalent to  $p(\tau)$ . Thus:  $p(b\tau) = C(b\tau)^{-\alpha} = Cb^{-\alpha}p(\tau)$ . It represents that the form of the function is the same (i.e. as a power law with an exponent ( $\alpha$ ) as variables  $\tau$  multiplied by any factor  $b$ , only the constant has changed proportionally from  $C$  to  $Cb^{-\alpha}$ . The power-law distribution is the only function that satisfies the scale-invariant property [153], which is unique, because the sharp of the function can be zoomed in or out by any factor, in contrary to the sharps of other distributions varying by looking at different scales.

## Periodicity

Power-law distribution is not the only right-skewed distribution for natural phenomena. There are still other systems such as Poisson processes, which would result in a more rapid decrease of the likelihood of elapsed times as the elapsed times increase. Or perhaps, least of all, it would be a chaotic process, which would result in an uncorrelated white noise. A better insight in the periodicity can help us to characterize the fluctuations of natural behaviors.

In this section, I compared different  $1/f^\beta$  noises by following Gisiger’s review on “scale invariance in biology” [79], including the white noise ( $\beta = 0$ ), Brown noise ( $\beta = 2$ ), pink noise ( $\beta = 1$ ) and violet noise ( $\beta = -2$ ). I first generate random time series that follow the characteristics of different color noises (Figure 1.4). The occurrences of the white noise values follow a Gaussian process. It gives values alternating around an expected mean value ( $\mu = 0$ ) within an expected boundary ( $\sigma$ ). Most importantly, the white noise represents an uncorrelated signal: the values are totally independent of any other value on the time series (Figure 1.4 A). If the next value is produced by adding a random value by following the Gaussian distribution to a previous value, the process of integrating white noises will generate a Brownian motion, the time series is so-called Brown noise. By this process,

the values are strongly correlated in a short range of time (Figure 1.4 B). A short-range fluctuation results in a significantly different time series in comparison with white noise, although both time series follow the Gaussian process. In the Brown noise time series, the values keep increasing or decreasing for a certain range of time, as we can see on the curve (Figure 1.4 B). On the other hand, the time series of pink noise appears to be similar to Brown noise, except that pink noise is noisier with a lot of high frequency modulation along the whole time series (Figure 1.4 C). The pink noise process may be equivalent to the Brown noise process plus the white noise process, just as if the behavior of the Brown noise process is observed with measurement errors. Therefore, the pink noise can be observed in many physical and biological systems [173, 79]. In visual appearance, the violet noise seems to be similar to the white noise, because the values of the violet noise are also positively and negatively alternating around an expected value 0. Since the violet noise is the result of the differentiation of the white noise, the positive and negative values are also alternating around an expected value, resulting in a time series that appears similar to those of the white noise (Figure 1.4 D). In contrast to the white noise time series, the positive and negative values of violet noises are correlated. Such short-range of fluctuation requires a power-spectrum analysis to be distinguished from the white noise.

The periodogram or the power spectrum  $S(f)$  of the signal estimates the contributions of each frequency ( $f$ ) to the signal [79]. By analyzing the periodograms, we are able to differentiate between these color noises more clearly than looking at different time series. The white noise should have equal density at every frequency as expressed by the noise function:  $p(f) \sim 1/f^\beta$  with  $\beta = 0$ , as when plotting with the double logarithmic scale, the power spectrum of white noise signal shows an almost straight line with a slope equal to 0 (Figure 1.4 A). Brown noise is the consequence of integrating the Gaussian process, by which positive values are likely to be closed together in time, so as the negative values are likely to be closed together. Therefore, fluctuation between many successions of positive and many successions of negative values over time will result in more possibilities of lower frequency than faster frequency. The frequency is the fraction of time ( $f = \frac{1}{T}$ ). The lower the frequency, the longer the duration of fluctuation time. We modeled the various noises with the mean and the variance equal to 0 and 1, respectively. The time series of Brown noise appears to have large fluctuations over time. The power spectrum shows that the abundance of longer durations of fluctuation is much higher than shorter durations of fluctuation (Figure 1.4 B). For the pink noise, as there are more shorter durations of fluctuations,  $S(f)$  decreases slower than for the Brown noise (Figure 1.4 C). It is worth noticing that, by comparison to the white noise with  $\beta = 0$ , both, in Brown and pink noise power density decreases with increasing frequency. The pink noise with  $\beta = 1$  has equal power in all frequency ranges that are proportionally wide, whereas the power of Brown noise with  $\beta = 2$  decreases much faster than the pink noise by increasing frequency. The power spectrum of the violet noise  $\beta = -2$  falls on a straight line on a double logarithmic plot with a slope equal to 2. Further, the periodogram reveals a quantitative feature of the color noise signal. The power of the violet noise increases by increasing frequency, which indicates that higher frequency fluctuations occur more often than lower frequency



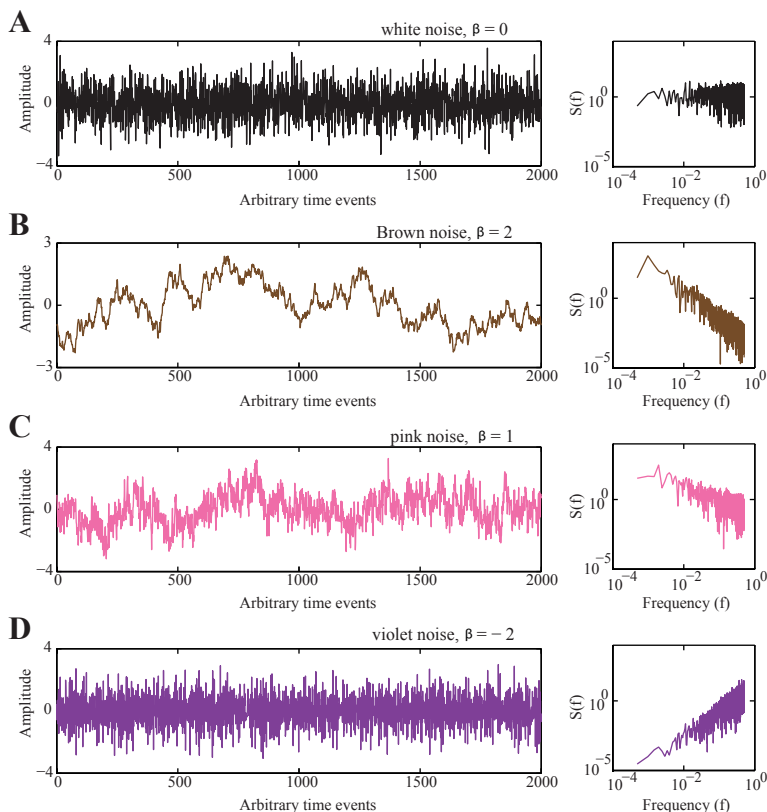


Figure 1.4: Time series and periodograms of different color noises. (A) White noise: The positive and negative values of white noise are alternating in time, but the values are not correlated. (B) Brown noise: The values are increasing and decreasing in a short period of time, which results in a long-range fluctuation. (C) Pink noise: that is equivalent to the brown noise, but with more high frequency modulations. (D) Violet noise: The positive and negative values are correlated and alternating in time.

fluctuations (Figure 1.4 D). Compared with the values of white noise, the values of violet noise are correlated, but these values are organized such that positive and negative values are alternating in time [165].

### 1.4.3 Statistic of power-law

Since the power-law descriptions have been discovered in the empirical world, many mechanistic models arose to explain power-law functions that seem to be omnipresent in biological and non-biological systems. More and more reports claim that certain behaviors follow power-law dynamics, but often lack sufficient data or statistical support [206]. A frequently encountered problem is that many empirical data are non-negative values and right-skewed, in which a tail will be seen by most of the right-skewed distributions.

Moreover, a straight line appearing on the log-log plot is a hint, but that is very far from statistical support of power-law, because the tail can be stretched in different shapes by binning the data with different scales [153] (Figure 1.3). Although in many people's views the data fitting is not an exact science compared with analytic mathematics, we will first acquire knowledge of whether or not a certain behavior is quantifiable. With this, we may further derive the underlying basis for this behavior. If statistical tests support that a behavior expresses power-law characteristics, the scaling features of power-law can help us to explore the biological relevance of this behavior, such as by using perturbation experiments (see Manuscript 2.1).

In this section, I introduced the fitting procedure described by Clauset et. al. [39]. Although a perfect fitting can neither exist nor be realistic, their method has so far been the best algorithm to fit power-law to empirical data in respect to its mathematical features. Among others, the most important aspect is that power-law scaling is not valid for the full range of data. Lower values saturation ( $\tau_{min}$ ) and higher degree cut off ( $\tau_{max}$ ) are often observed in real data. With the constraint of the ranges between and , we will have a clean region that fits to the power-law (Figure 1.3 B). In contrast to the scaling properties of network sciences, in which the  $\tau_{max}$  has a clear meaning with the analytical support, the  $\tau_{max}$  does not have a clarified meaning in vocal interaction yet. In this thesis, I speculated that the  $\tau_{max}$  could be the largest inter-call interval at which it is still biological informative for one vocal event related to the previous one. In contrast to the  $\tau_{max}$ , which can go to infinity for a power-law, the  $\tau_{min}$  must be bigger than 0, not only for practical reasons, but also, most importantly, because of the mathematical reason derived from power-law properties. For the practical reason, there would be very large numbers of possible fits to the empirical data if all the lower values were considered for the fitting. For the mathematical reason, there must be a lower bound value of  $\tau_{min}$  by which the power-law makes senses, because the integral of power-law function  $\tau^{-\alpha}d\tau$  over a certain range of value will be convergence only if the range is between  $\tau_{min}$  and  $+\infty$ , whereas the power-law function will be divergence when the values are approaching 0. Taken together, I fitted the data over a range from  $\tau_{min}$  to the possible largest value in observations.

To avoid confusions, I used “*min*  $\tau$ ” and “*max*  $\tau$ ” to indicate the dynamics ranges of the data, “ $\tau_{min}$ ” and “ $\tau_{max}$ ” to indicate scale regions for power-law. The fitting procedure has the aim to estimate the exponent from a discrete set of data points (i.e. the inter-call intervals,  $\tau$ ) from  $\tau_{min}$  to any possible largest value in observations (about 1000 seconds for our longest observation period: 10 hours during the day). The fitting algorithm follows Clauset et. al. by estimating the lower saturation for the scaling region [39, 8].

The fitting procedure aimed to estimate the exponent from a discrete set of inter-event intervals ( $\tau_i, i = 1, \dots, n$ ) from a lower bound:  $\tau_{min}$  to any possible largest value in a continuous recording. The steps of the fitting procedure are:

1. Estimate the candidate exponents  $\alpha'$  the candidate  $\tau'_{min}$  between the shortest and the longest inter-call intervals:

$$\alpha' = 1 + n \left[ \sum_{i=1}^n \ln \frac{\tau_i}{\tau'_{min} - \frac{1}{2}} \right]^{-1}$$

2. Many  $(\alpha', \tau'_{min})$  pairs will be obtained from step 1. Assuming that the discrete power-law distribution is:  $p(\tau) = \frac{1}{\zeta(\alpha, \tau_{min})} \tau^{-\alpha}$ , thus the cumulative density function (CDF) will be:

$$P'(\tau) = 1 - \frac{\zeta(\alpha, \tau)}{\zeta(\alpha, \tau_{min})}$$

3. Calculate the distance (D) between the empirical CDF and the estimated CDF:

$$D = \max_{\tau \geq \tau_{max}} |eCDF(\tau) - P'(\tau)|$$

4. Repeat steps 1 – 3 until we have all possible  $(\alpha', \tau'_{min})$  pairs calculated by whole range of  $\tau'_{min}$  from the shortest (i.e.  $\min \tau$ ) and the longest inter-event intervals (i.e.  $\max \tau$ ). By looking at the plot of distance (D) against  $\tau'_{min}$ , we will obtain the minimum distance with corresponding  $\tau_{min}$  and exponent  $\alpha$ .
5. Calculate the p-value to test the similarity (i.e. the distance (D) in step 3) between the empirical and the simulated CDFs. The Kolmogorow-Smirnov statistic was used to quantify the distance between the empirical CDF and the synthetic CDF with corresponding parameters. The closer the p-value is to 1, the more the simulated data with corresponding  $\alpha$  and  $\tau_{min}$  are similar to the real data.

#### 1.4.4 Vocal interaction in the short dynamic range

For theoretical reasons, power-law scaling does not cover the whole dynamic range (see Introduction 1.4.2.). The distribution of inter-call intervals deviates from the power-law characteristic below a minimum value  $\tau_{min}$ . In this subsection, I introduced the call-based vocal interactions in the short dynamic range from 0 up to 2 seconds (Table 2.1.).

As mentioned above, the dynamics of call-based vocal interactions can be defined as a mixture of the self-contained and the reactive dynamics. The self-contained and reactive behaviors in the short dynamic range can be revealed by the auto- and cross-correlations. Many studies have reported that the vocal activity of either individual was potentially contingent upon what the partner did over a window of up to 5 seconds for zebra finch dyads [52, 211, 77, 166, 205] and up to 60 seconds for most human dyads (Jaffe et al., 2001). These events that fall into the short dynamic range are often observed as the low-degree saturation in the range between 0 and  $\tau_{min}$  on the log-log plot of empirical data (Figure 1.3). This indicates that some reactive and self-contained events that occur faster than

$\tau_{min}$ , do not follow the power-law. In contrast to the plasticity of the calling activity that may give rise to the power-law behavior in the long dynamic range, the observation of the low-degree saturation in the range between 0 and  $\tau_{min}$  may due to the certainty of some calling activity. As the cross-correlation analyses showed, the zebra finches responded more contingently to their mates than to others. This indicates some higher degree of certainty in the vocal coordination between paired zebra finches than between unfamiliar zebra finches.

## 1.5 Neurophysiology of call-based vocal communications

In Manuscript 2.2, we studied the neurophysiology of the sensorimotor system for call-based vocal interactions in the short dynamic range. In this chapter, I will introduce the neurophysiological background of call-based vocalizations in birds.

Vocal communication is a complex behavior that allows animals to interact with each other by using sound signals. Interactions depend on the ability to control motor outputs while analyzing sensory inputs [174]. Previous studies showed that perceiving and producing sound signals requires both, sensory and motor systems [23, 227, 169]. In part, own sound signals and sound signals of others are monitored continuously in the auditory cortex of squirrel monkeys [50, 51]. The motor system and sensorimotor system both pertain to active and passive feedback [95, 30, 192]. Although sensorimotor systems are critical to task-specific motor performances and the motor learning [102, 217, 240], it is still unknown how sensorimotor systems contribute to natural vocal interactions. In the following sections, I will first introduce the motor control, then the auditory processing and the sensorimotor control of calls in birds. Subsequently, I will introduce the neurophysiology of vocal interactions in birds.

### 1.5.1 Motor control of bird calls

In contrast to insects and fish, calls of birds and mammals depend on respiration. The neural control of vocalizations of songbirds consists of two distinctive forebrain circuits to produce learned and unlearned vocalizations [157]. The pathways converge into a common pathway in the brainstem [233, 234]. The medullary vocal-respiratory circuit, including the syringeal motonucleus (XIIIts), nucleus retroambiguialis (RAm) and rostroventral respiratory group (rVRG), is common to all vocalizing birds [234, 177, 67] (Figure 1.5). All these nuclei of the vocal-respiratory circuit are defined by the mRNA expression of androgen receptors [69]. In addition to the vocal-respiratory circuit, anatomical studies and electrical stimulation showed that the avian midbrain, dorsomedial nucleus (DM) of the intercollicular complex (ICo) are a gating apparatus for initiating vocalization in chicken

[167], pigeons and zebra finches [235]. DM innervates RA and XIIIts that control vocal productions [27, 231, 232, 234, 235]. Electrical stimulation studies suggest that the mid-brain area receives a descending vocal command from the diencephalon in quails [193] and from the archistriatum in pigeons [248].

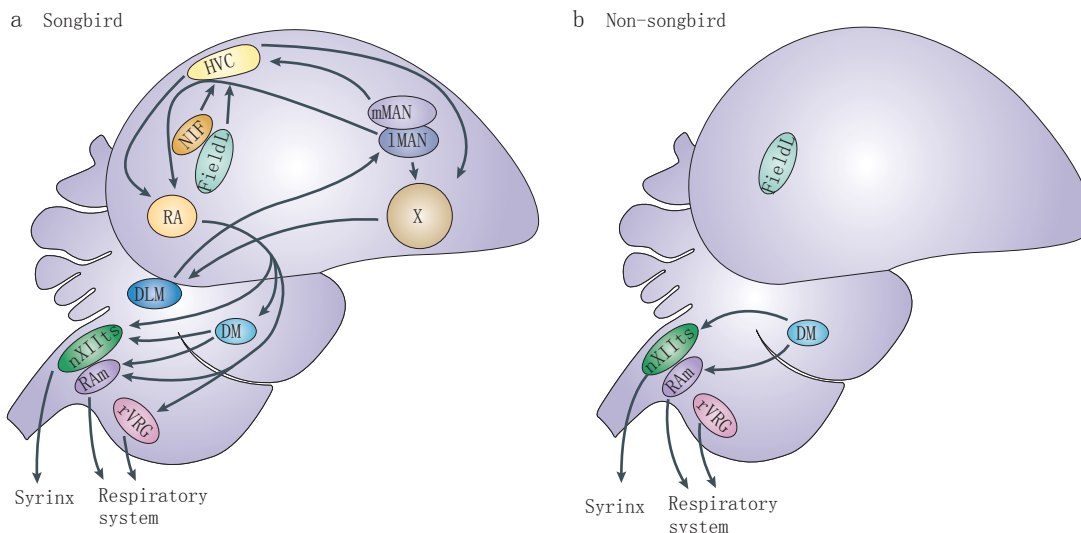


Figure 1.5: The song systems of songbirds and non-songbirds. Songbirds and non-songbirds share a similar motor system in the brainstem. The song system with interconnected nuclei in the forebrain is presented in songbirds as compared with non-songbirds. Compared with a non-songbird, a songbird has an elaborate interconnected nuclei in the forebrain. DLM, nucleus dorsolateralis anterior, pars medialis; DM, dorsomedial nucleus of the midbrain nucleus intercollicularis; HVC, higher vocal center; lMAN, lateral magnocellular nucleus of the anterior nidopallium; mMAN, medial magnocellular nucleus of the anterior nidopallium; NIF, nucleus interface of the nidopallium; nXIIIts, tracheosyringeal portion of the nucleus hypoglossus; RA, robust nucleus of the arcopallium; RA<sub>m</sub>, nucleus retroambigualis; rVRG, rostro-ventral respiratory group; X, Area X. Modified from [24] with permission.

Male zebra finches sing and call while females only call. Simpson and Vicario explored the sexual dimorphic vocal control of calls in zebra finches. They found out that male songs and female calls maintained largely unaffected after DM lesions. RA lesions however, impaired song performance in males dramatically. These results indicate that, in zebra finches, the vocal command of songs begins in the archistriatum [233]. But as common to all birds, instead of as a central pattern generator for vocalizations, the DM may be an integral part for gating songs and calls which are driven by an integration mechanism involving intrinsic and extrinsic inputs from cortex-like and subcortex-like forebrain areas. Although some in-vitro studies have been done to explore the physiological mechanism of DM [112] and the vocal-respiratory axis [183, 182], the behavioral correlation of these circuits is unknown in-vivo due to the difficulty of deep brain recoding in behaving

animals. Female zebra finches seem to produce calls even without RA projection to DM [235, 203]. Hence, ICo seems to serve as vocal command to drive call productions in both, male and female zebra finches. However, the neuronal mechanism underlying vocal gating of ICo/DM is unknown.

The forebrain vocal control system of songbirds consists of two pathways, the motor pathway and the anterior forebrain pathway, each of which emerges from the sensorimotor nucleus HVC [156]. The sensorimotor system plays an essential role in song learning and production [157, 139, 118, 4, 124, 146]. Recent studies showed that a specific type of HVC neuron reflects the feedback monitoring between sensory and motor information for song communication in swamp sparrows [172]. However, multiunit recordings showed that the neuronal activity in the forebrain nucleus HVC is also associated with the production of harmonic calls, although unspecified by the author [138]. More specifically, a subset of RA-projecting HVC neurons is associated with the production of “Distance” calls in zebra finches [85]. Furthermore, two other motor-related forebrain nuclei, the NIF and RA showed a neuronal correlation with the production of calls [138, 132, 211]. The first evidence that forebrain vocal areas are involved in call interactions comes from the work of Ter Maat et al. in RA [211]. Premotor activity of HVC related to innate calls such as the “Stack” call in a social context is unknown.

### 1.5.2 Auditory representations of bird calls

Vocal controls are tightly linked to auditory representations (Figure 1.6). Many behavioral studies showed that birds (whether they learn sounds or not) are able to discriminate sounds based on their structure and their semantics [160, 75, 71, 212, 24, 53]. Noticeably, zebra finches are able to distinguish individual “Distance” calls, which differ very little in their acoustic parameters [221]. For the auditory representation, most studies were dedicated to the neurophysiological basis of auditory learning, in particular, the song learning and memory in birds (reviews [71, 170, 212, 24]). Few were done for the auditory representations of bird calls [37, 15, 16, 53].

Field L is the auditory cortex analog forebrain area of birds [120, 81, 2]. It is subdivided into three sub-regions L1, L2 and L3 that are interconnected with secondary auditory areas in the caudal nidopallium (NCM) and caudal mesopallium (CM). A connection between the auditory system and the song system of songbirds emerges from the shelf region of the sensorimotor nucleus (HVC) [109, 133]. Concerning the auditory representations of natural sounds of songbirds, the midbrain nucleus mesencephalicus lateralis dorsalis (MLd) and the primary auditory region (Field L) are not selectively tuned to particular semantic sounds, but sensitive to the spectral and temporal structure of natural sounds [212, 243]. However, guinea fowls (*Numida meleagris*) showed auditory selectivity in Field L to species-specific calls [26]. Auditory selectivity to semantic sounds is commonly found in the secondary auditory system and the sensorimotor nucleus HVC. Neurons of NCM and CM of zebra

finches show a distributed pattern of auditory sensitivity to “Distance” calls [53]. The HVC neurons are selective for the birds’ own songs (BOS) [220], whereas the secondary auditory areas, including NCM and CM, are not selective for the BOS, but showed selectivity for conspecific songs [74, 75]. Lewicki and Arthur suggested a hierarchical processing in the avian auditory system by showing the gradients of an auditory context-dependent sensitivity between Field L and HVC [121].

During life-history, zebra finches produce variable versions of calls ranging from “Begging” calls in juveniles, “Nesting” calls during nesting periods, “Wsst” calls under stress and “Distance” calls for daily routines. Elie and Theunissen analyzed the neural representations of these semantic categories of calls in primary and secondary auditory areas. They found that the neurons in these auditory areas are highly selective for different semantic categories of calls [53] but are distributed throughout the auditory forebrain.

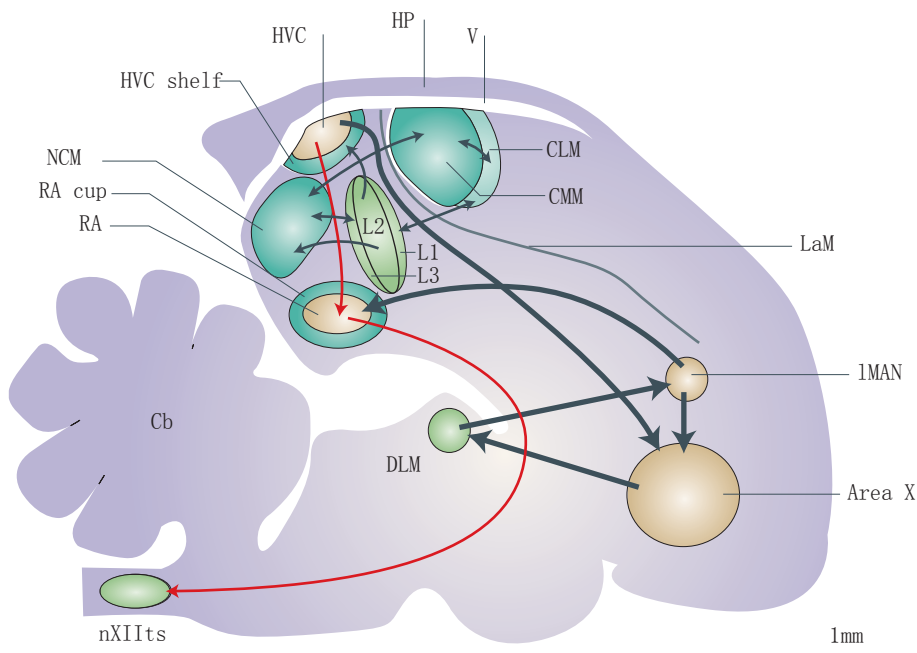


Figure 1.6: The forebrain auditory system and its connectivity with the song system. Abbreviations: Cb, cerebellum; CLM, caudal lateral mesopallium; CMM, caudal medial mesopallium; DLM, nucleus dorsolateralis anterior; HP, hippocampus; HVC, higher vocal center; L1, L2, L3, subdivisions of field L; LaM, lamina mesopallialis; IMAN, lateral magnocellular nucleus of the anterior nidopallium; NCM, caudal medial nidopallium; nXIIts, tracheosyringeal portion of the nucleus hypoglossus; RA, robust nucleus of the arcopallium; V, ventricle. Adapted from [24] with permission.

### 1.5.3 Neurophysiology of vocal interactions

To this day, very few studies have investigated the neurophysiology of inter-agent vocal interaction in a social context [211, 60]. Recently, a lightweight radio transmitter and a wireless neurophysiological system have been developed for recording individual sounds and neuronal signals simultaneously. With this approach, it has been shown that a bilateral interaction pattern of “Stack” call communication emerges between paired-bonded zebra finches [211, 77], and the premotor nucleus RA is related to call productions during calling interactions [211].

The sensorimotor nucleus HVC is widely credited with song learning and performance, but little is known on how the HVC contributes to the coordination of ongoing vocal interactions. In duet-singing birds, such as in plain-tailed wrens (*Pheugopedius euophrys*), males and females produce alternating singing to coordinate with each other. The sensorimotor nucleus HVC is associated with both, the production of duet songs [63] and the coordination of duet singing [60]. By analyzing the sequences of the duet songs and comparing to the solitary songs, Fortune et. al. showed that the duet songs of plain-tailed wrens are not fixed action patterns, which could be sang separately and yet precisely fell into each other’s song gaps [60]. On the contrary, the duet singing of plain-tailed wrens is very variable since singing errors occur quite often during ongoing duet singing [63].

Although the production of “Stack” calls in zebra finches per se does not need the forebrain pathway [199], the plastic usage of “Stack” calls may require forebrain pathways to cope with changing situations in a social context. In zebra finches, vocal interactions with “Stack” calls can be as short as 250 ms. If the response time exceeds the time of auditory processing, two possibilities may allow rapid interactions. First, they may learn a series of action patterns that timely interleave with each other, so that the resulting gaps between the actions of each performer are very short. Second, they may use the expectation, which is known to benefit the response time to stimuli in humans [86]. Studies in humans [100, 114] and monkeys [131] reported that the sensorimotor system and context-based interactions with the experimental objects play an important role in the predictive coding of upcoming events, which constitutes an essential prerequisite to social interactions. However, the predictive coding during a natural social interaction remains largely unknown.

## 1.6 Neurophysiology of expectations

Neurophysiological research of expectation dates back to the 1960s. Kornhuber and Deecke found neural activity, so-called readiness-potential prior to the onset of self-initiated movements in the motor cortex of humans [113]. Walter et. al. discovered that human brains associated with expectation of sensory events by measuring the contingent negative variation [226]. This contingent negative variation consists of two components: the re-



sponse to the initial stimulus and the response to the terminal stimulus [226]. The interval between the initial and the onset of terminal stimulus was about 1 second. Both, initial and terminal stimulus alone, elicited a post-stimulus activity in the brain. Only if two stimuli were paired, the readiness-potential appeared prior to the onset of the stimulus, no matter whether the latter stimulus was passively attentive or actively prompted by pressing a button [226]. Furthermore, Walter et.al. showed that the readiness-potential was not only associated with volitional movements, but also associated with involuntary movement such as the corneal reflex [226]. Therefore, the readiness-potential occurs only if it is contingent upon a cue that is given prior to an operant event, no matter whether the operant event is either sensory or motor.

The relationships between sensation, movement and decision making have been debated heatedly since the readiness-potential and expectancy was reported in the 1960s [123]. In relation to the expectation of upcoming auditory events in context-based communications, here I focused on introducing some recent neurological observations of anticipated sensations in the sensorimotor or premotor cortices. The first question about the sensation in the sensorimotor or premotor cortices is their functional relevance since the sensory cortices may be totally capable of mediating the perceptual discrimination of the stimuli. For this question, Romo and his colleagues showed that the sensory representation in the cortical motor area contributes to decision making and guides motor behavior. In their study, when two paired stimuli at different frequencies (F1 and F2) were presented to monkeys (*Macaca mulatta*), the movement decision was made based on the discrimination between F1 and F2 [184]. Further, they found that the cells in the premotor cortex and the supplementary motor area fired sustainably during the gap period between F1 and F2. By contrast, the primary and secondary sensory cortices only fired transiently to F1 and F2 [184]. Although they did not further elaborate the importance of cued stimulus (F1) in the auditory representation of operant stimulus (F2), their data and analyses are similar to Walter's study and further suggest, that the decision to act according to the operant stimulus was made by matching the expectancy to the operant stimulus (F2), when the cued stimulus (F1) is encoded in working memory during the gaps between F1 and F2. Indeed, motor performance requires sensory information, and sensorimotor and premotor areas play a crucial role in motor planning and decision making [180, 40, 131]. Sensory information not only guides movement via feedback control, but also relates to feed-forward control regulating the accuracy and the timing of movements [241, 101]. If movements were only guided by sensory feedback, the movements would be slow and clumsy, since the brain would need to update the sensory consequence given by the last movements before performing the next. The solution to this problem is the inverse model of the sensorimotor integration, by which particular movements are expected to achieve certain desired sensory consequences by experiences. On the other hand, the forward model of the sensorimotor integration will anticipate the sensory consequence by actions [241]. In other words, a forward model anticipates what is going to happen by giving it a try. Despite the importance of sensory information to calibrate self-generated actions, in this thesis, I asked how the sensory information is coordinated in the sensorimotor system during ongoing vocal interactions. In

particular, does the sensorimotor system in birds generate neuronal activity to anticipate upcoming vocalization during context-based vocal interactions?

In Manuscript 2.2, we focused on the sensorimotor association in male HVC for the short dynamic range (0 – 4 s) of call-based vocal interaction in zebra finches. Since pair-bonded zebra finches are able to reliably react to their partners in less than 250 ms, a similar mechanism underlying expectation may exist in zebra finches as described above in humans.

## 1.7 Sex differences of auditory object formation

Previously, I introduced the dynamics of call-based vocal interaction in the short dynamic range between 0 and  $\tau_{min}$  and in the long dynamic range between  $\tau_{min}$  and  $+\infty$ . In this chapter, I will introduce further works in the auditory coding of natural sounds in male and female zebra finches' auditory forebrain. Here, I compared the auditory responses to natural sounds between the male and female auditory forebrain. I further elaborated a putative mechanism of a sensorimotor feedback that may cause the sexual difference of auditory responses to natural sounds in the auditory forebrain.

Resolving the acoustic information from a natural acoustic environment depends crucially on the temporal structures in addition to spectral structures [22]. In order to investigate the spectro-temporal regularity of sound elements in a complex environment, auditory scene analysis provides an ideal paradigm when compared with introducing a mixture of natural sounds that may induce more noises than information in a decoding scheme. In a classic auditory scene analysis, a sequence of syllables is presented to a listener. Each syllable consists of two tones at different frequencies (“A” tone and “B” tone). “A” and “B” tones are separated with a silent gap (inter-tone interval). While the repeated syllables are presented at increasing rates (i.e. the inter-syllable intervals are decreasing), a listener is likely to report hearing two separated streams, if the frequency difference between “A” tone and “B” tone is large [31]. With this approach, studies in human [31] and European starlings [18] reported that the inter-tone interval (the gap between “A” and “B”), the inter-syllable interval (the gap between two consecutive “AB”) and the spectral difference between tones are crucial to auditory stream segregation.

Previous studies by using operant conditioning and discrimination tasks reported that female canaries exhibited a higher probability of courtship solicitation displays (CSD) to song syllables that were uttered with high speed and were broad-banded [32, 44]. European starlings and zebra finches showed more attention to specific features in conspecific songs in discrimination tasks [125, 72, 73]. Furthermore, HVC-lesioned female canaries lost the acoustic preference for conspecific songs over conspecific poor-stimulating songs and heterospecific songs [32, 44]. Lesions to HVC affected the discrimination of conspecific songs

in European starlings [73]. Lesions of the anterior forebrain pathway of the song system and of HVC also affected the conspecific and heterospecific song discrimination in zebra finches [125, 187]. In summary, the forebrain nuclei of the song control system appear linked to the acoustic discrimination of male and female songbirds. However, the neuronal mechanism underlying the sexual difference of auditory perception and the sensorimotor feedback in maintaining auditory sensitivity to conspecific sounds is unknown. Field L has bi-directional connections with the secondary auditory areas of NCM (caudal nidopallium) and CM (caudal mesopallium) and a connection with the shelf region of HVC [62]. Although the physiological importance of these routes between sensory and sensorimotor system is unknown [212], given the anatomical and behavioral evidence mentioned above, I examined the auditory properties in male and female Field L to investigate putative mechanism of sensorimotor feedback underlying auditory sensitivity to conspecific sounds.

In Manuscript 2.3, we investigated the neuronal mechanism underlying auditory response to trains of syllables. The train of syllables was composed of repeated syllables including natural syllables and mechanic sounds. The natural sounds were: 1) conspecific song syllables that subdivided into the bird’s own song (BOS) syllable for males, i.e. the familiar song syllable for females, and the not bird’s own songs (NBOS) syllable for males, which was unfamiliar song for females. 2) heterospecific song syllables were extracted from a song of a common blackbird (*Turdus merula*). 3) mechanic gallop sounds consisted of two tones at different frequencies with 17.8 semitone difference (2800 Hz for “A” tone and 1000 Hz for “B” tone). In Manuscript 2.3, we examined the temporal coding (spike timing) and the spectral coding (spike rate) of male and female Field L neurons in relation to different stimuli with varied lengths of inserted gaps. Further, we employed an auditory scene paradigm to examine the temporal and spectral regulation of natural sounds in a playback train.

Unfortunately, the data of GABA treatments do not provide sufficient support for the impact of the pre-motor activity of HVC on the sensory properties of field L. In the discussion of this thesis and in the supplementary text of the Manuscript 2.3, I provide pilot data related to this question and discuss the putative mechanism of sensorimotor feedback in maintaining auditory object formation of natural sounds in the auditory forebrain.

## 1.8 A telemetric approach for simultaneously measuring vocal and neuronal activity of freely interacting animals

Telemetric approaches provide opportunities for advancing our understanding of how animals interact vocally with each other in their environment [189, 77]. Many group-living social animals, such as zebra finches, communicate with each other by using acoustically

rather similar vocalizations. The on-bird audio-transmitters and electrophysiological transmitters enable to gain temporal precision of behavioral and neuronal activities from behaving individuals simultaneously (Figure 1.7).

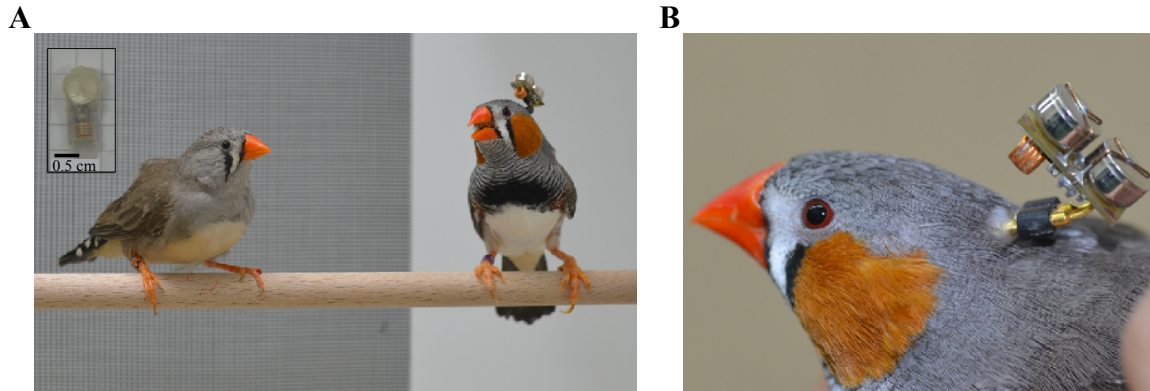


Figure 1.7: A telemetric approach for measuring vocal and neuronal activity of freely interacting animals. (A) Both male and female zebra finches were equipped with backpack audio-transmitters (invisible on-birds in this figure, but shown in zoom image). Male was equipped with an electrophysiological transmitter. (B) The male zebra finch was equipped with an electrophysiological transmitter.

Since the on-bird radio transmitter faces tightly toward the surface of the animal's body, the on-bird microphone records a higher power in the low frequencies as compared to the external recording and the recording of non-focal animals (Figure 1.8 A) [77]. This power-spectral characteristic enables us to differentiate individual vocalizations easily. Further, the same animal can be equipped with multiple transmitters to monitor other variables such as neuronal signals and vocalizations simultaneously. E.g., we have successfully recorded single-units in the male HVC (Figure 1.8 B). These singing-related units exhibited stereotyped bursting patterns that tightly locked to the song syllables during active singing (Figure 1.8 C).

The telemetric infrastructure consists of three parts: transmitter, receiver and data storage. We separate these parts into three individual boards to ensure real-time monitoring and big-data storage. We used two types of transmitters in this thesis: a wireless microphone for vocal activity and a wireless neurophysiological transmitter for neuronal activity (Figure 1.7). The wireless microphone weighs 0.6 g, including one battery (Typ 10, PR 70). The battery lasts for 12 – 14 days of continuous transmitting. The transmitting range is 5 meters. The wireless neurophysiological transmitter weighs 0.9 g, including two batteries (Typ 10, PR 70). The batteries lasted for 7 days of continuous transmitting. The transmitting range is 10 meters. The transmitted signals (transmission frequency: 375 – 380 MHz for audio signals and 270 – 290 MHz for neurophysiological signals) were picked up by respective antennas, which were positioned outside of the sound boxes. Each

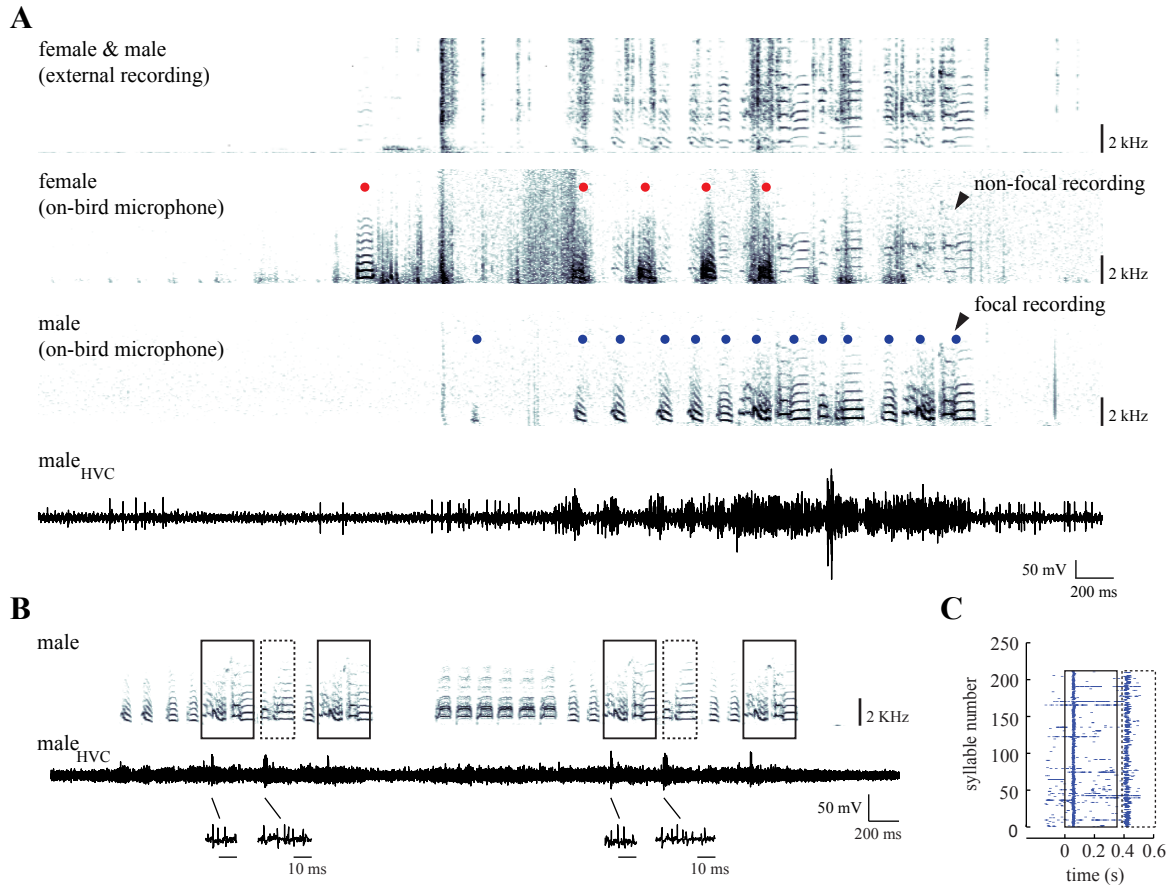


Figure 1.8: Simultaneous recordings of vocal and neuronal activity in freely interacting zebra finches. (A) Comparing spectrogram between external and on-bird recordings. Example multi-units recording in male HVC that synchronizes with microphone recordings. Color-coded circles indicate focal recording from the on-bird microphones (red: female, blue: male). (B) Example single-unit recording in male HVC. (C) The neurophysiological recording in HVC showed the singing-related unit exhibited stereotyped bursting patterns.

antenna was connected to the communications receiver (AR8600, AOR, USA). All signals coming from the respective communications receivers (AR8600) were digitized simultaneously by an 8-channel audio interface (M-audio fast track Ultra 8R, USA). We sampled multi-channel signals at a rate of 22050 Hz by using a customized program (ASIO Rec by M. Abels) and stored them as uncompressed files (WAV). The sounds were extracted from an uncompressed WAV file by setting a trigger level. A customized program (Sound Explorer Pro by R.F. Jansen) was used to compare sounds automatically with respect to a set of temporal-spectral parameters: the duration, the mean frequency and its standard deviation, the model frequency and its standard deviation, the wiener entropy and its standard deviation, the fundamental frequency and its standard deviation, the time of zero crossing, the time of maximum positive peak and the time of maximum negative peak.

The sounds were subsequently sorted by a k-means clustering algorithm according to the distance matrix generated by the temporal-spectral comparison. After manual refinements by visual inspection of the spectrograms of the sounds, data are saved as: 1) a text file that contains information of the timestamps and the cluster identities. 2) a set of bitmap files that illustrate all sounds in each cluster. 3) a set of binary files that contain the spectral information of the images. As the result, a duration of 8 hours recording ( 1.18 TB, sample rate: 22050 Hz) can be reduced to several hundred KB of text file that contains a precise time series. To process the neurophysiological data, the data were down-sampled to 11050 Hz and subsequently high-pass filtered with a cutoff frequency 300 Hz. In order to have an overview of synchronized data, the vocalization data were down-sampled to 8 KHz and fed into Spike 2 (Cambridge Electronic Design, UK) together with the neurophysiological data. Spikes were extracted and sorted with spike 2. Spike-forms and spike-times were saved as a mat file for further analysis in Matlab (MathWorks, US).



# Chapter 2

## Manuscripts

This chapter consists of three manuscripts. The first and the second manuscripts cover the dynamics of call-based vocal interaction in the long dynamic range between  $\tau_{min}$  and  $+\infty$  and short dynamic range between 0 and  $\tau_{min}$ , respectively. The first manuscript was designed to identify the behavioral characteristics, whereas the second manuscript contributed to the neurophysiological base of call-based vocal interactions in the short dynamic range. The third manuscript was to examine the sexual differences of auditory responses to complex auditory scene in auditory forebrain of zebra finches. In this chapter, I kept the original tags and letterings for figures with respect to each manuscript. I added additional tags for figures in this chapter in order to line up for the tags of this thesis. The first figure of this chapter will be “Figure 2.1”, the second figure will be “Figure 2.2” and so on. Further, I added all the references as part of the bibliography in the end of this thesis.

### 2.1 Power-law scaling of vocal activity in zebra finches

#### 2.1.1 Abstract

Social mammals and birds have a rich repertoire of communication calls. Some call types are uttered rarely but in specific contexts while others are produced in large numbers but are not linked to a specific context. An example for the latter is the “stack” call that male and female zebra finches utter thousands of times per day in a seemingly erratic manner. We quantified this calling activity of captive zebra finches by using on-bird telemetric microphones that permitted a precise temporal resolution. We separated the calling interactions into the reactive and the self-contained callings. Despite a large dynamic range in the succession of calling events, the temporal distribution of the reactive and the self-contained calling was characterized by a power-law with exponents ranging between 2 and 3, which implies that all calls in that scale have similar dynamic patterns. This evidence of power-law scaling provides an accurate description of the calling dynamics. This quantitative description shows that calling dynamics of mated zebra finches depend transiently



on physiological (water availability) and social (separation from the reproductive partner) situations of the individuals. The calling dynamics may inform about an individual’s state, even though a single “stack” call has no predetermined meaning.

### 2.1.2 Introduction

Many animals, in particular group-living birds and mammals communicate with repertoires of species-specific but individualized sounds such as calls that might be uttered in very large numbers during the course of a day [135, 15, 152]. Individual calls can be uttered either with a specific context [134, 56, 35, 194, 222, 215, 54] or without a discernable context [136, 93]. The latter have no obvious predetermined signaling value and often occur in diverse circumstances such as group progression and social interaction [137, 93, 126]. Previous works on vocal communication have focused mainly on context-specific calls. Those studies examined either the relationship between form and function of such calls [135, 54] or the linguistic-like characteristics [161, 162]. Context-specific calls account, however, only for a small part of the emitted sounds. Some group-living birds produce thousands of calls per day that are not answered by conspecifics and that are not answers to calls from other birds [211, 77]. “Stack” calls of captive zebra finches (*Taeniopygia guttata*) for example appear to be self-contained without specific context, in particular those that are not contingent upon mates or other group-members. Both self-contained callings (own call is uttered after an own call) and reactive callings (own call is uttered after the call of others) of mammals and birds occur with a large dynamic range [52, 209]. The problems of characterizing these events in previous studies are due to the fact that the production of these calls appears not to follow any discernible pattern.

Insights into behavioral patterns are often gained by using mathematical models that describe complex behaviors. In a stream of vocalizations for example, some callings might be self-contained action patterns and others are reactions during vocal interactions, while both actions are performed identically in terms of muscle activity and observed movements. One method to discriminate between self-contained and reactive callings is to use a mixture model that divides the behavior into self-dynamics and pairwise reactions [11]. Moreover, traditional time series analysis mainly employs auto- and cross-correlations to analyze self-contained and reactive behaviors but their results only reveal a short-range (1 minute or less) of temporal correlations in bird and human dyads [106, 52, 211, 77, 166, 205]. However, similar to human activities such as email communication [7] and mobility [200], the “stack” calling of zebra finches has a much larger dynamic range [52, 209]. Time series analytical tools that cover large dynamic ranges (hours to several days) such as power-law distributions showed indeed that human activities follow a power-law distribution [7]. That is, the frequency of interaction between two individuals is approximated by an inverse square power exponent ( $\alpha$ ) of inter-event intervals of interactions [7]. The power-law description confers strong predictive power and has been used to reveal the orderly scale-free pattern in many complex behaviors [91, 165, 198]. For example, healthy heart activity in

humans displays a scale-free pattern and changes of scaling exponents reflect a pathologic state [165]. We speculated that the dynamics of the self-contained and of the reactive callings might be described by power-law distributions and vary with the physiological states of the zebra finches.

Debates about the dynamics of natural behaviors concern whether the bursts of activities [52] and the distribution of their inter-event intervals are exponential, lognormal or power-law [128]. In this study, we quantified the calling activity of zebra finch pairs by using on-bird telemetric microphones that permit a high temporal precision of the measurements while the animals interact freely with their reproductive partners. We examined the statistical support [153, 39, 206] for various mathematical models for the self-contained and reactive callings, which represent self-dynamics and reactions, respectively. Subsequently, we probed the models of self-contained and reactive callings by perturbation experiments to directly access the plasticity and homeostasis of the behaviors [11]. We applied biologically meaningful perturbations by either limiting social-sensory cues or the access to water in order to explore the dynamic changes of the reactive and self-contained calling activity [202, 166]. In many species including the zebra finch, audio-visual integration is a common feature [25, 90, 221, 166]. Hence, separating auditory from other sensory cues emitted by the mate may change the reactive calling. Furthermore, the reproductive activity of zebra finches depends on the availability of water. Previous studies showed that water-restricted zebra finches reduced their singing activity [175] but otherwise adapted rapidly to that condition, reflecting their evolutionary origin as birds of arid zones. Hence in this study we removed water for a short period from one individual (male or female) of a mated pair at a time but allowed social (audio and visual) contacts between them. We examined dynamic changes of self-contained and reactive callings in response to the above treatments.

### 2.1.3 Results

#### Quantification of calling activity

To quantify the calling activity, we used a telemetric audio recording system to identify the vocal events generated by 22 male-female pairs of zebra finches with high temporal precision. Zebra finches do not utter sounds during the night. Thus, we excluded the long silent pauses corresponding to the night from the quantification of vocal dynamics. The daily patterns of successive calling events exhibited characteristic bursts of rapid vocal exchanges separated by periods of silence (Fig. 1a & Supplementary Fig. S1a – S1o). For both male and female zebra finches, the numbers of calls accumulated in a nearly linear relationship with time (Fig. 1b [i.]). About 90% of measurements (108 and 106 of 119 days of measurement for males and females, respectively. 22 pairs) showed high coefficients of linear relationships ( $R^2 > 0.9$ ) between the numbers of accumulated calls and time (Fig. 1c [i.]). If we consider that the change of calling activity ( $\Delta N$ ) with respect to time ( $\Delta t$ ) corresponds to the call rate ( $\gamma$ ), we obtain:  $\frac{\Delta N_i}{\Delta t_i} = \gamma_i$  (1), where  $i$  indicates the individual

male or female. The call rates ( $\gamma$ ) of males and females were not constant over time but appeared to vary appreciably with 86% of the regression coefficients being  $< 0.1$  (Fig. 1b [ii.] & Fig. 1c [ii.]). Furthermore, changes in call rate ( $\Delta\gamma$ ) varied in a range between  $-2$  and  $2$  calls per second squared (Fig. 1b [iii.] & Fig. 1c [iii.]). The average min are  $-0.66$  and  $-0.55$  and the average max are  $0.74$  and  $0.60$ , for females and males, respectively. The periodograms show a positive correlation between the power spectra ( $S$ ) and the frequency ( $f$ ) of  $\Delta\gamma$  (see example in Fig. 1b [iv.]). This indicates that positive and negative values of call rate changes are close to each other and alternate in time. The average slopes of the periodograms are  $0.86$  and  $0.82$  for female and male birds, respectively (Fig. 1c [iv.]). This analysis showed a short-range fluctuation in the changes of call rates. Subsequently, we explored the potential power-law behavior of the inter-event intervals.

### Power-law scaling of calling interactions

It is uncertain whether the burst-like vocal activity and the inter-event intervals between single calls are distributed exponentially, log-normally or according to power-law [153, 39]. To study the dynamics of calling interactions, we considered a model consisting of self-contained and reactive callings to describe calling interactions. For a fixed recording time window  $[0, T]$ , let the initial event be  $t_0$  at  $0$ , the times of calling events for female and male be  $t_{f_i}$  and  $t_{m_j}$ , respectively. Self-contained events follow vocal events of the same bird ( $t_{f-f}$  and  $t_{m-m}$ ), whereas reactive events follow events of the partner ( $t_{m-f}$  and  $t_{f-m}$ , Fig. 2a). An interaction can be characterized by a transition between self-contained and reactive callings in male and female zebra finches (Fig. 2b). To determine the burst dynamics of calling interactions, we measured the inter-event intervals ( $\tau$ ), including the female self-contained inter-event intervals ( $\tau_{f-f} = t_{f_{i+1}} - t_{f_i}$ ), the male self-contained inter-event intervals ( $\tau_{m-m} = t_{m_{j+1}} - t_{m_j}$ ), the female reactive inter-event intervals ( $\tau_{m-f} = t_{f_i} - t_{m_j}$ ) and the male reactive inter-event intervals ( $\tau_{f-m} = t_{m_j} - t_{f_i}$ ) of the 22 pairs of zebra finches. By plotting the distribution of inter-event intervals, we found that the empirical cumulative distribution function (eCDF) of the self-contained inter-event intervals (i.e.  $\tau_{f-f}$  and  $\tau_{m-m}$ ) and the reactive inter-event intervals ( $\tau_{m-f}$  and  $\tau_{f-m}$ ) is approximated by a power-law distribution:  $P_{Power-Law}(\tau) = \left(\frac{\tau}{\tau_{min}}\right)^{1-\alpha}$  (2), where  $\alpha$  indicates the exponent and min the lower bound (Fig. 2c, the example is taken from the female-male pair ‘‘a’’, see supplementary Fig. S1).

The power-law scaling is not valid for the full range of data. Lower value saturations ( $\tau_{min}$ ) and higher value cut offs ( $\tau_{max}$ ) are often observed in real data. In the mathematics of power-law, values of  $\tau_{max}$  can go to infinity and the values of  $\tau_{min}$  must be larger than  $0$ . For this reason, there would be very large numbers of possible models fitting to the empirical data if all the lower values were considered for the fitting, i.e., there must be a lower bound value of min below which the power-law does not make sense. The integral of power-law function  $\tau^{-\alpha}d\tau$  over a certain range of values will converge only if the range is between  $\tau_{min}$  and  $+\infty$ , whereas the power-law function will diverge when the values are

approaching 0. Taken together, we fitted our data over a range from  $\tau_{min}$  to the largest observed values. In contrast to power-law functions, the exponential functions are not constrained by a dynamic range.

The exponential cumulative distribution functions (Exp CDF):  $P_{Exp}(\tau) = \frac{\tau}{\mu}$  (3) with the corresponding measured average inter-event intervals ( $\mu_\tau$ ) did not capture the distribution of the data (Fig. 2c). Furthermore, the measured standard deviations of inter-event intervals ( $\sigma_\tau$ ) were significantly larger than the estimated values (Table 1 & Supplementary Table S1). In order to quantify statistical significance between the empirical data and the models, we fitted the data to the power-law cumulative distribution function (equation (2)) by using the fitting algorithms 32. Log-log plots of the eCDF of display a heavy tail decreasing in the fashion of a straight line (Fig. 2c). The Exp CDF decreased much faster than the eCDF and the Kolmogorov-Smirnov test between the Power-law CDFs and the Exp CDFs rejected the hypothesis of exponential distribution ( $p < 0.001$ ). In contrast, the Power-Law CDF fitted well to the data ( $p_{f-f} = 0.065$ ,  $p_{m-m} = 0.605$ ,  $p_{m-f} = 0.490$ ,  $p_{f-m} = 0.029$ , Kolmogorov-Smirnov test). 83% of our data (73 of 88 modes of callings, i.e. 4 modes in each pair) could be fitted significantly to the power-law distribution ( $p > 0.01$ , Supplementary Fig. S1 & Supplementary Table S1). Data that failed fitting to the power-law distribution also failed fitting to the exponential distribution. Taken together, our analysis showed that the dynamics of calling interactions were described by power-law distributions with  $2 < \alpha < 3$  (Table 1) in which the mean value of calling activity is finite, whereas the variance is divergent. A power-law function has a well-defined mean only if the exponent  $\alpha > 2$  and has a finite variance only if  $\alpha > 3$  (the mathematical explanations and further examples can be found in Supplementary Text S2 and Supplementary Fig. S4).

	min $\tau$	max $\tau$	$\mu_\tau$	$\sigma_\tau$	$\tau_{min}$	$\alpha$
$f - f$	0.120 (0.100)	879.5 (197.6)	4.045 (2.438)	21.624 (11.718)	$3.290 \pm 3.270$	$2.34 \pm 0.35$
$m - m$	0.100 (0.002)	603.0 (526.7)	4.535 (2.989)	30.652 (17.992)	$2.389 \pm 1.148$	$2.24 \pm 0.25$
$m - f$	0 (0)	371.9 (160.9)	3.331 (1.436)	14.474 (7.384)	$2.202 \pm 1.625$	$2.41 \pm 0.36$
$f - m$	0 (0)	338.0 (191.9)	2.175 (1.501)	11.100 (6.786)	$2.070 \pm 1.103$	$2.42 \pm 0.31$

Table 2.1: Table 1. Measured values and estimated parameters of the calling behavior. Abbreviations:  $f - f$ : female self-contained callings;  $m - m$ : male self-contained callings;  $m - f$ : female reactive callings;  $f - m$ : male reactive callings; min  $\tau$ : minimum inter-event interval (in seconds); max  $\tau$ : maximum inter-event interval (in seconds);  $\mu_\tau$ : mean inter-event interval (in seconds);  $\sigma_\tau$ : standard deviation of inter-event interval (in seconds);  $\tau_{min}$ : lower bound for fitting algorithms (in seconds);  $\alpha$ : exponent of power-law distribution. (measured value: median (25% quantile); estimated parameter: mean  $\pm$  s.d.,  $n = 22$ )

Previous analysis of human activities suggested that power-law dynamics were a simple consequence of circadian cycles and Poisson processes [129, 185], partially because the data included long periods of inactivity such as sleep during the night. To rule out this

possibility, we excluded the nightly period of inactivity and sorted the calling events with respect to their inter-call intervals in a descending order, and then plotted the sorted calling events against the day-times. Neither the occurrences of calling events after short inter-call intervals nor the occurrences of calling events after long intervals had a circadian bias (Supplementary Fig. S5).

### The power-law scaling changes transiently in response to perturbations

First, we consider that the burstiness of inter-event intervals could be the consequence of a fixed behavioral pattern. Especially the reactive events might not be causally related to answering mate’s calls. Second, although we define the states of events based on the observation of calls, some calls might follow other activity, e.g. nonverbal gestures. Thus, it is possible that 1) the reactive dynamics could be the consequence of a self-contained behavior of each individual instead of one depending on external cues such as the partner’s calls, and that 2) the predefined reactive state is irrelevant for a reaction. In order to rule out these possibilities, we performed a sensory perturbation experiment by allowing zebra finch mates to interact only via microphones and loudspeakers. The results from one such pair are shown in Fig. 3: The power-law dynamics persisted during separation in which the reproductive partners could communicate acoustically (Fig. 3b & 3d, “sep”) and showed a similar mathematical function to that during cohabitation (Fig. 3a & 3d, “coh”) in all four modes of callings. Moreover, we measured the calling behavior during isolation (Fig. 3c & 3d, “iso”) in which the reproductive partners could not communicate acoustically or visually. After isolation, the exponents of the self-contained callings stayed at around 2 (Fig. 3d), whereas the exponents of the “reactive” callings decreased strongly below 0.5 during isolation (Fig. 3d, Supplementary Fig. S6). In summary, the power-law scaling of the calling dynamics revealed that the self-contained calling behavior is distinct from the reactive calling behavior.

Subsequently, we tested the power-law scaling in response to biological relevant perturbations (1. reproductive partner is out of sight but in hearing range; 2. water is transiently not available) in eight zebra finch pairs (Fig. 4). The heavy-tailed dynamics were maintained unchanged during separation with acoustical interconnection ( $\langle eCDF \rangle_{sep}$ ) as compared to that during cohabitation ( $\langle eCDF \rangle_{coh}$ ) in all four modes of callings. The exponents of the self-contained female ( $p_{f-f} = 0.578$ ) and male ( $p_{m-m} = 0.371$ ) callings persisted unchanged as indicated by a two-sided Wilcoxon signed-rank test ( $n = 8$ ), whereas the exponents of the reactive female ( $p_{m-f} = 0.027$ ) and male ( $p_{f-m} = 0.039$ ) callings decreased during separation with acoustical interconnection as indicated by a right-sided Wilcoxon signed-rank test (Fig. 4a,  $n = 8$ ). This indicates that the self-contained callings are indeed self-contained and not affected by the social-sensory environment, whereas the reactive callings depend on multiple-sensory cues. However, we could not exclude the possibility that some reactive callings are by chance. Such incidental “reactive” callings can be considered as an experimental noise, as any other experimental

noise. Nevertheless, the perturbation experiments suggest that the incidental noise is not the main factor in this system.

### 2.1.4 Discussion

Zebra finches, like many other group living vertebrates produce large numbers of sounds per day [135, 152, 54, 219]. Calls that are answered by others or calls with a predetermined meaning, e.g. advertising calls, are generally thought to be biologically meaningful [56, 135, 222, 54]. However, calls without identified meanings that are not answered by conspecifics are a mystery. Next to answered “stack” calls, zebra finches of both sexes utter thousands of un-answered “stack” calls that are acoustically identical to the answered ones [211]. We show that the time intervals of both the self-contained “stack” callings and the reactive (answered and answers) “stack” callings of zebra finch pairs are characterized by a power-law distribution. This power-law model provides an accurate statistical description of the calling activity that changed transiently with varying physiological states. In particular, the calling activity during separation and water-removal experiments remained power-law distributed with exponents ranging between 2 and 3, indicating that the zebra finches maintained calling rates within bounds during altered environmental conditions and recovered to pre-perturbation values after the birds were re-supplied with water. Although the power-law distribution was frequently suggested for large data sets, there are actually very few good biological examples [206], including allometric scaling of metabolic rate [230] and power-law distribution of biological taxa [238]. The present study reports the first power-law scaling of vocal communications of vertebrates and extends the analysis of vocal time series that were previously limited to short time ranges of vocal contingency [211, 77]. Next to the impact of social interactions [166, 205], calling dynamics were regulated by the physiological states in zebra finches.

How could the power-law properties be biologically informative? Due to the scale-free properties of the power-law behavior, a receiver such as the mate may just need to analyze a fraction of all inter-call intervals to compute the expected distribution of daily calling intervals. Since this distribution is stable under normal conditions, a conspecific could learn the activity pattern of a focal individual, e.g. of the reproductive partner. The actual signature (exponent) of the calling pattern of an individual would inform the pre-informed receiver about deviations that reflect social or physiological disturbances of the calling individual. In relation, since the acoustic structure of “stack” calls is almost invariant [246, 54], the sequential hearing of several “stack” calls contains very limited information. However, information could be encoded in varying the inter-call intervals, which is reflected in the exponent of the power-law function. Adjusting the exponent of the callings may extend the information capacity limited by invariant calls such as the “stack” calls.

The power-law characteristics of “stack” calls allowed us to study the biological dynamics of this behavior. We tested its stability under different environmental conditions: 1)

visual separation of the reproductive partner, 2) auditory and visual separation of the reproductive partners (isolation), and 3) temporarily restricted access to water (see Materials and Methods) [202, 166]. We chose these conditions to simulate important situations in the life of zebra finches assuming that vocal communication of captive zebra finches is similar to their wild-living conspecifics. These birds live in the grasslands of Australia, mate for life and breed opportunistically following sporadic rainfall [247]. Vocal contact with the mate during foraging or dispersal in grassland is certainly part of their pair-maintenance behavior. Rainfall is essential to initiate breeding periods although the animals are well adapted to survive without drinking water for very long periods [202]. Each of the simulated conditions affected the exponent of the power-law function of either the self-contained calling, the reactive calling, or of both. The direction of change differed between experiments. A decrease of the exponent was observed following isolation and an increase of the exponent was observed following water removal. Furthermore, the distribution of the call intervals of the visually deprived or water-restricted males and females relaxed to the pre-perturbation values once the situation normalized. Thus, the daily calling intervals (ranging from 50 ms to 20 minutes) of mated male and female zebra finches are characterized by individual signatures that are stable over days and that are sensitive to varying sensory and physiological states. This long-term stability of calling pattern may indicate an important role for pair maintenance and success. Another role of the power-law behavior of calling activity might concern the situations in which the mate is out of visual and auditory detection range [166]. To know that long call intervals occur regularly might prevent zebra finches immediately searching the reproductive partner once it is out of sight. Because zebra finches mate for life and are opportunistic breeders, information about the mate and therefore well-synchronized vocal behaviors could be important for breeding success [77], as suggested based on the temporal correlations of reactive callings.

Some data that are approximated by a power-law distribution can also be modeled by the lognormal distribution (e.g. Gibrat's proportionate growth model) [145]. These are feasible models for natural phenomena and have comparable distribution properties, which result in very similar regressions over a range of inter-event interval values ( $\tau_i$ ) between  $\tau_{min}$  and  $\tau_{max}$  [145]. Gibrat's law may also hold for our data when assuming a growth process [76, 66, 145], where two zebra finches start to communicate with an initial set of 0. At each step, they may accelerate or reduce the rate of response times. After sufficient steps, their inter-event intervals ( $\tau$ ) change proportionately, independent of initial  $\tau_0$  and would be lognormally distributed. Besides and lognormal models, it is plausible that Weibull [201], bimodal distributions [244], and time-scale segmenting [128] may capture the dynamic patterns of daily calling interactions. However, all these models such as Gibrat's law explain only 50% of our data sets (44 of 88 modes of callings based on 22 pairs and 4 modes of calling per pair) while the power-law distribution explains 83% of our data sets. Our statistical tests and empirical measurements (Table 1, Fig. 1, Supplementary Table S1) showed that the daily moment-by-moment calling interactions of zebra finches, consisting of self-contained and reactive calling activity are best approximated by a power-law model.

Our data suggest two types of calling behavior in zebra finches: the self-contained and the reactive calling. This distinction is supported by the independent changes of the signatures of these two behaviors, e.g. after separation of the mates only the reactive calling behavior changes. It remains, however, to be seen if self-contained callings and reactive callings are controlled by separate vocal circuits, as suggested for primates [164], and how internal states could influence the homeostasis of calling behavior. Nevertheless, our study suggests that the utterance of large numbers of calls as seen in many birds [137, 135, 54] and mammals [93, 92, 152, 219] is a characterization of the physiological state of the sender, even though a single call has no predetermined meaning.

### 2.1.5 Materials and Methods

#### Animals

Animal care and experiments were approved by the Government of Upper Bavaria (Az. 55.2-1-54-231-25-09). All further animal husbandry or handling was conducted according to the directives 2010/63/EU of the European parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes. Experimental animals were adult male and female zebra finches obtained from our breeding facility. Zebra finches were kept in a 14/10 (Light/Dark) cycle, food and water *ad libitum*. For pair-formation, one female and one male bird were force-paired by keeping them pairwise in cages (size: 54x40x28 cm), placed in custom-made, sound-attenuated boxes. Each sound box was equipped with a microphone (C2, Behringer, Willich-Münchheide II, Germany), a speaker (FRS 8, 30w, 8 $\Omega$ , VISATON, Germany) and a telescopic antenna for wireless recordings. Further, 18 pairs were studied in the frame of the perturbation experiments (see below).

#### Sound recording

Vocalizations of birds were recorded for 4 to 10 hours per day between 05:10 a.m. and 23:20 p.m. (Supplementary Table S1). Each bird was equipped with a custom-made wireless microphone transmitters (0.6 g including a battery) [211]. The microphone was placed on the back and fixed with an elastic band around the upper thighs of the bird. The frequency modulated radio signals were received with communication receivers (AOR5000, AOR Ltd., Japan). Audio signals were either fed into an eight channel audio A/D converter (Fast Track Ultra 8R, Avid Technology, Inc. U.S.A.) and recorded with custom-made software or registered on a data recorder (DASH8X, Astro-Med Inc., RI, U.S.A.). The audio signals were recorded continuously and written to WAV files. Vocalizations were extracted from audio files by using custom-written software (sound explorer, available at <https://github.com/ornith>). The vocalizations of zebra finches were clustered by analyzing their sound features [211].



### Perturbation experiments

1. Separation: Eight zebra finch pairs were used for separation by limiting visual cues emerging from their mates (ID: c, i, j, k, l, m, n, o corresponding to Supplementary Table S1). Male and female zebra finches were separated into two sound-attenuated boxes for four hours. Animals were communicating via wired microphones and speakers, which delayed the communication by less than 1 millisecond (ms) (Supplementary Fig. S3).
2. Isolation: The calling activity of the male and female partner was tested in isolation (visual and acoustic separation) by separating the partners into two sound-boxes without acoustic and visual interconnections for eleven hours ( $N = 3$  pairs).
3. Temporarily restricted access to water: Seven zebra finch pairs (ID: p – v) were used for the water-removal experiment. Each pair was first put into a recording chamber for cohabitation during 5 – 7 days prior to the treatment. The procedure of the treatment was as follows: 1 day of water-removal from the male; 2 – 3 days of recovery; 1 day of water-removal from the female; 2 – 3 days of recovery. Vocalizations of both male and female birds were recorded through backpack microphones from 06:30 to 21:30 (light/dark cycle: 7:00/21:00). During the water-removal experiments, mates were separated with a transparent plastic board, so that the male and the female had auditory and visual contact, but only one mate had access to the water. Food was provided *ad libitum* for both males and females during the procedure.

### Data analysis

The periodogram  $S(f)$  estimates the contributions of each frequency ( $f$ ) to the signal. Let us consider the change of call rates ( $\Delta\gamma$ ) in time ( $t$ ). We obtained a time series  $\Delta\gamma_1, \Delta\gamma_2, \dots, \Delta\gamma_i$  at times  $t_1, t_2, \dots, t_i$ . The periodogram  $S_{\Delta\gamma}(f)$  of a time series  $\Delta\gamma(t)$  is defined as the contribution of each frequency  $f$  to the time series  $\Delta\gamma(t)$ . An uncorrelated white noise will appear as a flat line (i.e. slope = 0) on the periodogram. If a time series has a high power at low frequency, this time series has a long-range fluctuation. By contrast, if a time series has a high power at high frequency, this time series has a short-range fluctuation.

The fitting procedure [39, 10] aimed to estimate the exponent from a discrete set of inter-event intervals from a lower bound:  $\tau_{min}$  to any possible largest value in a continuous recording (about 1,000 seconds for our longest observation period: 10 hours during the Light cycle). The steps of the fitting procedure in our study were as follows:

1. Let the inter-event intervals be  $\tau_i, i = 1, \dots, n$ . We estimate the candidate exponents  $\alpha'$  and the candidate  $\tau'_{min}$  between the shortest and the longest inter-call intervals. The larger  $n$  we have, the more precise the estimated  $\alpha'$  is approximated by the true

$\alpha$  [39].

$$\alpha' = 1 + n \left[ \sum_{i=1}^n \ln \frac{\tau_i}{\tau_{min}' - \frac{1}{2}} \right]^{-1}$$

2. Many  $(\alpha', \tau_{min}')$  pairs will be obtained from step 1. Assuming that the discrete power-law distribution is:  $p(\tau) = \frac{1}{\zeta(\alpha, \tau_{min})} \tau^{-\alpha}$ , thus the cumulative density function (CDF) will be:

$$P'(\tau) = 1 - \frac{\zeta(\alpha, \tau)}{\zeta(\alpha, \tau_{min})}$$

3. Calculate the distance (D) between the empirical CDF and the estimated CDF:

$$D = \max_{\tau \geq \tau_{max}} |eCDF(\tau) - P'(\tau)|$$

4. Repeat steps 1 – 3 until we have all possible  $(\alpha', \tau_{min}')$  pairs calculated by whole range of  $\tau_{min}'$  from the shortest (i.e.  $\min \tau$ ) and the longest inter-event intervals (i.e.  $\max \tau$ ). By looking at the plot of distance (D) against  $\tau_{min}'$ , we obtain the minimum distance with corresponding  $\tau_{min}$  and exponent  $\alpha$ .
5. Calculate the goodness of the fit of the exponential and lognormal distributions to the real data. Kolmogorov-Smirnov statistic was used to quantify the distance between the empirical CDF and the synthetic CDF with corresponding parameters. The closer the p-value is to 1, the more the simulated data with corresponding  $\alpha$  and  $\tau_{min}$  are similar to the real data.

Maximum likelihood estimation (mle, in MATLAB, Mathworks) was used to estimate the parameters ( $\theta$ ) of the exponential model ( $\mu$ ) and the lognormal model ( $\mu, \sigma$ ). Let the inter-event intervals be  $\tau_i, i = 1, \dots, n$ . The parameters are estimated by maximizing the log-likelihood function:  $\log \mathcal{L}(\theta; \tau_1, \tau_2, \dots, \tau_n) = \sum_{i=1}^n \log(\tau_i | \theta)$ , where  $p(\tau_n | \theta)$  is the likelihood probability function. The parameters  $\theta$  can be one parameter ( $\mu$ ) for exponential function and multiple parameters ( $\mu, \sigma$ ) for the lognormal function. All data points (i.e. form any possible smallest to any possible largest value in a continuous recording) were used for the maximum likelihood estimation. To estimate the goodness of the fits of exponential and lognormal distributions to the data, we used Kolmogorov-Smirnov statistic to test the distance between the empirical CDF and the estimated CDF.

The empirical cumulative distribution function (eCDF) is shown by a plot of events against the sorted inter-event intervals ( $\tau$ ) with corresponding indices. We used the One-sample Kolmogorov-Smirnov test to test whether the empirical data follow the exponential distribution. The Power-law distribution was estimated using fitting algorithms as described by Clauset et. al. [39]. The decreases of exponents in all variables were analyzed with a right-sided Wilcoxon signed-rank test by testing whether the data (e.g.  $\alpha_{cohabitation} - \alpha_{separation}$ ) came from a distribution with a median greater than 0. The increase of exponents was analyzed with a left-sided Wilcoxon signed-rank test by testing whether the

data (e.g.  $\alpha_{base\ line} - \alpha_{wr\ male}$ ) originated from a distribution with a median smaller than 0. The two-sided Wilcoxon signed-rank test was used to analyze whether the data (e.g.  $\alpha_{cohabitation} - \alpha_{separation}$ ) came from a distribution with a median different than 0. The test results are considered for significance at a 5% confidence level. The one sided Wilcoxon signed-rank test was performed only after the two-sided Wilcoxon signed-rank test indicated significance.

## 2.1.6 Supplementary Information

### Supplementary Text S1

**Quantitative analysis of calling activity in response to water-removal.** We consider a system where the change of calling activity ( $\Delta N$ ) with respect to time ( $\Delta t$ ) corresponds to the rate ( $\gamma$ ). Thus we obtain:  $\frac{\Delta N_i}{\Delta t_i} = \gamma_i$ , where,  $i$  indicates individual males or females. This equation accounted for a linear dynamics and the number of vocal events grew positively with time. Our measurements showed that temporary water-removal did not change the calling activity, neither the linear accumulation of calls over time (Supplementary Fig. S2b [i.] & 2c [i.]), nor the diversity of instantaneous call rates (Supplementary Fig. S2b [ii.] & S2c [ii.]), nor the changes of call rates (Supplementary Fig. S2b [iii], S2b [iv.], S2c [iii.] & S2c [iv.]). It also suggests that the temporary water-removal does not cause severe abnormality in calling behaviors. Finally, we analyzed the probability distribution of the call rates. We found that the treatment of water-removal affected the distributions of the call rates in both male and female, although just one individual was treated at a time. Before these treatments, the likelihood to call with lower rates was higher than the likelihood to call with higher rates in the female and the male (Supplementary Fig. S2a [v.]). During the water-removal from the male (Supplementary Fig. S2b [v.]) and from the female (Supplementary Fig. S2c [v.]), the likelihood to call with lower rates changed in the female whereas the call rates of the male became irregular and were distributed nearly equally between minimum and maximum call rates. The distribution of call rates in the male and in the female returned to a similar pattern ones the animals had again access to water (Supplementary Fig. S2d [v.]). Although we did not find a consistent pattern for the call rates, we assume that 1) regularity of call rates and 2) maintaining the changes of call rates (i.e. the acceleration) within bounds during calling interactions are two important factors for the scaling property of interaction dynamics. However, we were unable to derive an accurate mechanistic model (e.g. no power-law function) that would analytically predicts the call rates in vocal interactions.

### Supplementary Text S2

**The moments and the periodicity of the power-law distribution.** We showed that the calling activity followed a power-law with exponents range between 2 and 3 in

which the mean value of calling activity is finite, whereas the variance is divergent. According to the moment in mathematics, the  $n$ th moment of a distribution is defined as:  $\langle \tau^n \rangle \sim \int_{\tau}^{\infty} p(\tau) d\tau$ , where the  $\tau$  is the inter-event interval and the  $p(\tau)$  is the probability function. The first moment:  $\langle \tau \rangle$  is the mean inter-event interval. The second moment:  $\langle \tau^2 \rangle$  is the variance of inter-event intervals. While there are a  $\tau_{min}$  and a  $\tau_{max}$  for a power-law distribution, the  $n$ th moment will be:  $\langle \tau^n \rangle = \frac{\tau_{max}^{-\alpha+n+1} - \tau_{min}^{-\alpha+n+1}}{-\alpha+n+1}$ . The value of  $\langle \tau^n \rangle$  depends on  $-\alpha + n + 1$  [10]. If  $n \leq \alpha - 1$ , the value of  $\langle \tau^n \rangle$  will be finite as  $\tau_{max}$  goes to  $+\infty$ . If  $n > \alpha - 1$ , the value of  $\langle \tau^n \rangle$  will be infinity as  $\tau_{max}$  goes to  $+\infty$ . For empirical data, the  $\tau_{max}$  always exists. The measured variance of the data that follow power-law must be significantly larger than the measured mean values and the variance of the data that follow lognormal or exponential distributions. By contrast, the tail distribution of this “reactive” behaviors (Fig. 3c) have exponents much smaller than 2 (Supplementary Fig. S6,  $n = 3$ ). As predicted by the moment  $\langle \tau^n \rangle$  neither the mean nor the variance will be finite. The measured mean ( $\mu_{f-m} = 208.5$  s and  $\mu_{m-f} = 333.2$  s) and variance ( $\sigma_{f-m} = 392.1$  s and  $\sigma_{m-f} = 750.6$  s) are large.

Cricket songs are considered as one of the fixed action patterns [94]. We recorded two song bouts from a field cricket in isolation that contained 150 and 133 trills, respectively. Each trill contained 3 – 4 syllables. We measured the inter-trill intervals of these two song bouts. The exponents of these two song bouts are 4.01 ( $\alpha_1 = 0.33$  s,  $\sigma_1 = 0.14$  s,  $n_1 = 150$  trills) and 4.22 ( $\alpha_2 = 0.38$  s,  $\sigma_2 = 0.18$  s,  $n_2 = 133$  trills), respectively. The measured means and variances agree with the prediction of the moment in which  $n < \alpha - 1$ , hence both mean and variance will be finite. Furthermore, the periodograms show that the trill rates ( $\gamma$ ) of the cricket’s songs have a spectrum with higher power ( $S_{\gamma}(f)$ ) at lower frequencies (i.e. a long-range of fluctuation), and the changes of the trill rates ( $\Delta\gamma$ ) have a spectrum with equal power ( $S_{\Delta\gamma}(f)$ ) at every frequency (i.e. slope  $\sim 0$ ) (Supplementary Fig. S4). These analyses support a Brownian characteristic, i.e. a stochastic process in generating trills of the cricket’s songs. In contrast to the self-similar patterns (i.e. the normal calling activity of zebra finches) and the fixed action patterns (i.e. cricket’s song), the signals of uncorrelated behaviors should have equal density at every frequency as expressed by the periodograms of both  $\gamma$  and  $\Delta\gamma$ . As predicted, the “reactive” calling of the isolated zebra finch mates ( $\alpha \leq 2$ , depicted in Fig. 3d) exhibited an uncorrelated (i.e. slope  $\sim 0$ ) and a week long-range of fluctuations in  $\gamma$  (i.e. the  $\gamma$  has a spectrum with small power in the low frequency that constitutes a large part of the signal), suggesting an uncorrelated behavior (Supplementary Fig. S4).

Table 2.2: Supplementary Table S1. Estimated parameters and statistical tests for different zebra finch pairs during cohabitation before treatments.

ID	Peri	Symb	N	min $\tau$	max $\tau$	$\mu_{\tau}$	$\sigma_{\tau}$	$\tau_{min}$	$\alpha$	$P_{pl}$	$\mu_{logN}$	$\sigma_{logN}$	$P_{logN}$
a	16:07-	f-f	2245	0.04	1418	4.67	32.96	3.66	2.21	0.07	0.52	1.14	0.04

Table 2.2: Supplementary Table S1. Estimated parameters and statistical tests for different zebra finch pairs during cohabitation before treatments.

ID	Peri	Symb	N	min $\tau$	max $\tau$	$\mu_\tau$	$\sigma_\tau$	$\tau_{min}$	$\alpha$	$P_{pl}$	$\mu_{logN}$	$\sigma_{logN}$	$P_{logN}$
	2:07	m-m	281	0.10	543	6.67	34.2	1.49	2.05	0.61	0.76	1.19	0.25
		m-f	850	0	278	2.13	10.89	1.72	2.26	0.49	-0.44	1.36	0
		f-m	850	0	1478	5.81	62.19	1.78	2.21	0.03	0.06	1.32	0.36
b	06:36-16:36	f-f	802	0.45	469	5.81	21.69	3.36	2.36	0.68	1.10	0.86	0.01
		m-m	2359	0.17	2329	5.30	48.87	3.58	2.36	0.92	0.84	0.99	0.01
		m-f	2100	0.08	2150	4.26	51.66	2.79	2.43	0.14	0.15	1.26	0.05
		f-m	2101	0	2004	2.79	44.04	1.93	2.29	0.04	-0.45	1.37	0
c	05:22-15:22	f-f	2922	0.19	666	3.46	21.56	2.13	2.58	0	0.44	0.90	0.04
		m-m	266	0.12	677	19.45	60.76	4.00	1.84	0.70	1.57	1.48	0.43
		m-f	706	0	1042	7.94	53.47	1.69	1.97	0.53	0.23	1.54	0.30
		f-m	706	0	343	5.04	24.11	1.91	1.92	0.74	-0.19	1.62	0
d	05:12-23:12	f-f	3837	0.13	753	5.72	34.70	2.61	2.10	0	0.16	1.45	0
		m-m	1218	0.15	2044	12.86	90.60	5.96	2.06	0.05	0.86	1.33	0
		m-f	543	0	1376	13.38	73.01	2.71	1.79	0.71	0.41	1.97	0.65
		f-m	543	0	996	9.89	58.40	3.99	1.95	0.37	0.24	1.91	0.29
e	06:11-22:11	f-f	7197	0.04	72	1.48	2.09	3.52	3.12	0.68	0.01	0.85	0.09
		m-m	2690	0.12	2011	9.64	78.68	2.36	2.10	0	0.85	1.16	0
		m-f	3268	0	1255	1.59	22.76	2.10	2.63	0.10	-0.85	1.35	0
		f-m	3267	0	2600	2.59	46.49	1.57	2.42	0.20	-0.48	1.42	0
f	07:00-22:00	f-f	17973	0.09	1490	2.43	14.34	1.86	2.44	0.02	0.20	0.94	0
		m-m	661	0.13	618	4.43	27.10	1.81	2.36	0.09	0.55	0.97	0.05
		m-f	1158	0	82.08	1.44	4.06	1.66	2.66	0.34	-0.60	1.37	0
		f-m	1159	0	61.63	1.54	4.62	1.38	2.47	0.05	-0.53	1.32	0.34
g	14:00-22:00	f-f	6934	0.10	1501	1.82	19.19	3.40	2.20	0.80	-0.52	1.23	0
		m-m	1025	0.12	1140	3.93	40.33	2.29	2.03	0.05	-0.09	1.07	0
		m-f	2632	0	3762	3.28	77.20	0.97	2.92	0	-0.80	1.14	0
		f-m	2633	0	335	1.20	6.79	1.40	3.21	0.01	-0.37	1.06	0
h	18:00-22:00	f-f	3170	0.09	189	2.39	7.71	1.51	2.40	0.01	0.25	0.87	0
		m-m	240	0.30	712	9.48	54.38	1.60	2.08	0	0.68	1.13	0.14
		m-f	738	0.00	349	2.49	15.88	1.02	2.30	0	-0.43	1.40	0.04
		f-m	737	0	263	2.23	12.54	1.13	2.22	0.27	-0.37	1.34	0.14
i	18:00-22:00	f-f	631	0.06	184	5.37	14.23	5.37	2.32	0.70	0.74	1.34	0.04
		m-m	1210	0.10	527	4.52	18.02	3.02	2.64	0.20	0.90	0.87	0.15
		m-f	809	0.01	161	4.21	11.29	2.71	2.39	0.65	0.51	1.40	0
		f-m	809	0.01	61	2.53	4.53	3.09	2.62	0.64	0.10	1.35	0.32
j	18:00-22:00	f-f	672	0.15	1420	8.49	82.21	0.72	1.95	0.34	-0.08	1.22	0
		m-m	2865	0.06	217	1.97	6.76	1.62	2.33	0.69	-0.07	1.05	0.04

Table 2.2: Supplementary Table S1. Estimated parameters and statistical tests for different zebra finch pairs during cohabitation before treatments.

ID	Peri	Symb	N	min $\tau$	max $\tau$	$\mu_\tau$	$\sigma_\tau$	$\tau_{min}$	$\alpha$	$P_{pl}$	$\mu_{logN}$	$\sigma_{logN}$	$P_{logN}$
		m-f	1143	0	178	1.36	7.38	0.58	2.10	0.47	-0.93	1.33	0.01
		f-m	1143	0	164	1.05	5.81	1.18	2.37	0.03	-1.02	1.24	0
k	18:00-22:30	f-f	999	0.14	1811	5.73	77.24	1.06	2.20	0.03	-0.13	1.08	0.07
		m-m	2221	0.08	567	2.14	15.00	1.67	2.46	0.60	-0.07	0.97	0.01
		m-f	922	0	29	0.68	1.72	1.07	2.62	0.77	-1.09	1.12	0.12
		f-m	921	0	1115	1.85	36.72	1.06	2.68	0.91	-1.10	1.19	0.02
l	18:00-22:00	f-f	2419	0.12	1281	4.33	40.81	1.57	1.91	0.70	-0.12	1.15	0
		m-m	1729	0.12	36	0.68	1.53	0.87	2.81	0	-0.85	0.80	0
		m-f	2059	0	45	0.61	2.00	0.60	2.67	0.11	-1.22	1.14	0.05
		f-m	2060	0	407	0.69	9.25	0.43	2.42	0.02	-1.54	1.15	0
m	18:00-22:00	f-f	620	0.14	198	3.68	12.85	1.24	1.98	0.25	0.23	1.19	0.19
		m-m	1772	0.11	588	4.57	20.10	3.36	2.42	0.18	0.80	0.94	0.01
		m-f	1128	0	64	1.39	3.76	1.34	2.46	0.47	-0.51	1.25	0.08
		f-m	1129	0	231	2.11	10.90	1.84	2.57	0.01	-0.41	1.35	0.38
n	18:00-22:00	f-f	17973	0.09	1490	2.44	14.34	1.86	2.44	0.02	0.20	0.94	0
		m-m	661	0.13	618	4.43	27.10	1.81	2.36	0.09	0.55	0.97	0.05
		m-f	1158	0	82	1.44	4.06	1.66	2.66	0.34	-0.60	1.37	0
		f-m	1159	0	62	1.54	4.62	1.38	2.47	0.05	-0.53	1.32	0.34
o	18:00-22:00	f-f	950	0.10	58	1.32	3.15	1.58	2.99	0.05	-0.26	0.95	0.05
		m-m	1786	0.11	564	3.67	17.99	2.80	2.16	0.14	0.33	1.13	0.01
		m-f	2129	0	26	0.59	0.98	1.30	3.47	0.40	-1.12	1.13	0.07
		f-m	2129	0	192	0.94	6.50	1.28	3.01	0	-1.02	1.22	0
p	06:30-18:08	f-f	886	0.12	302	2.77	9.21	1.75	2.33	0.32	0.26	1.08	0
		m-m	9789	0	1411	5.87	51.09	1.24	1.81	0.20	-0.73	1.73	0
		m-f	1110	0.15	1528	4.91	52.62	7.75	2.35	0.91	0.19	1.20	0
		f-m	1111	0	256	1.88	10.21	1.93	2.25	0.68	-0.43	1.34	0.11
q	06:30-14:08	f-f	869	0.18	1006	8.85	52.85	2.54	2.10	0.23	0.96	1.17	0.07
		m-m	807	0	268	4.55	21.04	2.49	1.91	0.75	-0.43	1.68	0
		m-f	364	0.10	1822	14.07	74.66	4.85	2.05	0.52	1.28	1.43	0.25
		f-m	363	0.01	1051	12.55	66.11	3.70	2.13	0.02	0.97	1.60	0.04
r	06:30-14:08	f-f	607	0.10	2399	8.50	68.99	4.60	2.08	0.34	0.80	1.19	0
		m-m	1591	0	264	1.38	8.27	1.35	2.33	0.92	-0.86	1.29	0
		m-f	1207	0.13	649	6.48	38	1.54	2.23	0.22	0.62	1.08	0.02
		f-m	1207	0	632	5.08	31	1.12	2.13	0	-0.14	1.44	0.01
s	06:30-14:08	f-f	3031	0.12	297	3.26	11.72	3.30	2.10	0.71	0.14	1.19	0
		m-m	2574	0	224	1.48	6.02	1.31	2.33	0.16	-0.59	1.28	0
		m-f	2776	0.16	395	2.6	9.63	1.80	2.63	0.46	0.44	0.78	0

Table 2.2: Supplementary Table S1. Estimated parameters and statistical tests for different zebra finch pairs during cohabitation before treatments.

ID	Peri	Symb	N	min $\tau$	max $\tau$	$\mu_\tau$	$\sigma_\tau$	$\tau_{min}$	$\alpha$	$P_{pl}$	$\mu_{logN}$	$\sigma_{logN}$	$P_{logN}$
		f-m	2777	0	145	1.80	4.95	2.14	2.57	0.13	-0.30	1.37	0
t	14:08- 21:30	f-f	655	0.11	1845	10.57	70.48	5.69	2.42	0	1.28	1.16	0.02
		m-m	1485	0	1529	8.72	85.86	2.76	2.19	0.31	0.03	1.63	0.54
		m-f	417	0.12	246	4.43	13.06	3.56	2.29	0.43	0.59	1.22	0.81
		f-m	417	0	398	6.41	24.59	4.86	2.22	0.62	0.54	1.57	0.97
u	06:30- 14:08	f-f	4049	0.14	1902	13.28	77.51	16.58	2.00	0.67	1.17	1.29	0
		m-m	1117	0	1178	3.49	45.97	1.77	2.44	0.03	-0.58	1.31	0
		m-f	1708	0.15	695	4.53	18.97	2.43	2.34	0	0.69	0.99	0
		f-m	1708	0	186	1.50	7.71	2.16	2.69	0.18	-0.57	1.21	0
v	06:30- 14:08	f-f	1493	0.13	1137	3.76	24.81	3.12	2.72	0.07	0.62	0.93	0.13
		m-m	2815	0	557	2.99	19.31	2.60	2.46	0.01	-0.35	1.51	0
		m-f	1797	0.15	184	3.39	8.65	3.27	2.43	0.22	0.49	1.10	0.67
		f-m	1797	0	341	2.44	11.30	3.50	2.50	0.08	-0.21	1.40	0.53

Abbreviations: **ID**: identities of the tested zebra finch pairs; **Peri**: the start and the end of a recording; **Symb**: f-f: female self-contained callings, m-m: male self-contained callings, m-f: female reactive callings, f-m: male reactive callings; **N**: number of events; **min**  $\tau$ : minimum inter-event interval (in seconds); **max**  $\tau$ : maximum inter-event interval (in seconds);  $\mu_\tau$ : average inter-event interval (in seconds);  $\sigma_\tau$ : standard deviation of inter-event interval (in seconds);  $\tau_{min}$ : lower bound for fitting algorithms (in seconds);  $\alpha$ : exponent of power-law distribution;  $P_{pl}$ : p-value indicates the statistical significance of the fit to the power-law, the closer p-value is to 1, the more likely that the empirical data fit the distribution. The model accepts  $p > 0.01$  for a significant fitting.  $\mu_{logN}$ : indicates the mean value (location) of the data's natural logarithm.  $\sigma_{logN}$ : indicates the standard deviation (scale) of the data's natural logarithm.  $P_{logN}$ : p-value indicates the statistical significance of the fit to the lognormal distribution model, the closer p-value is to 1, the more likely that the empirical data fit to the model. The model accepts  $p > 0.01$  for a significant fitting. N = 22 pairs.

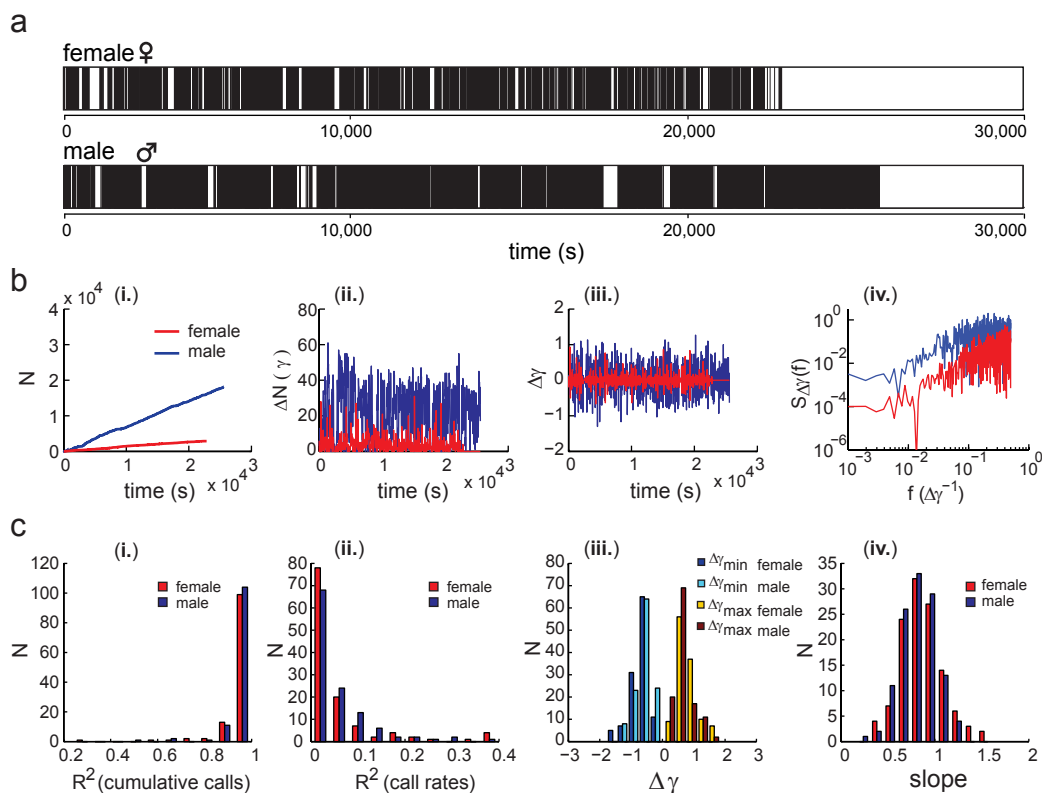


Figure 2.1: Figure 1. Calling activity of zebra finch pairs. (a) Example of a succession of calling interactions between a female (upper panel) and a male (lower panel) zebra finch. Vertical black lines corresponding to calling events are plotted as a function of time (in seconds). (b) Quantitative measurements of calling activity of one zebra finch pair (male in blue, female in red). (b [i.]) Calling events were accumulated in a linear relationship over time; (b [ii.]) the call rates ( $\gamma$ ) corresponds to the number of calls ( $\Delta N$ ) in a unit of time ( $\Delta t = 30$  s); (b [iii.]) The changes of call rates ( $\Delta \gamma$ ) were plotted as the function of time; (b [iv.]) The power spectra ( $S$ ) and the frequency ( $f$ ) of were positively correlated. (c) Statistical summary of all 22 tested zebra finch pairs including 119 days of calling activity. (c [i.]) The density counts of the coefficient of determination ( $R^2$ ) of cumulative calls (bin: 0.1). The values of  $R^2$  are in range from 0 to 1. With increasing linearity of the data,  $R^2$  approaches 1. (c [ii.]) The density counts of the  $R^2$  of call rates (bin: 0.04). With decreasing linearity of the data,  $R^2$  approaches 0. (c [iii.]) The density counts of the changes of call rates ( $\Delta \gamma$ ) (bin: 0.35). (c [iv.]) The density counts of the slopes for the power spectrum of  $\Delta \gamma$  (bin: 0.14)



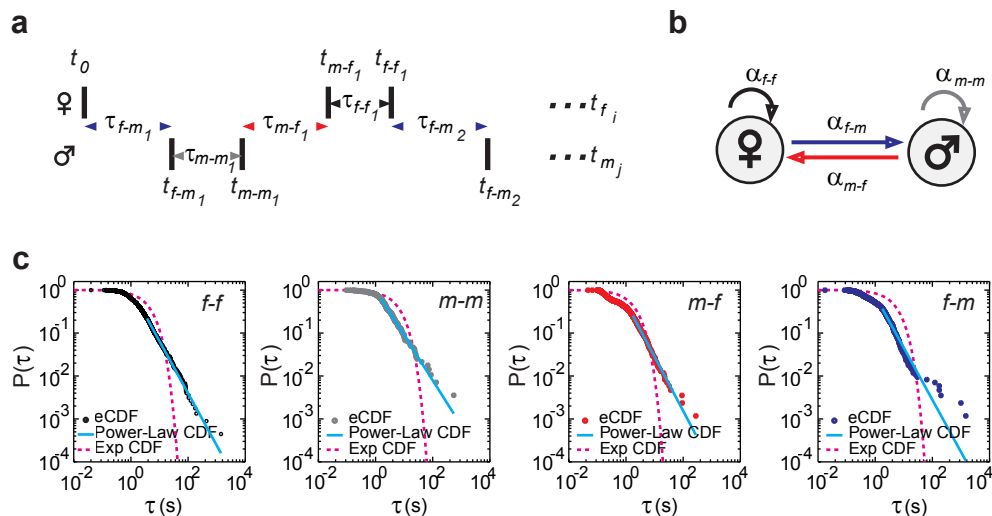


Figure 2.2: Figure 2. The self-contained (female: “ $f - f$ ”; male: “ $m - m$ ”) and the reactive (female “ $m - f$ ”; male: “ $f - m$ ”) callings exhibit power-law dynamics. (a) Schematic of calling events ( $t$ ) and inter-event intervals ( $\tau$ ). (b) Calling interactions are described as a mixture of self-contained (“ $f - f$ ”; “ $m - m$ ”) and reactive (“ $m - f$ ”; “ $f - m$ ”) callings of male and female zebra finches. (c) The empirical cumulative distribution functions (eCDF, dotted line, dark blue) of the inter-event intervals ( $\tau$ ) were plotted in a log-log scale. Estimated power-law cumulative distribution function (Power-Law CDF, solid lines, light blue) were modeled by using the maximum likelihood method [39]. Estimated exponential cumulative distribution functions (Exp CDF, dashed lines, magenta) were modeled with corresponding average inter-event intervals ( $\mu_\tau$ ). Data are shown from the female-male pair “a”.  $\alpha$ : the exponent of a power-law distribution;  $P(\tau)$ : the cumulative distribution of the inter-event intervals  $\tau$ ; (s): the inter-event intervals in seconds. (for estimated parameters, see Supplementary Table S1)

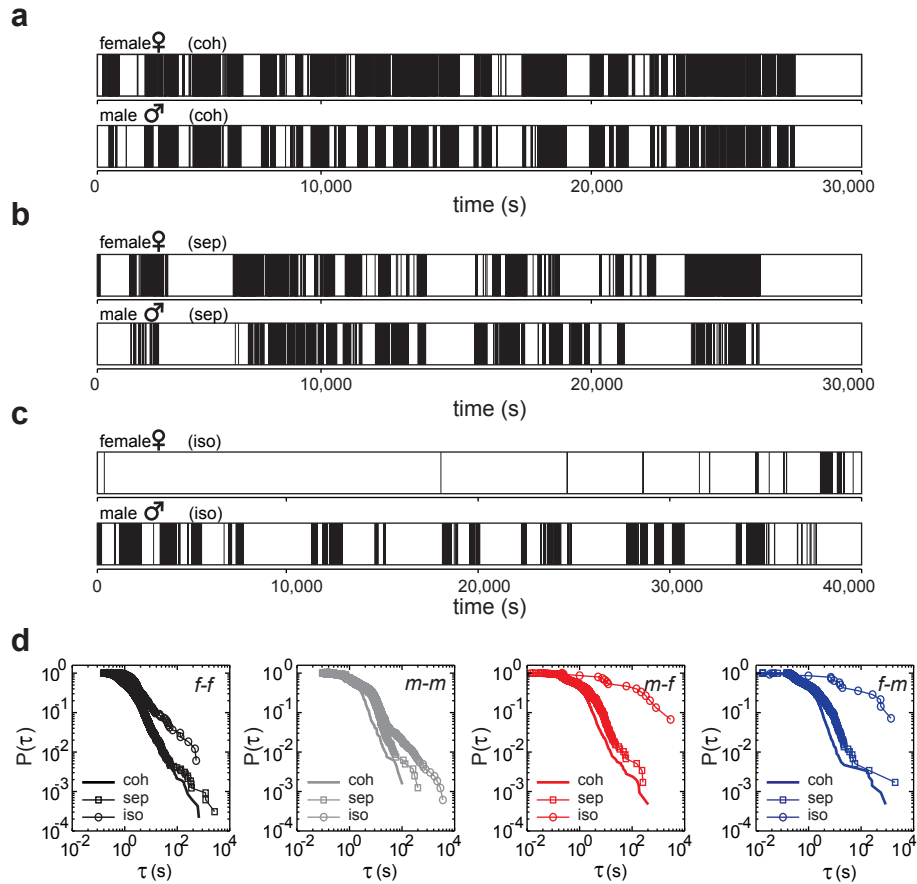


Figure 2.3: Figure 3. The power-law scaling of the self-contained (“ $f - f$ ”, “ $m - m$ ”) and of the reactive (“ $m - f$ ”, “ $f - m$ ”) callings change independently. a: Successions of calling events of a female and a male during cohabitation (coh). b: Successions of calling events of a female and a male after separation (sep) of the birds into two acoustically interconnected chambers allowing only acoustic interactions. c: Successions of calling events of a female and a male after separating the birds into two chambers without any interconnection, neither visual nor auditory (iso = isolation). d: Comparison of empirical cumulative distribution functions (eCDFs) between the inter-event intervals ( $\tau$ ) of cohabitation (coh), separation (sep), (sep) and isolation (iso).

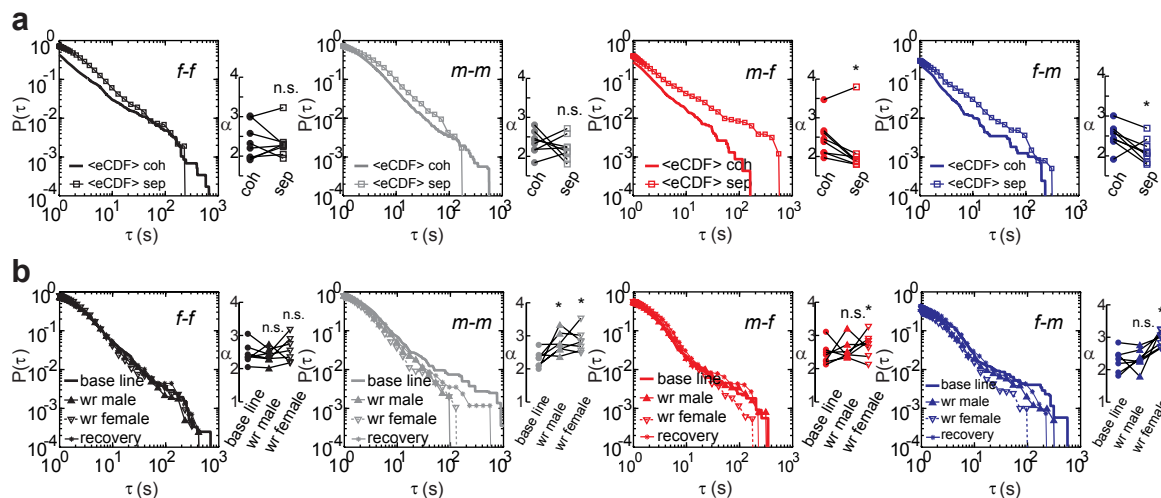


Figure 2.4: Figure 4. The exponents of different callings changed transiently in response to the social context (a) and the access to water (b). (a) Comparison of median empirical cumulative distribution functions ( $\langle \text{eCDF} \rangle$ ) between the inter-event intervals ( $\tau$ ) of “stack” calls during cohabitation (coh) and separation (sep). After limiting the social input to the auditory cues, the exponents of “ $m - f$ ” ( $p_{m-f} = 0.039$ ,  $n = 8$ ) and “ $f - m$ ” ( $p_{f-m} = 0.027$ ,  $n = 8$ ) decreased while this did not change the exponents of “ $f - f$ ” and “ $m - m$ ” ( $p > 0.7$ ,  $n = 8$ ). (b) Comparison of  $\langle \text{eCDFs} \rangle$  between the situations before, during and after removal of water. Removing water from males (wr male) increased the exponents of “ $m - m$ ” ( $p_{m-m} = 0.023$ ,  $n = 7$ ). Removing water from females (wr female) increased the exponents of “ $m - m$ ” ( $p_{m-m} = 0.016$ ,  $n = 7$ ), of “ $m - f$ ” ( $p_{m-f} = 0.016$ ,  $n = 7$ ) and of “ $f - m$ ” ( $p_{f-m} = 0.008$ ,  $n = 7$ ). The other modes of callings remained unchanged ( $p > 0.050$ ,  $n = 7$ ). The exponents of base line and recovery were similar for all four modes of callings ( $p > 0.100$ ,  $n = 7$ , recovery data are not shown on the ladder plots). n.s. and \* indicate non-significant and significant, respectively.  $P(\tau)$ : the cumulative distribution of the inter-event intervals;  $\tau(s)$ : the inter-event intervals in seconds.

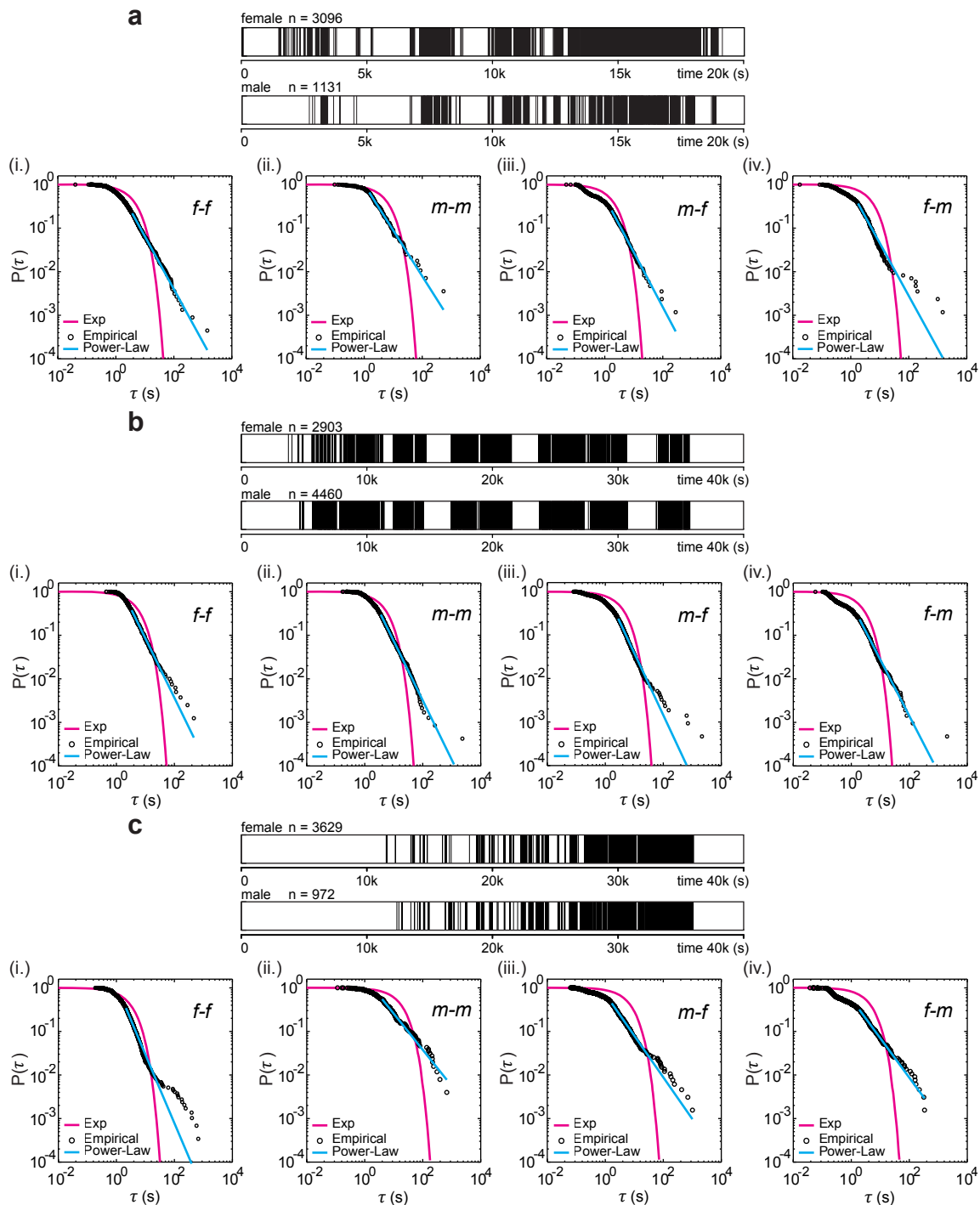
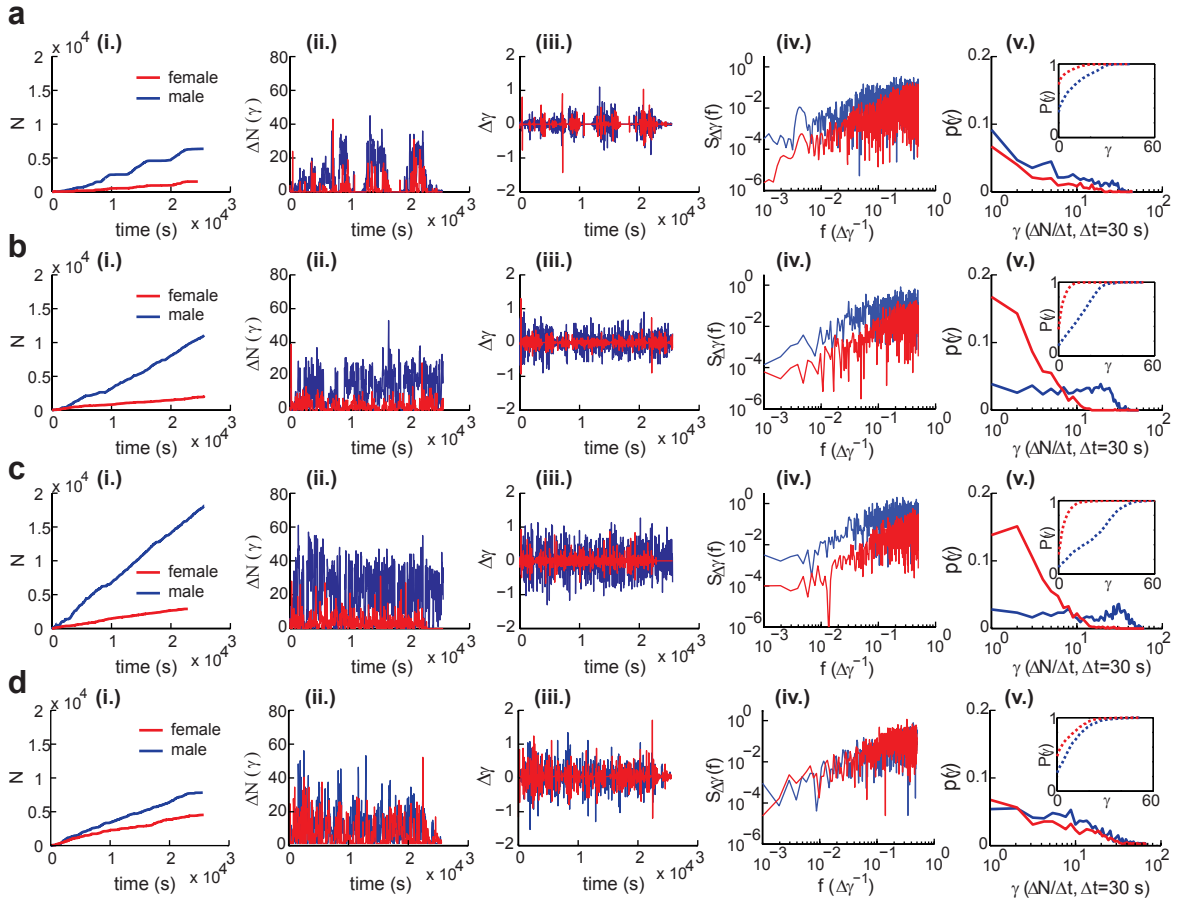


Figure 2.5: Supplementary Fig. S1a – S1c: Successions of calling interactions between females and males of all zebra finch pairs (pairs a – c) not exposed to any treatment. Comparisons between empirical data and exponential cumulative density models with estimated parameters (Supplementary Table S1) for the self-contained female callings (i.), the self-contained male callings (ii.), the reactive female callings (iii.) and the reactive male callings (iv.).  $P(\tau)$ : the cumulative distribution of the inter-event intervals;  $\tau(s)$ : the inter-event intervals in seconds.



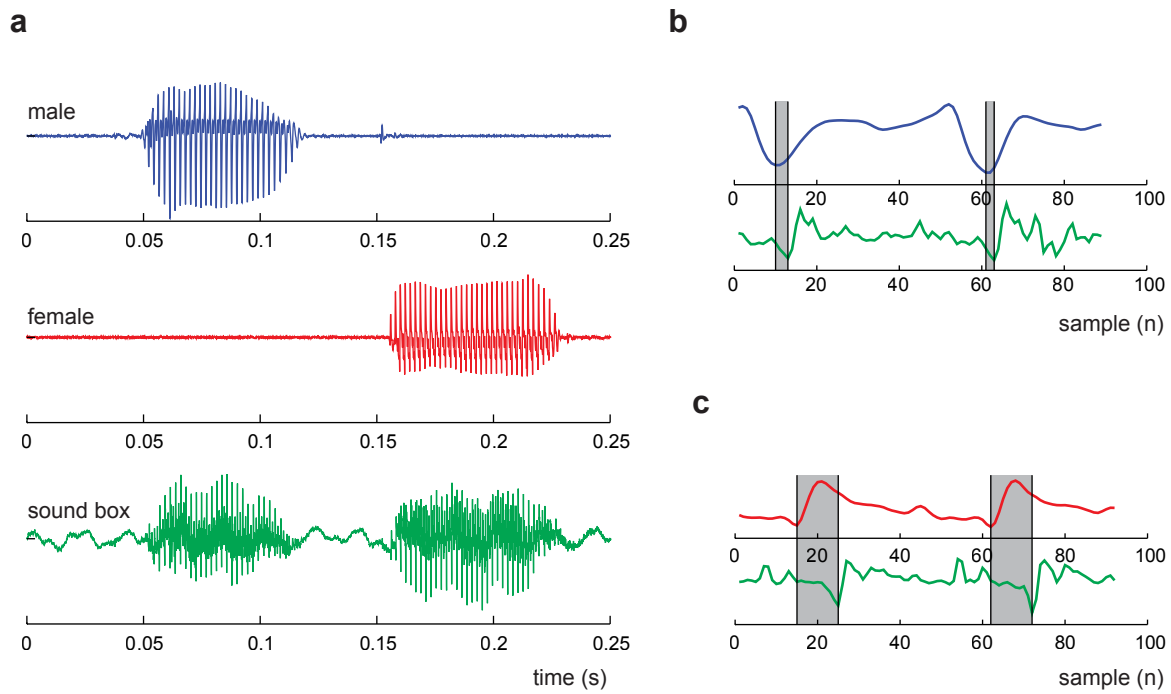


Figure 2.7: Supplementary Fig. S3. Synchronous recording of vocal interactions during sensory perturbation. (a) illustrates the segments of male (blue) and female (red) backpack recordings. Both male and female vocalizations were recorded by a microphone in the male-hosted sound box (green). b: By comparing the time shifts between the sound waves, we found a size of 3 samples (0.1 ms, sample rate: 22050) discrepancy between the male backpack recording and the male-hosted sound box. c: a size of 10 samples (0.5 ms, sample rate: 22050) discrepancy between the female backpack recording and the male-hosted sound box.

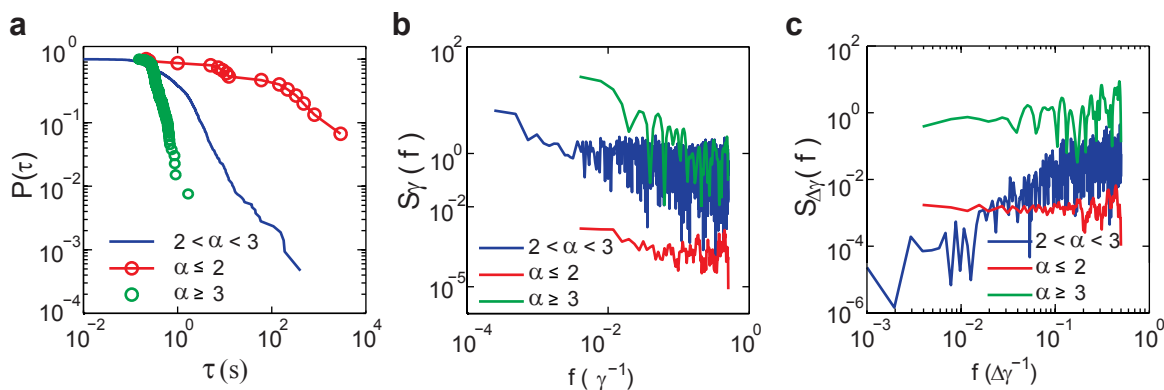


Figure 2.8: Supplementary Fig. S4. Comparing the tail distributions of different behavioral categories. (a) The tail distributions of inter-event intervals that fall into three different behavioral categories: (1.) “reactive” callings of isolated zebra finches ( $\alpha \leq 2$ ), (2.) the natural calling activity of zebra finches ( $2 < \alpha < 3$ ), (3.) the trills of cricket songs ( $3 \leq \alpha$ ). (b) The periodograms of the instantaneous rates ( $\gamma$ ) of the three different behaviors depicted in (a). (c) The periodograms of the change of the instantaneous rates ( $\Delta\gamma$ ) of three different behaviors depicted in (a). Note that the signal has a long-range fluctuation if the signal has a high power at low frequency (i.e. slope  $< 0$ ). The signal has a short-range fluctuation if the signal has a high power at high frequency (i.e. slope  $> 0$ ). The white noise signal has equal power at every frequency (i.e. slope  $\sim 0$ ).

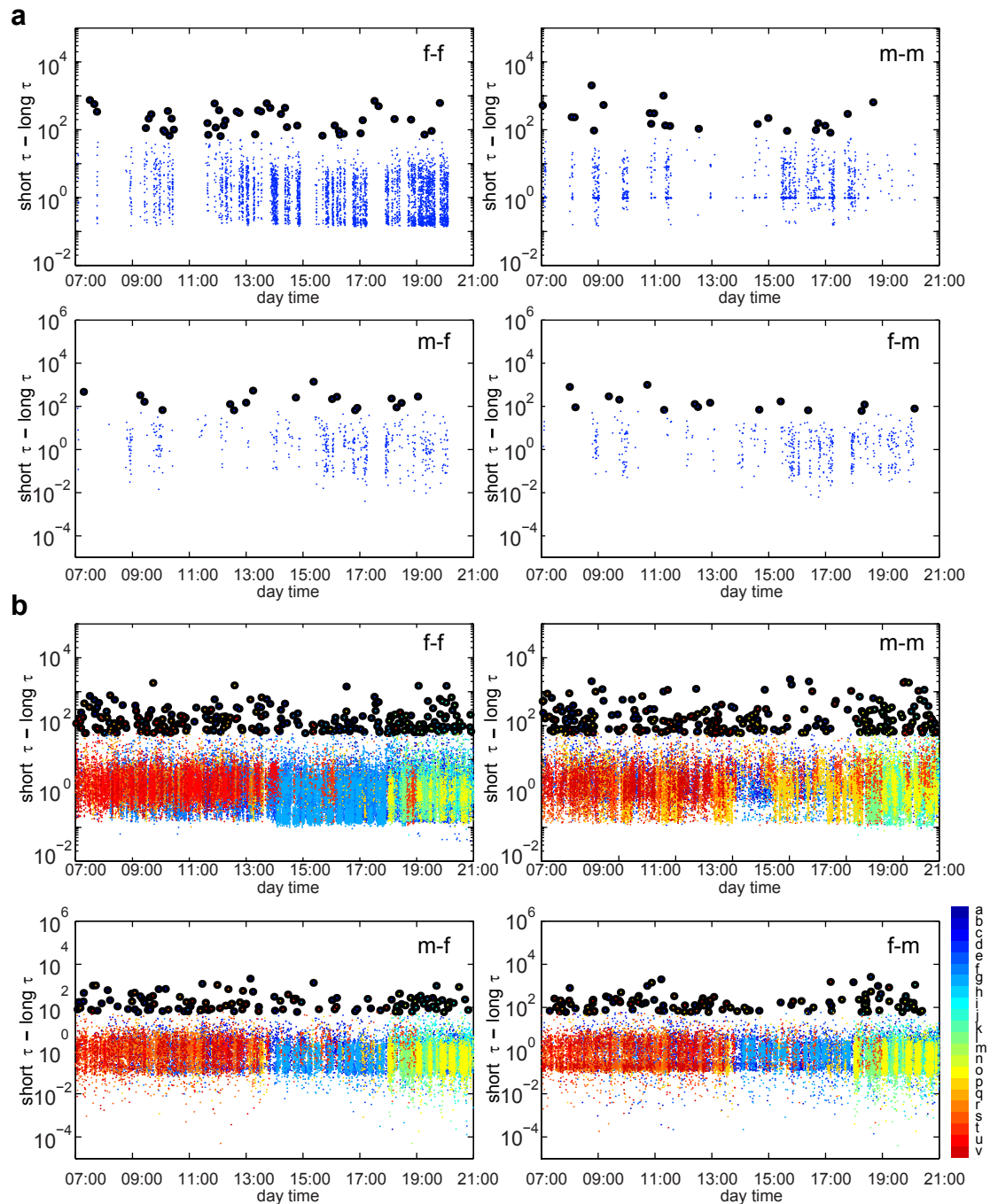


Figure 2.9: Supplementary Fig. S5. Power-law dynamics of calling intervals is not affected by the daily circadian cycle. (a) Events of self-consistent ('f-f', 'm-m') and of reactive ('m-f' and 'f-m') callings of one zebra finch pair are sorted with respect to their inter-event intervals ( $\tau$ ) and plotted against the day time of occurrence. The occurrences of calling events after short and long inter-event intervals ( $\tau$ ) distributed equally on the logarithmic scale of inter-event intervals ( $\tau$ ) during active periods. (b) In all studied pairs of zebra finches, the occurrences of calling events after short  $\tau$  and long  $\tau$  distributed equally on the logarithmic scale of  $\tau$ . Color assignment to 22 pairs of zebra finches (a - v). The events that occurred after long  $\tau$  ( $> 60$  s) are indicated by black circles.



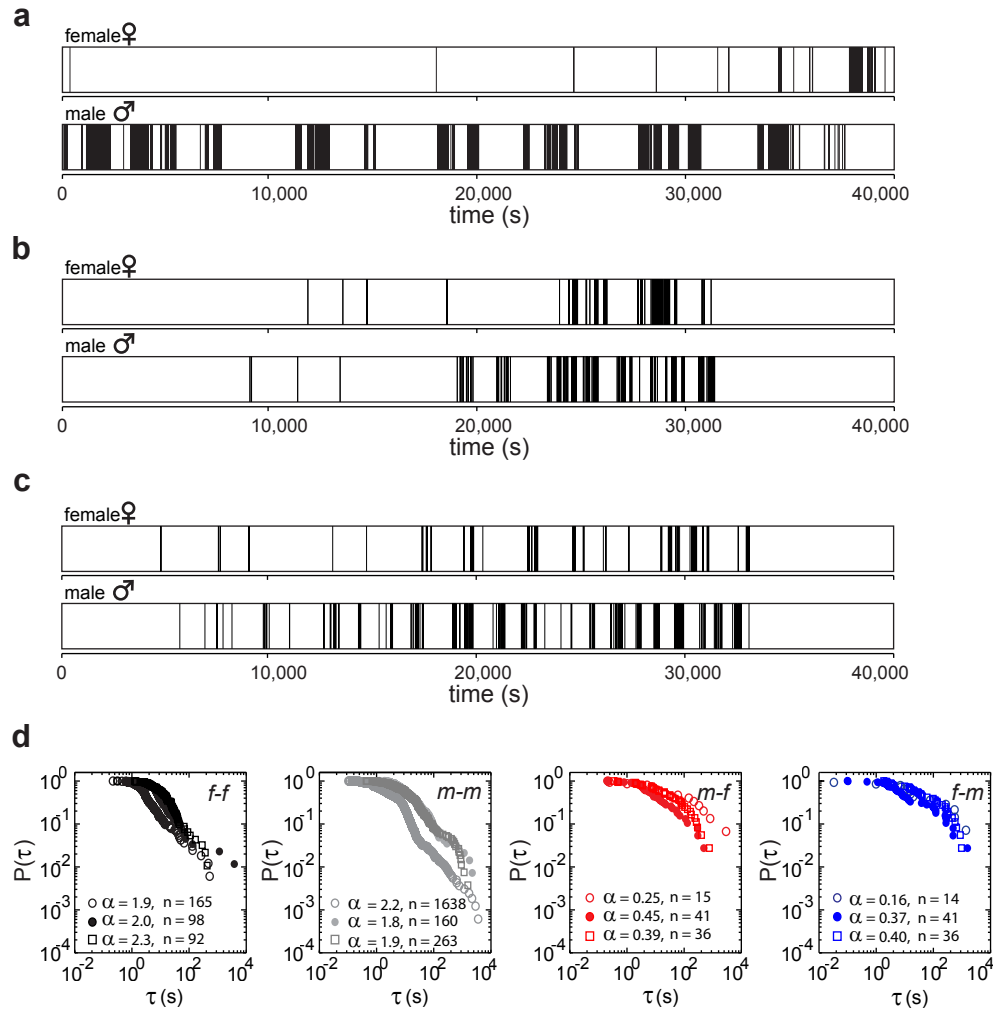


Figure 2.10: Supplementary Fig. S6. Isolation changes the exponents of “reactive” callings. (a - c): Successions of calling events of three zebra finches pairs after separation of the male and the female of each pair into separated chambers that did not allow visual or vocal interactions with others. (d) Comparison of the empirical cumulative distribution functions (eCDF) and the exponents ( $\alpha$ ) between  $\tau_{f-f}$ ,  $\tau_{m-m}$ ,  $\tau_{m-f}$  and  $\tau_{f-m}$  of these three zebra finches pairs. Symbols “open circle”, “closed circle” and “open square” represent the data of pair a, b and c depicted in (a), (b) and (c), respectively.

## 2.2 Sensorimotor expectation in call-based vocal communications of zebra finches

### 2.2.1 Abstract

Forebrain nucleus HVC (higher vocal center) is a cortex-like sensorimotor area essential for singing and song learning in zebra finches. However, the sensorimotor mechanism for calling interactions in a social context was unknown. To study the role of the HVC in calling interactions, we used a wireless telemetric system for simultaneous measurement of neuronal activity and vocalizations in birds that freely interact with each other. In a non-social context, the same male HVC neurons displayed stereotypic premotor activity related to own calling. In a social context, the HVC neurons showed calling related premotor activity and a predictive discharge to upcoming female calls and a responsive discharge to hearing these calls. The predictive discharge required either auditory or visual cues and was obligate for the elicitation of the auditory response to the females' call. Next to singing, HVC seems for the learning of call-based vocal communications in a mating context.

### 2.2.2 Introduction

Social animals such as zebra finches use a large number of calls to interact within their social group and in particular with their mates [246, 15, 77, 54]. During a natural vocal interaction, all birds are confronted with a mixture of sounds. Relevant information can only be retrieved and delivered when there is an effective system to process auditory inputs while controlling vocal outputs [172]. Furthermore, the interaction pattern of callings shows that vocal responses between zebra finch can be very fast ( $\sim 200$  ms) and that calling becomes vary mate-specific during the breeding cycle [211, 77, 20]. If the call productions were neither a consequence of fixed action patterns nor random, animals would profit from an anticipation of upcoming events to achieve such rapid responses. Studies in humans [100, 114], monkeys [131] and birds [20] suggest that the predictive coding in the premotor and the sensorimotor systems may benefit the response time in relation to experimental objects. However, the predictive coding for a natural social interaction remains unknown.

To study sensorimotor mechanisms underlying vocal interactions in a social context, we employed telemetric systems that enable a precise synchrony and the continuous monitoring of vocal activity and neurophysiological activity in the avian sensorimotor song control nucleus HVC. With this technique, it has been shown that zebra finches respond more contingently to their sexual partners than to other group members [211, 77] and that the premotor song control nucleus RA (robust nucleus of the archistriatum) plays a role in active calling [211]. The motor-cortex like RA is the afferent area of the sensorimotor cortex-like HVC in the descending song control pathway of songbirds [157, 24]. HVC seemed an appropriate target for this study since it is known to process individualized

auditory song units to control the production of song motor units and being important for the learning of vocal sequences [172, 171]. In addition, previous studies reported incidental observations that the song control nucleus HVC also pertains to active calling of isolated birds [138, 132, 85].

In this study, we analyzed the temporal organizations of calling interactions of mated zebra finches. Subsequently, we investigated the context-dependence of neuronal activity in male HVC in the social context to delineate the sensorimotor mechanisms underlying calling interactions in zebra finches.

### 2.2.3 Results

#### Stable calling interactions between pair-bonded zebra finches

Pair-living zebra finches produce multiple vocalizations to interact with each other (Fig. 1a). While song is only uttered by males, which accounts for a small subset of all vocalizations (Fig. 1b), male and female zebra finches share an extensive repertoire of calls (Fig. 1c) of which the “stack” calls are the most commonly uttered calls (Fig. 1b, Supplementary Fig. S1,  $66.7 \pm 22.1\%$  in male and  $85.3 \pm 8.4$  in female, mean  $\pm$  s.d.,  $n = 8$ ).

The cross-correlation histograms depicted in Fig. 1d & Fig. 1e are the outcome of mixture associations. To differentiate the elements of calling associations, we measured the occurrences of different possible associations between male and female mates. First, we divided the calling events into single events and associated events: If the calls are preceded or followed by other calls within 500 ms, these calls are defined as associated calls. The calls other than associated calls are defined as single calls. The silent gaps greater than or equal to 500 ms in duration are defined as pauses. Subsequently, we divided the associated calls into 16 types of possible associations with a set of elements (“M”, “F”, “-”) (see Materials and Methods), where “M”, “F”, “-”, corresponds to a male “stack” call, a female “stack” call and a pause, respectively (Fig. 2a & b). Our analysis showed that the associated male calls (“xMx”) were more frequent than the single male calls (“-M-”) in pair 2 ( $n = 12$  days) and in other tested male zebra finches (Fig. 2c,  $p = 0.02$ , right-sided Wilcoxon signed-rank test,  $n = 8$  pairs). In contrast, the associated female calls (“xFx”) were as abundant as the single female calls (“-F-”) (Fig. 2d,  $p = 0.20$ , two-sided Wilcoxon signed-rank test,  $n = 8$  pairs). Despite the permutation-based classification, the boundary to differentiate “MMF” and “FMM” as the repeated calls or the alternated calls is vague. Hence, we focused on four classifications: the repeated calls without pause (“MMM” in male; “FFF” in female), the repeated calls with pause (“MM-”, “-MM” in male; “FF-” and “-FF” in female), the injected calls without pause (“FMF” in male; “MFM” in female) and the alternated calls with pause (“-MF” and “FM-” in male; “-FM” and “MF-” in female) (Fig. 2e, f). Our results showed that males and females used alternated calls with pause more often than other call types ( $p < 0.01$ , right-sided Wilcoxon signed-rank test,  $n$

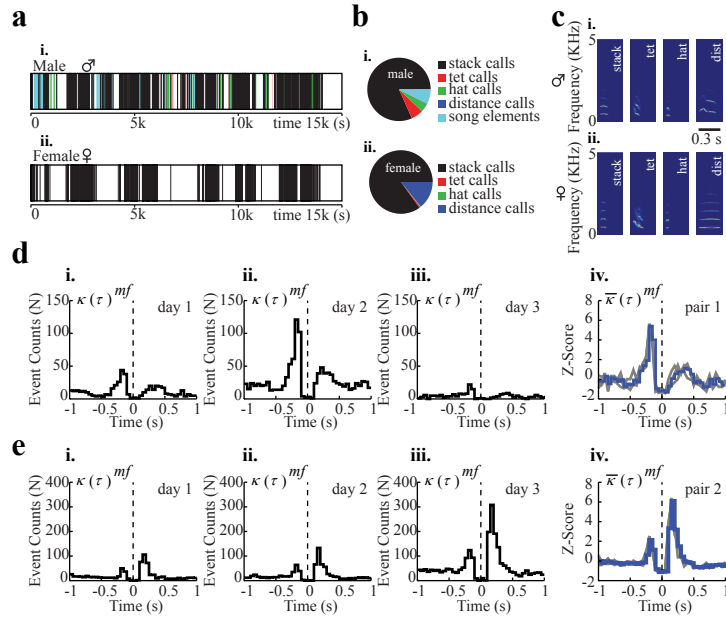


Figure 2.11: Figure 1. Vocal interactions between male and female zebra finches. (a) Successions of [i.] male and [ii.] female vocal events. The horizontal axis denotes time; each vertical line corresponds to a vocal event. (b) Pie charts illustrate the proportions of different call types in [i.] male and [ii.] female zebra finches. (c) The spectrograms of different call types in [i.] male and [ii.] female zebra finches. (d) A left asymmetric calling interaction patterns between male and female zebra finches of pair 1. (e) A right asymmetric calling interaction patterns between male and female zebra finches of pair 2. The accumulative counts of finding  $N$  male “stack” calls before and after female stack calls are plotted as function of time with a fixed time interval (50 ms) in [i.] day 1, [ii.] day 2 and [iii.] day 3, where the onset of female “stack” calls were aligned at zero. [iv.] The standard score (Z-Score) was calculated to compare calling interaction patterns between days. The gray dotted lines correspond to calling interaction patterns of individual days, whereas the blue line corresponds to average Z-Scores.

= 8).

### Neuronal activity of male HVC during calling interactions

We recorded the male HVC with a wireless telemetric device and put backpack radio-microphones on both male and female zebra finches to measure neural activity and vocalizations simultaneously in freely behaving animals that interact with each other (Fig. 3a). Since the HVC of female zebra finches is very small, we have so far been unsuccessful in recording female HVC activity in freely behaving birds. Our recordings showed that the calling-related units correlated with the self-produced calls and the perception of female calls. We tested males in isolation by plotting the spiking activity related to the onset

of the males' calls and the playbacks of female calls. The calling-related units displayed premotor activity related to their own "stack" calls (Fig. 3b [i.]). The same HVC units responded to the playbacks of the calls of their female mates (Fig. 3b [ii.]).

Next we studied the calling-related activity of male HVC neurons in a social context with their female mates. Like in the isolated male, HVC neurons displayed stereotypic premotor firing pattern related to their own calls (Fig. 3c [i.] in the males behaving in a social context. The calling-related (motor) activity of HVC neurons exhibited precise premotor activity at  $-17.5 \pm 13.9$  ms before calling. This premotor activity was preceded by pre-inhibition at  $-162.5 \pm 30.6$  ms and followed by a post-inhibition at  $+163.8 \pm 114.4$  ms (mean  $\pm$  s.d.  $n = 8$ ) relative to the onset of their own calls (Fig. 3c [i.]). However, the same male HVC units showed auditory related increased activity not only after ( $+162.5 \pm 53.9$  ms) but also before ( $-193.7 \pm 26.7$  ms) the onset of the female calls in the social context (Fig. 3c [ii.]). Units that showed an increased activity prior to a female call did not show call-production related activity.

We showed that calling interactions in zebra finches were carried out mainly by alternated calling in addition to the repeated calls (see above). We, thus, investigated neuronal activity in male HVC during the different vocal contexts defined above. In comparing the neuronal activity related to alternated calling versus repeated calling, HVC neurons reduced activity significantly in relation to repeated male calls (Fig. 3d [i.], comparing "xMx" versus "MM-",  $F(1,14) = 18.79$ ,  $p = 6.9 \times 10^{-4}$ ; comparing "xMx" versus "MMM":  $F(1,14) = 97.77$ ,  $p = 1.1 \times 10^{-7}$ ) and repeated female calls (Fig. 3d [ii.], comparing "xFx" versus "FF-":  $F(1,14) = 7.10$ ,  $p = 0.02$ ; comparing "xFx" versus "FFF":  $F(1,14) = 7.02$ ,  $p = 0.02$ ). By contrast, male HVC neurons displayed strong neuronal activity related to alternated calling. Although the numbers of the injected male calls without pause were as small as the repeated male calls without pause in males (Fig. 2e,  $p = 0.84$ , two-sided Wilcoxon signed-rank test,  $n = 8$ ), and the numbers of the injected female calls without pause were larger than the repeated female calls without pause (Fig. 2f,  $p = 0.02$ , right-sided Wilcoxon signed-rank test,  $n = 8$ ), male HVC neurons displayed neuronal activity related to their own injected calls without pause ("FMF") as strong as related to their own alternated calls with pause ("FM-") (Fig. 3e [i.],  $F(1,14) = 3.46$ ,  $p = 0.08$ ). Similarly, the injected female calls without pause ("MFM") elicited neuronal activity as strong as the alternated female calls with pause ("MF-") (Fig. 3e [ii.],  $F(1,14) = 0.02$ ,  $p = 0.90$ ) in male HVC.

### Male HVC released pre-sensory activity in relation to the upcoming calls of female mates

Above we showed that male HVC displayed neuronal activity before and after the onset of female calls during calling interactions. If auditory input provided by female calls was processed by HVC, this would explain the auditory-elicited discharge after the onset of

female calls. But what causes the activity of male HVC neurons before the onset of female calls?

We first tested whether the activity prior to the onset of female calls was induced by external cues. Males produced a great proportion of calls that were associated with their own calls and with those of their female mates. These associated calls could induce the activity prior to the onset of female calls. If it is so, the firing patterns of HVC units related to the single female calls (“-F-”) should be different from the pattern related to associated female calls (“xFx”), but our results showed no significant difference between them (Fig. 3f [ii.], comparing “xFx” versus “-F-”:  $F(1,14) = 0.23$ ,  $p = 0.64$ ). These results suggest that the neuronal activity in HVC prior to the onset of the hearing of female calls is not a premotor activity but is related to a pre-sensory activity that may resemble a predictive coding.

In order to predict upcoming calling events, zebra finches must have acquired conjunctive cues to foresee that their partners are about to vocalize. Conjunctive cues can either be call-context-dependent such as strings of related calls [99] or social-sensory (non-auditory)-dependent such as social gestures or postures [163]. For simplicity we refer to call-context as context cues and to social-sensory (non-auditory) as sensory cues. If both context and sensory cues were blocked, the single female calls would not be predictable, and HVC would not be able to generate predictive discharge related to upcoming female calls. To test this hypothesis, we limited the sensory cues by separating the female and male mate ( $n = 4$  pairs) into two distant boxes with acoustic interconnections, i.e. mates could only communicate acoustically. Either (1) limiting sensory cues (Fig. 4b,  $F(1,6) = 0.26$ ,  $p = 0.63$ ) or (2) limiting context cues by selecting those single female calls that occurred without any vocal association including other call types and songs for at least 1 second (Fig. 4c,  $F(1,6) = 2.71$ ,  $p = 0.15$ ) did not change the activity of HVC units prior to the onset of female calls. However, by limiting both sensory and context cues, the premotor activity prior to the onset of female calls decreased significantly (Fig. 4d,  $F(1,6) = 7.93$ ,  $p = 0.03$ ). The pre-sensory activity of male HVC neurons required cues that emerged from the female. In the social context, HVC neurons of male zebra finches may respond to the female calls only if they produce a pre-sensory activity (comparing supplementary Fig. S3c [ii.] versus supplementary Fig. S3e, g [ii.]).

Furthermore, all males increased their response times to answer their mates’ calls during separation (Fig. 5a,  $p = 0.002$ , right-sided Wilcoxon signed-rank test,  $n = 9$ ). In contrast, females did not change their response times significantly during separation (Fig. 5b,  $p = 0.730$ , two-sided Wilcoxon signed-rank test,  $n = 9$ ).

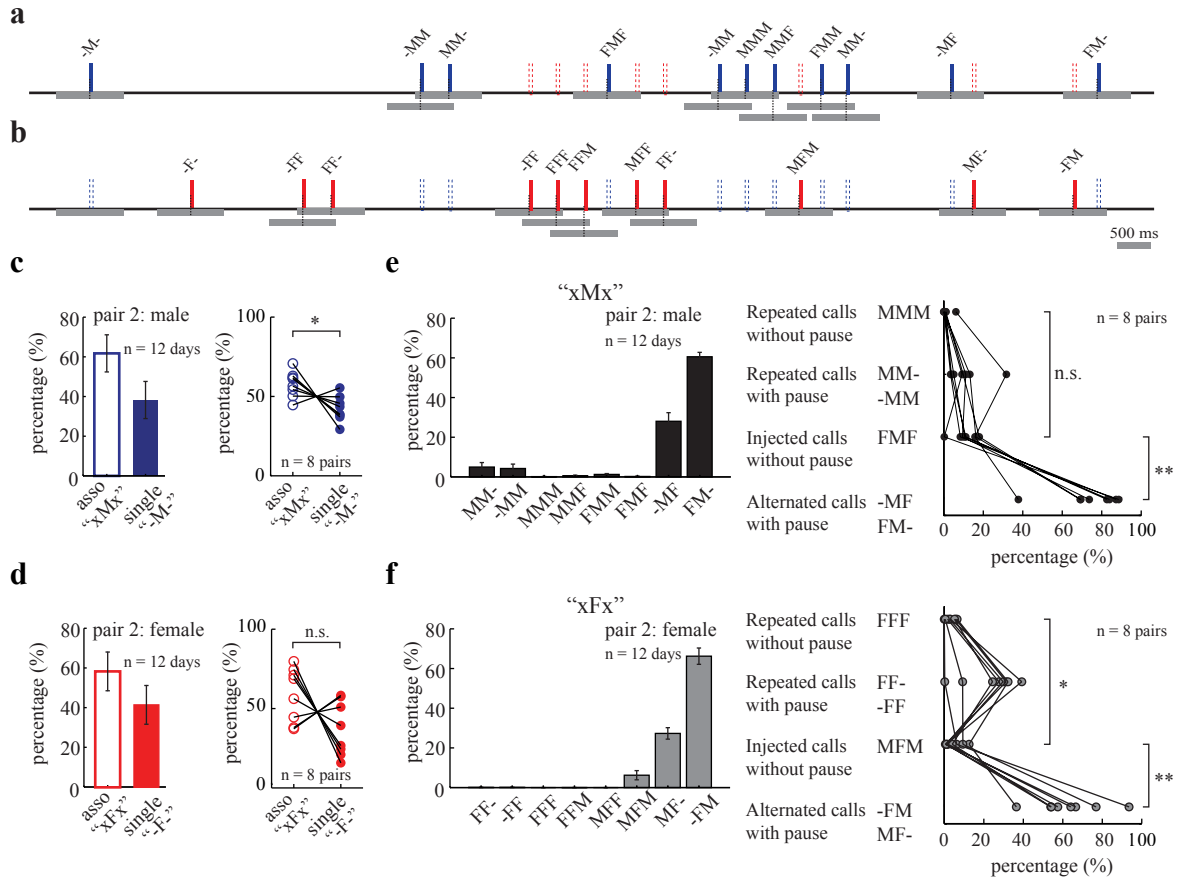


Figure 2.12: Figure 2. Different types of calling associations between male and female zebra finches. Diagrammatic representations of different possible calling associations in (a) male and (b) female. x-axis: time. Vertical bars: the calling events of male (blue) and female (red). The vertical bars with dashed lines represent the calling events of opposite sex at corresponding times. The horizontal bars: the 500 ms durations. (c) Comparing the proportions of the associated male calls (asso "xMx") versus single male calls (single "-M-") (left panel) in pair 2 and (right panel). (d) Comparing the proportions of the associated female calls (asso "xFx") versus the single female calls (single "-F-") (left panel) in pair 2 and (right panel). (e) Bar plot: the proportions of different male calling associations ("xMx") in pair 2, Ladder plot: comparing between the proportions of different male calling associations. (f) Bar plot: the proportions of different female calling associations ("xFx") in pair 2. Ladder plot: comparing between the proportions of different female calling associations.

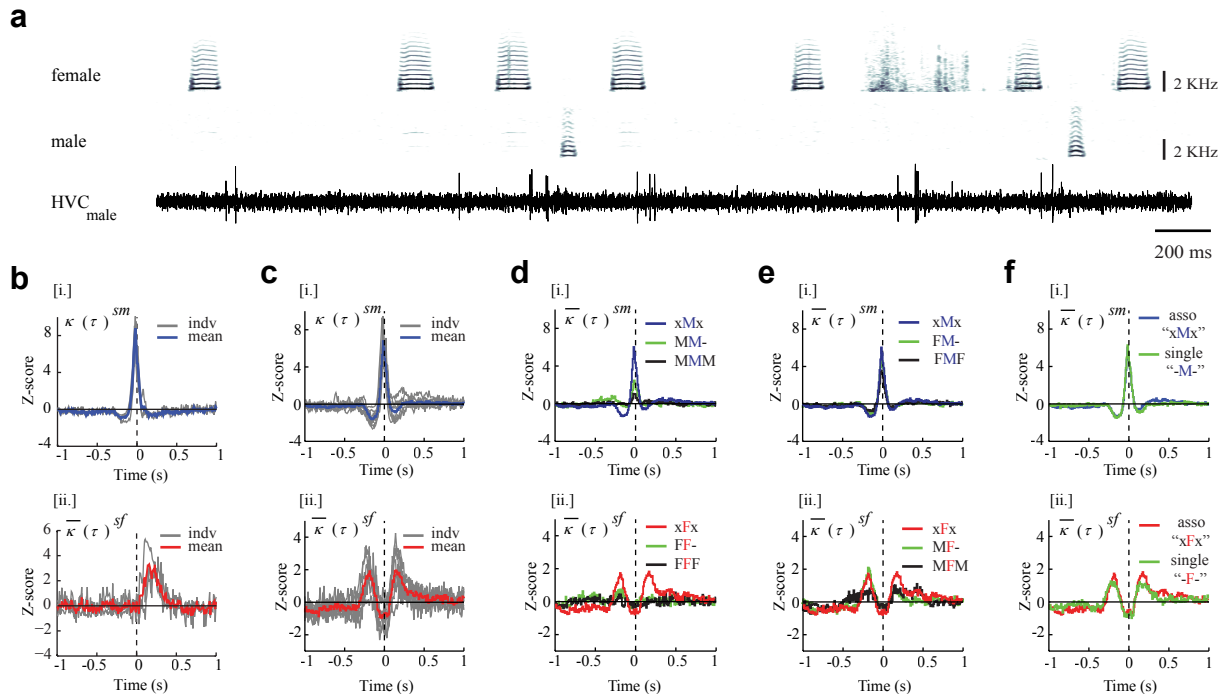


Figure 2.13: Figure 3. Neuronal activity of male HVC during calling interactions. (a) Male and female vocalizations were recorded simultaneously with extracellular activity of HVC neurons in a male zebra finch. (b) Calling-related unit of male HVC neurons exhibited [i.] premotor activity pattern in relation to his (own) calls and [ii.] activity pattern in relation to the playbacks of the calls of his female mate in isolation.  $n = 3$ . (c) Calling-related unit of male HVC neurons exhibited [i.] premotor activity pattern in relation to their own calls and [ii.] in relation to the calls of their female mates in a social context. Gray lines indicate the individual traces (indv). Blue and red lines are the mean traces (mean). ( $n = 8$ ). Premotor activity of male HVC neurons [i.] related to their own calls and [ii.] related to female calls in different contexts, i.e. (d) calling repetition; (e) calling alternations and (f) single callings.  $n = 8$ . Only mean traces were showed on (d – f). (See other examples in Supplementary figure S3 and see Supplementary Table S1 for all pairwise F-tests).



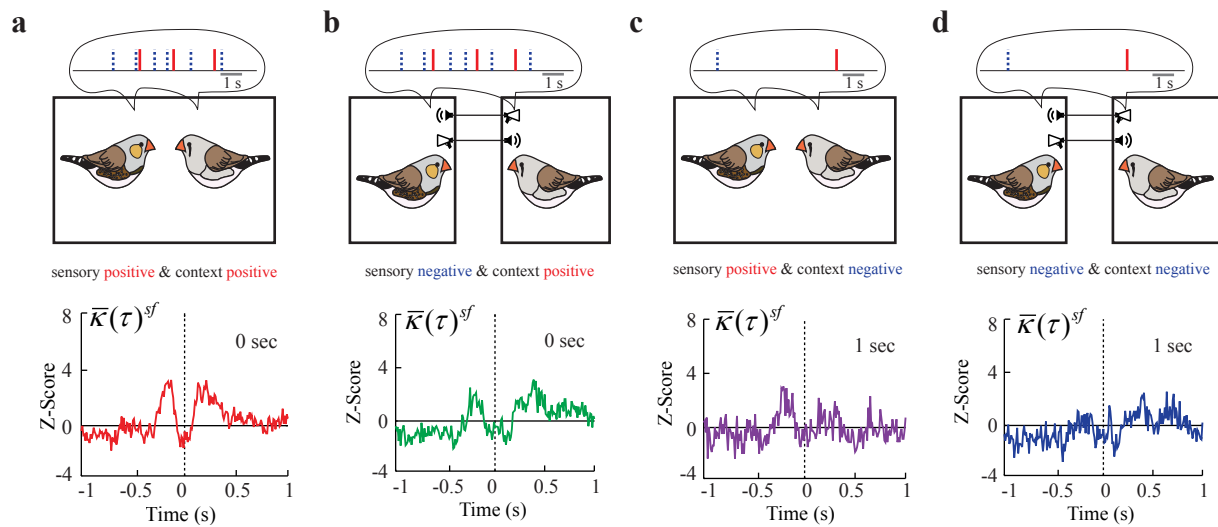


Figure 2.14: Figure 4. Premotor activity of male HVC neurons in relation to female calls are sensory and context dependent during calling interactions. (a) Control situation, where the sensory cues and the context cues were intact during cohabitation. Female calls were closely associated with male calls. (b) The sensory cues were limited by separating a male and a female zebra finch into two acoustically interconnected sound boxes while the context cues were intact. Female calls were also closely associated with male calls. (c) The context dependent cues were limited by selecting those single female calls that occurred without any vocal association for at least 1 second, while the sensory cues were intact during cohabitation. (d) Both the sensory and context dependent cues were limited. Upper panels illustrate different contexts, in which female calls (red vertical strikes) associated differently with male calls (blue vertical dashed strikes) as function of time. Lower panels display the neuronal activity of male HVC related to female calls in different contexts. Example of one zebra finches pair. See other examples in Supplementary figure S4.

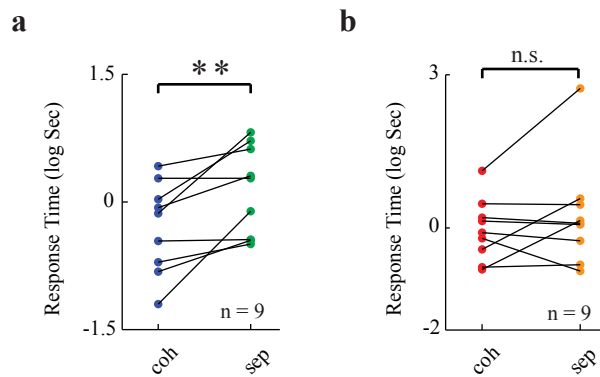


Figure 2.15: Figure 5. Limiting social-sensory cues affect the response times during calling interactions. (a) Comparison of response times between cohabitation (abbv. coh) and separation (abbv. sep) in male zebra finches. (b) Comparison of response times between cohabitation and separation in female zebra finches. \*\* and n.s. indicate significant and not-significant, respectively.  $n = 9$ .

### 2.2.4 Discussion

By using a telemetric system, we described neuronal mechanisms of song control nucleus HVC underlying natural calling interactions of zebra finch dyads. We showed that male HVC neurons displayed distinctive neuronal activities related to male-female calling interactions as compared with call repetitions. Second, we showed that the calling (motor production) related HVC neurons are as well involved in sensory processing of incoming female calls. In particular, we showed that the sensory processing of HVC neurons depends on the social context: Being in social context leads to auditory responses of HVC neurons only if a pre-sensory activity is produced. Thus, HVC neurons play distinctive roles in 1) initiating a call, 2) the decision to call in response to incoming female calls, and 3) the evaluation of the reliability of the mate's calling behavior. In contrast to HVC neurons, male RA neurons only show premotor activity for active calling, but exhibit neither auditory-elicited discharge to incoming female calls nor predictive discharge to upcoming female calls [211] and in consequence are likely to execute the HVC motor instructions for call production. Likewise, in song control RA neuronal activity is related to the production of motoric details of the syllables but not in the temporal organization of singing [124]. Although bilateral damage to RA and HVC have no effect on the production of harmonic "Stack" calls [199], the calling-related activity in HVC found in previous studies [138, 132, 85] and in this study may suggest that the telencephalic song pathway is associated with active callings. Many songbirds, like zebra finches are social animals. The calling-related activity in HVC may contribute to the plasticity of the calls that ensures a complex vocal interaction in a social group. Recent work has obtained evidence regarding the call timing associated with the telencephalic song nucleus RA, to which the RA projection neurons in HVC exhibited a premotor activity in relation to the harmonic "Stack" calls [20]. Such neuronal coordination between RA and HVC may encode shared information about call timing that facilitates vocal interactions in a social context.

The pre-sensory activity of HVC neurons further suggests that activity in this sensorimotor nucleus bring about an action and resemble a predictive discharge to evaluate the sensory states of self and others for better prediction of upcoming sensory events. The expectation could not be done without exogenous sensory and contextual cues. Limiting the sensory cues caused significant delays of response times in male but not female zebra finches. This might hint to the fact that female zebra finches develop only a rudimentary HVC [82] that might lack the circuitry necessary to produce and evaluate sensory expectations. Thus, for vocal communication, female zebra finches might rely only on sensory processing that takes place in their auditory cortex-like forebrain areas [178, 74] and that then inform HVC or other areas capable to initiate call production. Unfortunately, we were unable to record from female HVC in order to clarify if they differentiate sensorimotor neurons similar to those of male HVCs. The male-specific effect on call response times together with a sex-specific HVC circuitry, would be consistent with the idea that the sensory representation in sensorimotor cortex is a process of sensory evaluation driven by action [14], whereas the sensory representation without action can be accomplished

by sensory cortex [64]. Accumulating evidence show that the predictive representation requires mirror neuron systems [111] and contextual information [131], and substantially supports the internal model of sensory-motor coupling [190] in which the motor command serves an efference signal that captures the causal relationship between actions and sensory cues [241]. Similar to prediction of visual cues, auditory events are also perceived in a predictable manner [19]. In this respect, the neuronal activity of male HVC may build up a connection between the motor commands of calls and their consequences of upcoming female calls as sensory feedbacks, thereby facilitating cognition and bonding during a natural social interaction.

How could the predictive discharge of male HVC be behaviorally relevant? First, sensorimotor expectation might benefit response time in actions [100]. Second, sensorimotor expectation might help to strengthen the usage of calls for maintaining pair bonding. Zebra finches may need reinforcement learning to adjust call usage to the calling behavior of the mate during the breeding cycle. Zebra finches show a dynamic and plastic call usages in social groups that is in addition related to breeding success [77]. In addition to anticipate the perceptual effect of their own calls, male zebra finches may also use such action-effect [101] associations with their potential mates. Let's consider the female call is a positive reinforcer [99], the pair bonding by emerging a pattern of calling interactions would be enhanced, if the predictive discharge were rewarded by the upcoming female calls. By contrast, if a predictive discharge kept failing to predict upcoming female calls, the pattern of calling interactions would not emerge or would gradually stop occurring. Such processes underlying the enhancement of pair bonding may be driven by intrinsic sensorimotor output that coincides with extrinsic sensory input of female calls in a social context.

### 2.2.5 Materials and Methods

#### Animals

Animal cares and experiments were carried out according to the Government of Upper Bavaria (Az. 55.2-1-54-231-25-09). Experimental zebra finches were adult male and female zebra finches (*Taeniopygia guttata*) obtained from our breeding facility. The birds were kept in cages, placed in custom-made, sound-proofed boxes. Each sound box was equipped with a microphone, a speaker and a telescopic antenna for wireless recordings. Zebra finches were kept in a 14/10 Light/Dark cycle (fluorescent lamps), 24°C and 60-70 % humidity.

#### Sound recording

Custom-made wireless microphones (0.6 g including a battery) were used for wireless sound recording [211]. The wireless microphone was placed on the back and fixed with

elastic band around the upper thighs of the animal. The frequency modulated radio signals were received with AOR5000 communication receivers (AOR, Ltd., Japan). Audio signals were either fed into an eight channel audio A/D converter (Fast Track Ultra 8R, Avid Technology, Inc. U.S.A.) and recorded with custom-made software or registered on a DASH8X data recorder (Astro-Med, Inc., RI, U.S.A.).

### **Playback experiment**

Playback experiment was performed in isolation, where male was separated from female and was broadcasted by the recorded “stack” calls of his female mate via a speaker (FRS 8, 30w, 8 $\Omega$ , VISATON, Germany). Female “stack” calls were recorded by the in-box microphone (TC20, Earthworks, U.S.A.) prior the experiment. The playback sequences were uniformly randomized from the interval between 1 and 30 seconds. The total number of playback calls was  $n = 960$ .

### **Implantation of electrode and chronic recording of neuronal activity**

Male zebra finches were anesthetized using isoflurane inhalation (0.8-1.8% at 0.5l  $O_2$  /min). Animal was kept warm on a constant temperature pad and warped in a thin gauze blanket. The skin of the head was plucked, disinfected and treated with lidocaine (Xylocain Gel 2%, AstraZeneca). After a window on the skull was opened, the bifurcation of the midsagittal sinus was served as reference coordinate. A 2 M $\Omega$  tungsten electrode (FHC, Bowdoin, U.S.A.) was lowered to HVC using a micromanipulator (Lugis and Neumann, Feinmechanik & Elektrotechnik, GmbH, Germany). After verifying the signals of HVC by playing back BOS, NBOS and noise [133], the window on the skull and the pins for both reference and recording electrode was covered and fixed with dental cement (Tetric evoflow, Ivoclar Vivadent, GmbH, Germany). Animal was equipped with a custom-made neural telemetric device and placed back to sound box for chronic recordings of both vocal and neural signals [211]. After finishing experiment, a lesion was made at the recording site by using a linear stimulus isolator (WPI, Inc, U.S.A.).

### **Sorting vocalizations and neural data**

All vocal and neuronal signals were recorded continuously at 22050 Hz sampling rate and written to WAV files. Vocalizations were extracted from audio files by using custom-written software (sound explorer). The vocalizations of zebra finches were clustered by analyzing their sound features [211]. Spike sorting and clustering were done by using custom-written software based on C++ language [211].

### Analysis of behavioral data

We record the vocalizations of female and male calls simultaneously. To analysis the interaction pattern of calls between male and female zebra finches, we calculate the probability of finding  $N$  events of male calls that occur before and after the female calls as function of time with a fixed bin interval (50 ms). The onset of female and male calls denote  $t_i^f$  ( $1 \leq i \leq N^f$ ) and  $t_j^m$  ( $1 \leq j \leq N^m$ ), respectively. The relative time lags ( $\tau_{ji}^{mf}$ ) of male calls to female calls are the time differences between all male calls to each female call. Hence, we obtain  $\tau_{ji}^{mf} = t_{j \times N^f}^m - t_{N^m \times i}^f$ , where the counts of all time lags ( $\tau_{ji}^{mf}$ ) will be  $N_{ji}^{mf} = N^m \times N^f$ . The interaction pattern is approximated by a cross-correlation between the probability density functions of male and female calls. Let the probability density functions of male and female calls to be  $Pr^m$  and  $Pr^f$ , the discrete cross-correlation will be:  $\kappa(\tau)^{mf} = \sum_{a=-\infty}^{\infty} Pr^m[a]Pr^f[a+b]$ , where  $a$  and  $b$  sets of integers which indicate the discrete time sequence of  $Pr$  and  $\kappa(\tau)$ , respectively.  $\kappa(\tau)^{mf}$  indicates the calling interactions between male and female zebra finches. In Fig. 5, the absolute time lags ( $|\tau|$ ) were calculated to indicate the absolute response times, i.e. the time differences between two joint call events. Hence, the response times of female calls to male calls will be:  $|\tau^{mf}| = t_i^f - t_j^m$ , where the response times of male calls to female calls will be:  $|\tau^{fm}| = t_j^m - t_i^f$ . Two sided Wilcoxon signed-rank test was used for the null hypothesis that the response times  $|\tau_{separation}^{fm}| - |\tau_{cohabitation}^{fm}|$  or  $|\tau_{separation}^{mf}| - |\tau_{cohabitation}^{mf}|$  come from a distribution with median different than 0. If the result of two-sided Wilcoxon signed-rank test indicates significant. Right-sided Wilcoxon signed-rank test is used for the null hypothesis that the response times  $|\tau_{separation}| - |\tau_{cohabitation}|$  come from a distribution with median greater than 0. The test results are considered for significance at a 5% confident level.

### Analysis of neuronal data

Analyzing response pattern is similar as analyzing calling interaction pattern. We recorded the neuronal activity of male HVC simultaneously with vocalizations of female and male zebra finches. The spike times denote  $t_k^s$  ( $1 \leq k \leq N^s$ ). We calculate the probability of finding  $N$  events of spike that occur before and after the male calls as function of time with a fixed bin interval (10 ms) i.e.  $\kappa(\tau)^{sm}$  indicates the neuronal activity related to male calls, i.e. the probability of finding  $N$  events of spike that occur before and after male calls as function of time with a fixed bin interval (10 ms).  $\kappa(\tau)^{sf}$  indicates the neuronal activity of male HVC related to female calls, i.e. the probability of finding  $N$  events of spike that occur before and after female calls as function of time with a fixed bin interval (10 ms). Standardized Z-score was calculated to compare between individuals ( $n = 8$ ). The values corresponding to Z-scores of 1.96 or  $-1.96$  indicate the cumulative percentages of 95% or 5% accordingly. We used F-test to compare the neuronal activity patterns. For example, the average curve depicted in Fig. 3c [i.] was used as the reference to compare the curves between “xMx” and “-M-” depicted in Fig. 3f [i.]. The sum of squared distances between reference and tested curves (e.g. D1: “xMx” and D2: “-M-”) were calculated

separately. The Null hypothesis is that the difference between  $D_1$  and  $D_2$  yields a same variability as the difference within  $D_1$  and  $D_2$ . F-ratio between sum of squared differences between  $D_1$  and  $D_2$  and sum of squared difference within  $D_1$  and  $D_2$  were calculated. The degree of freedom for between groups ( $df_b$ ) and within groups ( $df_w$ ) is 1 and 14 ( $n = 8$ ), respectively. The p value for the test is given by:  $1 - f_{cdf}(x, df_b, df_w)$ , where x denotes F-ratio.

### Association of calling events

Since the likelihood of the calling responses accounts for a peak within a time window of 500 ms, we use this time window to differentiate single and associated events. If any (self and other) call event occurred 500 ms before and after a vocal event, this event will be defined as an associated call event. If not, it will be defined as a single call event. Those associated call events were divided into different groups to cover all the possibilities of associations, which were resulted by combination with repetition of an element set “male call (M)”, “female call (F)” and “neither (-)”. There were 27 possible associations, if repetition was allowed. We left out 9 associations, where the element “neither (-)” was placed in the middle (e.g. “F-M”, “M-F and “—”). Thus we obtained all 18 possible associations: “-F-”, “-FF”, “-FM”, “-M-”, “-MF”, “-MM”, “FF-”, “FFF”, “FFM”, “FM-”, “FMF”, “FMM”, “MF-”, “MFF”, “MFM”, “MM-”, “MMF”, “MMM”.

### 2.2.6 Supplementary information

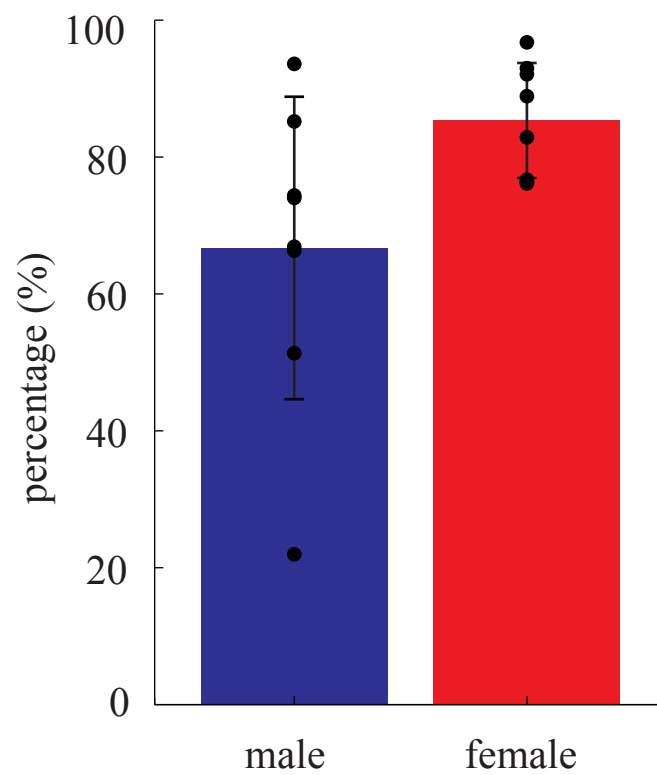


Figure 2.16: Supplementary Fig. S1. The proportions of “stack” calls in males  $66.7 \pm 22.1\%$  and females  $85.3 \pm 8.4\%$ . (mean  $\pm$  s.d.,  $n = 8$ ).



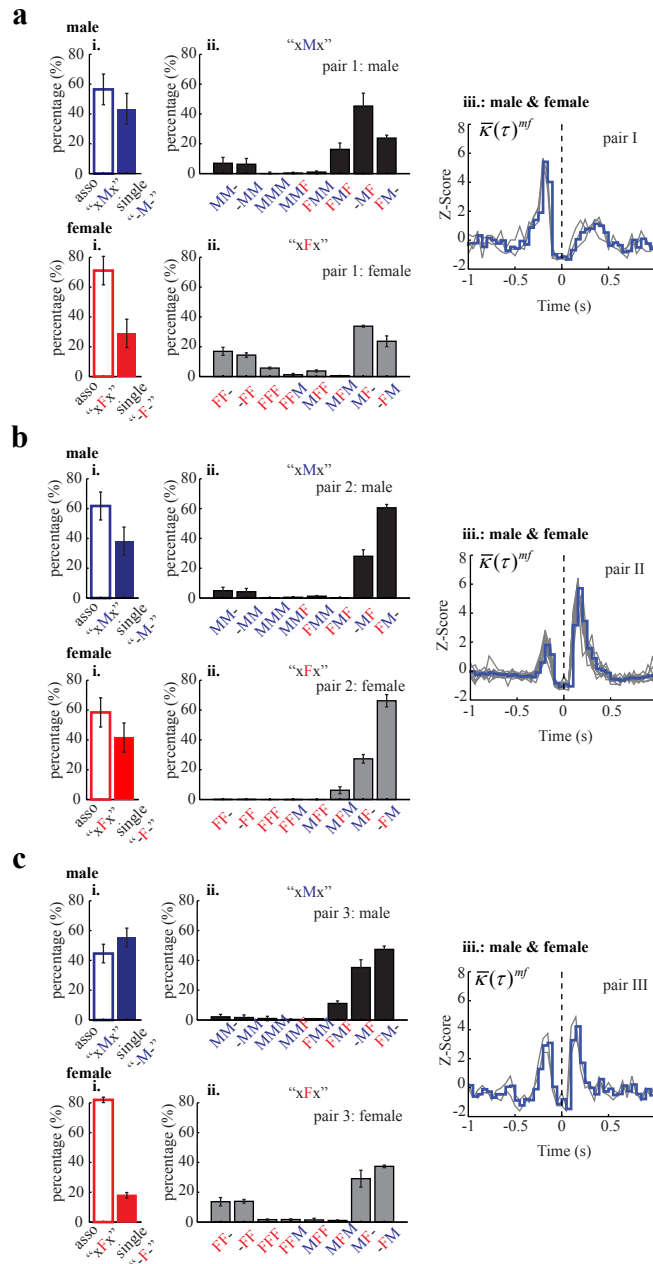


Figure 2.17: Supplementary Fig. S2. Different types of calling associations between of male and female zebra finches. Eight zebra finches pairs were tested (a – c). The panel male [i.] compares between the proportions of the associated male calls (“xMx”) and the single male calls (“-M-”). The panel female [i.] compares between the proportions of the associated female calls (“xFx”) and the single female calls (“-F-”). The panel [ii. “xMx”] compares the proportions of different calling associations in male. The panel [ii. “xFx”] compares the proportions of different calling associations in female. The panel [iii. male & female] illustrates the standard score (Z-Score) of calling interactions between male and female. The gray lines correspond to different calling interaction patterns in different days, (a)  $n = 3$ ; (b)  $n = 12$ ; (c)  $n = 2$ , whereas the blue line corresponds to average calling interaction pattern.

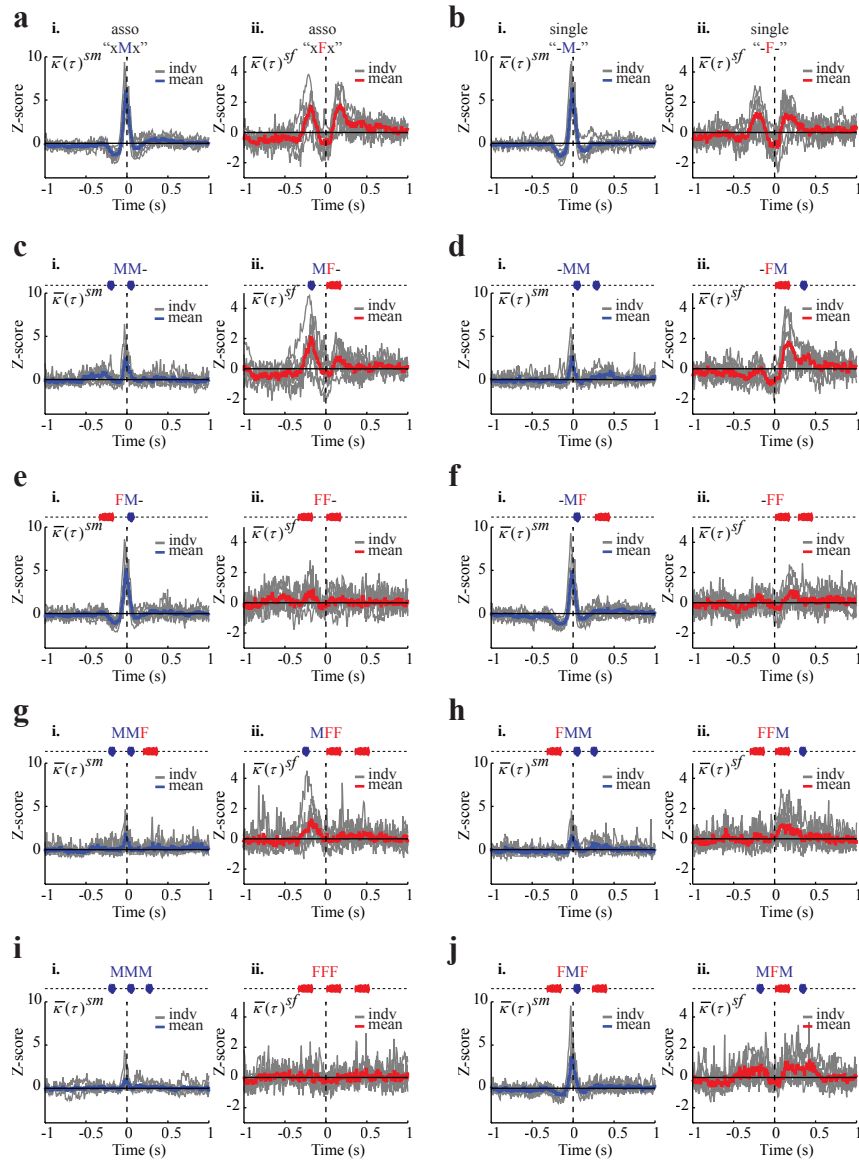


Figure 2.18: Supplementary Fig. S3. Neuronal activity of male HVC related to different types of calling associations. (a) Neuronal activity of male HVC related to [i.] the associated male (“xMx”) and [ii.] the associated female (“xFx”) calls. (b) Neuronal activity of male HVC related to [i.] the single male (“-M-”) and [ii.] the single female (“-F-”) calls. (c – j) Neuronal activity of male HVC related to different types of associated male and female calls. To schematically illustrate different types of calling associations, the waveforms of the male calls (blue) and female calls (red) are showed on the upper panels (Gray line indicates individual trace. Blue and red lines indicate average traces of neuronal activity [i.] related to their own calls [ii.] related to female calls, respectively.).

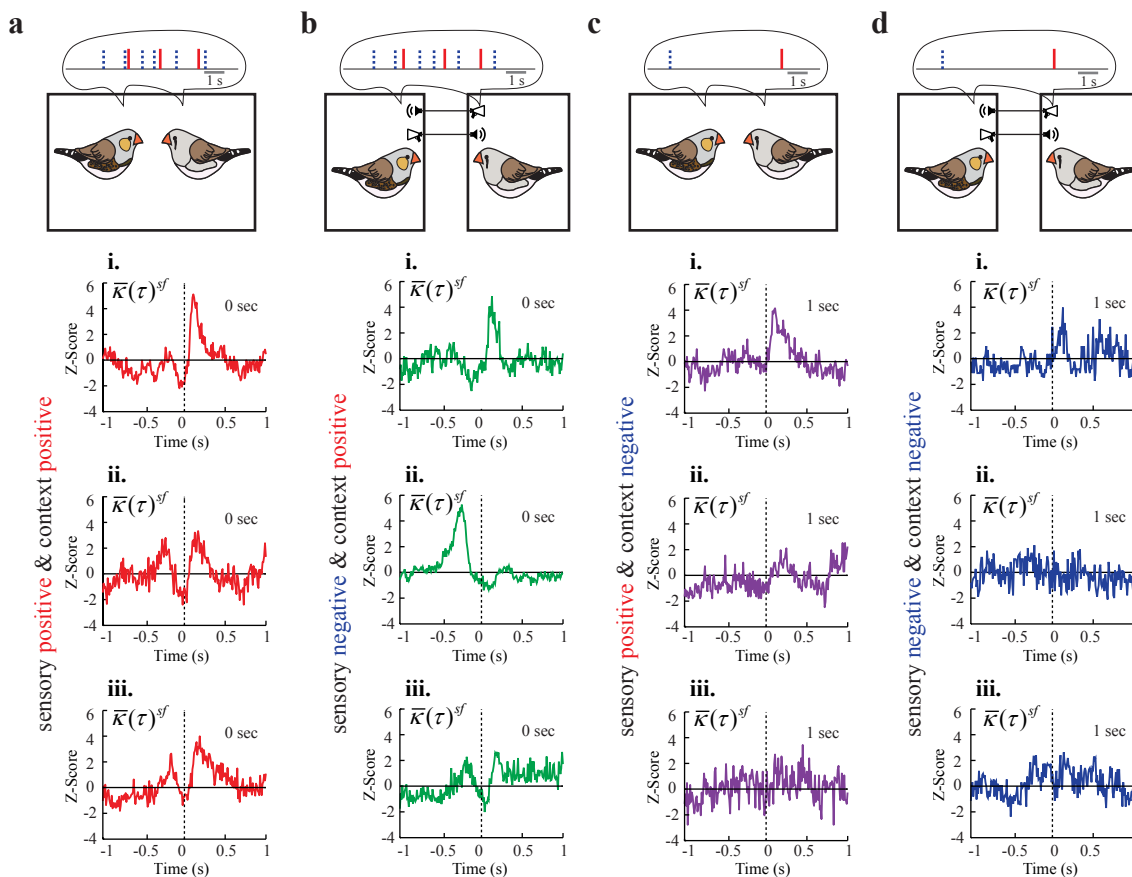


Figure 2.19: Supplementary Fig. S4. Premotor activity of male HVC are sensory and context dependent during calling interactions. (a) Control situation, where the sensory cues and the context cues were intact during cohabitation. Female calls were closely associated with male calls. (b) The sensory cues were limited by separating a male and a female zebra finch into two acoustic interconnected sound boxes while the context cues were intact. Female calls were also closely associated with male calls. (c) The context dependent cues were limited by selecting those instant female calls that occurred without any vocal association for at least 1 second, while the sensory cues were intact during cohabitation. (d) Both the sensory and context dependent cues were limited. The neuronal activity of male HVC related to female calls in different contexts was plotted against time. [i.], [ii.] and [iii.] indicate three different individuals.

	value	-M-	xMx	MM-	-MM	MMM	MMF	FMM	FMF	-MF	FM-
-M-	F	0									
	p	1									
xMx	F	3.1E-3	0								
	p	0.96	1								
MM-	F	19.21	18.79	0							
	p	6.2E-4	6.9E-4	1							
-MM	F	10.97	10.75	0.06	0						
	p	5.1E-3	0.01	0.81	1						
MMM	F	103.87	97.77	0.22	0.48	0					
	p	7.4E-8	1.1E-7	0.64	0.50	1					
MMF	F	45.80	44.43	0.35	0.61	0.07	0				
	p	9.1E-6	1.1E-5	0.57	0.45	0.79	1				
FMM	F	30.65	29.88	0.11	0.31	4.8E-3	0.06	0			
	p	7.3E-5	8.3E-5	0.74	0.59	0.95	0.80	1			
FMF	F	10.40	9.84	4.81	2.48	19.20	12.50	8.35	0		
	p	6.1E-4	0.01	0.05	0.14	6.3E-4	3.3E-3	0.01	1		
-MF	F	3.14	2.87	10.49	5.90	43.75	24.84	16.96	2.10	0	
	p	0.10	0.11	0.01	0.03	1.2E-5	2.0E-4	1.0E-3	0.17	1	
FM-	F	1.43	1.28	12.32	7.07	49.60	28.26	19.53	3.46	0.21	0
	p	0.25	0.28	3.5E-4	0.02	5.8E-6	1.1E-4	5.8E-4	0.08	0.65	1

Table 2.3: Supplementary Table S1. Comparison of neuronal activity of male HVC in relation to their own calls between different types of calling associations. F-tests for pairwise comparisons of different types of calling associations. “F” and “p” in the “value” column indicate the F-values and the p-values, respectively.

	value	-F-	xFx	FF-	-FF	FFF	FFM	MFF	MFM	-MF	FM-
-F-	F	0									
	p	1									
xFx	F	0.23	0								
	p	0.64	1								
FF-	F	3.27	7.10	0							
	p	0.09	0.02	1							
-FF	F	2.46	7.68	0.31	0						
	p	0.14	0.01	0.59	1						
FFF	F	3.11	7.02	0.01	0.23	0					
	p	0.10	0.02	0.94	0.64	1					
FFM	F	2.18	7.03	0.42	0.01	0.33	0				
	p	0.16	0.02	0.53	0.91	0.58	1				
MFF	F	2.37	7.80	0.39	4.8E-3	0.29	2.4E-3	0			
	p	0.15	0.01	0.54	0.95	0.60	0.96	1			
MFM	F	1.63	4.91	0.51	0.05	0.51	0.02	0.03	0		
	p	0.22	0.04	0.49	0.82	0.53	0.90	0.86	1		
MF-	F	1.51	3.65	0.25	2.6E-3	0.19	1.3E-3	2.4E-7	0.02	0	
	p	0.24	0.08	0.62	0.96	0.67	0.97	1	0.9	1	
-FM	F	1.96	5.89	0.41	0.02	0.31	2.6E-3	3.9E-3	0.01	2.3E-3	0
	p	0.18	0.03	0.53	0.90	0.58	0.99	0.95	0.92	0.96	1

Table 2.4: Comparison of neuronal activity of male HVC in relation to female calls between different types of calling associations. F-tests for pairwise comparisons of different types of calling associations. “F” and “p” in the “value” column indicate the F-values and the p-values, respectively.

## 2.3 Spectrotemporal coding of auditory trains in the auditory forebrains of zebra finches

### 2.3.1 Abstract

Many animals live in social groups. Hearing and communication in a complex acoustic environment require an ability to segregate contextual sounds from environmental noises. Field L is primary auditory forebrain area in birds that is highly selective to natural sounds over synthesized sounds. It is however unknown how the neuronal sensitivity to contextual sounds is maintained in a complex auditory scene. Based on previous studies on auditory scene analysis in humans and birds, I investigated the neuronal mechanisms underlying the spectrotemporal precision in neuronal response to single sounds and the maintenance of sound elements in a stimulus train. I showed that Field L neurons responded equally well and tightly locked to the spectrotemporal structures of single sounds with inserted gaps and sounds without gaps. Furthermore, distinctive Field L neurons exhibited specific spectrotemporal coding of sound elements in response to stimulus trains. The activity of Field L neurons was attenuated faster in response to stimulus trains of natural sound elements than trains of artificial sound elements. Moreover, the habituation rate in response to trains of natural sound elements was higher in males than in females. The results provide further support for the temporal encoding of natural sounds in primary auditory forebrain. The sexual specificity in response to stimulus trains may provide a valuable framework for comparative studies in the auditory system.

### 2.3.2 Introduction

Sounds are rarely heard in isolation. Social mammals and birds that live in a complex acoustic environment must develop a functional mechanism to discriminate contextual sounds from environmental noises. Field L is a primary auditory forebrain area in birds that are highly selective to natural sounds [120, 81, 2]. While discriminating natural sounds from artificial noise, animals must simultaneously organize and segregate different types of natural sounds to extract contextual information. For the neural substrate to resolve the acoustic information from a natural auditory scene, encoding of temporal and spectral structure of the sound is crucial [22]. In order to investigate the spectrotemporal regularity of sound elements in a complex scene, auditory scene analysis provides an ideal paradigm as compared with introducing a mixture of natural sounds that may induce more noise than information in a decoding scheme.

In a classic paradigm of auditory scene analysis, a sequence of syllables is presented to a listener. Each syllable consists of two pure tones with different frequencies (“A” tone and “B” tone). “A” and “B” tones are separated by a silent gap (inter-tone interval). While the sequence is presented at increasing rates (i.e. the inter-syllable intervals are decreasing), a listener is likely to report hearing two separated streams, if the frequency difference

between “A” tone and “B” tone is large [31]. With this approach, studies in human [31], monkeys [59] and birds [18] reported that the inter-tone interval (the gap between “A” and “B”), the inter-syllable interval (the gap between two consecutive “AB”) and the spectral difference between tones are crucial for auditory stream segregation. Furthermore, psychoacoustic studies in human reported that auditory experience such as lexicon affected the possibility to identify speech sounds in playback trains [21], suggesting a neural mechanism underlying auditory selectivity for familiar sounds in the auditory system.

In line with previous studies, here I investigated the neuronal mechanism underlying spectrotemporal coding of auditory trains in the avian auditory forebrain. I adapted the paradigm of auditory scene analysis and manipulated two coefficients: the spectral structure of the syllable and the inter-syllable interval. The durations of the inter-syllable intervals is decreasing exponentially from 535 ms to 35 ms in the course of the stimulus train. For the spectral coefficient, I compared the auditory sensitivity in relation to artificial tones, heterospecific (blackbird) and conspecific syllables. For the temporal coefficient, I separated an intact syllable into two segments (“A” and “B”) and inserted a duration of 35 ms gap between these two segments in line with previous behavioral and physiological studies, which showed that the intervals of zebra finches’ songs ranged between 20 ms and 100 ms and the mean interval of songs was approximated by 36 ms [186, 3]. In this study, I first examined the neuronal activity in male and female Field L in relation to single stimuli with varied lengths of inserted gaps. Further, I employed an auditory scene paradigm to examine the temporal and spectral regulations of natural sounds in playback trains. Next, I examined the neuronal activity in male and female Field L in relation to the stimulus trains.

### 2.3.3 Results

#### Comparison of auditory sensitivity to gaps in male and female auditory forebrain

To examine the temporal precision of auditory responses to stimuli, I compared the auditory sensitivity to single syllables with gaps of different lengths between syllable segments (duration: 35 ms, 70 ms and 100 ms). Figure 1 shows three isolated neurons in male Field L (Fig. 1 a) and three isolated neurons in female Field L (Fig. 1 b) that responded to auditory stimuli. These male and female Field L neurons exhibited robust firing patterns that were tightly locked to the spectrotemporal structure of the stimulus sounds regardless of the lengths of inserted gaps. To compare the response precision in relation to single stimuli with different lengths of gaps, I excluded responses that occurred during the “gap” and shifted the responses to the segment 2 (S2) leftward to align with the responses to the first segment (S1). I show that both male and female Field L neurons exhibited precise timed auditory responses to the stimuli regardless of the length of inserted gaps and the type of sounds (Fig. 1c, d). In male ( $n = 8$ ), 88 % (127 out of 144 pairwise comparisons) of data yielded a high degree of correlation (average Pearson coefficient:  $r = 0.49$ ,  $p <$

0.001) between the responses to intact syllables and the responses to the syllables with inserted gaps, whereas 12 % (17 out of 144 pairwise comparisons) of data did not show significant correlation (average Pearson coefficient:  $r = 0.06$ ,  $p \geq 0.001$ ); In female ( $n = 6$ ), 82 % (98 out of 120 pairwise comparisons) of data yielded a high degree of correlation (average Pearson coefficient:  $r = 0.54$ ,  $p < 0.001$ ) between the responses to intact syllables and the responses to the syllables with inserted gaps, whereas 18 % (22 out of 120 pairwise comparisons) of data did not show significant correlation (average Pearson coefficient:  $r = 0.06$ ,  $p \geq 0.001$ ). Furthermore, I used multi-way ANOVA analysis to test the effects of the spectral parameter (stimulus), the temporal parameter (gaps) and the sexes on the response strengths in Field L. The results show that the effects of the spectral parameter (stimulus) contributed significantly to the variance of the response strengths ( $p_{df(3,200)} = 0.037$ ). The temporal parameter (gaps) ( $p_{df(3,200)} = 0.458$ ) and the sexes ( $p_{df(3,200)} = 0.127$ ) did not show significant effect on the variance of the response strengths.

### **Comparison of auditory responses to dynamic stimuli in male and female auditory forebrain**

The present study shows precise spike timing in response to individual stimuli that depends on the spectrotemporal structure of the stimulus. I next examined the auditory responses to trains of stimuli in male and female Field L. In each trial of the playback, a synthesized syllable was repeated 48 times. The inter-segment interval (35 ms) was kept constant except for the second last syllable (70 ms), whereas the inter-syllable intervals were decreasing exponentially in the course of the trial (Fig. 2a). I show that the same isolated neurons in male Field L ( $N_m1$ ,  $N_m2$  and  $N_m3$  depicted in Figure 1a) maintained a precise temporal spiking in response to initial syllables of the trains (Fig. 2b). The response rates of the Neuron  $N_m1$  in relation to BOS, conspecific (NBOS) and heterospecific (blackbird) stimuli were attenuated in the later course of the stimulus trains, whereas the response strengths of the Neuron:  $N_m1$  in relation to pure tone stimuli maintained constant throughout the playback (Fig. 2c). The response rates of the Neurons 2 and 3 ( $N_m2$  and  $N_m3$ ) exhibited different degrees of cohabitation in relation to different types of stimulus trains (Fig. 2c). The same isolated neurons in female Field L (as depicted in Figure 1b) exhibited distinctive activity patterns in relation to different types of stimulus trains (Fig. 2d, e). I have identified a neuron ( $N_f1$ ) that did not respond to stimulus trains of pure tones, but showed a robust response to the spectrotemporal structure of natural sounds (Fig. 2d, e). Further, auditory responses of this neuron ( $N_f1$ ) to unfamiliar conspecific and heterospecific sounds were attenuated in the later course of the stimulus trains. but increased in the later course of the playback trains of the familiar conspecific sounds (Fig. 2d, e). Taking together, the results showed that field L neurons displayed distinctive coding schemes to the dynamic stimulus.

In order to compare the sexual differences in auditory responses to stimulus trains of conspecific, heterospecific syllables and pure tones, I measured the multiunit activity in



male and female Field L. I first employed an ANOVA analysis to test the effects of the stimulus and the sexes on the habituation rates in Field L. The results showed that both effects contributed significantly to the variance of the habituation rates (Fig. 3, stimulus:  $p_{df(3,51)} = 0.040$ , sex:  $p_{df(1,51)} = 0.006$ ). Further, I compared the habituation rates in relation to different playback trains in male and female, respectively. In both male and female Field L, auditory responses (units) showed neither significant difference in discrimination of BOS syllables (i.e. familiar syllables) from conspecific syllables (NBOS and unfamiliar syllables, respectively) ( $p_{male} = 0.926$ ,  $n = 8$ ,  $p_{female} = 0.834$ ,  $n = 6$ ) nor from heterospecific syllables (blackbird) ( $p_{male} = 0.067$ ,  $n = 8$ ,  $p_{female} = 0.789$ ,  $n = 6$ ). However, both male and female Field L units showed significant difference in discrimination of natural sounds (zebra finches' and blackbird's song syllables) from pure tones (Fig. 3,  $p_{male} = 0.039$ ,  $n = 8$ ,  $p_{female} = 0.016$ ,  $n = 6$ ). Subsequently, I compared the habituation rates in relation to different playback trains between male and female Field L. The paired-sample test yielded a significantly higher habituation rate in relation to conspecific (BOS and NBOS) trains and mechanic gallop trains in male than in female (Fig. 3,  $p_{BOS\ trains} = 0.038$ ,  $p_{NBOS\ trains} = 0.032$ ,  $p_{gallop\ trains} = 0.021$ ,  $n = 6$  pairs). But male and female showed no significant difference in the habituation rates in relation to heterospecific trains (Fig. 3,  $p_{blackbird\ trains} = 0.366$ ,  $n = 6$  pairs).

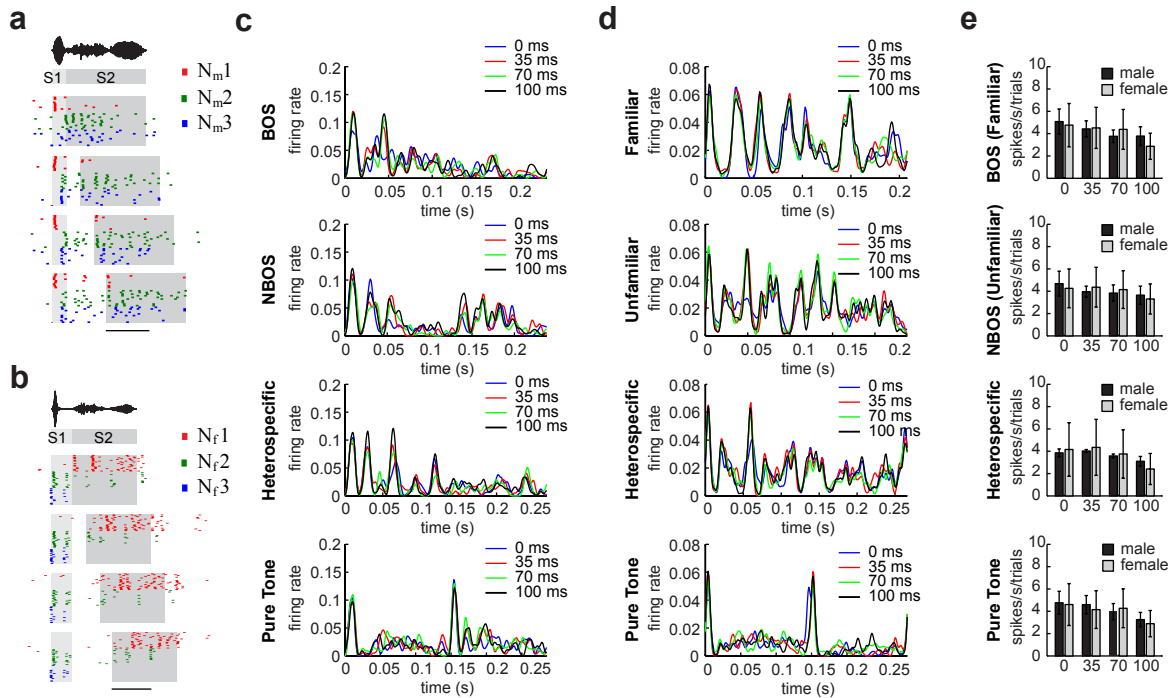


Figure 2.20: Figure 1: Auditory sensitivity to single stimuli with inserted gaps in male and female auditory forebrain. (a) An illustration of three neuron types in male Field L ( $N_m1$ ,  $N_m2$ ,  $N_m3$ ) that responded to BOS syllables with different lengths of inserted gaps (35 ms, 70 ms and 100 ms). Scale bar: 200 ms. (b) An illustration of three neuron types in female Field L ( $N_f1$ ,  $N_f2$ ,  $N_f3$ ) that responded to familiar song syllables with different lengths of inserted gaps (35 ms, 70 ms and 100 ms). Scale bar: 200 ms. The stimulus with 0 ms gap indicates an intact stimulus. (c) Comparing multi-unit responses to stimuli with different lengths of gaps in male. The response profile to BOS is showed in (a). PSTHs of firing rates were smoothed by the FIR filter with a 11-point Gaussian window. (d) Comparing multi-unit responses to stimuli with different lengths of gaps in female. The response profile to the familiar song syllable is showed in (b). (e) Comparing sexual difference of auditory responses in Field L. The error bars show standard error.

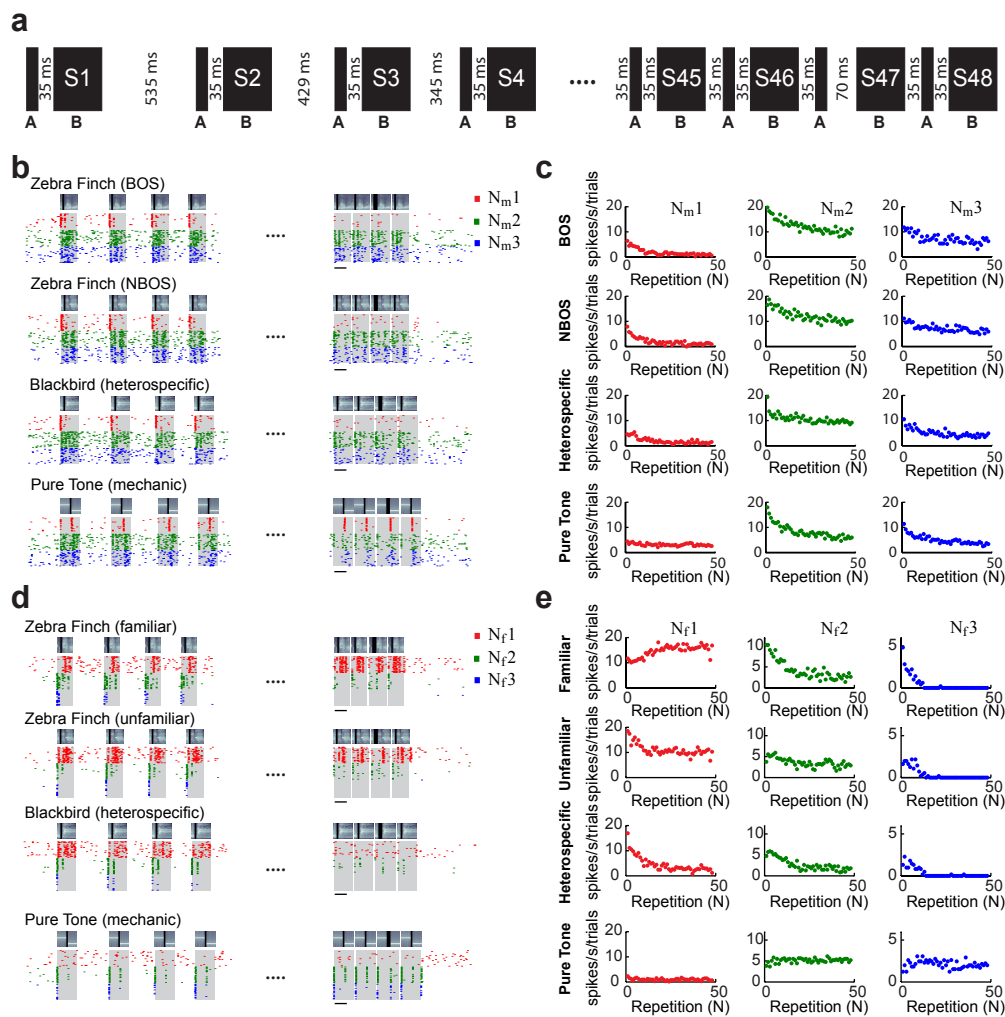


Figure 2.21: Figure 2: Multi-unit responses to stimulus trains in male and female Field L. (a) Schematic illustration of the playback paradigm, in which 48 syllables were played in a repetitive sequence. The inter-segment interval was kept constant while the inter-syllable interval decreased exponentially. (b) Examples of three multi-unit activity in male Field L ( $N_{m1}$ ,  $N_{m2}$  and  $N_{m3}$ ) in relation to BOS syllables, NBOS syllables, blackbird (heterospecific) syllables and pure tones. Scale bar: 200 ms. (c) Response strengths of three neurons shown in (b) as a function of syllable repetition. (d) Examples of three multi-unit activity in female Field L in relation to familiar song syllables, unfamiliar song syllables, blackbird (heterospecific) syllables and pure tones. Scale bar: 200 ms. (e) Response strengths of shown in (d) as a function of syllable repetition.

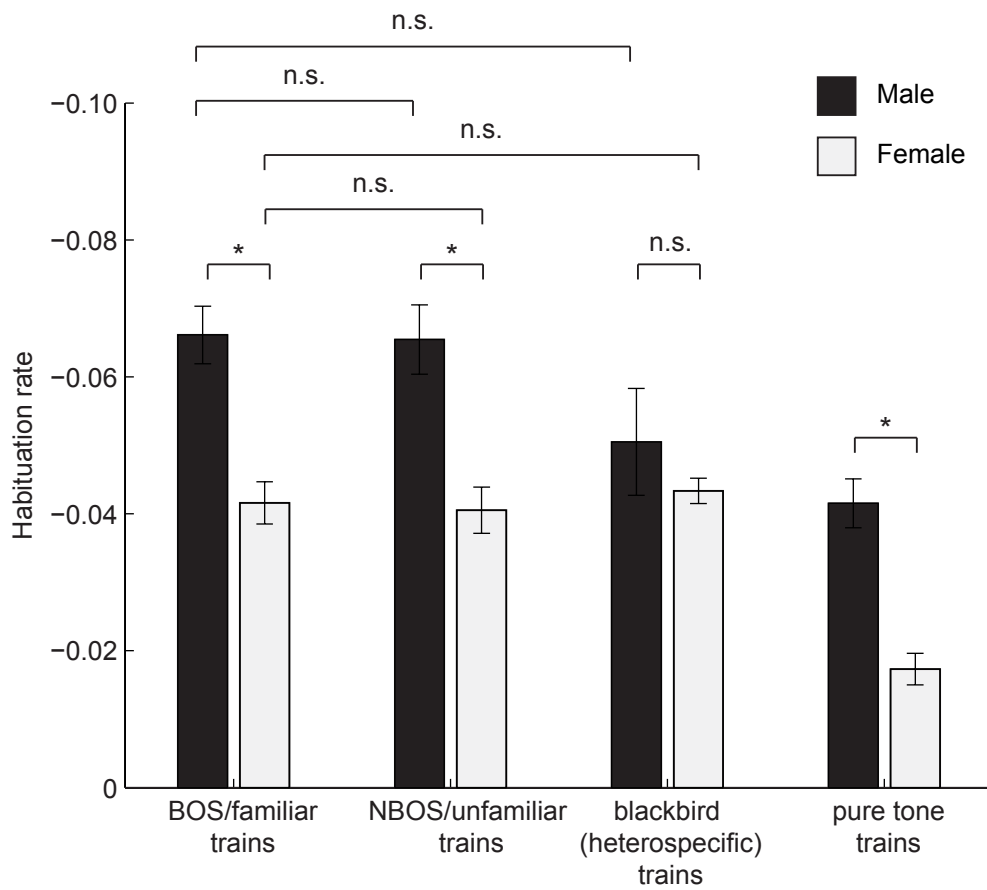


Figure 2.22: Figure 3: Pair-wise comparison of habituation rates in relation to different stimulus trains.  $n = 6$  pairs. The error bars show standard error. \* and n.s. indicate significance and non-significance, respectively.

### 2.3.4 Discussion

In present study, I show that Field L neurons in zebra finches responded precisely to the spectrotemporal structure of sounds regardless of inserted gaps. In line with previous studies that compared the neuronal responses between forward and reversed songs in Field L [121, 2], I provided further evidence that auditory responses in Field L depend on the spectrotemporal structure of the stimulus. This is consistent with the idea that tuning for spectrotemporal structures of sounds in the primary auditory forebrain is the neuronal basis for song selectivity in birds [212]. Moreover, the present study showed that Field L neurons did not discriminate between conspecific and heterospecific song syllables, but discriminated natural from artificial sounds. This result agrees with previous studies on auditory selectivity for natural sounds over synthetic sounds in Field L [213, 2].

The present study advances the knowledge about temporal and spectral encoding of stimulus trains with complex sound elements in Field L. In classical studies on auditory scene analysis in primary auditory cortex of mammals and birds, pure tones with certain semitone separations were presented at different rates [59, 17]. These studies showed that neurons in the primary auditory system responded equally well to pure tones with different frequencies that were presented at low repetition rates, but showed preference toward one tone over another when the representation rate was increased. This suggested that frequency selectivity plays an important role in auditory stream segregation [59, 17, 18]. In present study, I found Field L neurons exhibited spectral encoding of natural sounds. These spectral encoded neurons were insensitive to the spectral structure of pure tone but fired exclusively to the natural sounds with spectrotemporal characteristics. However, when the stimulus trains of natural sounds were played in a sequential train with increasing rates, I observed different degrees of habituation rates. Thus, I speculate that the auditory representation of natural sounds in a complex acoustic scene results from the interplays between heterogeneous neurons that attribute for the functional organization in the auditory forebrain.

This study further showed that the neuronal representation of stimulus trains was sexually dimorphic in zebra finches. Although both male and female zebra finches showed indifference to familiar (his own song syllable for male) and unfamiliar song syllable, the auditory responses to stimulus trains were attenuated faster in male than in female zebra finches. The selectivity for familiar song as opposed to unfamiliar song may not be encoded in Field L, but in secondary auditory forebrain CM [75]. I speculate that the sexual difference of auditory representation in the primary forebrain may link to the sexual dimorphic song system in zebra finches. In fact, convergent evidence by using operant conditioning, discrimination tasks and HVC-lesions support that the forebrain nuclei in song system affect the acoustic discriminations in songbirds [32, 44, 125, 187, 72, 73]. Furthermore, anatomical studies in birds have revealed horizontal and hierarchical inter-connected auditory forebrain/areas [133, 62]. Furthermore, physiological studies and gene-expressions further support for neuronal plasticity in auditory forebrain [143, 75]. Accumulated evi-

dence may suggest that inter- and intra- cortical connections may allow neural trainings in auditory forebrain for perceptual sensitivity. However, the neuronal mechanism underlying the sexual difference of auditory perception and the sensorimotor feedback in maintaining auditory sensitivity to species-specific sounds is unknown. Since the developed song system may contribute to the singing behavior in male as compared with female zebra finches, Further I inhibited the neuronal activity of HVC by infusing an inhibitory neurotransmitter GABA into male HVC and compared the auditory responses to playback trains in male Field L before ( $GABA^{HVC-}$ ) and after ( $GABA^{HVC+}$ ) the GABA treatment in male HVC (Supplementary Fig. S4). Subsequently, I compared the auditory responses to auditory trains in male Field L with  $GABA^{HVC-}$ , male Field L with  $GABA^{HVC+}$  and female Field L. Unfortunately, the data of GABA treatments do not provide sufficient support for the impact of the pre-motor activity of HVC on the sensory properties of field L. In the supplementary text S1, I provide pilot data related to this question and discuss the putative mechanism of sensorimotor feedback in maintaining auditory object formation of natural sounds in the auditory forebrain.

Zebra finches are group-living socially monogamous birds. Despite the importance of signature songs for individual recognition, zebra finches communicate with each other by using acoustically rather similar vocalizations. Interestingly, although the identical structure of harmonic calls in both sexes, there are on average about 14 semitones separation (Supplementary Fig. S2,  $n = 10$ ) between male and female harmonic calls. I speculate that, when a mixture of harmonic male and female calls is presented in a complex acoustic environment, the auditory stream segregation in primary auditory forebrain may play an important role in segregating acoustic information. Furthermore, electrophysiological and corticographic studies in human patients and macaque monkeys (*Macaca fascicularis*) showed that the high habituation rate was associated with the unattended stream and the non-best frequency [59, 144]. Accordingly, the higher habituation rate in male than in female zebra finches may reflex the maintenance of auditory sensitivity to conspecific song that is higher in female than in male zebra finches. In other words, it could be that female birds pay more attention to conspecific songs than male birds do.

### **2.3.5 Materials and Methods**

#### **Animals**

8 male and 6 female adult zebra finches were tested in this study. 6 of male birds were kept pairwise together with 6 female birds. Before each experiment, male and female birds were kept pairwise in sound-attenuated box (size: 54x40x28 cm) for 1 – 2 weeks. Each sound box was equipped with a microphone (C2, Behringer, Willich-Münchheide II, Germany). Zebra finches were kept in a 14/10 Light/Dark cycle (fluorescent lamps), 24°C and 60-70 % humidity.

## Stimulus

Songs were recorded from each pair and high-pass filtered with cutoff frequency at 300 Hz. The BOS syllables were extracted from the songs. The BOS syllable is the longest syllable with most spectrotemporal complexity in the song bout. The familiar syllable for the female was the BOS syllable of her cohabitated male (Supplementary Fig. S1 a – h). A NBOS syllable was selected from previous data that is unfamiliar to all tested birds (Supplementary Fig. S1 i). Blackbird (heterospecific) song was recorded in the field (Starnberg, Germany) (Supplementary Fig. S1 j). For the single stimulus, an intact syllable was separated into two segments (“A” and “B”) and inserted lengths of 35 ms, 70 ms and 100 ms gaps (the inter-segment interval) between these two segments (Supplementary Fig. S1 k). The BOS syllables were separated into two segments at varied semitone differences (Supplementary Fig. S2 a – h, 8.2 semitones difference in average,  $n = 8$ ). The semitone difference between “A” and “B” segment of NBOS syllable was 6 (Supplementary Fig. S2 i). Blackbird syllable consists of “A” and “B” segment at 11.1 semitone difference (Supplementary Fig. S2 j). The pure tone consists of “A” and “B” tone at 17.8 semitone difference (Supplementary Fig. S2 k, 2800 Hz for “A” tone and 1000 Hz for “B” tone) was generated by Audacity 2.0.2 (Audacity team, Carnegie Mellon University, freely available at: <http://audacity.sourceforge.net/>). The semitone difference between female and male “stack” calls was 14.9 in average (Supplementary Fig. S2 l,  $n = 10$ ). The playback paradigm consists of 48 repeated syllables (with a fixed 35 ms inter-segment interval in each syllable) and decreasing lengths of inter-syllable intervals from 535 ms to 35 ms (i.e. an increase of playback rate). The stimulus was delivery by a speaker (Vifa 10 BGS 119/8, Acoustic Systems Engineering, Germany) that was placed 40 cm in front of the animals. The intensity of the playback sounds ( $\sim 65$  dB) were calibrated by a Datalogging sound level meter (HD600, Extech, US).

## Extracellular recordings

Zebra finches were anesthetized using isoflurane inhalation (0.8-1.8% at 0.5l  $O_2$  /min). Animal was kept warm on a constant temperature pad and warped in a thin blanket. The head was immobilized and positioned in the surgical stereotaxic apparatus. The skin of the head was plucked, disinfected and treated with lidocaine (Xylocain Gel 2%, AstraZeneca). After a window on the skull was opened, the bifurcation of the midsagittal sinus was served as reference coordinate. A 1 M $\Omega$  glass tetrode (multi channel systems MCS GmbH, Germany) was lowered to Field L using a micromanipulator (Lugis and Neumann, Feinmechanik & Elektrotechnik, GmbH, Germany). The recoding sites associated with the position of Field L were as follows: male Field L ( $n = 8$ ): lateral (right):  $1009 \pm 71 \mu\text{m}$ ; rostral:  $904 \pm 98 \mu\text{m}$ ; ventral:  $1364 \pm 161 \mu\text{m}$ ; female Field L L ( $n = 6$ ): lateral (right):  $995 \pm 84 \mu\text{m}$ ; rostral:  $995 \pm 130 \mu\text{m}$ ; ventral:  $1232 \pm 252 \mu\text{m}$ . Extracellular signals were amplified via a customized amplifier (designed by M. Abels) and digitized by an 8-channel audio interface (M-audio fast track Ultra 8R, USA) and stored at sample rate 22050 Hz in

a PC.

## Data analysis

All electrophysiological data were analyzed off-line in Matlab (MathWorks, US). Multi-units and single-units were extracted by using spike 2 (Cambridge Electronic Design, UK). The single-units were showed as raster plots. The PSTHs of multi-units in response to individual syllables were binned with 1 ms time window. To compare the precision of spiking time between intact syllable and the syllable with inserted gaps, PSTH (peristimulus time histogram) was filtered with a 11-point Gaussian window (equivalent to 11 ms associated with 1 ms bin size for the PSTH). The delay induced by the FIR filter as a function of frequency was compensated by shifting the average delay (5 ms). Single-units were first sorted by PCA analysis (spike 2) and further verified by comparing the waveforms and using the cross-correlation to compare the time lags between each unit (Matlab). Raster plots of single-unit illustrate the accurate spiking times in relation to the onset of stimulus. The multi-units response to each syllable in a playback train was averaged over the responses to each syllable (including the responses during the inter-segment interval), but the responses to the inter-syllable intervals were excluded. The habituation rate (the slope) was estimated by the least square regression, in which the multi-units responses are the variables as function of repetitions. Person correlation, one-way and multiple-way ANOVA were performed in Matlab (MathWorks, US) with a criterion of  $\alpha = 0.05$  to determine statistical significance.

### 2.3.6 Supplementary Information

#### Supplementary text S1

I used tetrode to record the HVC neurons in response to single stimuli (Supplementary Fig. S3 a). HVC neurons responded selectively to the bird own songs (BOS), in which a particular song syllable elicited a robust firing pattern (Supplementary Fig. S3 b). In order to investigate the temporal precision of HVC neurons in response to single stimuli with inserted gaps, I presented a set of synthesized syllables to the anesthetized male zebra finches (Supplementary Fig. S3 c). I isolated two types of neurons that responded to the playback stimuli (Supplementary Fig. S3 d). The raster plots and the PSTHs showed that both neurons responded strongly to the spectral structure of the second segments of BOS syllables as compared with the responses to NBOS syllables. While increasing the durations of silent gaps between the first and the second segment to 100 ms, the same neurons 1 and 2 started to respond to the beginning of the second segments, whereas both units were insensitive to the spectrotemporal structure of NBOS syllables (Supplementary Fig. S3 e).



A solution with 250 mM GABA + 5% dextran [13] was infused into male HVCs. The multi-unit activity of Field L neurons in response to stimulus trains was measured and quantified as the habituation rate. The data showed that the neuronal activity of male Field L was decreasing in the course of stimulus trains. The habituation rates of neuronal activity in male Field L in response to stimulus trains were slower after the infusion of GABA, whereas the base-line activity did not change significantly as compared with the change of habituation rates, indicating that GABA treatment did not influence the neuronal activity in Field L via diffusion into Field L (Supplementary Fig. S4). However, although changes of habituation rates were observed by GABA treatment, the statistical test did not show significant effect of GABA treatment in 4 males. The statistical tests indicated that the factor of sexual difference contributed significantly to the variance of the habituation rates (Supplementary Fig. S4,  $p_{df(1,66)} = 0.010$ ), whereas the GABA treatment did not influence the habituation rates significantly ( $p_{df(1,66)} = 0.058$ ) and the stimulus had marginally significant effect on the habituation rates ( $p_{df(3,66)} = 0.049$ ). The habituation rates in response to 4 types of playback trains did not change after the GABA treatment in male HVC ( $p > 0.27$ , two-sample t-test,  $n = 4$ ).

Here I provided some preliminary results of functional relation between sensorimotor nucleus HVC and primary auditory complex Field L. The results in the sexual differences of auditory responses to stimulus trains may provide a valuable framework for integrating how sensorimotor feedback may influence the auditory processing playback trains of natural sounds. To further investigate the sensorimotor feedback on the formation of auditory objects in auditory forebrain, future experiments including tracing and simultaneous recordings of HVC and Field L may allow detailed descriptions of the synaptic connections between the sensorimotor system and the auditory system.

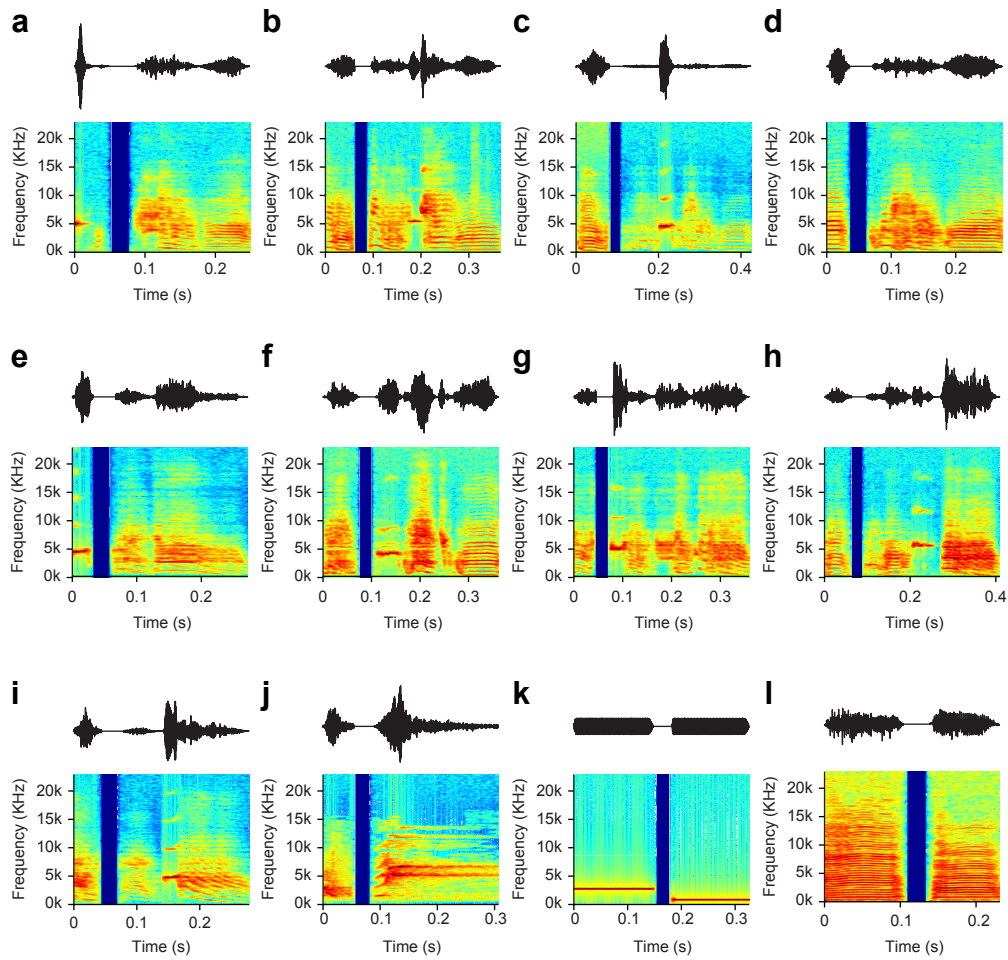


Figure 2.23: Supplementary Fig. S1: Waveforms and spectrograms of the stimuli with 35 milliseconds of silent gaps. (a – h) Eight BOS (familiar) syllable; (i) NBOS syllable; (j) heterospecific (blackbird) song syllable; (k) pure tones; (l) male and female calls.

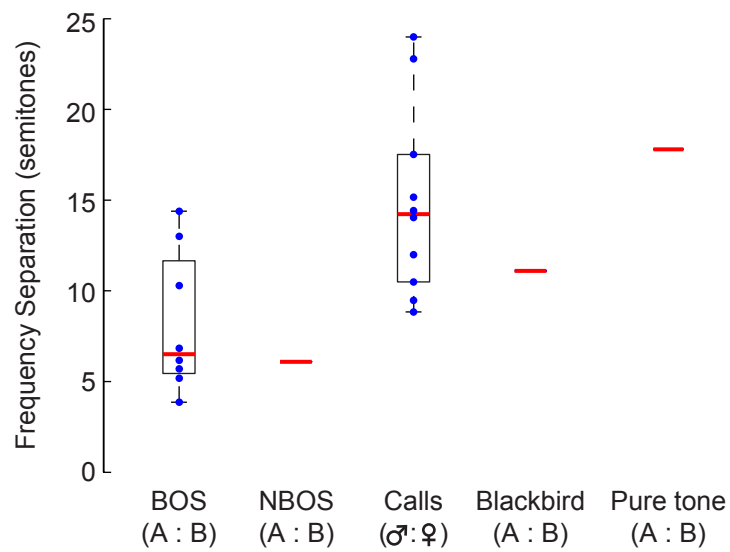


Figure 2.24: Supplementary Fig. S2: Semitone differences between “A” and “B” segments of stimuli.

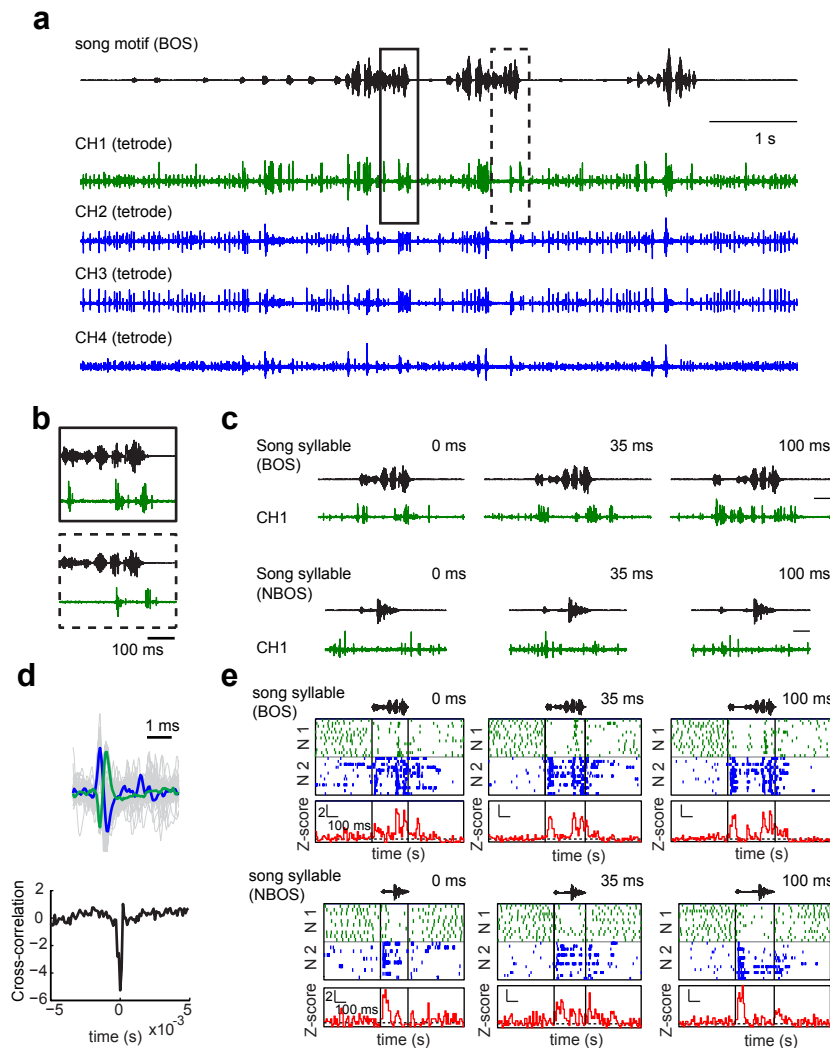


Figure 2.25: Supplementary Fig. S3. The auditory responses to song syllables with different length of inserted gaps in male HVC. (a) Tetrode recordings of HVC neurons in response to the playback of bird own song (BOS). (b) Neuronal activity in response to two syllables in a song motif. (c) Comparison of neuronal activity in response to BOS syllables and NBOS syllables with different lengths of inserted gaps (35 ms and 100 ms). Intact syllable (0 ms gap). (d) Two clusters of single unit were acquired by spike sorting. Cross-correlation between the spike times in these two clusters of single unit. Bin width is 1 ms. (e) Raster plots of two isolated neurons 1 and 2 in response to 13 repeated playbacks of BOS syllables and NBOS syllables with different lengths of inserted gaps. PSTHs showed the summed activity of these two units in response to 13 repetitions of the stimuli with inserted gaps (35 ms and 100 ms). 10 repetitions of the intact NBOS syllable (0 ms) were tested. Bin width is 10 ms. The dash horizontal lines indicate the Z-scores at value 0. The solid vertical lines indicate the onset and the offset of the stimulus.

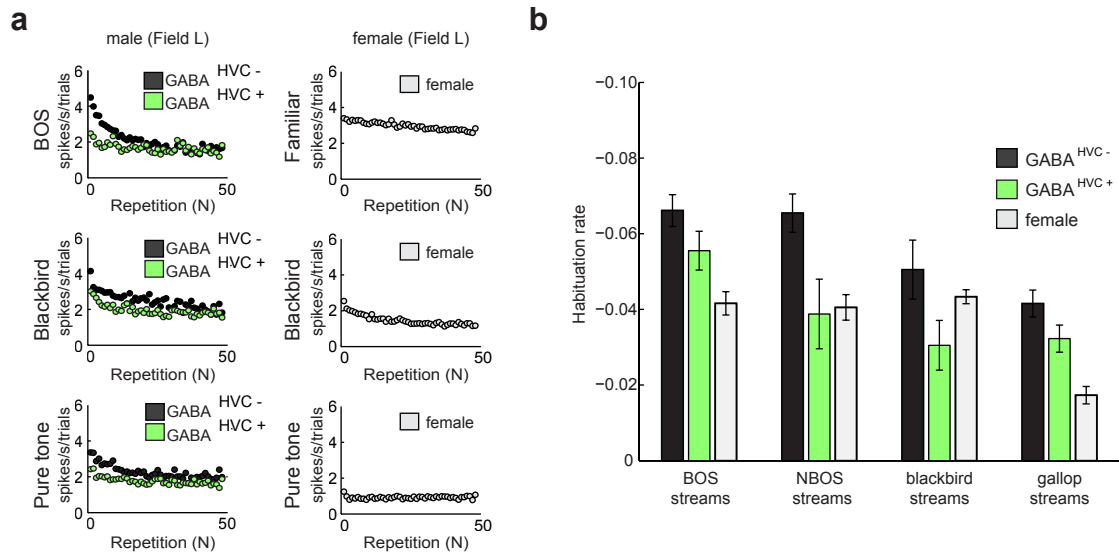


Figure 2.26: Supplementary Fig. S4. Inactivation of HVC decreased the habituation rates in response to stimulus trains. (a) GABA infusion in HVC affected the habituation rate of neuronal activity in male Field L in response to the stimulus trains. The base-line activity was not affected by the treatment of GABA infusion.  $GABA^{HVC+}$  indicates the GABA treatment in HVC as contrast to  $GABA^{HVC-}$  without GABA treatment in HVC. (b) GABA infusion in HVC influenced the habituation rates of male Field L neurons in response to the stimulus trains.

# Chapter 3

## Discussion

Vocal signaling is a form of animal communication: Senders emit sounds, either single or sequences of sounds, with particular spectro-temporal structures. Information might be coded in the sound structure per se or the sequential organization of the sounds and silent intervals or both. Claude Shannon has linked two most important concepts in communication: the information and the structural variance of signals [195]. The vocal signal is detectable by the receiver in the communication system, because the vocal signal is deviant from acoustic background scenes, and the receivers are able to identify the vocal deviances [218]. In addition to information coding of sound structures, this thesis studies information coding in inter-sound intervals, since many animal species communicate with acoustically invariant sounds [29].

In the early 20th century, Paul Lévy found out that the probability distribution of interval times in a random walk has infinite moments [197]. Such random walk trajectories are described by a power-law distribution. Many fields, including Chemistry, Physics and Biology, developed tools based on Lévy flights to study system dynamics. Most notably, contemporary network science has grown rapidly, since Barabási and Albert reported the invariant scaling property in the World Wide Web with a generic mechanistic support [9]. They found out that many network properties depend on the exponent of the power-law function [10].

Power-law as an analytic tool is used to measure complex behaviors by a one dimensional parameter, the exponent. The exponent is the scale to compare the similarity and difference between systems and compare the changes of a complex behavior under different circumstances. As mentioned in the introduction, power-law is a large-scale measurement of complex systems. It is only valid for the values larger than a lower bound ( $\tau_{min}$ ), i.e. power-law cannot describe the dynamics of calling events that occur shorter than the lower bound.

Power-law scales confer a strong predictive power, only if the candidate behavior is proven by statistical supports. However, despite the attractive scaling features of power-

law, it is remarkable that neglected statistical tests have led to questionable interpretations [206]. As stated in the introduction, a burst-like pattern or a straight line appearing on the log-log plot are hints but not a proof of a power-law function [153]. In addition, noise and incomplete data can affect the examination of power-law [207, 206]. E.g., a previous study found out that a power scale of “2” corresponds to the most optimal foraging pattern in many animal species [224]. However, the conclusion is problematic, since the data do not follow power-law [48]. Very few studies of power-law behaviors are actually supported by statistical supports and mechanistic explanations [206]. The most noticeable “good” examples of power-law functions are the allometric scaling for plants and animals [230], Yule-Simon scaling for speciation [245] and power-law scaling of the World Wide Web networks [9].

Examination of statistical support for power-law support is not a nontrivial task. In order to prove power-law behavior of call-based vocal communications we need adequate data. To achieve this, I used on-bird microphone transmitters that permitted a precise temporal resolution of vocal timings. With this, I recorded vocal activity continuously during the day and avoided possible data manipulation such as data segmenting. Then, I applied the maximum likelihood methods to compare the power-law model and the candidate data. At last, I probed the exponents of power-law system by using biological meaningful perturbations to exclude experimental noise, such as incidental “reactive” calls. We concluded that call-based vocal communication follow power-law with the exponent ranges between 2 and 3.

In the following discussions, I focused first (3.1) on the information coding in animal communication systems, and how the power-law dynamics is related to information coding in vocal communication systems. Then I summarized (3.2) the study of sensorimotor expectation and how the male HVC neurons detected expected and unexpected acoustic-invariant female calls. At last, I discussed (3.3) the top-down process of auditory decoding of acoustic information in male and female primary auditory forebrains.

## 3.1 Power-law scales of information in vocal communication system

### 3.1.1 Vocal communication system in birds and humans

Animal communication is a process, by which animals provide information to others. There are many ways to provide information in animal communication systems such as making sounds, postures or releasing chemical trails. Both, animals and humans, share a common trait in vocal communications, in which information is exchanged based on shared experiences [46] and self-similar patterns [141].

Human language is one particular form of vocal communications. The usage and the comprehension of languages deeply depend on shared experiences, learned through social interactions [46]. Further, statistical analysis of human language showed some basic pattern that might encode information: the frequency of words in human languages is consistent with the inverse square power-law distribution. This is the so-called Zipf's law for human languages, describing that all languages have a self-similar pattern. Letters, phonemes, words or spoken English, German, Chinese follow the same power-law principle and can be predicted by an exponent approximated by 1 [249]. A power-law pattern is also called a scale-free pattern, because that pattern is the same on whatever scale (phonemes, letters, words, sentences, paragraphs). The power-law feature allows for optimized usages of words in human communications and prevents repetitive or random usage of words in messages. For example, a message is stereotyped by repetitive usages of words, whereas a message is diverse by random usage of words. Thus, the balance between these two extremes (predictability and flexibility) in a power-law system allows for a high capacity of information transfer in vocal communications. Moreover, the change of the exponent of Zipf's law during language acquisition has been observed in humans [141]. Human babies undergo a babbling phase. The sounds produced by babies are likely random [142] and the frequencies of babbling sounds are distributed equally. By the time human babies have learned new languages, the distribution of sound repertoires, and the organizational patterns of sound sequences converges toward the characteristics of Zipf's law [142].

Birdsongs, as analogy to human language, share the same properties as mentioned above. The frequencies of song syllables are not equally distributed, since some syllables are used more often than others. For example, Indigo and Lazuli buntings have over 120 song syllable types. There is a strong bias towards certain types of syllables in the buntings' songs [6]. In zebra finches, the juvenile birds produce first sub-song at about 30 days. The acoustic structures of these sub-songs are noisy and distinct from adult songs. The zebra finches develop stereotyped song motifs between 70 – 90 days of age [104]. As showed in figure 3.1, the frequency of song syllables in adult zebra finches decreases in relationship to their syllable ranks e.g. from the most frequent syllable (rank 1) to the less frequent syllable (rank 9).

Next to language, birdsongs are considered as analogy to human music, since some songbird species sing in harmonic series [45]. Similar to human language, human music consists of sequential notes organized in temporal sequences, which links to the rhythm and tempo. In particular, the rhythm reflects the quality of musical performance and plays an essential role in emotional expressions. The birdsongs with sequential combinations of notes agrees with the property of human music. However, the inter-note intervals of birdsongs appear to be less variable than human music. The inter-syllable intervals in zebra finch songs followed an exponential distribution [5, 159], indicating that the interval times of birdsongs are constraint around an expected mean value and varied very little. By contrast, the interval times between the notes of human music are more variable. E.g., I examined the interval times of a musical piece "Chair de lune" composed by Claude De-



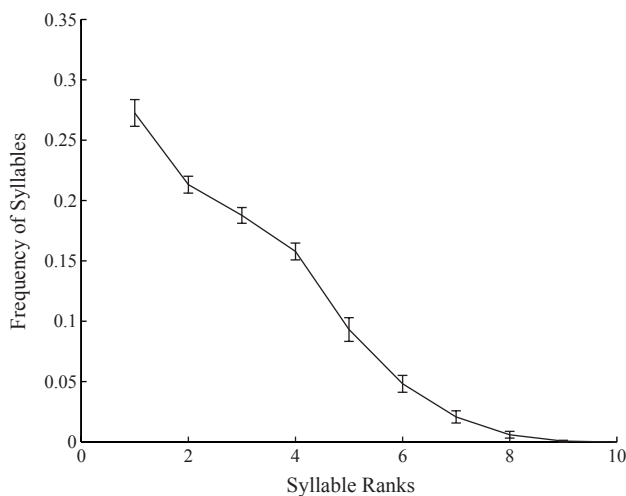


Figure 3.1: The relationship between the frequency of syllables and the syllable ranks in zebra finches.  $N = 30$  songs. Error bar indicates the standard error of the mean. These 30 song sequences were collected from the male zebra finches in Manuscript 2.1, 2.2 and 2.3.

bussy (played by Caela Harrison [43]). The exponent of this musical piece is approximated by 2.5. Since this is within the exponent range of power-law scales of 2 and 3, it may indicate a scale-free pattern in human music.

### 3.1.2 Measurements of information in vocal communications

#### Measurement of information

Despite the similarities and differences of vocal communications in animals and humans, an important question is, whether there is a fundamental trait in all forms of vocal communications. As mentioned above, the variation of notes and times plays an important role in human language, music and birdsong. Claude Shannon stated that the measure of information is the measure of possibility [195]. In Shannon’s theory of communication [195], logarithmic measurements provide a quantitative scale for various communicational channels, where  $H(x)$  is the information entropy of  $x$  and  $p(x)$  is the probability of  $x$ :

$$H(x) = -\log_2 p(x)$$

. For example, English texts have 27 symbols (26 letters and 1 space). If the occurrences of each letter are equally possible, each letter has a quantity of information ( $H_0$ ):  $-\log_2 \frac{1}{27} = 4.76$ . However, the frequencies of English letters are unequal in a real text, in which letter “e” occurs more often than letter “z”. Hence, the average amount of information ( $H_1$ ) in English text is 4.03 according to [196], according to formula:

$$H_i = -\sum_{i=1}^n p_i \log_2 p_i$$

, where  $p_i$  is the probability of letter  $i$  in the text. This equation is used to quantify how many information is contained in an English text and in any other communication system [196, 141]. For example in birds, the vocal repertoire of zebra finches comprises songs and calls. Songs in zebra finches are male specific and contain 5 – 15 distinct song syllables, whereas male and female share a call repertoire of approximate 10 types of monosyllabic sounds [237, 54]. If the occurrences of each syllable in zebra finches are equally possible, zebra finches' calls convey the maximum quantity of information:  $H_{calls} = -\log_2 \frac{1}{10} = 3.32$ , whereas the zebra finches' songs convey maximum quantity of information between 2.32 and 3.91. However, the frequencies of vocal types in animal communication system are not equal (Figure 3.1). If the calls are independent and the entropy depends on the frequencies of the calls, the first-order approximation:  $H_1$  is approximated by  $0.79 \pm 0.09$  ( $n = 25$ ). This indicates a large bias toward a certain type of call(s) over others in zebra finches.

### Measurements of informational degree of organization

The “order approximations” provide quantitative data of how much information is contained in a sequence of animal vocalizations [196, 141]. Figure 3.2 shows a song sequence of a male zebra finch that contains 8 syllable types and appears to be arranged in an organized order. The zero-order ( $H_0$ ) approximation (estimated by choosing syllables independently with equal probability) and the first-order ( $H_1$ ) approximation (estimated by choosing syllables independently with respect to the frequencies of syllables) do not seem to be able to capture the natural sequence of bird song in zebra finches. In general, a series of entropic approximations  $H_1$  (monogram),  $H_2$  (digram),  $H_3$  (trigram),  $H_4$  (tetragram),  $H_5$  (pentagram) are used to measure the organizational complexity of animal vocalizations [141]. The  $n$ -order of entropy is given by [196]:

$$H_N = - \sum_{i,j} p(b_i, j) \log_2 p_{b_i}(j)$$

,  $H_N$  is the conditional entropy.  $H_N$  is obtained by subtracting entropy of  $H_{N-1}$  from the joint entropy between  $b_i$  and  $j$ , where  $j$  is an arbitrary syllable following  $b_i$ . According to Channon's paper [195, 196],  $H_N$  is associated with the probability of  $N$  when previous  $N - 1$  syllabic orders are known. For example, there are 7 syllables following “a b c” in the song depicted in figure 3.2 These 7 syllables are all “a”. In other words, it is with 100% certainty to predict next syllable, if a sequence of three preceding syllables is known. In this case, the information entropy of  $H_4$  is 0.

The information entropy of  $N$  entropic orders showed that birds' songs had a high regression slope as compared to birds' calls in zebra finches (Fig. 3.2.B). Decreasing information entropy in increasing entropic orders indicates that the song syllable is more predictable when a longer preceding sequence is known. By contrast, the sequence of bird calls is less predictable than bird songs in zebra finch. By giving 4 zebra finch' calls in a sequence, it is still hard to predict, which call type will appear next. In other words, the predictability by giving 4 calls ( $H_5$ ) does not differ much from the predictability by independent estimation

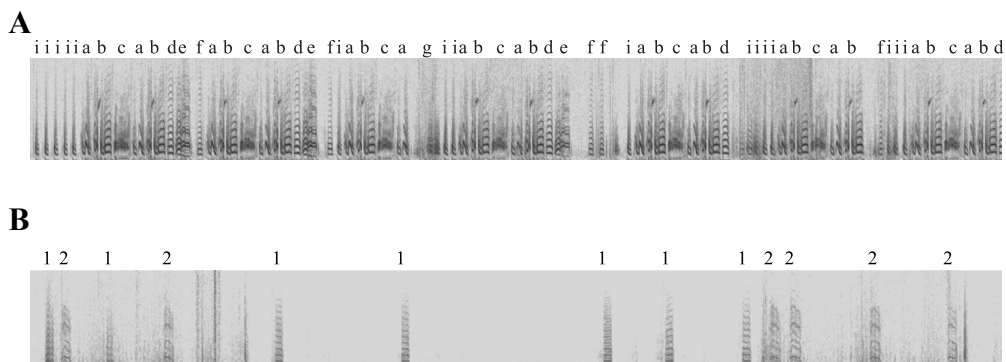


Figure 3.2: Song and call sequences of a male zebra finch. (A) The song syllables are labeled with letters. “i i i i a b . . .” represents the sequence of the song, “i” being the introductory syllables, “a – f” being song syllables. (B) The four call syllables are labeled with numbers “1” to “4”. “1 2 1 . . .” represents the sequence of two calls (calls 3 and 4 are not shown).

of the frequency of each call type ( $H_1$ ). The low predictability of zebra finch calls is not caused by small number of call types. E.g., DNA sequences that consist of 4 nucleotides (C, G, A and T) contain biological information to encode proteins. By increasing the entropy order, the information entropy of DNA sequences of T7 phage and HSV (herpes simplex virus) decreased more dramatically than the information entropy of zebra finch’ calls (Fig. 3.3. C). The frequencies of each nucleotide in both T7 and HSV are approximated by 25%, which corresponds to 2 bits of  $H_1$  information. By knowing a sequence of 10 nucleotides (AATACGACTC), it is 100% predictable that the next nucleotide will be “A” in T7 phage. The measure of the informational degree of organization confirms that the sequence of zebra finch’ calls is hardly predictable and shows very little organizational rule.

### 3.1.3 Increase in information content by varying inter-call intervals

Biological communication systems evolved by increasing the capacity of information transfer [168]. To achieve this with a finite number of phonemes (or syllables), one solution is to combine phonemes into words. Such word formations and sequential combinations are widely observed in human languages, birdsongs and whistles of marine mammals [46, 168, 140].

As above mentioned, word formation extends the capacity of information beyond the limit of single use of monosyllabic sounds. However, suppose there is a set of identical monosyllabic sounds in a sequence or a musical piece is composed by a monotonic drum. In this case, the maximum information is 1 bit, because there are only two notes (i.e. a monotonic sound and a pause) that comprise the message. That is a common problem to

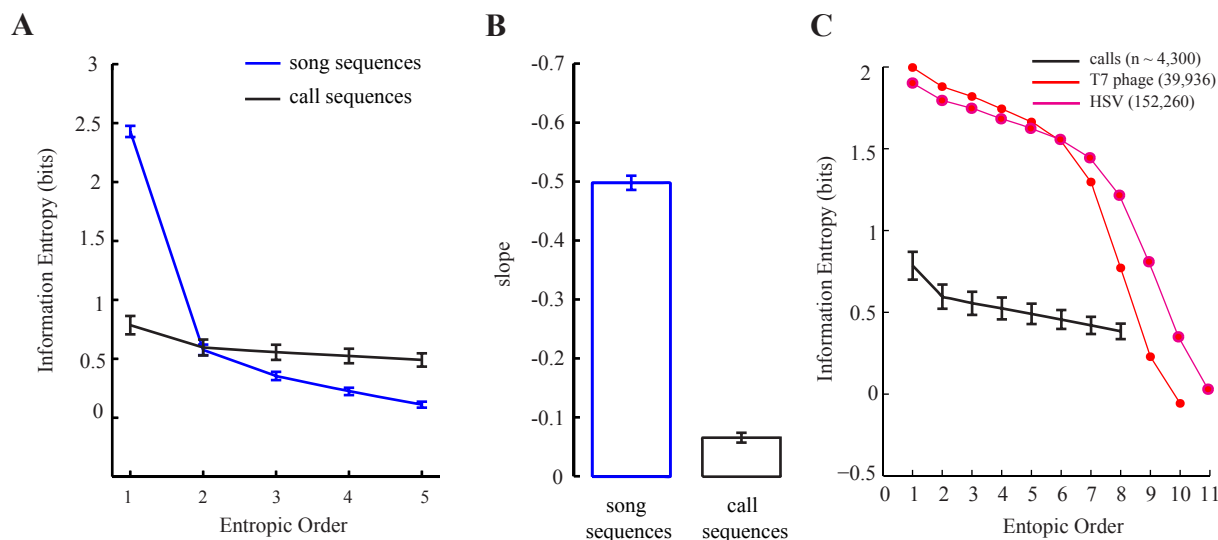


Figure 3.3: The information entropy of call and song sequences with respect to different entropic orders. (A) Comparing of the information entropy with respect to different entropic orders between call and song sequences. (B) Comparing the slopes of the information entropy between call and song sequences.  $n_{\text{song sequences}} = 30$ ,  $n_{\text{call sequences}} = 25$ . Error bar indicates the standard error of the mean. (C) Comparing the slopes of the information entropy between call and DNA sequences of T7 Phage and HSV. T7 and HSV sequences consist of 39936 and 152260 nucleotides, respectively. The sequences of nucleotides of T7 phage and HSV are available online at: <https://www.ncbi.nlm.nih.gov>. The average number of calls of the 25 recordings was 4300.

the communication system of many animal species without or with very little phonemic variation.

The information capacity is defined as repertoire size of signals ( $S$ ) to the power of signal rate ( $r$ ) [195]. Animals, such as birds, can either increase the song syllable size and repetition rate to extend the capacity of information. But the rate of sound in animal communication systems is always limited by the length of syllables. Animals frequently use combinations of less distinctive sounds to convey, possibly, large amounts of information [168]. However, the information coding by an increase of word length is likely limited by the capacity of memory in animal communication systems [168]. Some animal species, such as the zebra finches, produce large numbers of almost identical calls (“stack” calls) throughout a day. Even though the call (= signal) rate ( $r$ ) can reach up to 10 Hz (i.e. less than 100 ms of syllable length), since the call repertoire (= signal) size ( $S$ ) of “stack” calls is 1, it is impossible to increase the information capacity by using identical calls in regard to  $S^r$ . One could argue that increasing the amplitude levels would increase the size of signals ( $S$ ) of a call. But the receiver would have difficulties to decode the signals of a call with different amplitude levels due to the environmental noise and the decoding

system of the receivers. Thus, the acoustic signature of calls is the reliable solution to increase the information capacity of calling. In Manuscript 2.1, we introduce another type of information coding, namely the variation of interval times, by which animals adjust the tempo of acoustic-invariant vocalizations to provide information to others, even though single “stack” call has no predetermined meaning. Information seems encoded in varying the inter-call intervals, which is reflected in the exponent of the power-law function. Adjusting the exponent of the callings may extend the information capacity limited by invariant calls such as the “stack” calls, and human music that is composed by monotonic instruments. The similarity between the scale-free property of zebra finch calling and human music [97] might indicate that zebra finch also use rhythm and tempo to convey information.

### 3.1.4 Power-law scaling of vocal activity and Barabasi’s model for choice of actions

The queuing model based on human decision making [7] is so far the best analytic model to predict bursty nature of human behaviors. That is: people have a list of tasks with different priorities in their minds. The unevenness of timing to execute those tasks is caused by attributed priorities to different tasks. Although identifying the priorities of the tasks on an imaginary list in one’s mind seems to be impossible experimentally, the attractive feature of Barabasi’s view on human dynamics brings the study of conscious choice of action in cognitive system and unconscious appraisal system in line with the studies of behavioral dynamics, which may indicate two major forms: an explicit choice of action and an implicit system that limits the options from which to choose [105].

Indeed, zebra finches have heterogeneous behaviors including vocal and non-vocal gestures that could block each other’s way. As being vividly described by Richard Zann, an Australian ornithologist: *“The male leads the female in search of a suitable nest site he gives Stack Calls much of the time. As he lands at a potential site, these switch to Kackle Calls, and if the female shows interest he hops to and from her to the site giving Ark Calls each time he lands on the site itself. If she approaches he performs the Head-down Tail-fan display in which he bows down and fans the tail and mandibulates at the female while slowly pivoting the body from side to side; Ark Calls are emitted continuously. At an old nest he may pause in the entrance and fan his tail at the female, his white rump highlighted like a beacon. ... If she is interested in a site she may approach in a Head Tail Twist, ...he will Nest Whine and start nibbling or pushing any nest material nearby. Male often give Undirected Songs....”* [246].

Clearly, “Stack” calling is not the only behavior during social interactions. The heterogeneity of animal behaviors leads to a question: whether the definitions of self-contained and reactive calls are solid, because some of reactive calls can be accidental since they may follow other vocal or non-vocal gestures. In this thesis (Manuscript 2.1), we could

not exclude the possibility that some reactive callings are by chance. Such incidental “reactive” callings can be considered as an experimental noise, as any other experimental noise. Nevertheless, the perturbation experiments suggest that the incidental noise is not the main factor in this system. The interval statistics of “reactive” callings was socially and physiologically dependent. One possible solution to this question is to perform video recordings that permit precise description all observable vocal and non-vocal gestures in succession. If there is any dependency between the decision making and information provided by self- and/or other-generated actions, we can measure the informational degree of organization as described above to get a better understanding of whether a decision is made based on preceding actions. More precisely, since vocal sequences and the interval times might contain information and can be measured by power-law function, one potential analysis would be to identify, if certain transitions between syllables are imperative. For example, in figure 3.2, it is with 100% certainty to predict that syllable “a” follows “a b c”. The transition time between “a” and “a b c” should be less variable. Although all inter-syllable intervals in zebra finch songs are distributed exponentially, it is worth testing the information content of song intervals in other animal species and to test the successions of heterogeneous behaviors.

In addition, an imperative transition between syllables seems to have significant neurophysiological meanings in song systems. Alison Doupe found that the LMAN (a song system region) neurons responded more strongly to the imperative transition between “a b c d” and “e f” than the non-imperative transitions between “a b c” and “d” or between “e” and “f” (Doupe, 1997). Identifying the degrees of organizational rules may help us to understand the neurophysiological mechanisms underlying sequential behaviors.

## 3.2 Sensorimotor expectation in call-based vocal communications

In this thesis, the term “expectation” refers to a stance of behavioral or mental properties. Accordingly, anticipative activity refers to readiness-potential, since this thesis focuses on the neuronal activity of male HVC in processing and anticipating the auditory events of female calls during call-based vocal interactions. In addition to the call-based vocal interactions, duet singing is a cooperative behavior in two-way interaction of birds. Fortune et. al. have observed that plain-tailed wrens were able to catch up the pace of their partners after missing one or two notes occasionally during duet singing [60]. Their observation strongly suggests that there must be a cue for the birds to anticipate upcoming notes during duet singing. In zebra finches, we observed overlapping of singing in a social group, but never observed duet singing. One possible explanation is that the auditory mode of HVC neurons may be gated off during active singing [188]. However, the auditory mode of HVC neurons may still be active in a context-based call interaction, when the

self-contained and the reactive vocal events can be very flexible in times (Manuscript 2.1). For the rapid vocal responses that are less than 2 seconds in zebra finches, there may exist a time window to allow an expectation for an operant response.

There is a match between the 2 seconds time windows of call-based vocal communications and the lower bound of power-law dynamics ( $\tau_{min} \sim 2$  seconds, Table 2.1.) in zebra finches. The time scale of vocal exchanges between zebra finches' pairs is largely less than 2 seconds. As introduced before (Introduction 1.5.3.), mechanisms related to 'experience' and to 'expectation' are particularly important for achieving an efficient vocal communication within this time frame. The experience dependent call-based vocal communication is consistent with the idea that zebra finches' calls are positive reinforcers [99]. With this, a pattern of calling interactions would emerge, if the calls are rewarded by the mate's calls. It is, however, very difficult to record single units long enough to track the emergence of reinforced anticipative activity during pair bonding. An implantable multi-channel electrode array may help to answer this question.

The neurophysiology of expectation in call-based vocal interactions was studied in this thesis. In a non-social context, the male sensorimotor nucleus (HVC) displayed neuronal activity related to own calling and to the hearing of play-backed female calls. In a social context, male HVC neurons not only showed premotor activity to own callings, but also showed predictive discharge to upcoming female calls. Although the auditory mode of HVC neurons is gated off during active singing [188], our results suggest that the auditory mode of HVC neurons is active during call-based vocal communications in a social context. Instead of pure sensation or motor control, this thesis showed that HVC plays a crucial role in linking sensory information and vocal performance during vocal interactions. HVC may not only guide vocal performance via sensory feedback, but via anticipative activity to time of vocal output during social interactions. However, although our study suggests that the auditory mode of HVC neurons may be gated-on during social interactions, the underlying mechanism is not known. Whether the gating is driven by neuronal interactions within the HVC or is driven by afferent areas of HVC, needs to be seen.

Further, since female HVC is very small, we were unsuccessful to implant electrodes in female HVC. We have observed that female zebra finches could respond as fast as male zebra finches, but could not adjust the response time after relocating female birds in a separated but acoustic interconnected sound box, which only allowed sound-based communications between male and female. The results suggest that vocal responses in female zebra finches rely on the sensory experience (inverse model, see Introduction 1.6.), but does not rely on the feed-forward control. Nonetheless, it is not known whether female zebra finches have feed-forward control to regulate call timings during vocal interactions. Thus, it would be desirable to record HVC neurons of females in future experiments.

In Manuscript 2.2, I introduced evidence of the sensorimotor expectation in HVC for call-based vocal communications in a social context. Because neurophysiological record-

ings in freely behaving birds limit the precision to identify cellular mechanisms underlying expectation, further studies should design an experimental paradigm specifically for this question. E.g. our study suggests that the auditory mode of HVC neurons may be gated-on during social interactions. But whether the gating is driven by the cellular interactions within the HVC or it is driven by external innervations, needs to be seen. Further, our results suggest that the anticipative activity prior to the “operant” stimulus (F2, see Introduction 1.6., i.e. the female calls) requires a cued stimulus (F1), it is not known how the link between F1 and F2 is established in HVC. Further questions are: what are the F1s? Are they self-generated or the mates’ vocalizations? Does F1 also include non-vocal gesture? If the operant action is the call (saying “stack” calls), zebra finches should be trained to elicit “stack” calls in response to the cue stimulus (F1). Under this experimental paradigm, we could manipulate different parameters that have been shown important to evoke the anticipative activity in humans [226, 113], such as the interval times between F1 and F2, changes of cue stimulus to F1’ after conditioning between F1 and F2 has been established and so on. With the setup of experimental paradigms, we could identify how neurons make link between F1 and F2, how neurons integrate the information from F1 and maintain before the release of F2. Once the neuronal property has been determined, the next step might be to eliminate the anticipative activity by using optogenetic techniques. Finally, if the anticipate activity is important to the pair forming in zebra finches or other cooperative behaviors such as dance and duet, we could reinforce or perturb such behaviors by neuronal manipulations in behaving animals.

### 3.3 Acoustic representation in the auditory forebrains

In previous sections, I discussed if and how the song structure and inter-call intervals encode information. Transmission of information from senders to receivers is the key in animal communications. Even if we measure the information quantity of the emitted signals, we still need to understand how the signals are represented in the receiving animal. Such neural decoding processes in animals involve both bottom-up and top-down processes [33]. The bottom-up process is driven by sensory/perceptual filters to select allocated information [108], whereas the top-down process is driven by memory/experience-filters, goal setting and attentional modulation [33].

In avian auditory systems, the higher level of bottom-up and top-down processes emerges in the primary auditory forebrain (field L complex) [212, 24]. Field L consists of L, L2a, L2b, L1 and L3 sub-areas [133]. These sub-areas exhibited different response profiles to auditory stimulus, in which L2 showed more robust and unselective to auditory stimuli, whereas other subareas such as L1 and L3 responded more selectively [122, 15, 3]. Therefore, if L1 and L3 receiving auditory inputs from L2 is the bottom-up process, the emerging of auditory selectivity in L1 and L3 can be the result of hierarchical processes via sensorimotor and secondary auditory systems. Since L1 and L3 are closely adjacent to



secondary auditory areas and connect to HVC [133], the selectivity of L1 and L3 neurons may be: (1) directly mediated by secondary auditory forebrain; (2) directly mediated by HVC; (3) indirectly mediated by HVC through motor-to-auditory pathway via secondary auditory forebrain: CMM [181]; least of all, (4) mediated by the intrinsic properties underlying sex difference in male and female Field L.

In zebra finches, females don't sing but only produce calls. Nevertheless, female zebra finches showed auditory preference for familiar songs over unfamiliar conspecific songs [179]. Zenk-expression and electrophysiological studies confirmed that female have well-developed primary and secondary auditory forebrains for detecting the variations of songs and calls [212, 242, 15, 16]. Females lack a differentiated HVC [82] and thus likely lack a HVC to Field L projection. The experiments in Manuscript 2.3 were originally designed to test the hypotheses (1 – 3) as above mentioned. First, the results in Manuscript 2.3 showed that the activity of male Field L neurons attenuated faster than female Field L neurons in response to stimulus trains, suggesting that the selectivity of Field L neurons is not mediated by secondary auditory forebrain. If it were, the auditory representations of stimulus trains in male and female Field L neurons would be the same.

Unfortunately, I could not demonstrate whether HVC was directly or indirectly via secondary auditory forebrain (CMM) responsible for the stimulus-dependent response properties of field L of males and females. I infused a solution with 250 mM GABA + 5% dextran into male HVC. I found decreases of habituation rates in male Field L after the infusion of GABA (Manuscript 2.3, Figure 2.26). However, the statistical test did not provide sufficient support for the impact of the pre-motor activity of HVC on the sensory properties of field L.

Nevertheless, Manuscript 2.3 may provide a valuable framework for investigating how a train of continuous sounds is encoded in the auditory forebrain and how sensorimotor feedback may influence the auditory process of natural sounds. In this manuscript, the auditory scene analysis was used for investigating the neuronal representations of complex acoustic scenes in the auditory forebrain. Although the auditory scene analysis is an ideal paradigm to achieve an optimized signal to noise ratio as a decoding scheme, the limitation inherent in the auditory scene analysis are due to the synthesized unnatural sounds [31]. In fact, different methodological constraints more or less limit comprehensive understanding of neural correlations underlying neuronal representation of sounds from natural sources [22]. Like our own experiment, previous studies attended to decompose a natural acoustic scene into small components, then investigated the neural correlations of these components that contains a part of acoustic information in a natural scene. In fact, a natural scene likely provides a mixture of messages. Considering that single sounds [53] and sequences of sounds [121, 2] comprise messages, the information of these components is quantifiable according to Shannon's theory. In addition, studies have used sound masking to investigate the limit (capacity) of auditory system to decode signals in various species [82, 155, 47]. With this, an information map of natural sounds can be reconstructed by combining the

information contents of these acoustic components.

By studying the sexual differences of auditory responses to continuous sound trains in the primary auditory forebrain, we found that both male and female Field L encoded similarly to single sounds. If the sounds were played continuously as trains, the neuronal responses to the stimuli attenuated faster in male Field L than in female Field L. Electrophysiological and corticographic studies in human patients and macaque monkeys (*Macaca fascicularis*) showed that the high habituation rate was associated with the unattended stream and the non-best frequency [59, 144]. Accordingly, the higher habituation rate in male than in female zebra finches may reflect the maintenance of auditory sensitivity to conspecific song that is higher in female than in male zebra finches. In other words, it could be that female birds pay more attention to conspecific songs than male birds do. The sexual difference of auditory representation may be co-evolutionarily related to the vocal assortments in the social living zebra finches. Future behavioral studies may focus on how female zebra finches select for male partners based on their song features on a group level to investigate the regulation of social feedback to shape the development of male song. Future neurophysiological studies may focus on the sexual difference of anatomical and physiological mechanisms underlying their sexual difference in the auditory system.



# Chapter 4

## Summary

Information coding and decoding is essential to animal communication. Increasing the repertoire size and the sequencing of syllables facilitate the information coding, but this information coding is limited by the biological constraints such as the energetic cost [154] and the capacity of memory [115, 83]. I propose that information coding by adjusting the inter-call interval durations may extend the information capacity in animal communication systems. The information coding of interval durations might play a particularly important role for the communication of those animal species that produce large amounts of acoustically invariant calls throughout a day.

The deviance of sounds in a sequence contains information. It is important for the animals to capture attention by unexpected events. This phenomenon is known as “mismatch negativity” in human [148] and deviance detection in birds [16]. Similar, the deviance of the inter-call interval durations might contain information, in particular, if the calls per se are invariant. Since female zebra finches produce mostly acoustic-invariant calls, the self-similarity (scale-free) feature of call timings may facilitate the prediction of their partners to minimize surprise or free energy [65] during call-based vocal communications.

Further, I showed that male HVC exhibited anticipative activity to upcoming female calls, even though the female calls are acoustically invariant. Future studies should identify the neurophysiological mechanism underlying the prediction error (surprise) of hearing female calls that deviate from recent statistics. The measurement of information is the measurement of probability, the surprise [195, 216, 65]. Since this study provides evidence for the information coding of interval times, future study may design an experimental paradigm to investigate how the information is gained and represented in the brain by measuring the differences between the probability of the interval times for expected female calls and the probability of the actual interval times for hearing female calls.

Furthermore, auditory coding in the brain is essential to the auditory representation of information coming from the senders during communications. The top-down process that is driven by memory/experience-filters [33] can affect the information coding of au-

ditory representation. I compared the auditory representation of stimulus trains between male and female primary auditory forebrain (Field L). The results showed that male and female field L did not differ in responding to single stimulus, but showed different auditory activity in response to stimulus trains. However, I could not demonstrate that HVC was responsible for the stimulus-dependent response properties of field L of males and females.

Several limitations in this thesis affect the generalization of my findings to the vocal interactions of other songbird species. I used zebra finch' pairs as animal model. Zebra finches are monogamous birds that pair for life while pairs are transient in many other songbird species. The application of the methods described here to study the vocal activity of different species should inform, whether call usage in such species (e.g. those that bond only briefly) is similar to zebra finches. It is known that some nonhuman primates and marine mammals combine different types of calls in a sequence to encode information. Thus, information coding based on inter-call intervals might be a general feature for animals that produce large numbers of call that are acoustically invariant.

# Appendix A

## List of abbreviations and symbols

Table A.1: List of abbreviations

Amb	Nucleus ambiguus
Area X	Area X in the basal ganglia
BOS	Bird own songs
Cb	Cerebellum
CLM	Caudal lateral mesopallium
CMM	Caudal medial mesopallium
CDF	Cumulative distribution function
DLM	Nucleus dorsolateralis anterior
DM	Dorsomedial nucleus of the intercollicular complex
eCDF	Empirical cumulative distribution function
F	Female
HP	Hippocampus
HVC	Higher vocal center
LaM	Lamina mesopallialis
lMAN	Lateral magnocellular nucleus of the anterior nidopallium
M	Male
min	Minute
mMAN	Medial magnocellular nucleus of the anterior nidopallium
ms	Milliseconds
NBOS	Not bird own songs
NCM	Caudal medial nidopallium
NIF	Nucleus interface of the nidopallium
Nr.	Number of
nXIIts	Tracheosyringeal portion of the nucleus hypoglossus
PAG	Caudal periaqueductal gray
PDF	Probability density function

Table A.1: List of abbreviations

RA	robust nucleus of the archistriatum
RAm	Retroambigualis
rVRG	Rostro-ventral respiratory group
s	Seconds
V	Ventricle
wr	Water-removal

Table A.2: List of symbols

$\alpha$	Exponent
$\tau$	Inter-event intervals
$\gamma$	Rate
$\sigma$	Standard deviation
$\infty$	Infinite
$\Delta$	Change of quantity

# Bibliography

- [1] E. Akutagawa and M. Konishi. New brain pathways found in the vocal control system of a songbird. *J Comp Neurol*, 518(15):3086–100, 2010.
- [2] N. Amin, A. Doupe, and F. E. Theunissen. Development of selectivity for natural sounds in the songbird auditory forebrain. *J Neurophysiol*, 97(5):3517–31, 2007.
- [3] M. Araki, M. M. Bandi, and Y. Yazaki-Sugiyama. Mind the gap: Neural coding of species identity in birdsong prosody. *Science*, 354(6317):1282–1287, 2016.
- [4] D. Aronov, A. S. Andalman, and M. S. Fee. A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science*, 320(5876):630–4, 2008.
- [5] D. Aronov, L. Veit, J. H. Goldberg, and M. S. Fee. Two distinct modes of forebrain circuit dynamics underlie temporal patterning in the vocalizations of young songbirds. *The Journal of neuroscience*, 31(45):16353–68, 2011.
- [6] M. C. Baker and J. T. Boylan. A catalog of song syllables of indigo and lazuli buntings. *Condor*, 97(4):1028–1040, 1995.
- [7] AL. Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–11, 2005.
- [8] AL. Barabasi. *The scale-free property*, book section 4. 2012.
- [9] AL. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–12, 1999.
- [10] AL. Barabási. *Network Science*. Cambridge University Press, United Kingdom, 2016.
- [11] B. Barzel and AL. Barabasi. Universality in network dynamics. *Nat Phys*, 9(10):673–681, 2013.
- [12] A. H. Bass, E. H. Gilland, and R. Baker. Evolutionary origins for social vocalization in a vertebrate hindbrain-spinal compartment. *Science*, 321(5887):417–21, 2008.
- [13] E. E. Bauer, M. J. Coleman, T. F. Roberts, A. Roy, J. F. Prather, and R. Mooney. A synaptic basis for auditory-vocal integration in the songbird. *The Journal of neuroscience*, 28(6):1509–22, 2008.



- [14] P. M. Bays and D. M. Wolpert. Computational principles of sensorimotor control that minimize uncertainty and variability. *J Physiol*, 578(Pt 2):387–96, 2007.
- [15] G. J. L. Beckers and M. Gahr. Neural processing of short-term recurrence in songbird vocal communication. *PLoS One*, 5(6):e11129, 2010.
- [16] G. J. L. Beckers and M. Gahr. Large-scale synchronized activity during vocal deviance detection in the zebra finch auditory forebrain. *Journal of Neuroscience*, 32(31):10594–10608, 2012.
- [17] M. A. Bee and G. M. Klump. Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J Neurophysiol*, 92(2):1088–104, 2004.
- [18] M. A. Bee and G. M. Klump. Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. *Brain Behav Evol*, 66(3):197–214, 2005.
- [19] A. Bendixen. Predictability effects in auditory scene analysis: a review. *Front Neurosci*, 8:60, 2014.
- [20] J. I. Benichov, S. E. Benezra, D. Vallentin, E. Globerson, M. A. Long, and O. Tchernichovski. The forebrain song system mediates predictive call timing in female and male zebra finches. *Curr Biol*, 26(3):309–18, 2016.
- [21] A. J. Billig, M. H. Davis, J. M. Deeks, J. Monstrey, and R. P. Carlyon. Lexical influences on auditory streaming. *Curr Biol*, 23(16):1585–9, 2013.
- [22] J. K. Bizley and Y. E. Cohen. The what, where and how of auditory-object perception. *Nat Rev Neurosci*, 14(10):693–707, 2013.
- [23] D. A. Bodnar and A. H. Bass. Temporal coding of concurrent acoustic signals in auditory midbrain. *The Journal of neuroscience*, 17(19):7553–64, 1997.
- [24] J. J. Bolhuis and M. Gahr. Neural mechanisms of birdsong memory. *Nat Rev Neurosci*, 7(5):347–57, 2006.
- [25] J. J. Bolhuis, D. P. Van Mil, and B. B. Houx. Song learning with audiovisual compound stimuli in zebra finches. *Anim Behav*, 58(6):1285–1292, 1999.
- [26] D. Bonke, H. Scheich, and G. Langner. Responsiveness of units in the auditory neostriatum of the guinea fowl (*Numida meleagris*) to species-specific calls and synthetic stimuli .1. tonotopy and functional zones of field-1. *Journal of Comparative Physiology*, 132(3):243–255, 1979.
- [27] S. W. Bottjer and A. P. Arnold. Afferent neurons in the hypoglossal nerve of the zebra finch (*Poephila guttata*): localization with horseradish peroxidase. *J Comp Neurol*, 210(2):190–7, 1982.

- [28] S. W. Bottjer, K. A. Halsema, S. A. Brown, and E. A. Miesner. Axonal connections of a forebrain nucleus involved with vocal learning in zebra finches. *J Comp Neurol*, 279(2):312–26, 1989.
- [29] J. W. Bradbury and S. L. Vehrencamp. Principles of animal communication. 1998.
- [30] M. S. Brainard and A. J. Doupe. Auditory feedback in learning and maintenance of vocal behaviour. *Nat Rev Neurosci*, 1(1):31–40, 2000.
- [31] A. S. Bregman. *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.
- [32] E. A. Brenowitz. Altered perception of species-specific song by female birds after lesions of a forebrain nucleus. *Science*, 251(4991):303–5, 1991.
- [33] L. B. Buck and C. I. Bargmann. *Preception*, book section 5, pages 449 – 744. The McGraw-Hill Companies, Inc., New York, 5th ed. edition, 2015.
- [34] L. Cassotta, S. Feldstein, and J. Jaffe. The stability and modifiability of individual vocal characteristics in stress and nonstress interviews. *Research Bulletin*, (2), 1967.
- [35] I. Charrier, N. Mathevon, and P. Jouventin. Mother’s voice recognition by seal pups. *Nature*, 412(6850):873, 2001.
- [36] D. L. Cheney and R. M. Seyfarth. Constraints and preadaptations in the earliest stages of language evolution. *Linguistic Review*, 22(2-4):135–159, 2005.
- [37] S. J. Chew, D. S. Vicario, and F. Nottebohm. A large-capacity memory system that recognizes the calls and songs of individual birds. *Proc Natl Acad Sci U S A*, 93(5):1950–1955, 1996.
- [38] C. P. Chow, J. F. Mitchell, and C. T. Miller. Vocal turn-taking in a non-human primate is learned during ontogeny. *Proc. R. Soc. B*, 282(1807), 2015.
- [39] C.R. Newman M.E.J. Clauset, A. Shalizi. Power-law distributions in empirical data. *SIAM Review S1: 661-703. ArXiv e- prints*, 2009.
- [40] D. J. Crammond and J. F. Kalaska. Prior information in motor and premotor cortex: activity during the delay period and effect on pre-movement activity. *J Neurophysiol*, 84(2):986–1005, 2000.
- [41] S. R. Crowley and R. D. Pietruszka. Aggressiveness and vocalization in the leopard lizard (*Gambelia wislizenii*): the influence of temperature. *Animal Behaviour*, 31(4):1055–1060, 1983.
- [42] A. S. Dave and D. Margoliash. Song replay during sleep and computational rules for sensorimotor vocal learning. *Science*, 290(5492):812–6, 2000.

- [43] C. Debussy. *clair de lune*, 2009.
- [44] C. Del Negro, M. Gahr, G. Leboucher, and M. Kreutzer. The selectivity of sexual responses to song displays: effects of partial chemical lesion of the hvc in female canaries. *Behav Brain Res*, 96(1-2):151–9, 1998.
- [45] E. L. Doolittle, B. Gingras, D. M. Endres, and W. T. Fitch. Overtone-based pitch selection in hermit thrush song: unexpected convergence with scale construction in human music. *Proc Natl Acad Sci U S A*, 111(46):16616–21, 2014.
- [46] A. J. Doupe and P. K. Kuhl. Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci*, 22:567–631, 1999.
- [47] A. R. Dykstra, E. Halgren, A. Gutschalk, E. N. Eskandar, and S. S. Cash. Neural correlates of auditory perceptual awareness and release from informational masking recorded directly from human cortex: A case study. *Front Neurosci*, 10:472, 2016.
- [48] A. M. Edwards. Overturning conclusions of levy flight movement patterns by fishing boats and foraging animals. *Ecology*, 92(6):1247–57, 2011.
- [49] S. E. Egnor and M. D. Hauser. A paradox in the evolution of primate vocal learning. *Trends Neurosci*, 27(11):649–54, 2004.
- [50] S. J. Eliades and X. Wang. Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cereb Cortex*, 15(10):1510–23, 2005.
- [51] S. J. Eliades and X. Wang. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, 453(7198):1102–6, 2008.
- [52] J. E. Elie, H. A. Soula, N. Mathevon, and C. Vignal. Dynamics of communal vocalizations in a social songbird, the zebra finch (*Taeniopygia guttata*). *J Acoust Soc Am*, 129(6):4037–46, 2011.
- [53] J. E. Elie and F. E. Theunissen. Meaning in the avian auditory cortex: neural representation of communication calls. *Eur J Neurosci*, 41(5):546–67, 2015.
- [54] J. E. Elie and F. E. Theunissen. The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. *Anim Cogn*, 19(2):285–315, 2016.
- [55] M. Enc. The semantics of specificity. *Linguistic Inquiry*, 22(1):1–25, 1991.
- [56] C. S. Evans, L. Evans, and P. Marler. On the meaning of alarm calls - functional reference in an avian vocal system. *Animal Behaviour*, 46(1):23–38, 1993.
- [57] S. Feldstein. Temporal patterns of dialogue. *Studies in dyadic communication*, 7:91, 1972.

- [58] S. Feldstein and J. Welkowitz. A chronography of conversation: In defense of an objective approach. *Nonverbal behavior and communication*, pages 329–378, 1978.
- [59] Y. I. Fishman, D. H. Reser, J. C. Arezzo, and M. Steinschneider. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res*, 151(1-2):167–187, 2001.
- [60] E. S. Fortune. Starting the conversation with a songbird duet: mechanistic and evolutionary perspectives on the emergence of language. Abstract, Society for Neuroscience, 2015.
- [61] E. S. Fortune and D. Margoliash. Cytoarchitectonic organization and morphology of cells of the field-l complex in male zebra finches (*Taeniopygia guttata*). *J Comp Neurol*, 325(3):388–404, 1992.
- [62] E. S. Fortune and D. Margoliash. Parallel pathways and convergence onto hvc and adjacent neostriatum of adult zebra finches (*Taeniopygia guttata*). *J Comp Neurol*, 360(3):413–41, 1995.
- [63] E. S. Fortune, C. Rodriguez, D. Li, G. F. Ball, and M. J. Coleman. Neural mechanisms for the coordination of duet singing in wrens. *Science*, 334(6056):666–70, 2011.
- [64] K. Friston. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci*, 360(1456):815–36, 2005.
- [65] K. Friston. The free-energy principle: a unified brain theory? *Nat Rev Neurosci*, 11(2):127–38, 2010.
- [66] X. Gabaix. Zipf’s law for cities: An explanation. *The Quarterly Journal of Economics*, 114(3):739–767, 1999.
- [67] M. Gahr. Neural song control system of hummingbirds: Comparison to swifts, vocal learning (songbirds) and nonlearning (suboscines) passerines, and vocal learning (budgerigars) and nonlearning (dove, owl, gull, quail, chicken) nonpasserines. *J Comp Neurol*, 426(2):182–196, 2000.
- [68] M. Gahr and H. R. Guttinger. Functional-aspects of singing in male and female *Uraeginthus bengalus* (*Estrildidae*). *Ethology*, 72(2):123–131, 1986.
- [69] M. Gahr and J. M. Wild. Localization of androgen receptor mrna-containing cells in avian respiratory-vocal nuclei: An in situ hybridization study. *J Neurobiol*, 33(7):865–876, 1997.
- [70] N. Geberzahn and M. Gahr. Undirected (solitary) birdsong in female and male blue-capped cordon-bleus (*Uraeginthus cyanocephalus*) and its endocrine correlates. *PLoS ONE*, 6(10):e26485, 2011.

- [71] T. Q. Gentner. Neural systems for individual song recognition in adult birds. *Ann N Y Acad Sci*, 1016:282–302, 2004.
- [72] T. Q. Gentner and S. H. Hulse. Female european starling preference and choice for variation in conspecific male song. *Animal Behaviour*, 59:443–458, 2000.
- [73] T. Q. Gentner, S. H. Hulse, G. E. Bentley, and G. F. Ball. Individual vocal recognition and the effect of partial lesions to hvc on discrimination, learning, and categorization of conspecific song in adult songbirds. *J Neurobiol*, 42(1):117–33, 2000.
- [74] T. Q. Gentner, S. H. Hulse, D. Duffy, and G. F. Ball. Response biases in auditory forebrain regions of female songbirds following exposure to sexually relevant variation in male song. *J Neurobiol*, 46(1):48–58, 2001.
- [75] T. Q. Gentner and D. Margoliash. Neuronal populations and single cells representing learned auditory objects. *Nature*, 424(6949):669–674, 2003.
- [76] R. Gibrat. Les inégalités économiques. *Paris, France, Libraire du Recueil Sirey*, 1931.
- [77] L. F. Gill, W. Goymann, A. Ter Maat, and M. Gahr. Patterns of call communication between group-housed zebra finches change during the breeding cycle. *Elife*, 4:e07770, 2015.
- [78] N. Ginsburg. Conditioned vocalization in the budgerigar. *J Comp Physiol Psychol*, 53:183–6, 1960.
- [79] T. Gisiger. Scale invariance in biology: coincidence or footprint of a universal mechanism? *Biol Rev Camb Philos Soc*, 76(2):161–209, 2001.
- [80] J. M. Gottman. *Time series analysis: A comprehensive introduction for social scientists*. Cambridge University Press, 1981.
- [81] J. A. Grace, N. Amin, N. C. Singh, and F. E. Theunissen. Selectivity for conspecific song in the zebra finch auditory forebrain. *J Neurophysiol*, 89(1):472–87, 2003.
- [82] M. E. Gurney and M. Konishi. Hormone-induced sexual differentiation of brain and behavior in zebra finches. *Science*, 208(4450):1380–3, 1980.
- [83] H. Haarmann and M. Usher. Maintenance of semantic information in capacity-limited item short-term memory. *Psychon Bull Rev*, 8(3):568–578, 2001.
- [84] P. Haccou and E. Meelis. *Statistical analysis of behavioural data: an approach based on time-structured models*. Oxford university press, 1992.
- [85] R. H. Hahnloser, A. A. Kozhevnikov, and M. S. Fee. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, 419(6902):65–70, 2002.

- [86] M. M. Haith, C. Hazan, and G. S. Goodman. Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Dev*, 59(2):467–79, 1988.
- [87] M. L. Hall and R. D. Magrath. Temporal coordination signals coalition quality. *Curr Biol*, 17(11):R406–7, 2007.
- [88] K. Hammerschmidt, J. D. Newman, M. Champoux, and S. J. Suomi. Changes in rhesus macaque ‘coo’ vocalizations during early development. *Ethology*, 106(10):873–886, 2000.
- [89] K. Hammerschmidt, E. Reisinger, K. Westekemper, L. Ehrenreich, N. Strenzke, and J. Fischer. Mice do not require auditory input for the normal development of their ultrasonic vocalizations. *BMC Neurosci*, 13:40, 2012.
- [90] K. Hashiya and S. Kojima. Acquisition of auditory-visual intermodal matching-to-sample by a chimpanzee (*Pan troglodytes*): comparison with visual-visual intramodal matching. *Anim Cogn*, 4(3-4):231–9, 2001.
- [91] J. M. Hausdorff, C. K. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger. Is walking a random-walk - evidence for long-range correlations in stride interval of human gait. *J Appl Physiol*, 78(1):349–358, 1995.
- [92] M. D. Hauser. *The evolution of communication*. MIT press, 1996.
- [93] M. D. Hauser and P. Marler. Food-associated calls in rhesus macaques (*Macaca mulatta*) .1. socioecological factors. *Behavioral Ecology*, 4(3):194–205, 1993.
- [94] B. Hedwig. Control of cricket stridulation by a command neuron: efficacy depends on the behavioral state. *J Neurophysiol*, 83(2):712–22, 2000.
- [95] R. Held and A. Hein. Movement-produced stimulation in the development of visually guided behavior. *J Comp Physiol Psychol*, 56:872–6, 1963.
- [96] A. Heller. Probabilistic automata and stochastic transformations. *Theory of Computing Systems*, 1(3):197–208, 1967.
- [97] H. Hennig. Synchronization in human musical rhythms and mutually interacting complex systems. *Proc Natl Acad Sci U S A*, 111(36):12974–9, 2014.
- [98] L. Henry, A. J. Craig, A. Lemasson, and M. Hausberger. Social coordination in animal vocal interactions. is there any evidence of turn-taking? the starling as an animal model. *Front Psychol*, 6:1416, 2015.
- [99] A. M. Hernandez, E. C. Perez, H. Mulard, N. Mathevon, and C. Vignal. Mate call as reward: Acoustic communication signals can acquire positive reinforcing values during adulthood in female zebra finches (*Taeniopygia guttata*). *J Comp Psychol*, 130(1):36–43, 2016.

- [100] A. Herwig, W. Prinz, and F. Waszak. Two modes of sensorimotor integration in intention-based and stimulus-based actions. *Q J Exp Psychol (Hove)*, 60(11):1540–54, 2007.
- [101] B. Hommel, W. Prinz, M. Beisert, and A. Herwig. Ideomotor action control: on the perceptual grounding of voluntary actions and agents. *Action science: Foundations of an emerging discipline*, pages 113–136, 2013.
- [102] J. F. Houde and M. I. Jordan. Sensorimotor adaptation in speech production. *Science*, 279(5354):1213–6, 1998.
- [103] F. Huber. Central nervous control of sound production in crickets and some speculations on its evolution. *Evolution*, pages 429–442, 1962.
- [104] K. Immelmann. Song development in the zebra finch and other estrildid finches. *Bird vocalizations*, 1969.
- [105] S. Iversen, I. Kupfermann, and E. R. Kandel. *Emotional States and Feelings*, volume 4, book section Emotional States and Feelings, pages 984–996. McGraw-Hill New York, 2000.
- [106] J. Jaffe, B. Beebe, S. Feldstein, C. L. Crown, and M. D. Jasnow. Rhythms of dialogue in infancy: coordinated timing in development. *Monogr Soc Res Child Dev*, 66(2):i–viii, 1–132, 2001.
- [107] J. Jaffe and S. Feldstein. *Rhythms of dialogue*. Academic Press, 1970.
- [108] D. C. Javitt. When doors of perception close: Bottom-up models of disrupted cognition in schizophrenia. *Annu Rev Clin Psychol*, 5:249–275, 2009.
- [109] D. B. Kelley and F. Nottebohm. Projections of a telencephalic auditory nucleus-field 1 in the canary. *J Comp Neurol*, 183(3):455–69, 1979.
- [110] L. A. Kelley, R. L. Coe, J. R. Madden, and S. D. Healy. Vocal mimicry in songbirds. *Animal behaviour*, 76(3):521–528, 2008.
- [111] J. M. Kilner, K. J. Friston, and C. D. Frith. Predictive coding: an account of the mirror neuron system. *Cogn Process*, 8(3):159–66, 2007.
- [112] S. Kojima and K. Aoki. Intrinsic and synaptic properties of the dorsomedial nucleus of the intercollicular complex, an area known to be involved in distance call production in bengalese finches. *Brain Res*, 966(1):84–94, 2003.
- [113] H. H. Kornhuber and L. Deecke. Hirnpotentialänderungen bei willkürbewegungen und passiven bewegungen des menschen: Bereitschaftspotential und reafferente potentiale. *Pflüger’s Archiv für die gesamte Physiologie des Menschen und der Tiere*, 284(1):1–17, 1965.

- [114] S. Kuhn, A. Keizer, S. A. Rombouts, and B. Hommel. The functional and neural mechanism of action preparation: roles of eba and ffa in voluntary action control. *J Cogn Neurosci*, 23(1):214–20, 2011.
- [115] D. Laberge and S. J. Samuels. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6(2):293–323, 1974.
- [116] H. Lane. Temporal and intensive properties of human vocal responding under a schedule of reinforcement. *J Exp Anal Behav*, 3:183–92, 1960.
- [117] H. Lane. Operant control of vocalizing in the chicken. *J Exp Anal Behav*, 4(2):171–177, 1961.
- [118] A. Leonardo and M. S. Fee. Ensemble coding of vocal control in birdsong. *Journal of Neuroscience*, 25(3):652–661, 2005.
- [119] A. Leonardo and M. Konishi. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature*, 399(6735):466–70, 1999.
- [120] H. J. Leppelsack and M. Vogt. Responses of auditory neurons in the forebrain of a songbird to stimulation with species-specific sounds. *Journal of comparative physiology*, 107(3):263–274, 1976.
- [121] M. S. Lewicki and B. J. Arthur. Hierarchical organization of auditory temporal context sensitivity. *The Journal of Neuroscience*, 16(21):6987–6998, 1996.
- [122] D. Lim and C. Kim. Emerging auditory response interactions to harmonic complexes in field l of the zebra finch. *Auris Nasus Larynx*, 24(3):227–32, 1997.
- [123] S. G. Lisberger and W. T. Thach. *Voluntary Movement: The Parietal and Premotor Cortex*, book section 38, pages 866–892. McGraw Hill Professional, New York, 2015.
- [124] M. A. Long and M. S. Fee. Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature*, 456(7219):189–194, 2008.
- [125] S. A. MacDougall-Shackleton, S. H. Hulse, and G. F. Ball. Neural bases of song preferences in female zebra finches (*Taeniopygia guttata*). *Neuroreport*, 9(13):3047–52, 1998.
- [126] J. M. Macedonia and C. S. Evans. Variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology*, 93(3):177–197, 1993.
- [127] P. Maciej, I. Ndao, K. Hammerschmidt, and J. Fischer. Vocal communication in a complex multi-level society: constrained acoustic structure and flexible call usage in guinea baboons. *Front Zool*, 10(1):58, 2013.
- [128] R. D. Malmgren, D. B. Stouffer, A. S. Campanharo, and L. A. Amaral. On universality in human correspondence activity. *Science*, 325(5948):1696–700, 2009.



- [129] R. D. Malmgren, D. B. Stouffer, A. E. Motter, and L. A. Amaral. A poissonian explanation for heavy tails in e-mail communication. *Proc Natl Acad Sci U S A*, 105(47):18153–8, 2008.
- [130] K. Manabe, T. Kawashima, and J. E. Staddon. Differential vocalization in budgerigars: towards an experimental analysis of naming. *J Exp Anal Behav*, 63(1):111–26, 1995.
- [131] M. Maranesi, A. Livi, L. Fogassi, G. Rizzolatti, and L. Bonini. Mirror neuron activation prior to action observation in a predictable context. *The Journal of neuroscience*, 34(45):14827–32, 2014.
- [132] D. Margoliash. Functional organization of forebrain pathways for song production and perception. *J Neurobiol*, 33(5):671–693, 1997.
- [133] D. Margoliash and E. S. Fortune. Temporal and harmonic combination-sensitive neurons in the zebra finch’s hvc. *The Journal of neuroscience*, 12(11):4309–26, 1992.
- [134] P. Marler. Specific distinctiveness in the communication signals of birds. *Behaviour*, 11(1):13–38, 1957.
- [135] P. Marler. Bird calls: their potential for behavioral neurobiology. *Ann N Y Acad Sci*, 1016(1):31–44, 2004.
- [136] P. Marler, A. Dufty, and R. Pickert. Vocal communication in the domestic chicken .1. does a sender communicate information about the quality of a food referent to a receiver. *Animal Behaviour*, 34:188–193, 1986.
- [137] P. Marler, A. Dufty, and R. Pickert. Vocal communication in the domestic chicken .2. is a sender sensitive to the presence and nature of a receiver. *Animal Behaviour*, 34:194–198, 1986.
- [138] J. S. McCasland. Neuronal control of bird song production. *The Journal of neuroscience*, 7(1):23–39, 1987.
- [139] J. S. McCasland and M. Konishi. Interaction between auditory and motor activities in an avian song control nucleus. *Proc Natl Acad Sci U S A*, 78(12):7815–9, 1981.
- [140] B. McCowan, L. R. Doyle, J. M. Jenkins, and S. F. Hanser. The appropriate use of zipf’s law in animal communication studies. *Animal Behaviour*, 69(1):F1–F7, 2005.
- [141] B. McCowan, S. F. Hanser, and L. R. Doyle. Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Anim Behav*, 57(2):409–419, 1999.
- [142] B. McCowan and D. Reiss. *Vocal learning in captive bottlenose dolphins: a comparison with humans and nonhuman animals*, pages 178–207. Cambridge University Press, Cambridge, UK, 1997.

- [143] C. V. Mello and D. F. Clayton. Song-induced zenk gene expression in auditory pathways of songbird brain and its relation to the song control system. *J Neurosci*, 14(11 Pt 1):6652–66, 1994.
- [144] N. Mesgarani and E. F. Chang. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397):233–6, 2012.
- [145] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, 1(2):226–251, 2004.
- [146] R. Mooney. Neurobiology of song learning. *Curr Opin Neurobiol*, 19(6):654–60, 2009.
- [147] D. Morris. The reproductive behaviour of the zebra finch (*Poephila guttata*), with special reference to pseudofemale behaviour and displacement activities. *Behaviour*, 6(4):271–322, 1954.
- [148] R. Naatanen, M. Tervaniemi, E. Sussman, P. Paavilainen, and I. Winkler. ‘primitive intelligence’ in the auditory cortex. *Trends Neurosci*, 24(5):283–288, 2001.
- [149] K. I. Nagel and A. J. Doupe. Organizing principles of spectro-temporal encoding in the avian primary auditory area field l. *Neuron*, 58(6):938–55, 2008.
- [150] J. L. Nation. Courtship behavior and evidence for a sex attractant in the male caribbean fruit fly, *Anastrepha suspensa*. *Ann Entomol Soc Am*, 65(6):1364–1367, 1972.
- [151] D. A. Nelson and P. Marler. Selection-based learning in bird song development. *Proc Natl Acad Sci U S A*, 91(22):10498–10501, 1994.
- [152] J. P. Neunuebel, A. L. Taylor, B. J. Arthur, and S. E. Egnor. Female mice ultrasonically interact with males during courtship displays. *Elife*, 4:e06203, 2015.
- [153] M. E. J. Newman. Power laws, pareto distributions and zipf’s law. *Contemporary Physics*, 46(5):323–351, 2005.
- [154] J. E. Niven and S. B. Laughlin. Energy limitation as a selective pressure on the evolution of sensory systems. *J Exp Biol*, 211(11):1792–1804, 2008.
- [155] I. C. Noirot, E. F. Brittan-Powell, and R. J. Dooling. Masked auditory thresholds in three species of birds, as measured by the auditory brainstem response (1). *J Acoust Soc Am*, 129(6):3445–8, 2011.
- [156] F. Nottebohm. The neural basis of birdsong. *PLoS Biol*, 3(5):e164, 2005.
- [157] F. Nottebohm and A. P. Arnold. Sexual dimorphism in vocal control areas of the songbird brain. *Science*, 194(4261):211–3, 1976.

- [158] F. Nottebohm, D. B. Kelley, and J. A. Paton. Connections of vocal control nuclei in the canary telencephalon. *J Comp Neurol*, 207(4):344–57, 1982.
- [159] T. M. Otchy, S. B. Wolff, J. Y. Rhee, C. Pehlevan, R. Kawai, A. Kempf, S. M. Gobes, and B. P. Olveczky. Acute off-target effects of neural circuit manipulations. *Nature*, 528(7582):358–63, 2015.
- [160] M. J. Owren, J. A. Dieter, R. M. Seyfarth, and D. L. Cheney. Vocalizations of rhesus (*Macaca mulatta*) and japanese (*M. fuscata*) macaques cross-fostered between species show evidence of only limited modification. *Dev Psychobiol*, 26(7):389–406, 1993.
- [161] M. J. Owren and D. Rendall. Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology*, 10(2):58–71, 2001.
- [162] M. J. Owren, D. Rendall, and M. J. Ryan. Redefining animal signaling: influence versus information in communication. *Biol Philos*, 25(5):755–780, 2010.
- [163] S. R. Partan and P. Marler. Issues in the classification of multimodal communication signals. *Am Nat*, 166(2):231–45, 2005.
- [164] RE Passingham. Two cortical systems for directing movement. *Motor areas of the cerebral cortex*, 132:151–161, 1987.
- [165] C. K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, 5(1):82–7, 1995.
- [166] E. C. Perez, M. S. A. Fernandez, S. C. Griffith, C. Vignal, and H. A. Soula. Impact of visual contact on vocal interaction dynamics of pair-bonded birds. *Animal Behaviour*, 107:125–137, 2015.
- [167] R. E. Phillips and O. M. Youngren. A brain pathway for thalamically evoked calls in birds. *Brain Behav Evol*, 9(1):1–6, 1974.
- [168] J. B. Plotkin and M. A. Nowak. Language evolution and information theory. *J Theor Biol*, 205(1):147–59, 2000.
- [169] J. F. Prather. Auditory signal processing in communication: perception and performance of vocal sounds. *Hear Res*, 305:144–55, 2013.
- [170] J. F. Prather and R. Mooney. Neural correlates of learned song in the avian forebrain: simultaneous representation of self and others. *Curr Opin Neurobiol*, 14(4):496–502, 2004.
- [171] J. F. Prather, S. Nowicki, R. C. Anderson, S. Peters, and R. Mooney. Neural correlates of categorical perception in learned vocal communication. *Nat Neurosci*, 12(2):221–8, 2009.

- [172] J. F. Prather, S. Peters, S. Nowicki, and R. Mooney. Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature*, 451(7176):305–U2, 2008.
- [173] W. H. Press. Flicker noises in astronomy and elsewhere. *Comments on Astrophysics*, 7:103–119, 1978.
- [174] W. Prinz. *A common coding approach to perception and action*. Springer, 1990.
- [175] M. E. Rashotte, E. V. Sedunova, F. Johnson, and I. F. Pastukhov. Influence of food and water availability on undirected singing and energetic status in adult male zebra finches (*Taeniopygia guttata*). *Physiol Behav*, 74(4-5):533–41, 2001.
- [176] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–5, 2002.
- [177] H. Reinke and J. M. Wild. Identification and connections of inspiratory premotor neurons in songbirds and budgerigar. *J Comp Neurol*, 391(2):147–163, 1998.
- [178] S. Ribeiro, G. A. Cecchi, M. O. Magnasco, and C. V. Mello. Toward a song code: Evidence for a syllabic representation in the canary brain. *Neuron*, 21(2):359–371, 1998.
- [179] K. Riebel. Early exposure leads to repeatable preferences for male song in female zebra finches. *Proc. R. Soc. B*, 267(1461):2553–2558, 2000.
- [180] A. Riehle and J. Requin. Monkey primary motor and premotor cortex: single-cell activity related to prior information about direction and extent of an intended movement. *J Neurophysiol*, 61(3):534–49, 1989.
- [181] T. F. Roberts, E. Hisey, M. Tanaka, M. G. Kearney, G. Chattree, C. F. Yang, N. M. Shah, and R. Mooney. Identification of a motor-to-auditory pathway important for vocal learning. *Nat Neurosci*, 2017.
- [182] T. F. Roberts, M. E. Klein, M. F. Kubke, J. M. Wild, and R. Mooney. Telencephalic neurons monosynaptically link brainstem and forebrain premotor networks necessary for song. *Journal of Neuroscience*, 28(13):3479–3489, 2008.
- [183] T. F. Roberts, J. M. Wild, M. F. Kubke, and R. Mooney. Homogeneity of intrinsic properties of sexually dimorphic vocal motoneurons in male and female zebra finches. *J Comp Neurol*, 502(1):157–169, 2007.
- [184] R. Romo, A. Hernandez, and A. Zainos. Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron*, 41(1):165–73, 2004.
- [185] G. J. Ross and T. Jones. Understanding the heavy-tailed dynamics in human behavior. *Phys Rev E Stat Nonlin Soft Matter Phys*, 91(6):062809, 2015.

- [186] C. Scharff and F. Nottebohm. A comparative-study of the behavioral deficits following lesions of various parts of the zebra finch song system - implications for vocal learning. *Journal of Neuroscience*, 11(9):2896–2913, 1991.
- [187] C. Scharff, F. Nottebohm, and J. Cynx. Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway. *J Neurobiol*, 36(1):81–90, 1998.
- [188] M. F. Schmidt and M. Konishi. Gating of auditory responses in the vocal control system of awake songbirds. *Nat Neurosci*, 1(6):513–8, 1998.
- [189] D. S. Schreger, A. W. Pieneman, A. Ter Maat, R. F. Jansen, T. J. Brouwer, and M. L. Gahr. A lightweight telemetry system for recording neuronal activity in freely behaving small animals. *J Neurosci Methods*, 155(1):62–71, 2006.
- [190] R. I. Schubotz. Prediction of external events with our motor system: towards a new framework. *Trends Cogn Sci*, 11(5):211–8, 2007.
- [191] T. Schwalger and B. Lindner. Theory for serial correlations of interevent intervals. *European Physical Journal-Special Topics*, 187(1):211–221, 2010.
- [192] S. H. Scott. Optimal feedback control and the neural basis of volitional motor control. *Nat Rev Neurosci*, 5(7):532–46, 2004.
- [193] T. J. Seller and S. E. Armitage. Diencephalic sites from which calling can be evoked with small currents in japanese quail. *Behavioural Brain Research*, 9(3):305–314, 1983.
- [194] R. M. Seyfarth and D. L. Cheney. Meaning and emotion in animal vocalizations. *Ann N Y Acad Sci*, 1000:32–55, 2003.
- [195] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.
- [196] C. E. Shannon. Prediction and entropy of printed english. *Bell System Technical Journal*, 30(1):50–64, 1951.
- [197] M. F. Shlesinger and J. Klafter. *Lévy walks versus Lévy flights*, pages 279–283. Springer, 1986.
- [198] M. F. Shlesinger, G. M. Zaslavsky, and U. Frisch. Lévy flights and related topics in physics. In *Lévy Flights and Related Topics in Physics*, volume 450, 1995.
- [199] H. B. Simpson and D. S. Vicario. Brain pathways for learned and unlearned vocalizations differ in zebra finches. *J Neurosci*, 10(5):1541–56, 1990.
- [200] C. Song, T. Koren, P. Wang, and AL. Barabasi. Modelling the scaling properties of human mobility. *Nat Phys*, 6(10):818–823, 2010.

- [201] A. Sorribes, B. G. Armendariz, D. Lopez-Pigozzi, C. Murga, and G. G. de Polavieja. The origin of behavioral bursts in decision-making circuitry. *PLoS Comput Biol*, 7(6):e1002075, 2011.
- [202] R. Sossinka. Langfristiges durstvermögen wilder und domestizierter zebrafinken (*taeniopygia guttata castanotis gould*). *Journal für Ornithologie*, 113(4):418–426, 1972.
- [203] J. E. Spiro, M. B. Dalva, and R. Mooney. Long-range inhibition within the zebra finch song nucleus ra can coordinate the firing of multiple projection neurons. *Journal of Neurophysiology*, 81(6):3007–3020, 1999.
- [204] O. Sporns, G. Tononi, and R. Kotter. The human connectome: A structural description of the human brain. *PLoS Comput Biol*, 1(4):e42, 2005.
- [205] D. Stowell, L. Gill, and D. Clayton. Detailed temporal structure of communication networks in groups of songbirds. *J R Soc Interface*, 13(119), 2016.
- [206] M. P. H. Stumpf and M. A. Porter. Critical truths about power laws. *Science*, 335(6069):665–666, 2012.
- [207] M. P. H. Stumpf, C. Wiuf, and R. M. May. Subnets of scale-free networks are not scale-free: sampling properties of networks. *Proc Natl Acad Sci U S A*, 102(12):4221–4224, 2005.
- [208] D. Y. Takahashi, A. R. Fenley, Y. Teramoto, D. Z. Narayanan, J. I. Borjon, P. Holmes, and A. A. Ghazanfar. Language development. the developmental dynamics of marmoset monkey vocal production. *Science*, 349(6249):734–8, 2015.
- [209] D. Y. Takahashi, D. Z. Narayanan, and A. A. Ghazanfar. Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr Biol*, 23(21):2162–8, 2013.
- [210] L. R. Taylor. Aggregation, variance and the mean. *Nature*, 189(4766):732–735, 1961.
- [211] A. Ter Maat, L. Trost, H. Sagunsky, S. Seltsmann, and M. Gahr. Zebra finch mates use their forebrain song system in unlearned call communication. *PLoS ONE*, 9(10):e109334, 2014.
- [212] F. E. Theunissen, N. Amin, S. S. Shaevitz, S. M. Woolley, T. Fremouw, and M. E. Hauber. Song selectivity in the song system and in the auditory forebrain. *Ann N Y Acad Sci*, 1016:222–45, 2004.
- [213] F. E. Theunissen, K. Sen, and A. J. Doupe. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *The Journal of neuroscience*, 20(6):2315–31, 2000.
- [214] F. E. Theunissen, S. M. Woolley, A. Hsu, and T. Fremouw. Methods for the analysis of auditory processing in the brain. *Ann N Y Acad Sci*, 1016:187–207, 2004.

- [215] S. W. Townsend and M. B. Manser. The function of nonlinear phenomena in meerkat alarm calls. *Biol Lett*, 7(1):47–9, 2011.
- [216] M. Tribus. *Thermostatistics and thermodynamics: an introduction to energy, information and states of matter, with engineering applications*. van Nostrand, 1961.
- [217] T. W. Troyer and A. J. Doupe. An associational model of birdsong sensorimotor learning i. efference copy and the learning of song syllables. *J Neurophysiol*, 84(3):1204–23, 2000.
- [218] N. Ulanovsky, L. Las, and I. Nelken. Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, 6(4):391–398, 2003.
- [219] H. Vester, K. Hammerschmidt, M. Timme, and S. Hallerberg. Quantifying group specificity of animal vocalizations without specific sender information. *Physical Review E*, 93(2), 2016.
- [220] D. S. Vicario and K. H. Yohay. Song-selective auditory input to a forebrain vocal control nucleus in the zebra finch. *J Neurobiol*, 24(4):488–505, 1993.
- [221] C. Vignal, N. Mathevon, and S. Mottin. Audience drives male songbird response to partner’s voice. *Nature*, 430(6998):448–451, 2004.
- [222] C. Vignal, N. Mathevon, and S. Mottin. Mate recognition by female zebra finch: Analysis of individuality in male call and first investigations on female decoding process. *Behavioural Processes*, 77(2):191–198, 2008.
- [223] A. S. Villain, I. C. A. Boucaud, C. Bouchut, and C. Vignal. Parental influence on begging call structure in zebra finches (*Taeniopygia guttata*): evidence of early vocal plasticity. *R Soc Open Sci*, 2(11):150497, 2015.
- [224] G. M. Viswanathan, S. V. Buldyrev, S. Havlin, M. G. da Luz, E. P. Raposo, and H. E. Stanley. Optimizing the success of random searches. *Nature*, 401(6756):911–4, 1999.
- [225] C. Voigt, S. Leitner, and M. Gahr. Repertoire and structure of duet and solo songs in cooperatively breeding white-browed sparrow weavers. *Behaviour*, 143:159–182, 2006.
- [226] W. G. Walter, R. Cooper, V. J. Aldridge, W. C. McCallum, and A. L. Winter. Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature*, 203:380–4, 1964.
- [227] X. Wang. On cortical coding of vocal communication sounds in primates. *Proc Natl Acad Sci U S A*, 97(22):11843–9, 2000.

- [228] M. M. Weiner. Proposed algorithm for sequential implementation of adaptive antennas and receivers. *Ieee Antennas and Propagation Magazine*, 51(5):161–162, 2009.
- [229] K. D. Wells. Territoriality in the green frog (*Rana clamitans*): vocalizations and agonistic behaviour. *Animal Behaviour*, 26:1051–1063, 1978.
- [230] G. B. West, J. H. Brown, and B. J. Enquist. A general model for the origin of allometric scaling laws in biology. *Science*, 276(5309):122–6, 1997.
- [231] J. M. Wild. Peripheral and central terminations of hypoglossal afferents innervating lingual tactile mechanoreceptor complexes in fringillidae. *J Comp Neurol*, 298(2):157–71, 1990.
- [232] J. M. Wild. The avian nucleus retroambigialis: a nucleus for breathing, singing and calling. *Brain Res*, 606(2):319–24, 1993.
- [233] J. M. Wild. The auditory-vocal-respiratory axis in birds. *Brain Behav Evol*, 44(4-5):192–209, 1994.
- [234] J. M. Wild. Neural pathways for the control of birdsong production. *J Neurobiol*, 33(5):653–70, 1997.
- [235] J. M. Wild, D. Li, and C. Eagleton. Projections of the dorsomedial nucleus of the intercollicular complex (dm) in relation to respiratory-vocal nuclei in the brainstem of pigeon (*Columba livia*) and zebra finch (*Taeniopygia guttata*). *J Comp Neurol*, 377(3):392–413, 1997.
- [236] R. H. Wiley and M. S. Wiley. Recognition of neighbors’ duets by stripe-backed wrens *Campylorhynchus nuchalis*. *Behaviour*, 62(1):10–34, 1977.
- [237] H. Williams and K. Staples. Syllable chunking in zebra finch (*Taeniopygia guttata*) song. *J Comp Psychol*, 106(3):278–86, 1992.
- [238] J.C. Willis and G. U. Yule. Some statistics of evolution and geographical distribution in plants and animals, and their significance. *Nature*, 109(2728):177–179, 1922.
- [239] P. Winter, P. Handley, D. Ploog, and D. Schott. Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour*, 47(3):230–9, 1973.
- [240] D. M. Wolpert, J. Diedrichsen, and J. R. Flanagan. Principles of sensorimotor learning. *Nat Rev Neurosci*, 12(12):739–51, 2011.
- [241] D. M. Wolpert, Z. Ghahramani, and M. I. Jordan. An internal model for sensorimotor integration. *Science*, 269(5232):1880–2, 1995.
- [242] S. C. Woolley and A. J. Doupe. Social context-induced song variation affects female behavior and gene expression. *PLoS Biol*, 6(3):e62, 2008.



- 
- [243] S. M. Woolley and J. H. Casseday. Processing of modulated sounds in the zebra finch auditory midbrain: responses to noise, frequency sweeps, and sinusoidal amplitude modulations. *J Neurophysiol*, 94(2):1143–57, 2005.
- [244] Y. Wu, C. Zhou, J. Xiao, J. Kurths, and H. J. Schellnhuber. Evidence for a bimodal distribution in human communication. *Proc Natl Acad Sci U S A*, 107(44):18803–8, 2010.
- [245] G. U. Yule. A mathematical theory of evolution, based on the conclusions of dr. j. c. willis, frs. *Philos Trans R Soc Lond B Biol Sci*, 213:21–87, 1925.
- [246] R. A. Zann. *The Zebra Finch*, volume 5, book section 10 Vocalisations, pages 196–213. Oxford University Press, Oxford, 1996.
- [247] R. A. Zann, S. R. Morton, K. R. Jones, and N. T. Burley. The timing of breeding by zebra finches in relation to rainfall in central australia. *Emu*, 95(3):208–222, 1995.
- [248] H. Zeier and H. J. Karten. The archistriatum of the pigeon: organization of afferent and efferent connections. *Brain Res*, 31(2):313–26, 1971.
- [249] G. K. Zipf. Human behavior and the principle of least effort. *J Clin Psychol*, 6(3):306–306, 1950.

# Acknowledgment

This thesis could not be finished without the help and support of many people who are gratefully acknowledge here.

I would firstly thank my parents for their encouragement and ongoing supports. Very special thanks go to my beloved wife Jianying who walks with me along this way and fills my life with joy and happiness. To my son, Shengrui, I am sorry when Papa disappeared into his “study” and for your love. To my grandma, I cannot begin adequately to express my gratitude for raising me and always being with me when I was little.

I am especially grateful to the Graduate School of Systemic Neurosciences, where I have learned in-depth interdisciplinary neurosciences from theory to methodology. My deepest gratitude is to my mentor, Prof. Manfred Gahr who made me feel warmly welcome in Seewiesen and provided me great opportunities for accessing deeply in the field of neurosciences. I am deeply grateful to him for all his helps and supports for this thesis.

I would like to express my deeply thanks to Andries Ter Maat, Lisa Trost, Susanne Seltmann, Markus Abels and Hannes Sagunsky for their excellent and patient helps for my experiments. They are always kind and generous in teaching me how to prepare experiments, how to do surgeries and helping me out of difficulty in every problem. I would like to thank Diana Werner for improving my English skill and thank Diana Werner and Nicole Hoiss for checking all spellings and grammars presented in this thesis. Thanks goodness. Basically, this thesis could not be finished without helps of many peoples. I would like to thank Meng-Ching Ko, Amanda Monte, Pietro D’Amelio and Nico Mauricio who were very nice and inspiring to be sharing an office. At last but not least, I like to thank René Quispe, Camilla Villavicencio, Amanda Monte and Meng-Ching Ko, I felt always happy and relaxing when spending time with them, in particular the house parties of Camilla and René, the little sweets and cakes of Amanda and the English class with Diana, those would be my happiest memories in Seewiesen.



# Curriculum Vitae

## Education

PhD, Neuroscience, Max Planck Institute for Ornithology, Seewiesen, Germany, expected 2017, Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München, Thesis: A telemetric approach for characterizing behavioral dynamics and neurophysiology of vocal interactions in Zebra Finches. Thesis Advisory Committee: Manfred Gahr, Andries Ter Maat, Boris Chagnaud and Benedikt Grothe.

MSc, Neuroscience and Biophysics, Universität Konstanz, Germany, 2010, Thesis: Analysis of somata and dendrites odor responses in *Drosophila* ORNs. Exam committee and Thesis Advisory: Giovanni Galizia, Hans-Jürgen Apell and Wolfram Kutsch.

BSc, Biological Sciences, Universität Konstanz, 2007, Thesis: A comprehensive study of receptor neuron response profiles and their implication in *Drosophila melanogaster*. Thesis Advisory: Giovanni Galizia.

BSc, Veterinary Medicine, South China agricultural University, 2001, Thesis: Myelography of injured spinal cords in small animals. Thesis Advisory: Yaoming Zhan.

## Other research and employment

Max Planck Institute for Ornithology, Radolfzell, Germany, 2010, Internship, Subjects: Data analysis and management of Movebank, Advisory: Martin Wikelski.

Universität Konstanz, Germany, 2008 – 2010, research assistant, Subjects: wide field, confocal and two-photon imaging of olfactory receptor neurons in *Drosophila*, Advisory: Giovanni Galizia.

Universität Konstanz, Germany, 2009 – 2010, research assistant, Subjects: developing light sheet wide field fluorescence microscopy and coherent anti-stokes Raman spectroscopy (CARS). Advisory: Ming Lei.

Taojin Animal Clinic, Guangzhou, China, 2001 – 2002, Diagnostic assistant.



# List of publications

Ma, S., Ter Maat, A., Gahr, M. Power-law scaling of vocal activity in zebra finches, in manuscript

Ma, S., Ter Maat, A., Gahr, M., Sensorimotor expectation in call-based vocal communications of zebra finches, in manuscript

Ma, S., Gahr, M., Spectrotemporal coding of auditory trains in the auditory forebrains of zebra finches, in manuscript

Galizia, C.G., Münch, C., Strauch, M., Nissler, M., Ma, S., 2010, (Cover issue), Integrating heterogeneous odor response data into a common response model: A DoOR to the complete olfactome, *Chem Senses*



# List of Author Contributions

## 1. Introduction

Shouwen Ma (S.M.) contributed Figure 1.1, Figure 1.2 A-B, Figure 1.3-1.4 and Figure 1.7-1.8.

## 2. Manuscripts

### 2.1. Power-law scaling of vocal activity in zebra finches

Shouwen Ma (S.M.), Andries ter Maat (A.T.M.), Manfred Gahr (M.G.) conception; S.M. designed, performed the experiment and analyzed the data; S.M. contributed data and analyses to Figure 2.1-2.10 and Table 2.1-2.2. S.M., A.T.M., M.G. wrote the paper.

### 2.2. Sensorimotor expectation in call-based vocal communications of zebra finches

Shouwen Ma (S.M.), Andries ter Maat (A.T.M.), Manfred Gahr (M.G.) conception, designed; S.M. performed the experiment and analyzed the data. S.M. contributed data and analyses to Figure 2.11-2.19 and Table 2.3-2.4. S.M., A.T.M., M.G. interpreted the data and wrote the paper.

### 2.3. Spectrotemporal coding of auditory trains in the auditory forebrains of zebra finches

Shouwen Ma (S.M.) and Manfred Gahr (M.G.) conception, S.M. performed the experiment and analyzed the data. S.M. contributed data and analyses to Figure 2.20-2.26. S.M., M.G. interpreted the data and wrote the paper. Susanne Hoffmann (S.F.) edited and revised manuscript.

## 3. Discussion

Shouwen Ma (S.M.) contributed Figure 3.1-3.3.

Signature of supervisor

Signature of doctoral candidate





# Permissions

Figure 1.2. C, D Publication: eLife. Title: Female mice ultrasonically interact with males during courtship displays, eLife 2015;4:e06203 Joshua P Neunuebel, Adam L, Taylor Ben, J Arthur, SE Roian Egnor by eLife Sciences Publications License: CC-BY license 4.0

Figure 1.5. Publication: Nature Reviews Neuroscience. Title: Neural mechanisms of birdsong memory, Volume 7, 2006, p 347-357, Johan J. Bolhuis and Manfred Gahr © by nature publishing group, 2006, with permission from nature publishing group License Number: 3946431403110 License date: Sep 12, 2016

Figure 1.6. Publication: Nature Reviews Neuroscience. Title: Neural mechanisms of birdsong memory, Volume 7, 2006, p 347-357, Johan J. Bolhuis and Manfred Gahr © by nature publishing group, 2006, with permission from nature publishing group License Number: 3947550077574 License date: Sep 14, 2016



# Eidesstattliche Versicherung/Affidavit

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation “A telemetric approach for characterizing behavioral dynamics and neurophysiology of vocal interactions in Zebra Finches,, selbständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittle bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation “A telemetric approach for characterizing behavioral dynamics and neurophysiology of vocal interactions in Zebra Finches” is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

After completion of the examining process, this work will be given to the library of the Ludwig-Maximilians-Universität München, where it will be accessible to the public for viewing and borrowing. As author of this work, I agree to this procedure.

München, den 19.6.2017

Shouwen Ma

Unterschrift