MAX PLANCK INSTITUTE
FOR PSYCHOLINGUISTICS

WWW.MPI.NL

Vocal Learning in the Pale Spear-Nosed Bat, *Phyllostomus discolor*

ELLA Z. LATTENKAMP

Vocal Learning in the Pale Spear-Nosed Bat, *Phyllostomus discolor*

Ella Z. Lattenkamp

The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond.

More information can be found at www.mpi.nl/imprs.

ISBN
978-94-92910-10-3

Layout and cover:
Mark D. Scherz & Ella Z. Lattenkamp

Printed by Ipskamp Printing, Enschede

The work presented in the current thesis was supported by a Human Frontiers Science Programme Research Grant (RGP0058/2016). The work was carried out at the Ludwig Maximilian University in Munich, Germany.

Vocal Learning in the Pale Spear-Nosed Bat, *Phyllostomus discolor*

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus Prof. Dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op

donderdag, 23 april 2020

om 12:30 uur precies

door

Ella Zoe Lattenkamp

geboren op 24 september 1988

te Berlijn

Promotoren:
Prof. dr. S. E. Fisher
Prof. dr. L. Wiegrebe (Ludwig Maximilian Universiteit, München)

Copromotor:
Dr. S. C. Vernes

Manuscriptcommissie:
Prof. dr. S. C. Levinson
Prof. dr. C. ten Cate (Universiteit Leiden, Leiden)
Dr. S. Hoffmann (Max Planck Instituut voor Vogelkunde, Seewiesen)

Paranimfen:
Ine Alvarez van Tussenbroek
Ksenia Lavrichenko

Vocal Learning in the Pale Spear-Nosed Bat, *Phyllostomus discolor*

Doctoral thesis to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus Prof. Dr. J.H.J.M. van
Krieken, according to the decision of the Council of Deans
to be defended in public on

Thursday, 23 April 2020

at 12:30 hours

by

Ella Zoe Lattenkamp

born on 24 September 1988

in Berlin, Germany

Supervisors:
Prof. Dr. S. E. Fisher
Prof. Dr. L. Wiegrebe (Ludwig Maximilian University, Munich)

Co-Supervisor:
Dr. S. C. Vernes

Doctoral Thesis Committee:
Prof. Dr. S. C. Levinson
Prof. Dr. C. ten Cate (University of Leiden, Leiden)
Dr. S. Hoffmann (Max Planck Institut for Ornithology, Seewiesen)

Paranymphs:
Ine Alvarez van Tussenbroek
Ksenia Lavrichenko

In memory of Lutz Wiegrebe

*28.05.1965
†29.11.2019

Table of Contents

Chapter 1

Introduction 1

Chapter 2

Vocal learning: a language-relevant trait in need of a broad cross-species approach 31

Chapter 3

General principles of loudness coding in bats 45

Chapter 4

The vocal repertoire of pale spear-nosed bats in a social roosting context........................ 69

Chapter 5

Volitional control of social vocalisations and vocal usage learning in bats 93

Chapter 6

Indications of vocal production learning in the pale spear-nosed bat, *Phyllostomus discolor* 111

Chapter 7

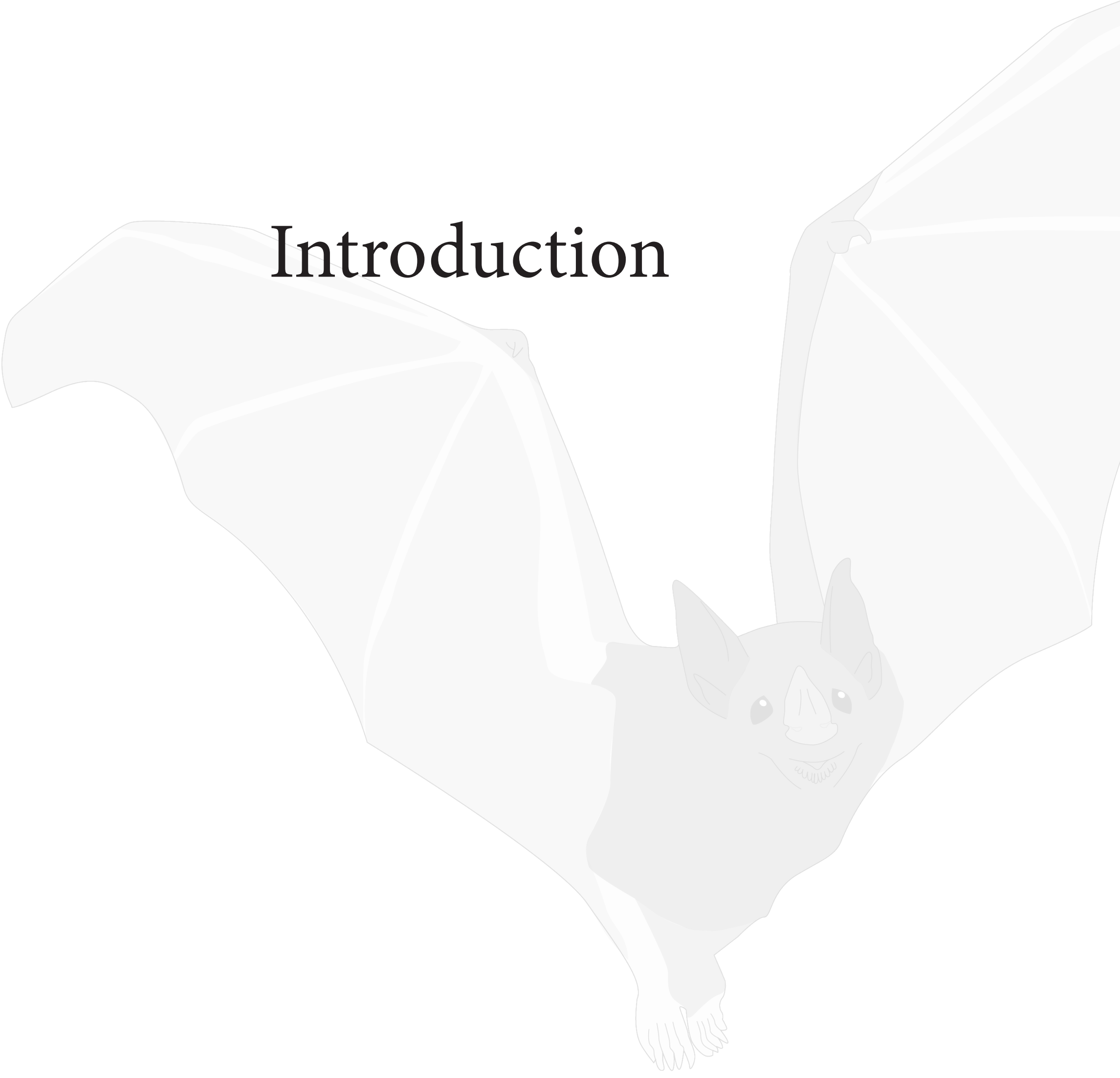
Discussion... 137

Appendix

English Summary.. 158
Nederlandse Samenvatting... 161
Research Data Management.. 164
Acknowledgements 165
Curriculum Vitae... 168
Publication List 169
MPI Series in Psycholinguistics... 170

Chapter 1

Introduction



1.1 Vocal communication

Communication, in its various forms, is a joint undertaking of two or more subjects imparting or exchanging information. For communication to function properly, sender and receiver must understand the information and intent of the exchanged signals; preferably through a shared and mutually understood signal repertoire (Bradbury and Vehrencamp, 2011). Communication can be accomplished using a single or multiple modalities, chosen from a wide array of possibilities. Well-known examples for different types of communication are olfactory: such as territorial claims via olfactory signals in wolves (Rothman and Mech, 1979); gestural: such as chimpanzee mothers gesticulating demands to visually attentive juveniles (Tomasello *et al.*, 1994); or acoustic: such as announcement of the rut through loud vocalisations in deer (Reby and McComb, 2003).

Acoustic communication is special among the different types of communication: it works unhindered by most obstacles, it is independent of light, and thus visual contact, and is possible over relatively large distances (especially in water). Although acoustic communication has these advantages over for example visual communication, it also bears the risk of being overheard, it may hint location cues to predators, information can be masked by other acoustic influences, and attenuation can be significant, especially for high frequency signals (Bradbury and Vehrencamp, 2011). The most studied organisms using acoustic signals for communicative purposes are mammals, birds, frogs, and insects. Already by comparing these four very different clades of organisms, it becomes clear that acoustic communication is not necessarily based on similar sound production or sound perception mechanisms. Insects alone have found at least five different ways for sound production, which mainly involve rubbing or vibrating different external structures, such as legs, wings, and tymbals (Alexander, 1957). Sound production in tetrapods (including mammals, birds, and frogs) occurs internally and usually includes the vibration of soft tissue (membranes) by air flow, and the subsequent sound filtration in the vocal tract. More specifically, the mechanisms involved in vocal communication generally comprise three physiological systems: The respiratory, phonatory, and filter system (**Fig. 1.1**; Fitch, 2000). The task of the respiratory system solely consists of lung regulation, which controls the exhalation process and drives the vocal organ oscillations. The vocal folds in mammals and frogs are located in the larynx, while in birds the oscillating tissue is located in the syrinx (Riede and Goller, 2010). In both, the larynx and syrinx, the oscillating tissues constitute the phonatory system (i.e. the origin of the produced sound) (**Fig. 1.1**). Following the myoelastic-aerodynamic sound source principle, the number of vocal fold oscillations within a certain timespan determines the fundamental frequency (or 'pitch') of a vocalisation (Riede and Goller, 2010; Elemans *et al.*, 2015). However, these vocal fold oscillations generate not only the fundamental frequency, but also its multiples (i.e. 'harmonics'). The produced complex sound then

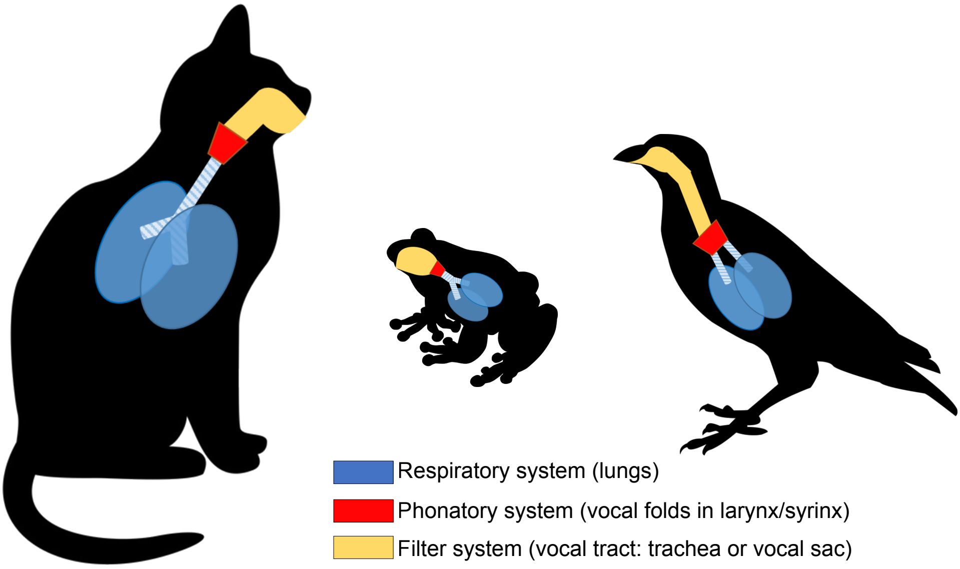


Figure 1.1: Similarities of the sound production systems of vocally communicating vertebrates. Three physiological systems - the respiratory, phonatory, and filter system – are working together during the production of vocalisations. The respiratory system (i.e. the lungs; indicated in blue) control the vibratory speed of the vocal folds, which are located in the syrinx (in birds) or larynx (most other tetrapods)(red). The speed of the vibrations determines the fundamental frequency of the produced vocalisation. The sound is then filtered, as it passes the vocal tract (yellow), and changes dependent on the resonance properties of the respective filter system. Figure inspired by Fitch and Hauser (2003).

passes the filter or articulatory system, where the resonance properties of the vocal tract determine the attenuation patterns of all generated frequencies. In frogs this vocal tract is the vocal sac, while in birds and mammals the vocal tract is the trachea following the larynx (**Fig. 1.1**). Movement of the articulators in the supralaryngeal tract (e.g. soft pallet, lips, and tongue) defines the 'formants', i.e. specific energy rich frequency ranges of the vocal output. Formants for example define different vowels in human speech (Fitch, 2000; Riede and Goller, 2010). Notable exceptions from these vibration driven forms of vocalisations are sounds generated through turbulent airflow, such as fricative consonants, or click consonants generated by tongue clicks in human spoken languages (e.g. Khoisan languages (Williams, 2003)). Interestingly, other vertebrates also make use of click sounds in vocal communication. For example, some frog species (from the family Pipidae) produce clicks by slipping two laryngeal cartilage discs against each other (Rabb, 1960). In this thesis, I focus on vibration driven vocal communication, specifically omitting click sounds.

All forms of acoustic communication require not only the capability of sound production by the sending side of the communication, but also the development of some form of acoustic perception in the receiver of the signal. 'Hearing' always involves the detection of some kind of vibration or particle motion, which can be in the form of pressure differences in air or water or the perception of vibrations in solid substances (Bradbury and Vehrencamp, 2011). The hair-cell-based reception of non-airborne sound is an evolutionary relatively old trait, which can be considered homologous across all jawed vertebrates. However, the evolution of tympanic ears, which allowed the perception of airborne sounds, occurred separately in mammals, sauropsids (e.g. birds and crocodiles), and some amphibians (i.e. modern frogs, order Anura) more than 300 million years ago (Grothe and Pecka, 2014). For vocally communicating animals, the perception of airborne sound is a vital requirement for the interpretation of acoustic signals with drastic impact on the chances of survival and reproduction. Vocal communication is thus an intricate interplay between acoustic emission and perception, where the vocal complexity of the signal emitter needs to be met by the hearing sensitivity of the signal receiver in order to fully convey the intended information.

1.2 Human speech and language capacity

Simple vocal communication, which is characterised by the use of only a limited number of innate, non-learned vocalisations, is relatively widespread in the animal kingdom and present in all major vertebrate groups (Bradbury and Vehrencamp, 2011). The development of very sophisticated, complex vocal communication, such as speech, and ultimately language, is a feature unique to humans. The capacity for speech and language acquisition, already existent in a babbling human baby, develops further until adulthood, allowing us (at least theoretically) to learn a vast number of languages, with an ever expanding repertoire, even at advanced ages (albeit with diminishing proficiency) (Oller *et al.*, 1976; review: Fisher and Marcus, 2006). No other extant species has shown anything comparable to the complexity with which humans make use of speech, syntax, and semantics to communicate their intentions or emotions (Fitch, 2005, 2010). This exceptional human capacity has intrigued scholars for hundreds of years and it is being intensely studied in several lines of research revolving around linguistics: phonetics, phonology, semantics, psycholinguistics, syntax, prosody, and others. But not only linguistics-focused fields of research have investigated this human capacity; other fields such as neurogenetics, neuroimaging, genetics of language, developmental and evolutionary, as well as behavioural biology have also put a focus on language and speech acquisition. Research in these diverse fields has provided us with a wealth of insights and a dynamic research community, which aims at combining these findings in order to ultimately explain the functional basis and evolution of this human-exclusive trait. For example, recent advances in neuroimaging techniques such as functional magnetic resonance imaging, diffusion-tensor imaging, and transcranial magnetic stimulation allow the non-invasive study of language-related activity in human brains (Hagoort and Indefrey, 2014; Willems and van Gerven, 2018). Among others, these methods were used to identify brain areas involved in speech learning tasks and semantic processing (Wang *et al.*, 2003; Price, 2012; Zatorre, 2013) and to investigate the connectivity between these brain areas during language processing (Sharoh *et al.*, 2019).

In the past, theories about the evolution of language were mainly focused on the uniqueness of humans and the innate properties of the human mind that enable the heritable human 'faculty of language' (Chomsky, 2006). Over time, the dynamic processes and complexity underlying human language acquisitions have been increasingly acknowledged and the diversity of language on all its organisational levels (sound, grammar, lexicon, meaning) has entered the spotlight (Evans and Levinson, 2009). The prevalent line of thought is now moving towards the idea that in order for humans to evolve a language faculty several subsystems need to be working together, which had independent evolutionary origins and came together in humans as this unique capacity. This proposed coevolutionary model (Levinson and Evans, 2010) can be applied to different levels of language evolution. It is applicable to the variation generated by the interaction between mind and cultural linguistic traditions (Levinson and Evans, 2010), but also to a more basic evolution of language originating in the coevolutionary development of morphological structures (e.g. brain and vocal production apparatus) necessary for speech and language acquisition. This more evolutionary perspective suggests that the human speech and language evolution was the result of a gradual and multi-layered process. The resulting neuronal system underlying language is likely made up of subsystems, rather than a single structure that adapted and allowed us to develop this trait (Poeppel, 2012; Poeppel *et al.*, 2012). The same holds true for genes: not a single 'language' gene is responsible for the human faculty of language, but rather complex genetic networks and their specific expression patterns enable humans to acquire speech and spoken language (Marcus and Fisher, 2003; Fisher and Vernes, 2015; Devanna *et al.*, 2019).

Persistent questions in the organismal investigation of language acquisition are concerned with the delimitation of human language in contrast to vocal communication in other species. In other words: 'What is it that makes us humans special?' In order to facilitate the distinction between different structures and processes involved in language acquisition, a division between the faculty of language in a broad and a narrow sense was proposed (Hauser *et al.*, 2002): The broad sense includes a sensory-motor system (e.g. allowing for vocal production and auditory perception), a conceptual-intentional system (incl. intentional signalling), and the computational mechanisms for recursion (which enable memory, learning, naming), while the narrow sense only entails recursion. While the faculty of language in the broad sense is shared with other animals, the capacity for recursion is proposed to be human specific (Hauser *et al.*, 2002). Although decades of research have provided fascinating insights, the biological basis and evolution of language still offers ample material for investigation. Open questions concern amongst other topics the uniqueness of the human capacity to use recursion for the generation of infinite communicative elements; the exact changes that forge a neuronal foundation for language acquisition; the genetic basis for speech/language and vocal-motor control; the impact of social and cultural feedback on speech; and the evolutionary drivers promoting language acquisition in humans (Christiansen and Kirby, 2003; Fitch, 2010).

1.3 The need for animal models

Research on the biological basis of human speech is often focused on the underlying behavioural, genetic, and neurobiological foundation of this trait. However, studying humans in any of these fields has its limitations. Due to the long developmental time, behavioural and especially vocal development studies in humans generally take a long time and are often encumbered by the entanglement of intrinsic and extrinsic influences. Social environment, culture, and even diet, can have a strong impact on the phenotypic expression of speech and language acquisition and are challenging factors to control for. Animal studies investigating these effects of nature vs. nurture are often conducted as cross-fostering experiments, which are not possible with human subjects. The need for animal model systems becomes even more pressing when invasive studies in the fields of genetic or neurobiology are concerned.

The in-depth investigation of the genetic basis of speech often requires invasive experiments, such as genetic manipulations. Genetic manipulations, e.g. the manipulation of expression patterns or structure of specific genes (i.e. gene-editing), can be performed *in vivo* for example via virus-mediated RNA interferences. Such gene manipulation studies allow us to draw functional conclusions between the expression levels of specific genes and the vocal learning phenotype. Researchers used RNA interferences for example to reduce levels of *Foxp2* (a gene associated with human speech and language) in the brain of juvenile zebra finches and showed that song imitation becomes less precise and their song generally more variable (Haesler *et al.*, 2007). In another experiment, gene-editing of *Foxp2* in mice affected procedural learning and striatal neuroplasticity (Schreiweis *et al.*, 2014). Such *in vivo* gene manipulation studies thus provide functional insights into gene expression patterns involved in learning, which are not possible with human subjects. The need for animal models is not only prominent when the genetic basis of speech is investigated, but also when its neurogenetic foundation is studied.

The general neurobiological basis for speech acquisition and language learning in humans is being successfully studied with the above-mentioned recent neuroimaging techniques (section 1.2). However, the in-depth study of precise neuronal pathways (i.e. on a single circuit level) and neuronal connectivity (i.e. on a single cell or synapse level) requires the use of invasive methods such as electrophysiological, optogenetic, tracing, or lesion experiments. These invasive methods are impossible to use in human studies, but applied to animal systems they provide important information about the biological basis of vocal production and perception. For example, the targeted optogenetic disruption of neural activity during song learning in zebra finches prevented song copying, identifying a neural premotor structure important for the encoding of tutor song (Roberts *et al.*, 2012). Tracing experiments can be used for example to investigate the neuroanatomical structures involved in vocal control. It involves the injection of a dye to functional brain areas of interest and the subsequent perfusion of the experimental animal. This method was for example used to demonstrate the shared design of the budgerigar and oscine vocal systems, which further uncovered the evolutionary background of neuronal structures involved in vocal control (Durand *et al.*, 1997). Experiments that utilise the induced death of specific neurons and brain lesions *in vivo* are also impossible to conduct on humans. However, these present useful techniques to experimentally test functionality of neuronal pathways. In zebra finches both methods were successfully applied to investigate brain regions and neuronal pathways contributing to their vocal development (e.g. Scharff and Nottebohm, 1991; Scharff *et al.*, 2000). Animal models are thus of great importance for the experimental approach to answering questions about the origins of human speech acquisition and language-related traits.

1.4 Vocal learning

Comparative studies are crucial to gain an understanding of the evolutionary basis of the human capacity for speech and language. As the capacity for speech and language is a uniquely human phenotype, comparative studies have focused on a trait that is believed to constitute the functional basis for speech and is shared by humans with a handful other vertebrate species: vocal learning. Vocal production learning (also referred to as vocal mimicry, vocal imitation learning, or simply vocal learning) is broadly defined as the ability to learn to produce new vocalisations (i.e. vocalisations not innately contained in a species' repertoire) or modify known vocalisations to a great extent through auditory experience (Janik and Slater, 1997, 2000; Boughman and Moss, 2003). Vocal learning species are of great importance for comparative studies investigating the evolution of human speech, but only few taxa have been shown to possess this capacity. Thus far, only a small selection of avian and mammalian taxa has been investigated successfully, including songbirds (Wilbrecht and Nottebohm, 2003), parrots (Pepperberg, 2010), cetaceans (Janik, 2014), pinnipeds (Reichmuth and Casey, 2014), elephants (Stoeger and Manger, 2014), and bats (Knörnschild, 2014). Research on potential vocal learning species has focused on three main areas: the behavioural assessment of species' vocal imitation capacity (section 1.4.1); the comparative evaluation of morphological structures required for sound production and neural pathways essential for controlling vocal emissions (section 1.4.2); and the evolutionary background of the studied species (section 1.4.3).

1.4.1 Behavioural assessment of vocal learning

The careful assessment of social and vocal behaviours is essential to identify and describe the vocal learning phenotype. These assessments usually include bioacoustic analyses detecting the production of novel vocalisations or vocal convergence towards an auditory target. The most high-profile evidence for vocal learning is the imitation of heterospecific vocalisations, especially mimicry of human speech. The imitation of human speech is widely known to be achieved by parrots (Pepperberg, 2010), but in isolated cases elephants (Stoeger *et al.*, 2012), pinnipeds (Schusterman, 2008), and cetaceans (Ridgway *et al.*, 2012; Janik, 2014; Abramson *et al.*, 2018) have also been shown to mimic human speech. The imitation of other heterospecific or artificial sounds also provides evidence for vocal learning. Bottlenose dolphins for example were shown to mimic novel, computer-generated whistle sounds (Richards *et al.*, 1984) and the superb lyrebird is an extreme example for vocal learning as it expertly mimics sounds in its environment and heterospecific vocalisations (Dalziell and Magrath, 2012). The most convincing evidence for vocal learning comes from isolation and deafening studies. These types of studies either interrupt the perception of acoustic targets (i.e. isolation studies) and thus hinder the formation of a memorised template or they interrupt the audio-vocal feedback (i.e. deafening studies), which prevents acoustic self-perception and directed adjustment of emitted vocalisations. With the help of such studies, the difference between innate and learned vocalisations can be distinguished. Vocal learning animals that develop their vocalisations while deprived of conspecific auditory input were shown to produce an incomplete, deteriorated adult vocal repertoire. This was shown for example for zebra finches, which produced abnormal song when raised in social and acoustic isolation (Fehér *et al.*, 2009). Furthermore, non-vocal learning species were shown to produce normal vocalisations despite being deafened (Winter *et al.*, 1973). Cross-fostering studies can also provide evidence for vocal learning, if the fostered animal adjusts its call parameters to heterospecific or artificial acoustic inputs. This was for example shown for a Risso's dolphin that was cross-fostered in a group of bottlenose dolphins and subsequently produced vocal signals more similar to those of its foster family than to those of wild conspecifics (Favaro *et al.*, 2016).

Indications for vocal learning can also be taken from the occurrence of vocal dialects (or local vocal variation) between different aggregations/populations of one species. For example, regional dialects were shown in the contact call of a parrot, the yellow-naped amazon (Wright, 1996) and in group-specific screech calls in greater spear-nosed bats (Boughman, 1998). This form of vocal divergence between individuals of one species due to geographic separation has previously been suggested as indication of, but not evidence for, vocal learning (Janik and Slater, 1997; Vernes and Wilkinson, 2019). Other factors, such as genetic drift due to geographic isolation, differences in arousal or stress due to population density, or even changes in vocal parameters due to habitat differences (e.g.

large or small cave, city or rural area, open or closed forest), could also be causative agents for the development of vocal dialects.

1.4.2 Morphology and neuronal control of sound production

In order to imitate a sound, the vocal apparatus needs to be shaped in a way that allows the reproduction of the acoustic target. Furthermore, vocal learning requires neuronal control over this vocal apparatus. As mentioned above, humans, which evidently possess the capacity for vocal learning, share the basic morphology of the sound production systems with several other vertebrates (section 1.1). However, our evolutionary closest relatives, who possess the same basic morphological setup of vocal production systems, are incapable of human speech imitation. This surprising inability of non-human primates to produce human speech sounds was long assumed to be caused by anatomical limitations and the position of their vocal tract (Lieberman *et al.*, 1969). However, this hypothesis has been refuted recently, as it has been demonstrated that the monkey vocal tract is in fact speech-ready (Fitch *et al.*, 2016). Specifically, Fitch and colleagues used x-ray videos to demonstrate that macaque vocal tracts are physiologically capable to generate a large acoustic space, including great overlap with the human vowel space (i.e. they are able to produce five easily discriminable vowels; Fitch *et al.*, 2016). Therefore, it was concluded that macaques in fact possess the morphological basis to generate human speech sounds, but that the appropriate neuronal control for speech production might be lacking (Fitch, 2018).

In contrast to simple vocal communication (e.g. using a repertoire of only a couple of innate vocalisations), vocal learning additionally requires a specialised neural basis, which allows the memorisation of an auditory target and the ability to match self-emitted vocalisations to this stored target (cf. cortico-striatal-thalamic loop for learning vocalisations, **Fig. 1.2a**). This also requires the repeated adjustment of the self-emitted vocal output in order to increase its similarity to the memorised target. Vocalisations in humans are thought to be controlled via two non-mutually exclusive neuronal pathways: an innate vocal production pathway (for involuntary vocal production) and a learned pathway (intentional, learned vocalisations, i.e. speech) (Petkov and Jarvis, 2012). The former involves neuronal structures such as anterior cingulate cortex, amygdala, periaqueductal grey, and orbito-frontal cortex (**Fig. 1.2a**), while the latter, among others, includes the laryngeal motor cortex, and regions in the thalamus and striatum (Jürgens, 2002, 2009; Jarvis, 2004; Simonyan and Horwitz, 2011). Humans are exceptional in their speech capacity, which is at least in part accredited to the location of the laryngeal motor cortex in the primary motor cortex and to the possession of direct projections from the laryngeal motor cortex to the nucleus ambiguus in the brainstem, where the laryngeal motor neurons are located (**Fig. 1.2a**; Jürgens, 2002; Simonyan and Horwitz, 2011; Fitch, 2018). These direct monosynaptic connections between neurons in the motor cortex

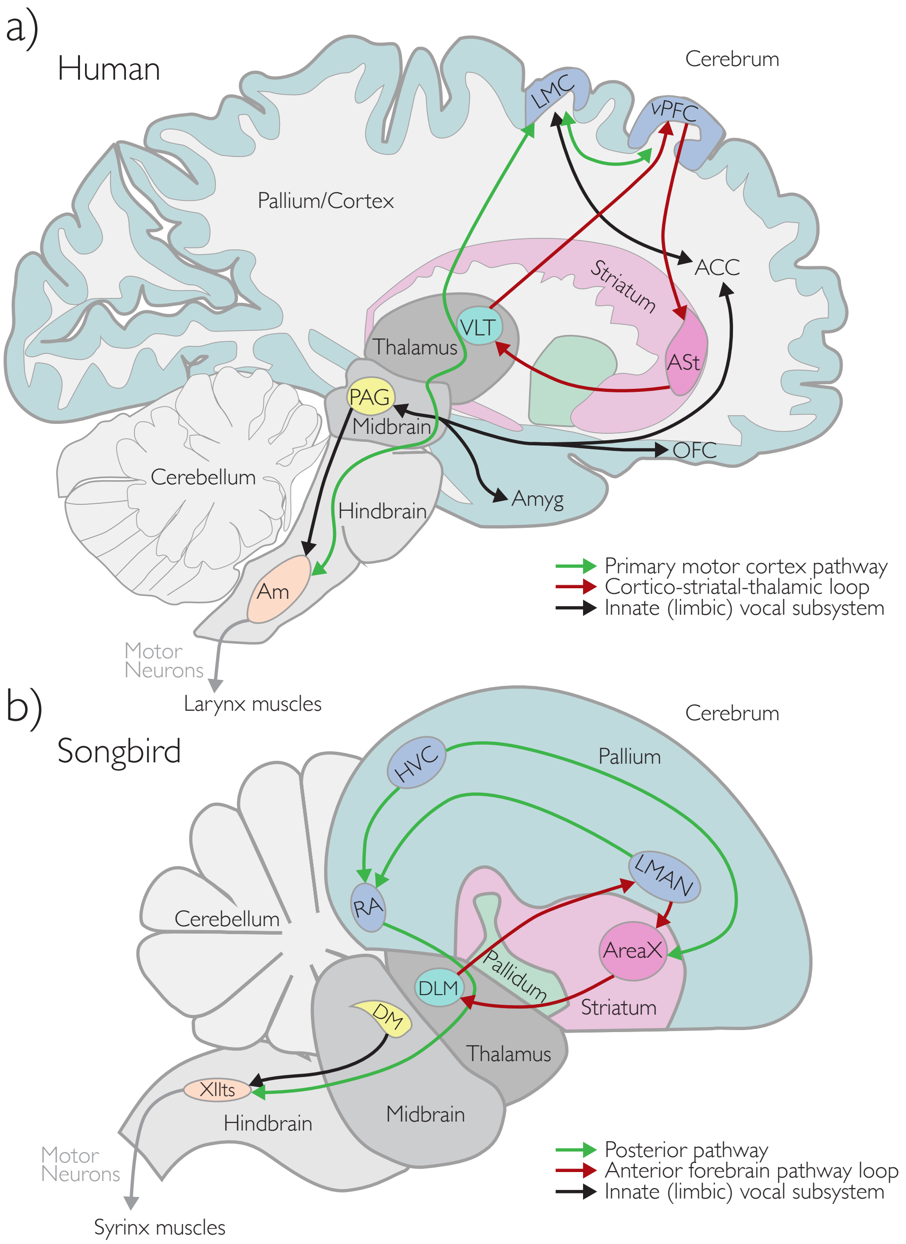


Figure 1.2: Schematic of vocal learning pathways in humans and songbirds. a) Schematic of the innate and learned neuronal pathways for human vocalisations. These pathways are mainly separate, but do overlap in some areas. The primary motor cortex pathway (indicated by green arrows) includes the cortico-striatal-thalamic loop for the learning of vocalisations (indicated by red arrows). Also shown is the limbic vocal subsystem involved in the production of innate vocalisations (black arrows). b) Schematic of the songbird vocal pathways, including the four major song nuclei (HVC, RA, AreaX, and LMAN). Green and red arrows indicate the posterior pathway and the anterior forebrain pathway loop, respectively. The limbic subsystem for innate vocalisations is shown by black arrows. Abbreviations: anterior cingulate cortex (ACC), nucleus ambiguus (Am), amygdala (Amyg), anterior striatum (ASt), dorsolateral nucleus of the medial thalamus (DLM), dorsal medial nucleus of the midbrain (DM), high vocal centre (HVC), lateral magnocellular nucleus of the anterior nidopallium (LMAN), laryngeal motor cortex (LMC), orbito-frontal cortex (OFC), periaqueductal grey (PAG), robust nucleus of the arcopallium (RA), ventro-lateral division of thalamus (VLT), ventral prefrontal cortex (vPFC), bird twelfth nerve nucleus (XIIts). Modified from Petkov and Jarvis (2012) and Mori and Wada (2015).

and the primary motor neurons controlling laryngeal musculature are hypothesised to be required for speech production ('Kuypers/Jürgens direct connections' hypothesis) (Jürgens, 2009; Fitch, 2011, 2018; Holstege and Subramanian, 2016).

The search for such direct connections in other vocal learning species is an important prerequisite for comparative approaches to the vocal learning trait, but investigations in other mammalian vocal learning species, such as seals, cetaceans, or elephants are as yet non-existent (Jarvis, 2006). Songbirds on the other hand, which have shown a remarkable capacity for vocal learning, have been investigated heavily on this subject. They share a general organisation of the nervous system with other vertebrates (five basic regions: the hindbrain, the midbrain, the thalamus, the cerebellum, and the cerebrum; **Fig. 1.2a**). However, in birds the cerebrum is not structured in a six-layered cortex as it is in mammals (reviews: Jarvis, 2004; Mori and Wada, 2015; Vernes and Wilkinson, 2019), but it is rather organised in large interconnected brain nuclei, the 'song system'. The song system includes four major song nuclei: the higher vocal centre, the robust nucleus of the arcopallium, the lateral magnocellular nucleus of the anterior nidopallium, and Area X (**Fig. 1.2b**; Mori and Wada, 2015). These nuclei are involved in an anterior and a posterior pathway allowing the necessary vocal control for vocal learning to be expressed. While the anterior forebrain pathway loop connects the pallium with the striatum and thalamus, the posterior pathway includes the RA, which directly projects onto the vocal motor neurons in the bird brainstem, which control the muscles of the syrinx (**Fig. 1.2b**; reviews: Jarvis, 2004; Bolhuis *et al.*, 2010; Petkov and Jarvis, 2012). Similar to the innate pathway in humans, all bird species studied to date get projections to the brainstem from the dorsal medial nucleus of the midbrain, which controls the production of innate vocalisations (**Fig. 1.2b**; Petkov and Jarvis, 2012). These parallels in specialised brain regions for song and speech between humans and songbirds enable the detailed functional study of vocal learning pathways (Pfenning *et al.*, 2014). The study of similarities and differences between the underlying morphology and neuronal motor-control pathways of vocal learning and non-learning species has been in the past and will in the future be an invaluable source of insight into the mechanics and evolution of the vocal learning trait.

1.4.3 Evolutionary considerations for the study of vocal learning

When using an animal model to study a human trait, it is important to consider the origin of the underlying similarity: Is the trait in question shared due to homologous or convergent evolution? Specifically, is the trait under investigation similar between the species because it originates from a common ancestor (e.g. human arm and whale flipper), or did it arise from an adaptive process to similar evolutionary pressures (e.g. development of wings in bats and flies)? Differentiating between convergent and homologous derived traits is important, as the questions that can be answered by different model systems are intrinsically different. Investigating a convergent trait allows insights into the pressures shaping the trait (e.g. reasons for the evolution of vocal learning in birds and bats) independent of the phylogenetic history of the species (e.g. independent evolution between birds and mammals for over 300 million years (Kumar and Hedges, 1998)). Studying homologous traits instead allows us to draw conclusions about the evolutionary history of the species directly (e.g. enables the investigation of the vocal learning trait within bats and answers the question if it evolved several times independently in bats or was secondarily lost in some species).

As vocal learning arose several times independently in vertebrates (probably three times independently in birds alone (Janik and Slater, 1997; Fitch, 2018)), this indicates evolutionary pressures in favour of this trait and speaks against the development as an evolutionary by-product. However, no singular evolutionary cause was identified in the species studied to date and the discussed reasons for the re-occurrence of vocal learning in the animal kingdom are manifold. In songbirds for example, vocal learning might have evolved in response to a number of selective pressures. (i) For example, a female preference for complex or varied song would support vocal learning (Fitch, 2005; Nowicki and Searcy, 2014). (ii) The imitation of vocalisations could also be beneficial, if a certain song is especially successful in a population or if its features make it so by e.g. carrying over long distances or evading background noises (Tyack, 2008; Nottebohm and Liu, 2010). (iii) An increase in vocal complexity has been hypothesised as causative agent for the development of vocal learning in other taxa as well (Sewall *et al.*, 2016). Other evolutionary pressures promoting vocal learning could have been (iv) group cohesion or (v) the display of group affiliation, which can result in the formation of vocal dialects (Tyack, 2008; Nottebohm and Liu, 2010). The current hypotheses for the evolution of vocal learning have been summarised as likely to be caused by local genetic adaptation, enhanced fitness of kin, more effective signalling, and/or enhanced social interactions (Nowicki and Searcy, 2014). Not only is the evolution of vocal learning in birds probably in response to a combination of several evolutionary drivers, but these evolutionary pressures are likely to be different for different species. The convergent evolution of vocal learning in different species provides us with the possibility to test evolutionary hypothesis in cross-species comparative analyses (Fitch, 2018).

1.4.4 Importance of comparative approaches to vocal learning

In order to understand the vocal learning capacity of any species in depth, a number of research questions need to be answered. These questions concern the fields of behavioural (WHAT makes a species a vocal learner?) and developmental biology (WHEN is vocal learning expressed?), as well as morphological, neuroanatomical, and genetic research (HOW is vocal learning expressed?). Furthermore, the phylogenetic (WHO else is capable of vocal learning?) and evolutionary basis (WHY did vocal learning evolve?) of the vocal learning trait need to be investigated (**Fig. 1.3**). Only by synergising within-species and comparative approaches will these different research areas be able to provide a comprehensive description of the vocal learning trait. However, it is also important to keep in mind that each vocal learning species only provides an approximation for the human capacity for speech and language acquisition. Moreover, no single species can provide all answers (e.g. about behaviour, neuronal control, genetic background) concerning vocal learning, but rather different species contribute different parts of the puzzle. Comparative studies, investigating several species on each of these levels are thus vital to gain an in-depth understanding of the general processes underlying vocal learning (**Fig. 1.3**; see chapter 2). While songbirds, and especially zebra finches, have been studied intensely in all of these fields in the past, studies of mammalian species to comparable depth are still lacking and thus wide gaps still remain in our knowledge about the underlying principles of mammalian vocal learning. In order to tackle this lack of knowledge, in this thesis I am focusing on the behavioural basis of vocal learning in bats, a mammalian model that in the past has repeatedly shown indications for vocal learning (Knörnschild, 2014).

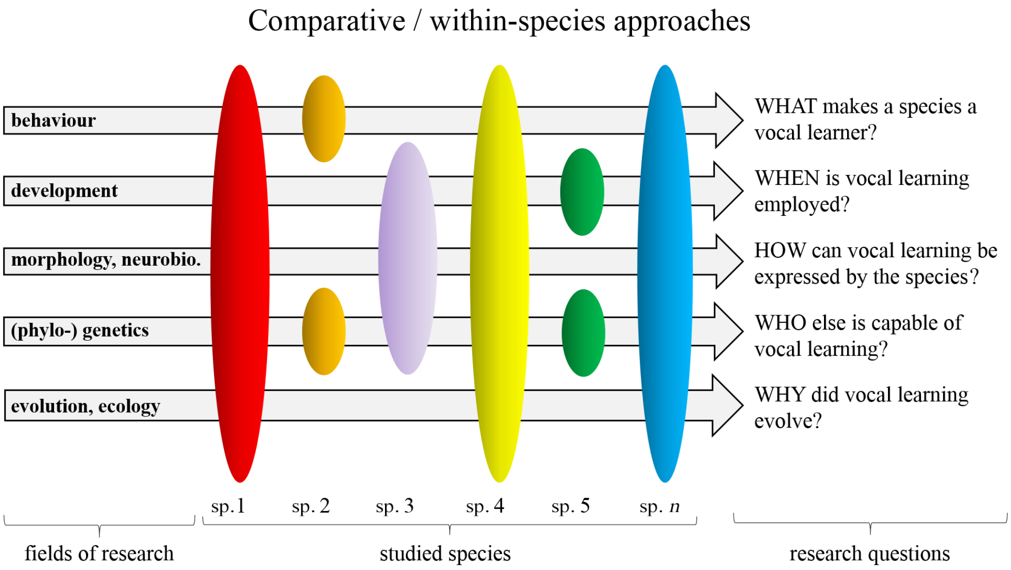


Figure 1.3: Schematic for comparative and within-species approaches to the study of vocal learning. In order to gain a comprehensive understanding of the complex vocal learning trait, a variety of different fields of research need to synergise. Each of these research areas focuses on its central research questions and cross-taxa comparative studies are essential to fully answer these. At the same time, it is important to combine research across fields in order to understand all different aspects leading to the expression of vocal learning in individual species. In zebra finches, for example, most aspects leading to vocal learning have been investigated to a greater or lesser extent. To date, no mammalian system has been studied in such depth, but in order to draw meaningful conclusions between vocal learning species, such a mammalian model system is urgently needed.

1.5 Bats

Bats (i.e. members of the order Chiroptera) are the only extant mammals capable of self-powered flight. Nevertheless, they are mostly known for their startling ability to navigate in complete darkness. Even though this phenomenon has been studied since the 18th century, the discovery of 'echolocation' as their main means of orientation, navigation, and foraging, took more than 100 years (review: Jones, 2005). Echolocation (also called 'biosonar') is an active process, in which the environment is probed by analysing echoes of self-emitted ultrasonic (i.e. above 20 kHz) vocalisations. Changes between the emitted call and the received echo, such as delay, intensity, and frequency of the calls, provide information about location, distance, and structure of the reflecting object (Neuweiler, 1990; Moss and Surlykke, 2010). With the aid of this active sensory system it is possible for bats to get detailed three-dimensional information about their surroundings without any visual input. Their ability to echolocate has made bats an especially interesting model system for the study of sensory motor-control of vocalisations, as they are highly dependent on effective and fast audio–vocal feedback. Echolocation is, however, not the only type of bat vocalisation. All bat species investigated to date make use of some form of social vocalisation and several species have even been shown to possess a broad and highly complex social call repertoire. Already in the late 1980s, some bat species were hypothesised to belong to the select club of animals that possess the capacity for vocal learning, starting with a study that showed that bat pups adjusted vocal parameters to their mothers' calls (Esser and Schmidt, 1989). Since then, evidence for vocal learning in several bat species has been accumulating. In the following sections, I will outline the factors that make bats intriguing models for the study of the vocal learning trait, focussing on the key aspects of behaviour, morphology, and genetics.

1.5.1 Vocal and social behaviour

The acquisition of new vocalisations in the course of vocal learning can lead to the attainment of an extensive and complex vocal repertoire. It must be said that a broad vocal repertoire is not evidence for a capacity for vocal learning, nor is a small vocal repertoire a definite disqualifier for this trait. However, a broad vocal repertoire can be used as an indicator for potentially promising candidate species. In bats, the size of the vocal repertoire is a species-specific feature and can range from only a couple of vocalisations (Barclay *et al.*, 1979) to more than 30 social calls (Kanwal *et al.*, 1994). To date, a number of bat social call repertoires have been described (e.g. Kanwal *et al.*, 1994; Bohn *et al.*, 2008, 2009; Gadziola *et al.*, 2012), but considering the diversity of the order Chiroptera with over 1300 extant species, comparatively little is known about their vocal communication. From the studies published so far, an encouraging image emerges: bat species were not only shown to have extensive social vocalisation repertoires containing complex and even combinatory syllables, but also to make use of them in specific behavioural contexts (Gillam and Fenton, 2016). Some bat species make use of their vocalisations in e.g. mating song (Behr and Von Helversen, 2004; Davidson and Wilkinson, 2004; Smotherman *et al.*, 2016; Knörnschild *et al.*, 2017), to signal group affiliation (Boughman and Wilkinson, 1998; Esser, 1998; Esser and Schubert, 1998; Knörnschild *et al.*, 2012), or identify offspring (Bohn *et al.*, 2004, 2007, 2013). This vocal complexity in itself does not render bats in general vocal learners, but it indicates them as a promising system for inquiry into the occurrence, origin, and expression of this trait.

Vocal complexity alone is not the sole indicator for a promising study system for vocal learning. Sociality and pronounced social behaviours are further interesting traits that could favour the emergence of vocal learning. Several social interactions are conceivable that would support the development of such a trait: One could imagine (i) that dispersing individuals might be better accepted in a new colony if they adjust to the local dialect; (ii) that mothers might identify their offspring if it vocally adjusted to the maternal call; or (iii) that groups might cohere better if they can converge on a group-specific vocalisation. In all of these hypothetical cases, the evolution of vocal learning would be supported by pronounced social behaviour, which can indeed be observed in many bat species. Most bat species (but not all) live in large colonies (up to millions of individuals). They thus have very frequent vocal interactions with related and unrelated individuals, which might support both the need to distinguish oneself (e.g. to develop an individual signature) and the need to blend in (e.g. to acquire group-specific calls). Furthermore, bats generally only have one or two offspring per year and exceptional longevity, which supports strong mother-offspring bonds (Ball *et al.*, 2018; Foley *et al.*, 2018). The close family and colony cohesions that can be observed in many bat species thus also support their position as a promising study system for vocal learning.

All three of the above-mentioned indications for vocal learning (i.e. vocal dialects, close mother-offspring bonds, and group cohesion calls) have been investigated in different bat species in the past. Promising behavioural evidence for bat vocal learning came for example from the discovery of vocal dialects in greater and pale spear-nosed bats (*Phyllostomus hastatus* and *P. discolor*, respectively), which were suggested to improve group cohesion especially during foraging (Boughman, 1998; Esser and Schubert, 1998). Unrelated greater spear-nosed bats from one roost were shown to share acoustic characteristics of their screech calls amongst each other, but to be different from individuals from a second roost (Boughman, 1998). Moreover, a transfer experiment mimicking natural dispersal found that these vocal characteristics were changed when individuals were translocated from one roost to another (Boughman, 1998). The greater sac-winged bat (*Saccopteryx bilineata*) was not only shown to have vocal dialects, but even differently sized vocal repertoires between different roosts (Davidson and Wilkinson, 2002). In order to show the effects of learning during development, several studies focused on bat vocal development: pups of three different bat species (specifically

P. discolor, *S. bilineata*, and *Rousettus aegyptiacus* (Egyptian fruit bat)) were shown to adjust their vocalisations to those of adults (mostly mothers) or even to playbacks (Esser and Schmidt, 1989; Knörnschild *et al.*, 2006, 2010, 2012; Prat *et al.*, 2017). Especially, *S. bilineata* pups were shown to undergo a vocal training phase similar to human babbling: during infancy, pups of both sexes produce renditions of all known adult vocalisation types, which appear to be independent of a distinct social context (Knörnschild *et al.*, 2006). The most convincing evidence for vocal learning in bats comes from experiments that demonstrated delayed or incomplete vocal development in acoustically isolated bat pups (*P. discolor* (Esser, 1994) and *R. aegyptiacus* (Prat *et al.*, 2015)). The development of vocal learning could thus be of great benefit for several bat species, as it supports their general vocal complexity, helps with group cohesion, and promotes mother-pup recognition. Further details on the evidence for vocal learning in bats will be given in the discussion (section 7.2), and specifically for *Phyllostomus discolor* also below in section 1.6.4.

1.5.2 Vocal tract morphology and neuroanatomy

The morphological structures for the generation and neural processing of sound are, at their core, shared between all mammalian species, including bats and humans. The underlying structure for the production of vocalisations shared among mammals is the larynx. This general mammalian vocal production system contains membrane tissue that, when vibrated by the air flow of the exhalation process, generates sound. While some mammalian species have modified the general bauplan of the vocal production system (e.g. baleen and toothed whales), most mammals still share this homologous morphology (Fitch, 2006; Smotherman, 2007; Saigusa, 2011; Janik, 2014). Different pitches of the produced vocalisations are controlled by the configuration (i.e. different states of tension) of the laryngeal muscles, which are innervated by the superior and inferior laryngeal nerves (Suthers and Fattu, 1982; Smotherman, 2007; Au and Suthers, 2014). Selective cutting of these laryngeal nerves in bats led to changes in emitted vocalisations, and even muteness (Suthers and Fattu, 1982; Rübsamen and Schäfer, 1990). This shared sound production system is an important similarity between mammals. Furthermore, this similarity is accompanied by a comparable mammalian neuroanatomy (e.g. existence of a layered cortex in humans and bats, but not in birds (**Fig. 1.2**; see section 1.4.2)) and similar neuronal motor-control pathways within the mammalian class (Metzner, 1996; Smotherman, 2007). This anatomical basis allows humans and bats, among others, to accurately control the spectro-temporal parameters of their vocalisations and underlines the suitability of bats as model system for the study of the mammalian vocal learning capacity.

In the past, several brain structures have been suggested to play a role in vocalisation control in bats. For example, electrical stimulation of the bat periaqueductal grey, the anterior cingulate cortex, and the paralemniscal area (i.e. an area located in a restricted region rostral and medial to the dorsal nucleus of the lateral lemniscus in the bat midbrain; cf. **Fig. 1.2a**) led to the emission of echolocation and/or social calls depending on the region of stimulation (Fenzl and Schuller, 2002, 2005, 2007). The fact that stimulation of different neuronal subpopulations of the periaqueductal grey leads to the emission of different types of vocalisations suggests separate, but probably overlapping, processing of vocal activity and discrete motor-control pathways (Gooler and Neill, 1987; Fenzl and Schuller, 2005, 2007; Au and Suthers, 2014; Vernes and Wilkinson, 2019). The neuronal motor-control of bat vocalisations has been researched intensely in the past decades, however the exact anatomical and physiological relationships of these pathways are far from understood (Vernes and Wilkinson, 2019). Although vocal motor-control and auditory pathways in bats have been understood in some detail, their involvement in vocal learning has not been investigated in the past. Future studies will help to illuminate the fundamental mechanisms and pathways involved in vocal production learning and will allow detailed comparative studies between vocal learning and non-learning species.

1.5.3 Genetics and evolution

Traditionally bats were categorised in two suborders depending on their use of laryngeal echolocation: Megachiroptera (which do not use laryngeal echolocation) and Microchiroptera (all species with laryngeal echolocation). This classification rested mainly on morphological and paleontological assessments and was supported by little molecular data. When genetic data became available, this historic concept was rejected (Teeling *et al.*, 2005) and two new suborders emerged: Yinpterochiroptera and Yangochiroptera. All non-echolocating Old World fruit bats (Pteropodidae) and four additional families of echolocating bats (Rhinolophidae, Megadermatidae, Rhinopomatidae, and Craseonycteridae) were pooled in the Yinpterochiroptera, whereas the Yangochiroptera contain all other families of echolocating bats (Teeling *et al.*, 2005; Jones and Teeling, 2006). This new classification presents a better understanding of the evolutionary relationships amongst bats and to their sister species, but it is not yet comprehensive enough to make distinct statements about the species-level phylogeny of bats. Thus, great research efforts are currently aimed at sequencing the complete genomes of representatives of all bat species (Bat1K project (Teeling *et al.*, 2018)). These genomes will then provide us with the opportunity to design testable hypotheses for evolutionary drivers of vocal learning in bats and will enable us to confirm or disprove its homologous origin in bats.

Aside from the ongoing genomic study of vocal learning in bats, a number of genetic studies have been performed on them in order to investigate the genetic basis of this trait. In recent years, genetic studies were for example able to map gene expression of human language-relevant genes in two different bat species (Rodenas-Cuadrado *et al.*, 2018) and furthermore identify transcriptomic networks in the bat periaqueductal grey (i.e. an important brain region for human spoken language and speech production), which were found to have a relationship to bat vocalisations (Rodenas-Cuadrado *et al.*, 2015). These studies are essential to bridge the fields of human speech research and vocal learning as they compare gene expression of language-related genes in different vocal learning species. To further probe the function of specific genes involved in the vocal learning process, genetic modifications are crucial. Current research (of the Neurogenetics of Vocal Communication lab (Max Planck Institute for Psycholinguistics) and others) is focused on studying such behavioural consequences of vocal-learning-relevant gene knock-downs in several bat species.

1.5.4 Suitability of bats as a study system for vocal learning

Bats show several characteristics that make them well-suited mammalian model systems for the study of vocal learning: their diverse family tree allows for a comparative approach to the evolutionary pressures shaping the vocal learning trait (reviews: Knörnschild, 2014; Vernes, 2017); their mammalian neuroanatomy enables us to identify neuronal pathways that might be shared with humans (review: Esser, 2003); their accessibility in large numbers and easy captive maintenance allow amongst others for genetic manipulations, electrophysiological and long-term studies (reviews: Konopka and Roberts, 2016a; b); and their highly developed ability to precisely produce and process acoustic signals predisposes them for the study of the biological basis of vocal learning and ultimately human speech. For these reasons, bats are an exceptionally well-suited animal model for the study of vocal learning. However, research on bat vocal learning so far has been spread over a variety of species and mainly focused on behavioural studies (Knörnschild, 2014; Vernes, 2017). No single bat species has been used to study all of the aspects of vocal learning in detail. On the one hand, this splitting up of study questions over a number of species leads to an increase of available information about bats as an order, but on the other hand, these findings are often difficult to transfer between species. Therefore, the overarching goal of my thesis research was to focus on a single species (i.e. the pale spear-nosed bat) and to expand the existing body of knowledge about this species' behaviour in order to establish it as a mammalian model system for the study of vocal learning.

1.6 Model system: the pale spear-nosed bat, *Phyllostomus discolor*

Phyllostomus discolor is a highly gregarious Neotropical bat species, which lives in large colonies with up to several hundred individuals (Kwiecinski, 2006). They are omnivorous and not strictly seasonal, and thus easy to keep and breed in captivity. They have one to two pups per year and are long lived (more than 14 years; personal communication: breeding facility at the Ludwig Maximilian University of Munich), which makes them well-suited for the study of developmental and long-term effects. Pale spear-nosed bats have been used in a variety of behavioural, psychoacoustic, and electrophysiological experiments researching their sound perception, neuronal vocal control, and echolocation system. Thus, much is known about the details of their auditory processing capacities, their neurophysiology and neuronal control over vocalisations, and their vocal development. I shortly summarise previous findings of these areas of research in the subsections below (sections 1.6.1–1.6.4).

1.6.1 Hearing in *Phyllostomus discolor*

The hearing capacity of *P. discolor* was first described in a behavioural audiogram reporting two sensitivity peaks in the hearing threshold, which seem to correspond in their bandwidth either to species-specific echolocation or communication calls (Esser and Daucher, 1996). Although these two peaks were confirmed by a second study using behavioural and neuronal audiograms (Hoffmann *et al.*, 2008a), they are likely a result of the acoustic filter features of external physiological structures, especially of the pinnae (De Mey *et al.*, 2008). This is supported by the absence of the two peaks in subsequent cochlear sensitivity measurements (specifically, distortion product otoacoustic emissions measurements; Wittekindt *et al.*, 2005). Recently, the development of hearing sensitivity in the first six post-natal months was studied using auditory brainstem response measurements. It was demonstrated that *P. discolor* is sensitive to a broad range of sounds already in the first week after birth, but also that its hearing capacity undergoes further sensitisation and spectral refinement during the first half year (Linnenschmidt and Wiegrebe, 2019). The detailed detection and discrimination thresholds for temporal and spectral features of acoustic stimuli were typically investigated with the help of psychoacoustic experiments employing two- or three-alternative-forced-choice setups. For example, the capacity of *P. discolor* for highly sensitive frequency discrimination (Esser and Kiefer, 1996) and for the detection of high spatial frequencies (Baier *et al.*, 2019) was ascertained with this method. Furthermore, *P. discolor* was shown to possess the ability to detect faint echo amplitude and temporal modulations, which are important for detection of prey-generated flutter (Baier and Wiegrebe, 2018; Baier *et al.*, 2018).

1.6.2 Neuroanatomy and neuronal control of vocalisations

In the past two decades, a number of neurophysiological studies have investigated the general neuronal anatomy of *P. discolor* and also specifically their neuronal control over vocalisations. It was found that acoustic stimulation mainly affects four anatomically distinguishable fields in the bat neocortex: a ventral and a dorsal part of the auditory cortex, which are each subdivided into an anterior and a posterior part (Hoffmann *et al.*, 2008b). The two ventral fields were distinguished by their tonotopic organisation with opposing frequency gradients, which is comparable to the primary auditory cortex and the anterior auditory field of other mammals. The two dorsal cortical fields contained neurons that were only responsive to high frequencies, often used in the echolocation calls of this species (Hoffmann *et al.*, 2008b). Neurons in the anterior region of the auditory cortex were even shown to rate code specific spectral properties of acoustic signals (i.e. echo roughness) (Firzlaff *et al.*, 2006). Moreover, cortical neurons in the posterior dorsal field of the auditory cortex were shown to preferably match the fast amplitude and frequency modulations of *P. discolor* social calls (Hörpel and Firzlaff, 2019). The neuronal control over vocalisations in *P. discolor* was shown to be influenced by several brain regions. Especially the periaqueductal grey matter and the region of the paralemniscal area were shown to contribute to vocalisation control (Fenzl and Schuller, 2002, 2005, 2007). While electrical microstimulation of the rostral periaqueductal grey region was able to elicit communication calls only, stimulation of the dorsal and ventral edges of the periaqueductal grey elicited echolocation and communication calls. Stimulation of the paralemniscal area only elicited echolocation calls (Fenzl and Schuller, 2002). This suggests different neuronal pathways for particular types of communication and echolocation calls, which then converge on the motorneurons used in the production of the vocalisations (Fenzl and Schuller, 2005, 2007). The demonstrated functional role of the periaqueductal grey in vocal pathways makes it an especially interesting brain area for the investigation of vocal control. Although auditory processing and neuronal pathways for vocal production have been well understood in *P. discolor*, nothing is known about the involvement of these structures in the vocal learning process.

1.6.3 Types of vocalisations and vocal development

Phyllostomus discolor, like most bat species, makes use of two forms of vocalisations: echolocation and social calls. Adult echolocation calls cover a frequency range of 40 to 100 kHz, are 0.5–3 ms long, and are mostly used for orientation and navigation purposes (Grunwald *et al.*, 2004; Goerlitz *et al.*, 2008; Heinrich *et al.*, 2011; Guarato *et al.*, 2013). Echolocation calls are spontaneously emitted only from postnatal day 10 onwards, when they are still long (i.e. 5–12 ms in first weeks of life) and low in minimum frequency. They then decrease in duration to the adult length of 1–3 ms (from fourth week onwards) and increase in minimum frequency (Rother and Schmidt, 1985; Luo *et al.*, 2017). The second form of vocalisations used by *P. discolor* is social calls. ‘Social calls’ is a collective noun for a variety of calls of longer duration (>3 ms) and with lower frequencies (fundamental frequency between ~4 and 20 kHz), which are used in different behavioural contexts. Until recently, only two types of *P. discolor* social vocalisations had been described in detail: infant isolation calls and maternal directive calls. Both, infant isolation calls and maternal directive calls, are quasi-sinusoidally frequency modulated vocalisations of 40–80 ms length, which are produced antiphonally during mother-pup reunions (Rother and Schmidt, 1985; Esser and Schmidt, 1989). Isolation calls are emitted already from postnatal day 0 to 45 and show marked individual differences, which persist throughout ontogenesis and allow mothers to discriminate between pups (Rother and Schmidt, 1985). A more extensive acoustic and behavioural assessment of the social call repertoire of this bat species has thus far only been attempted in an unpublished diploma thesis (Pistohl, 1998). However, being informed about the vocal repertoire of a species is crucial for the study of vocal learning, as knowledge about previously existing vocalisation types allows for the detection of novel or modified learned vocalisations. Heretofore, comprehensive information about the vocal complexity of *P. discolor* was largely lacking (see chapter 3).

1.6.4 Evidence for vocal learning

Although the earliest indications for vocal learning in *P. discolor* were revealed in the late 1980s, studies on the behavioural evidence for vocal learning are still sparse. Tentative indications for vocal learning come from the demonstration of persistent acoustic differences in the maternal directive calls from different captive *P. discolor* colonies (Esser and Schubert, 1998). While such vocal dialects are indicative for learned social modifications of vocalisations, they do not present sufficient evidence for the existence of the vocal learning trait (Janik and Slater, 1997; Vernes and Wilkinson, 2019). However, further evidence was provided by the demonstration of pup vocal adjustment towards maternal directive calls (Esser and Schmidt, 1989). Studying such closely-related associations (i.e. mother-pup pairs) it is difficult to exclude possible genetic effects. Therefore, a subsequent study was conducted, in which pups were hand-reared in isolation from adult conspecifics. Hand-reared pups were shown to structurally converge (i.e. in respect to the frequency modulation patterns and modulation depth) on playbacks of a maternal directive call. After 100 days, calls of the hand-reared pups resembled the playback significantly more than those of a control group, which was reared in acoustic isolation (Esser, 1994). This means that vocal convergence of pups towards adult vocalisations occurs independent of the level of genetic relatedness between pup and tutor. These different lines of evidence indicate *P. discolor* as a vocal learning species, however experiments with adult individuals have not been conducted as yet.

1.7 Open questions and thesis outline

How can we conclusively prove that a species is a vocal learner? To approach an answer to this question, in this thesis I studied the Neotropical bat species *Phyllostomus discolor* with the objective of addressing three main areas:

- Investigation of the basis of their vocalisations: How acoustically and behaviourally complex are the vocal emissions of the pale spear-nosed bat?
- Conditioning of vocalisations: Can *P. discolor* vocalisations be brought under operant control (i.e. is it possible to elicit vocal emissions in isolated individuals)?
- Study of their vocal imitation capacity: To what extent are these bats capable of controlling the characteristics of their vocal emissions and adjusting their vocalisations to auditory targets?

Answering these questions allowed me to gain a better understanding of the vocal learning capacities of *P. discolor*. Moreover, it led to the development of controlled behavioural paradigms, which will in the future enable the study of further developmental questions (e.g. 'Does *P. discolor* have a critical learning period? Which learning trajectories do bats follow during their development?'). Studying *P. discolor* also further highlighted their suitability as model system for the neurobiological and genetic study of vocal learning. The findings of my research on the vocal learning capacity of the pale spear-nosed bat are collected in this thesis.

In the **second chapter** of this thesis, I describe the challenges of studying vocal learning and present an approach for unifying the investigation of vocal learning across taxa. Incorporating previously suggested definitions, I propose a multi-layered study of related fields of research, which in combination provide an in-depth understanding of a species' vocal learning capacity and its placement along the vocal learning continuum. For such an approach, different fields (such as behavioural ecology, developmental biology, genetics, neurobiology, phylogenetics, and evolutionary biology) need to investigate the vocal learning capacity of a species and, by answering their respective research questions, define its capacity comprehensively. In this chapter, I again point out the importance of animal models for the study of vocal learning and discuss open questions in the field.

Having highlighted the importance of comparative approaches in the previous chapter, in the **third chapter** I focus on the significance of acoustic perception for vocal learning. Knowledge about sound perception is a prerequisite for understanding the ability to react to or imitate acoustic impressions. If the correct perception of a target sound is hindered (e.g. by a low temporal or spectral resolution of the animals' hearing), the attempted imitation of this acoustic target will always seem erroneous. Being informed about the acoustic perceptional abilities of a species is thus important for the evaluation of its imitation and learning success. In this chapter, I compare hearing thresholds of eleven Neotropical bat species and put their hearing thresholds in context with their species-specific echolocation and communication calls.

In order to show vocal learning, the acquisition of new vocalisations or modification of known vocalisations through auditory experience needs to be demonstrated (Janik and Slater, 1997, 2000; Boughman and Moss, 2003). To show the acquisition of such a novel vocalisation, the prior assessment of the vocal repertoire is required. For most vocally communicating species studied to date, the assessment of their vocal repertoire has either not even been attempted or not studied in its entirety. In the **fourth chapter**, I concentrate on my focal species, *P. discolor*. In order to better understand the vocal complexity of *P. discolor*, I assess its vocal repertoire and associated behaviours. By establishing an innovative new recording setup, I successfully recorded social vocalisations of adult *P. discolor* bats in an undisturbed social roosting context. The recorded calls were then classified and associated with behavioural contexts via video recordings. This chapter provides the knowledge base of *P. discolor* vocalisations, which can be used to assess novelty of learned vocalisations.

For the controlled study of individual vocal learning under laboratory conditions, it is essential to train animals to emit vocalisations in an isolated context. This way it is possible to separate the effect of individual vocal learning from social influences and differences in the animals' state of excitement. In the **fifth chapter** of this thesis, I present a new setup and training regime that, for the first time, allowed vocal conditioning of isolated bats and enabled us to demonstrate vocal usage learning and vocal plasticity in *P. discolor*. Inspiration for this training regime came from decades of successful operant conditioning of zebra finches and budgerigars (Manabe *et al.*, 1997, 2008). However, hitherto such operant vocal conditioning had not been established in bats.

Using a modification of the method and setup established in chapter five, in the **sixth chapter** I focus on the demonstration of the vocal learning capacity in these bats. Making use of a sophisticated multi-layered training regime, I demonstrate the bats' ability to change the spectral composition of their calls, which is an essential prerequisite for advanced forms of vocal learning. Specifically, I trained the bats to change the fundamental frequency of their vocalisations in a directional manner. Furthermore, I describe the attempt to use the same training paradigm to train the bats to structurally imitate complex acoustic targets and discuss the shortcomings of the setup for such an advanced task.

The **seventh chapter** serves as platform for the general discussion of my findings and allows me to frame my research in the context of current definitions of vocal learning. I elaborate on the insights my research provides for the central questions around vocal learning and defend my conclusions on the basis of the presented research and the existing body of knowledge on the subject. Lastly, I indicate persisting gaps in our understanding concerning vocal learning, specifically in bats, and suggest further promising research avenues.

References

Abramson, Z.; Herna, M.V.; Garcı, L.; Colmenares, F.; Aboitiz, F. and Call, J. (2018). Imitation of novel conspecific and human speech sounds in the killer whale (*Orcinus orca*). *Proceedings of the Royal Society B* 285, 1–9.

Alexander, R.D. (1957). Sound production and associated behavior in insects. *The Ohio Journal of Science* 57, 101–113.

Au, W.W.L. and Suthers, R.A. (2014). Production of biosonar signals: structure and form. In: A. Surlykke, P. E. Nachtigall, R. R. Fay, and A. N. Popper (Eds), *Biosonar - Springer Handbook of Auditory Research*. Springer-Verlag, New York, pp. 61–105.

Baier, A.L.; Stelzer, K.-J. and Wiegrebe, L. (2018). Flutter sensitivity in FM bats. Part II: amplitude modulation. *Journal of Comparative Physiology A* 204, 941–951.

Baier, A.L. and Wiegrebe, L. (2018). Flutter sensitivity in FM bats. Part I: delay modulation. *Journal of Comparative Physiology A* 204, 929–939.

Baier, A.L.; Wiegrebe, L. and Goerlitz, H.R. (2019). Echo-imaging exploits an environmental high-pass filter to access spatial information with a non-spatial sensor. *iScience* 14, 335–344.

Ball, H.C.; Levari-Shariati, S.; Cooper, L.N. and Aliani, M. (2018). Comparative metabolomics of aging in a long-lived bat: insights into the physiology of extreme longevity. *PLoS ONE* 13, 1–20.

Barclay, R.M.R.; Fenton, M.B. and Thomas, D.W. (1979). Social behavior of the little brown bat, *Myotis lucifugus*. *Behavioral Ecology and Sociobiology* 6, 137–146.

Behr, O. and Von Helversen, O. (2004). Bat serenades - Complex courtship songs of the sac-winged bat (*Saccopteryx bilineata*). *Behavioral Ecology and Sociobiology* 56, 106–115.

Bohn, K.M.; Boughman, J.W.; Wilkinson, G.S. and Moss, C.F. (2004). Auditory sensitivity and frequency selectivity in greater spear-nosed bats suggest specializations for acoustic communication. *Journal of Comparative Physiology A* 190, 185–192.

Bohn, K.M.; Schmidt-French, B.; Ma, S.T. and Pollak, G.D. (2008). Syllable acoustics, temporal patterns, and call composition vary with behavioral context in Mexican free-tailed bats. *The Journal of the Acoustical Society of America* 124, 1838–1848.

Bohn, K.M.; Schmidt-French, B.; Schwartz, C.; Smotherman, M. and Pollak, G.D. (2009). Versatility and stereotypy of free-tailed bat songs. *PLoS ONE* 4, 1–11.

Bohn, K.M.; Smarsh, G.C. and Smotherman, M. (2013). Social context evokes rapid changes in bat song syntax. *Animal Behaviour* 85, 1485–1491.

Bohn, K.M.; Wilkinson, G.S. and Moss, C.F. (2007). Discrimination of infant isolation calls by female greater spear-nosed bats, *Phyllostomus hastatus*. *Animal Behaviour* 73, 423–432.

Bolhuis, J.J.; Okanoya, K. and Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience* 11, 747–759.

Boughman, J.W. (1998). Vocal learning by greater spear-nosed bats. *Proceedings of the Royal Society B* 265, 227–233.

Boughman, J.W. and Moss, C.F. (2003). Social sounds: vocal learning and development of mammal and bird calls. In: A. M. Simmons, R. R. Fay, and A. N. Popper (Eds), *Acoustic Communication - Springer Handbook of Auditory Research*. Springer-Verlag New York, NY, pp. 138–224.

Boughman, J.W. and Wilkinson, G.S. (1998). Greater spear-nosed bats discriminate group mates by vocalizations. *Animal Behaviour* 55, 1717–1732.

Bradbury, J.W. and Vehrencamp, S.L. (2011). *Principles of animal communication*. 2nd ed. Sinauer Associates Oxford University Press, Oxford, UK.

Chomsky, N. (2006). *Language and mind*. 3rd ed. Cambridge University Press, New York, NY. Available from: www.cambridge.org/9780521858199.

Christiansen, M.H. and Kirby, S. (2003). Language evolution: consensus and controversies. *Trends in Cognitive Sciences* 7, 300–307.

Dalziell, A.H. and Magrath, R.D. (2012). Fooling the experts: accurate vocal mimicry in the song of the superb lyrebird, *Menura novaehollandiae*. *Animal Behaviour* 83, 1401–1410.

Davidson, S.M. and Wilkinson, G.S. (2002). Geographic and individual variation in vocalizations by male *Saccopteryx bilineata* (Chiroptera: Emballonuridae). *Journal of Mammalogy* 83, 526–535.

Davidson, S.M. and Wilkinson, G.S. (2004). Function of male song in the greater white-lined bat, *Saccopteryx bilineata*. *Animal Behaviour* 67, 883–891.

Devanna, P.; Dediu, D. and Vernes, S.C. (2019). The genetics of language: from complex genes to complex communication. In: S.-A. Rueschemeyer and M. G. Gaskell (Eds), *The Oxford Handbook of Psycholinguistics*. Oxford University Press, Oxford, UK, pp. 865–898.

Durand, S.E.; Heaton, J.T.; Amateau, S.K. and Brauth, S.E. (1997). Vocal control pathways through the anterior forebrain of a parrot (*Melopsittacus undulatus*). *The Journal of Comparative Neurology* 377, 179–206.

Elemans, C.P.H.; Rasmussen, J.H.; Herbst, C.T.; Düring, D.N.; Zollinger, S.A.; Brumm, H.; Srivastava, K.; Svane, N.; Ding, M.; Larsen, O.N.; Sober, S.J. and Vec, J.G.Š. (2015). Universal mechanisms of sound production and control in birds and mammals. *Nature Communications* 6, 1–13.

Esser, K.-H. and Daucher, A. (1996). Hearing in the FM-bat *Phyllostomus discolor*: a behavioral audiogram. *Journal of Comparative Physiology A* 178, 779–785.

Esser, K.H. (1994). Audio-vocal learning in a non-human mammal: the lesser spear-nosed bat *Phyllostomus discolor*. *NeuroReport* 5, 1718–1720.

Esser, K.H. (1998). Psychoacoustic studies in neotropical bats. In: S. Nielzén and O. Olsson (Eds), *Clinical Psychoacoustics*. Lund University Press, Lund, Sweden, pp. 45–59.

Esser, K.H. (2003). Modeling aspects of speech processing in bats - behavioral and neurophysiological studies. *Speech Communication* 41, 179–188.

Esser, K.H. and Kiefer, R. (1996). Detection of frequency modulation in the FM-bat *Phyllostomus discolor*. *Journal of Comparative Physiology A* 178, 787–796.

Esser, K.H. and Schmidt, U. (1989). Mother-infant communication in the lesser spear-nosed bat *Phyllostomus discolor* (Chiroptera, Phyllostomidae) - evidence for acoustic learning. *Ethology* 82, 156–168.

Esser, K.H. and Schubert, J. (1998). Vocal dialects in the lesser spear-nosed bat *Phyllostomus discolor*. *Naturwissenschaften* 85, 347–349.

Evans, N. and Levinson, S.C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *The Behavioral and Brain Sciences* 32, 429–492.

Favaro, L.; Neves, S.; Furlati, S.; Pessani, D.; Martin, V. and Janik, V.M. (2016). Evidence suggests vocal production learning in a cross-fostered Risso's dolphin (*Grampus griseus*). *Animal Cognition* 19, 847–853.

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

Fenzl, T. and Schuller, G. (2002). Periaqueductal gray and the region of the paralemniscal area have different functions in the control of vocalization in the neotropical bat, *Phyllostomus discolor*. *European Journal of Neuroscience* 16, 1974–1986.

Fenzl, T. and Schuller, G. (2005). Echolocation calls and communication calls are controlled differentially in the brainstem of the bat *Phyllostomus discolor*. *BMC Biology* 3, 1–12.

Fenzl, T. and Schuller, G. (2007). Dissimilarities in the vocal control over communication and echolocation calls in bats. *Behavioural Brain Research* 182, 173–179.

Firzlaff, U.; Schörnich, S.; Hoffmann, S.; Schuller, G. and Wiegrebe, L. (2006). A neural correlate of stochastic echo imaging. *Journal of Neuroscience* 26, 785–791.

Fisher, S.E. and Marcus, G.F. (2006). The eloquent ape: genes, brains and the evolution of language. *Nature Reviews: Genetics* 7, 9–20.

Fisher, S.E. and Vernes, S.C. (2015). Genetics and the language sciences. *Annual Review of Linguistics* 1, 289–310.

Fitch, T. (2006). Production of vocalizations in mammals. In: K. Brown (Ed), *Encyclopedia of Language & Linguistics*. Elsevier (incl. Pergamon), pp. 115–121.

Fitch, W.T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences* 4, 258–267.

Fitch, W.T. (2005). The evolution of language: a comparative review. *Biology & Philosophy* 20, 193–203.

Fitch, W.T. (2010). *The evolution of language*. first edit. Cambridge University Press, Cambridge, UK.

Fitch, W.T. (2011). The evolution of syntax: an exaptationist perspective. *Frontiers in Evolutionary Neuroscience* 3, 1–12.

Fitch, W.T. (2018). The biology and evolution of speech: a comparative analysis. *Annual Review of Linguistics* 4, 255–279.

Fitch, W.T.; de Boer, B.; Mathur, N. and Ghazanfar, A.A. (2016). Monkey vocal tracts are speech-ready. *Science Advances* 2, 1–7.

Fitch, W.T. and Hauser, M.D. (2003). Unpacking "honesty": vertebrate vocal production and the evolution of acoustic signals. In: A. Simmons, R. R. Fay, and A. N. Popper (Eds), *Acoustic Communication.* Springer Handbook of Auditory Research. Springer New York, NY, pp. 65–137.

Foley, N.M.; Hughes, G.M.; Huang, Z.; Clarke, M.; Jebb, D.; Whelan, C. V.; Petit, E.J.; Touzalin, F.; Farcy, O.; Jones, G.; Ransome, R.D.; Kacprzyk, J.; O'Connell, M.J.; Kerth, G.; Rebelo, H.; Rodrigues, L.; Puechmaille, S.J. and Teeling, E.C. (2018). Growing old, yet staying young: the role of telomeres in bats' exceptional longevity. *Science Advances* 4, 1–12.

Gadziola, M.A.; Grimsley, J.M.S.S.; Faure, P.A. and Wenstrup, J.J. (2012). Social vocalizations of big brown bats vary with behavioral context. *PLoS ONE* 7, e44550.

Gillam, E. and Fenton, M.B. (2016). Roles of acoustic social communication in the lives of bats. In: M. B. Fenton, A. D. Grinnell, A. N. Popper, and R. R. Fay (Eds), *Bat Bioacoustics*. Springer Handbook of Auditory Research. Springer New York, NY, pp. 117–139.

Goerlitz, H.R.; Hübner, M. and Wiegrebe, L. (2008). Comparing passive and active hearing: spectral analysis of transient sounds in bats. *The Journal of Experimental Biology* 211, 1850–1858.

Gooler, D.M. and Neill, W.E.O. (1987). Topographic representation of vocal frequency demonstrated by microstimulation of anterior cingulate cortex in the echolocating bat, *Pteronotus parnelli parnelli*. *Journal of Comparative Physiology A* 161, 283–294.

Grothe, B. and Pecka, M. (2014). The natural history of sound localization in mammals - a story of neuronal inhibition. *Frontiers in Neural Circuits* 8, 1–19.

Grunwald, J.-E.; Schörnich, S. and Wiegrebe, L. (2004). Classification of natural textures in echolocation. *PNAS* 101, 5670–5674.

Guarato, F.; Windmill, J. and Gachagan, A. (2013). A beam based method for target localization: inspiration from bats' directivity and binaural reception for ultrasonic sonar. *The Journal of the Acoustical Society of America* 133, 4077–4086.

Haesler, S.; Rochefort, C.; Georgi, B.; Licznerski, P.; Osten, P. and Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus area X. *PLoS Biology* 5, 2885–2897.

Hagoort, P. and Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience* 37, 347–362.

Hauser, M.D.; Chomsky, N. and Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579.

Heinrich, M.; Warmbold, A.; Hoffmann, S.; Firzlaff, U. and Wiegrebe, L. (2011). The sonar aperture and its neural representation in bats. *Journal of Neuroscience* 31, 15618–15627.

Hoffmann, S.; Baier, L.; Borina, F.; Schuller, G.; Wiegrebe, L. and Firzlaff, U. (2008)a). Psychophysical and neurophysiological hearing thresholds in the bat *Phyllostomus discolor*. *Journal of Comparative Physiology A* 194, 39–47.

Hoffmann, S.; Firzlaff, U.; Radtke-Schuller, S.; Schwellnus, B. and Schuller, G. (2008)b). The auditory cortex of the bat *Phyllostomus discolor*: localization and organization of basic response properties. *BMC Neuroscience* 9, 1–17.

Holstege, G. and Subramanian, H.H. (2016). Two different motor systems are needed to generate human speech. *Journal of Comparative Neurology* 524, 1558–1577.

Hörpel, S.G. and Firzlaff, U. (2019). Processing of fast amplitude modulations in bat auditory cortex matches communication call-specific sound features. *Journal of Neurophysiology* 121, 1501–1512.

Janik, V.M. (2014). Cetacean vocal learning and communication. *Current Opinion in Neurobiology* 28, 60–65.

Janik, V.M. and Slater, P.J.B. (1997). Vocal learning in mammals. *Advances in the Study of Behaviour* 26, 59–99.

Janik, V.M. and Slater, P.J.B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour* 60, 1–11.

Jarvis, E.D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences* 1016, 749–777.

Jarvis, E.D. (2006). Selection for and against vocal learning in birds and mammals. *Ornithological Science* 5, 5–14.

Jones, G. (2005). Echolocation. *Current Biology* 15, 484–488.

Jones, G. and Teeling, E.C. (2006). The evolution of echolocation in bats. *Trends in Ecology & Evolution* 21, 149–156.

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews* 26, 235–258.

Jürgens, U. (2009). The neural control of vocalization in mammals: a review. *Journal of Voice* 23, 1–10.

Kanwal, J.S.; Matsumura, S.; Ohlemiller, K. and Suga, N. (1994). Analysis of acoustic elements and syntax in communication sounds emitted by mustached bats. *The Journal of the Acoustical Society of America* 96, 1229–1254.

Knörnschild, M. (2014). Vocal production learning in bats. *Current Opinion in Neurobiology* 28, 80–85.

Knörnschild, M.; Behr, O. and von Helversen, O. (2006). Babbling behavior in the sac-winged bat (*Saccopteryx bilineata*). *Naturwissenschaften* 93, 451–454.

Knörnschild, M.; Blüml, S.; Steidl, P.; Eckenweber, M. and Nagy, M. (2017). Bat songs as acoustic beacons - Male territorial songs attract dispersing females. *Scientific Reports* 7, 1–11.

Knörnschild, M.; Nagy, M.; Metz, M.; Mayer, F. and Von Helversen, O. (2010). Complex vocal imitation during ontogeny in a bat. *Biology Letters* 6, 156–159.

Knörnschild, M.; Nagy, M.; Metz, M.; Mayer, F. and Von Helversen, O. (2012). Learned vocal group signatures in the polygynous bat *Saccopteryx bilineata*. *Animal Behaviour* 84, 761–769.

Konopka, G. and Roberts, T.F. (2016)a). Animal models of speech and vocal communication deficits associated with psychiatric disorders. *Biological Psychiatry* 79, 53–61.

Konopka, G. and Roberts, T.F. (2016)b). Insights into the neural and genetic basis of vocal communication. *Cell* 164, 1269–1276.

Kumar, S. and Hedges, S.B. (1998). A molecular timescale for vertebrate evolution. *Nature* 392, 917–920.

Kwiecinski, G.G. (2006). Phyllostomus discolor. *Mammalian Species*, 1–11.

Levinson, S.C. and Evans, N. (2010). Time for a sea-change in linguistics: response to comments on "the myth of language universals." *Lingua* 120, 2733–2758.

Lieberman, P.H.; Klatt, D.H. and Wilson, W.H. (1969). Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science* 164, 1185–1187.

Linnenschmidt, M. and Wiegrebe, L. (2019). Ontogeny of auditory brainstem responses in the bat, *Phyllostomus discolor*. *Hearing Research* 373, 85–95.

Luo, J.; Lingner, A.; Firzlaff, U. and Wiegrebe, L. (2017). The Lombard effect emerges early in young bats: implications for the development of audio-vocal integration. *The Journal of Experimental Biology* 220, 1032–1037.

Manabe, K.; Dooling, R.J. and Brittan-Powell, E.F. (2008). Vocal learning in budgerigars (*Melopsittacus undulatus*): effects of an acoustic reference on vocal matching. *The Journal of the Acoustical Society of America* 123, 1729–1736.

Manabe, K.; Staddon, J.E.R. and Cleaveland, J.M. (1997). Control of vocal repertoire by reward in budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology* 111, 50–62.

Marcus, G.F. and Fisher, S.E. (2003). FOXP2 in focus: what can genes tell us about speech and language? *Trends in Cognitive Sciences* 7, 257–262.

Metzner, W. (1996). Anatomical basis for audio-vocal integration in echolocating horseshoe bats. *The Journal of Comparative Neurology* 368, 252–69.

De Mey, F.; Reijniers, J.; Peremans, H.; Otani, M. and Firzlaff, U. (2008). Simulated head related transfer function of the phyllostomid bat *Phyllostomus discolor*. *Journal of the Acoustical Society of America* 124, 212–232.

Mori, C. and Wada, K. (2015). Songbird: a unique animal model for studying the molecular basis of disorders of vocal development and communication. *Experimental Animals* 64, 221–230.

Moss, C.F. and Surlykke, A. (2010). Probing the natural scene by echolocation in bats. *Frontiers in Behavioral Neuroscience* 4, 1–16.

Neuweiler, G. (1990). Auditory adaptations for prey capture in echolocating bats. *Physiological Reviews* 70, 615–641.

Nottebohm, F. and Liu, W.C. (2010). The origins of vocal learning: new sounds, new circuits, new cells. *Brain and Language* 115, 3–17.

Nowicki, S. and Searcy, W.A. (2014). The evolution of vocal learning. *Current Opinion in Neurobiology* 28, 48–53.

Oller, D.K.; Wieman, L.A.; Doyle, W.J. and Ross, C. (1976). Infant babbling and speech. *Journal of Child Language* 3, 1–11.

Pepperberg, I.M. (2010). Vocal learning in grey parrots: a brief review of perception, production, and cross-species comparisons. *Brain and Language* 115, 81–91.

Petkov, C.I. and Jarvis, E.D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience* 4, 1–24.

Pfenning, A.R.; Hara, E.; Whitney, O.; Rivas, M.V.; Wang, R.; Roulhac, P.L.; Howard, J.T.; Wirthlin, M.; Lovell, P.V.; Ganapathy, G.; Mouncastle, J.; Moseley, M.A.; Thompson, J.W.; Soderblom, E.J.; Iriki, A.; Kato, M.; Gilbert, M.T.P.; Zhang, G.; Bakken, T.; Bongaarts, A.; Bernard, A.; Lein, E.; Mello, C.V.; Hartemink, A.J. and Jarvis, E.D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* 346, 1256846.

Pistohl, D. (1998). Struktur und Funktion sozialer Laute bei der Kleinen Lanzennase, *Phyllostomus discolor* (Chiroptera, Phyllostomidae). *Unpublished diploma thesis*, Ulm, Germany.

Poeppel, D. (2012). The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology* 29, 34–55.

Poeppel, D.; Emmorey, K.; Hickok, G. and Pylkkanen, L. (2012). Towards a new neurobiology of language. *Journal of Neuroscience* 32, 14125–14131.

Prat, Y.; Azoulay, L.; Dor, R. and Yovel, Y. (2017). Crowd vocal learning induces vocal dialects in bats: playback of conspecifics shapes fundamental frequency usage by pups. *PLoS Biology*, 1–14.

Prat, Y.; Taub, M. and Yovel, Y. (2015). Vocal learning in a social mammal: demonstrated by isolation and playback experiments in bats. *Science Advances* 1, e1500019.

Price, C.J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage* 62, 816–847.

Rabb, G.B. (1960). On the unique sound production of the surinam toad (*Pipa pipa*). *Copeia*, 368–369.

Reby, D. and McComb, K. (2003). Vocal communication and reproduction in deer. *Advances in the Study of Behavior* 33, 231–264.

Reichmuth, C. and Casey, C. (2014). Vocal learning in seals, sea lions, and walruses. *Current Opinion in Neurobiology* 28, 66–71.

Richards, D.G.; Wolz, J.P. and Herman, L.M. (1984). Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus*. *Journal of Comparative Psychology* 98, 10–28.

Ridgway, S.; Carder, D.; Jeffries, M. and Todd, M. (2012). Spontaneous human speech mimicry by a cetacean. *Current Biology* 22, R860–R861.

Riede, T. and Goller, F. (2010). Peripheral mechanisms for vocal production in birds - differences and similarities to human speech and singing. *Brain and Language* 115, 69–80.

Roberts, T.F.; Gobes, S.M.H.; Murugan, M.; Ölveczky, B.P. and Mooney, R. (2012). Motor circuits are required to encode a sensory model for imitative learning. *Nature Neuroscience* 15, 1454–1459.

Rodenas-Cuadrado, P.; Chen, X.S.; Wiegrebe, L.; Firzlaff, U. and Vernes, S.C. (2015). A novel approach identifies the first transcriptome networks in bats: a new genetic model for vocal communication. *BMC Genomics* 16, 1–18.

Rodenas-Cuadrado, P.M.; Mengede, J.; Baas, L.; Devanna, P.; Schmid, T.A.; Yartsev, M.; Firzlaff, U. and Vernes, S.C. (2018). Mapping the distribution of language related genes FoxP1, FoxP2, and CntnaP2 in the brains of vocal learning bat species. *Journal of Comparative Neurology* 526, 1235–1266.

Rother, G. and Schmidt, U. (1985). Die ontogenetische Entwicklung der Vokalisation bei *Phyllostomus discolor* (Chiroptera). *Zeitschrift für Säugetierkunde* 50, 17–26.

Rothman, R.J. and Mech, L.D. (1979). Scent-marking in lone wolves and newly formed pairs. *Animal Behaviour* 27, 750–760.

Rübsamen, R. and Schäfer, M. (1990). Audiovocal interactions during development? Vocalisation in deafened young horseshoe bats vs. audition in vocalisation-impaired bats. *Journal of Comparative Physiology A* 167, 771–784.

Saigusa, H. (2011). Comparative anatomy of the larynx and related structures. *Japan Medical Association Journal* 54, 241–247.

Scharff, C.; Kirn, J.R.; Grossman, M.; Macklis, J.D. and Nottebohm, F. (2000). Targeted neuronal death affects neuronal replacement and vocal behavior in adult songbirds. *Neuron* 25, 481–492.

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

Schreiweis, C.; Bornschein, U.; Burguière, E.; Kerimoglu, C.; Schreiter, S.; Dannemann, M.; Goyal, S.; Rea, E.; French, C.A.; Puliyadi, R.; Groszer, M.; Fisher, S.E.; Mundry, R.; Winter, C.; Hevers, W.; Pääbo, S.; Enard, W. and Graybiel, A.M. (2014). Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proceedings of the National Academy of Sciences of the United States of America* 111, 14253–14258.

Schusterman, R.J. (2008). Vocal learning in mammals with special emphasis on pinnipeds. In: D. K. Oller and U. Gribel (Eds), *The Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication*. MIT Press, Cambridge, MA, pp. 41–70.

Sewall, K.B.; Young, A.M. and Wright, T.F. (2016). Social calls provide novel insights into the evolution of vocal learning. *Animal Behaviour* 120, 163–172.

Sharoh, D.; van Mourik, T.; Bains, L.J.; Segaert, K.; Weber, K.; Hagoort, P. and Norris, D.G. (2019). Laminar specific fMRI reveals directed interactions in distributed networks during language processing. *PNAS*, 1–6.

Simonyan, K. and Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist* 17, 197–208.

Smotherman, M.; Knörnschild, M.; Smarsh, G. and Bohn, K. (2016). The origins and diversity of bat songs. *Journal of Comparative Physiology A* 202, 535–554.

Smotherman, M.S. (2007). Sensory feedback control of mammalian vocalizations. *Behavioural Brain Research* 182, 315–326.

Stoeger, A.S. and Manger, P. (2014). Vocal learning in elephants: neural bases and adaptive context. *Current Opinion in Neurobiology* 28, 101–107.

Stoeger, A.S.; Mietchen, D.; Oh, S.; De Silva, S.; Herbst, C.T.; Kwon, S. and Fitch, W.T. (2012). An Asian elephant imitates human speech. *Current Biology* 22, 2144–2148.

Suthers, R.A. and Fattu, J.M. (1982). Selective laryngeal neurotomy and the control of phonation by the echolocating bat, *Eptesicus*. *Journal of Comparative Physiology A* 145, 529–537.

Teeling, E.C.; Springer, M.S.; Madsen, O.; Bates, P.; O'brien, S.J. and Murphy, W.J. (2005). A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 307, 580–584.

Teeling, E.C.; Vernes, S.C.; Dávalos, L.M.; Ray, D.A.; Gilbert, M.T.P. and Myers, E. (2018). Bat biology, genomes, and the Bat1K project: to generate chromosome-level genomes for all living bat species. *Annual Review of Animal Biosciences* 6, 23–46.

Tomasello, M.; Call, J.; Nagell, K.; Olguin, R. and Carpenter, M. (1994). The learning and use of gestural signals by young chimpanzees: a trans-generational study. *Primates* 35, 137–154.

Tyack, P.L. (2008). Convergence of calls as animals form social bonds, active compensation for noisy communication channels, and the evolution of vocal learning in mammals. *Journal of Comparative Psychology* 122, 319–331.

Vernes, S.C. (2017). What bats have to say about speech and language. *Psychonomic Bulletin and Review* 24, 111–117.

Vernes, S.C. and Wilkinson, G.S. (2019). Behaviour, biology, and evolution of vocal learning in bats. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 1–31.

Wang, Y.; Sereno, J.A.; Jongman, A. and Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience* 15, 1019–1027.

Wilbrecht, L. and Nottebohm, F. (2003). Vocal learning in birds and humans. *Mental Retardation and Developmental Disabilities Research Reviews* 9, 135–148.

Willems, R.M. and van Gerven, M.A.J. (2018). New fMRI methods for the study of language. In: S.-A. Rueschemeyer and M. G. Gaskell (Eds), *The Oxford Handbook of Psycholinguistics*. Oxford University Press, Oxford, UK, pp. 1–20.

Williams, N. (2003). Click to connect. *Current Biology* 13, R211–R212.

Winter, P.; Schott, D.; Ploog, D. and Handley, P. (1973). Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour* 47, 230–239.

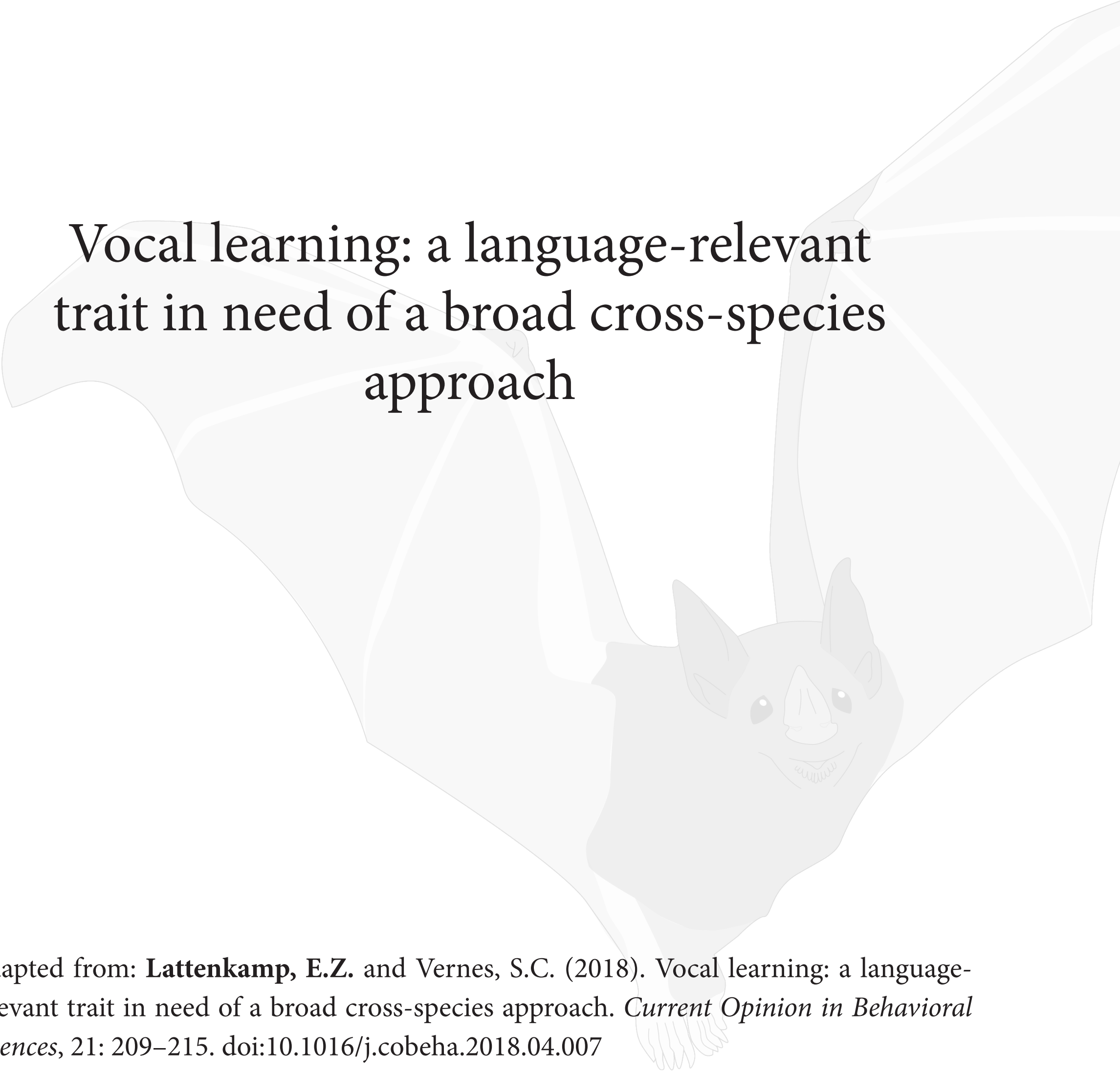
Wittekindt, A.; Drexl, M. and Kössl, M. (2005). Cochlear sensitivity in the lesser spear-nosed bat, *Phyllostomus discolor*. *Journal of Comparative Physiology A* 191, 31–36.

Wright, T.F. (1996). Regional dialects in the contact call of a parrot. *Proceedings of the Royal Society of London: Series B* 263, 867–872.

Zatorre, R.J. (2013). Predispositions and plasticity in music and speech learning: Neural correlates and implications. *Science* 342, 585–589.

Chapter 2

Vocal learning: a language-relevant trait in need of a broad cross-species approach



Adapted from: **Lattenkamp, E.Z.** and Vernes, S.C. (2018). Vocal learning: a language-relevant trait in need of a broad cross-species approach. *Current Opinion in Behavioral Sciences*, 21: 209–215. doi:10.1016/j.cobeha.2018.04.007

Abstract

Although humans are unmatched in their capacity to produce speech and learn language, comparative approaches in diverse animal models are able to shed light on the biological underpinnings of language-relevant traits. In the study of vocal learning, a trait crucial for spoken language, passerine birds have been the dominant models, driving invaluable progress in understanding the neurobiology and genetics of vocal learning despite being only distantly related to humans. To date, there is sparse evidence that our closest relatives, non-human primates have the capability to learn new vocalisations. However, a number of other mammals have shown the capacity for vocal learning, such as some cetaceans, pinnipeds, elephants, and bats, and we anticipate that with further study more species will gain membership to this (currently) select club. A broad, cross-species comparison of vocal learning, coupled with careful consideration of the components underlying this trait, is crucial to determine how human speech and spoken language is biologically encoded and how it evolved. We emphasise the need to draw on the pool of promising species that have thus far been understudied or neglected. This is by no means a call for fewer studies in songbirds, or an unfocused treasure hunt, but rather an appeal for structured comparisons across a range of species, considering phylogenetic relationships, ecological, and morphological constrains, developmental and social factors, and neurogenetic underpinnings. Herein, we promote a comparative approach highlighting the importance of studying vocal learning in a broad range of model species, and describe a common framework for targeted cross-taxon studies to shed light on the biology and evolution of vocal learning.

Human speech and the need for animal models

Human speech and language acquisition is unparalleled within the animal kingdom. Human speech is dependent on the integration of auditory perception and vocal production, both during early development for babies learning their first words (Lipkind *et al.*, 2013; Rasilo and Räsänen, 2017), and also later in life where audio-vocal feedback is essential for the maintenance of accurate speech production (Lee *et al.*, 2015; Mitsuya *et al.*, 2017). Although decades of research have elucidated many processes involved in human speech development, the full complexity of the mechanisms underlying human spoken language acquisition, like the integrated processes by which humans perceive, memorise, and compare auditory input to then guide vocal motor production, are still to be understood (Enard, 2011; Fitch, 2017). Deciphering the biological underpinnings of these abilities, such as the neuronal circuits or genetic mechanisms driving vocal learning, requires invasive *in vivo* studies that are largely technically and/or ethically impossible in humans. Such studies, including single cell electrophysiology, optogenetics, or genetic manipulations are essential to gain a thorough understanding of the biological processes underlying this trait. For this reason, animal models, in which controlled experiments can dissect the behavioural, neurological, and genetic components of language-relevant traits, are of great importance.

Vocal learning: a language-relevant trait

An essential component of human speech production and the acquisition of spoken language is the capacity for vocal production learning (herein referred to simply as 'vocal learning'). Vocal learning is defined as the ability to acquire novel vocalisations or modify existing vocalisations as a result of auditory experience (Janik and Slater, 2000; Boughman and Moss, 2003). Vocal learning is the basis for the capacity to learn sounds and words used for human speech, but has also been identified in a handful of non-human species. This seemingly straightforward definition actually entails a complex multilevel process including auditory experience, neuronal integration, and vocal production (**Fig. 2.1**).

Species-specific peculiarities of sound perception are a first limiting step for the ability to adjust vocalisations according to an auditory input. Auditory perception, including physical sound perception, signal filtering, and auditory processing, shapes and limits the input for subsequent neuronal processing of the sound (**Fig. 2.1**). Once perceived, vocal signals need to be processed at the neural level, first being memorised as target sounds, and in later recurrences compared to the auditory percept of self-emitted vocalisations. Neuro-motor control mechanisms are able to initialise sound production predominantly via the vocal apparatus. Sound generation can be achieved by vibrating the vocal folds (phonatory system) via exhalation (respiratory system), whereby the oscillation speed of the vocal folds determines the pitch of the vocalisation. The articulators in the vocal tract are used to filter the produced sound (filter system), which is a key step in the process of speech production (**Fig. 2.1**; Fitch, 2000). The accurate interplay between detailed sound perception, a precise neuronal interface, and a flexible vocal-motor production apparatus is essential to enable a dynamic adjustment of vocal emission to previously perceived and memorised auditory targets.

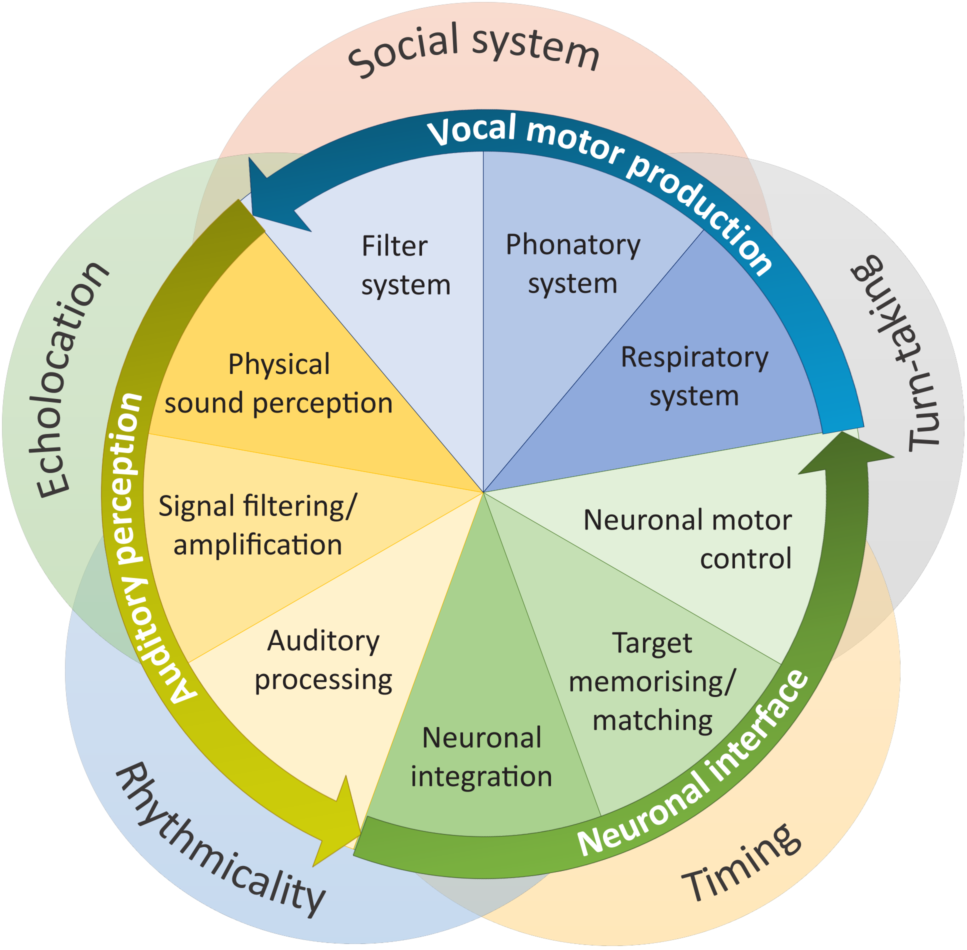


Figure 2.1: Conceptualisation of vocal learning. Three levels of organisation involved in vocal learning are illustrated (signal perception, neuronal integration, and vocal production) including their individual sublevels. The surroundings depict some of the hypothesised prerequisites for, or suggested traits correlated with, vocal learning — for which experimental evidence is still lacking or scarce.

Vocal learning animals

Vocal learning has so far been studied most extensively in passerines or 'songbirds' and this research has provided invaluable insights into the physiological pathways, evolutionary factors, and gene expression patterns underlying vocal learning (Wada *et al.*, 2004; Pfenning *et al.*, 2014; Wheatcroft and Qvarnström, 2015; Roberts *et al.*, 2017). Even though passerines are not the only vertebrate group, in fact not even the only order of birds, exhibiting vocal learning, they have some attributes that have made them a preferred model system. Songbirds are easy to keep, breed, and study and have a remarkable ability to copy detailed and complex song. Focusing on an easy and widely accessible study organism has certain benefits. The long-standing, combined focus of many labs around the world has advanced the study of vocal learning in songbirds immensely and provided detailed information about most levels of organisation from developmental processes and behavioural phenotypes to morphological structures, brain circuitry, and gene expression (e.g. Brainard and Doupe, 2002; Agate *et al.*, 2009; Elemans, 2014; Pfenning *et al.*, 2014; Elemans *et al.*, 2015). Ease, accessibility, amenability to study, as well as historic factors, have led to a strong scientific focus on birds, such that when we performed a survey of published papers on vocal learning from the last 25 years, we found that papers on birds represented ~84% (n = 378) of all original research articles (**Fig. 2.2a,b**). Despite vocal learning also being identified in bats, pinnipeds, cetaceans, and elephants, only 38 studies on vocal learning in these species (~8% of all articles) have been published over the same period (**Fig. 2.2c**). Encouragingly, the field now seems to be expanding to explore these and other less traditional models that also show intriguing abilities to imitate structure, spectral content, and timing of acoustic signals. Several mammalian species found across varied habitats have shown evidence for vocal learning. Aquatic mammals (cetaceans, baleen whales, toothed whales, and dolphins) have shown vocal learning abilities both in the wild and in captive studies. Evidence for cetacean vocal learning includes adoption of new vocalisations from parents or conspecifics, copying new whistle sounds, and novel song learning and synchronisation (Janik, 2014). In singular instances, pinnipeds (semiaquatic marine mammals), and elephants (terrestrial mammals) have demonstrated the ability to imitate human speech (Schusterman, 2008; Stoeger *et al.*, 2012; Ravignani *et al.*, 2016). Moreover, several species of bats (the only mammals capable of self-powered flight) have shown indications of vocal learning (Knörnschild, 2014; Prat *et al.*, 2015). Bat species have shown individual and group signatures in their social calls, imitation of maternal calls by pups, and adjustment of spectral call parameters after transfer of animals between social groups (Boughman, 1998; Knörnschild *et al.*, 2010, 2012).

Together these examples are already indicative for the existence of a broad spectrum of vocal learning species. However, only a small fraction of vertebrate species have thus far been tested, let alone recognised for their vocal learning abilities. For

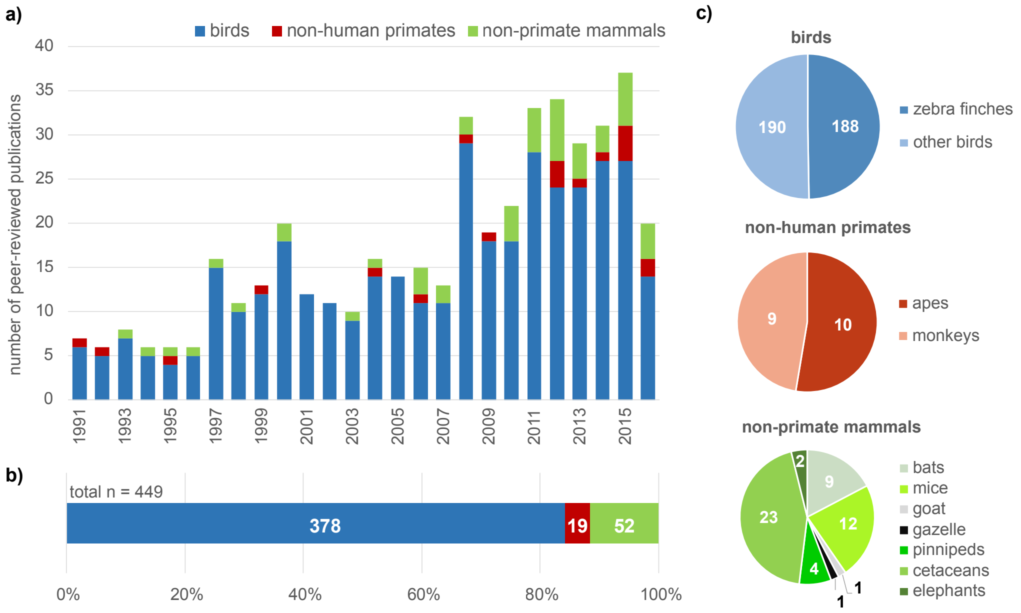


Figure 2.2: Primary research articles on 'vocal learning' from 1991 to 2016. (a) The presented numbers resulted from a search in the Web of Science Core Collection. The search term 'vocal learning' produced a total of 580 results in the section 'articles' (from 1970 to 2016), when searched for in 'topic'. Between 1970 and 1990 only 14 articles were published (precisely 2 studies on zebra finches, 8 on other bird species, 1 study on a monkey species, and 3 non-primary research papers), which were excluded from the presented results. We further manually excluded a total of 117 papers, of which 93 did not present primary research (methodological, modelling, reviews, opinion paper, etc.), 22 were studies on humans, and 1 study each was on frogs and cichlid fish. (b) This led us to a total of 449 peer-reviewed articles between 1991 and 2016. The articles were categorised into three vertebrate groups (birds, non-human primates, and non-primate mammals) and the total number of peer-reviewed articles for each category is given in the graphic. (c) The three categories were split to provide more detail about the studied taxa within each group. Note that different search platforms might produce slightly different results to these numbers, which were mined from Web of Science, however the same search performed using a different platform (Pubmed) returned very similar numbers.

some mammals only one or a handful of individuals were ever reported to exhibit vocal learning. Nevertheless, if the observation is meaningful, it can be enough to establish a whole species as vocal learners with value for further study, as was the case for elephants (Stoeger *et al.*, 2012). These incidental observations should raise awareness for the large number of species that are possibly overlooked for their potential.

The vocal learning continuum

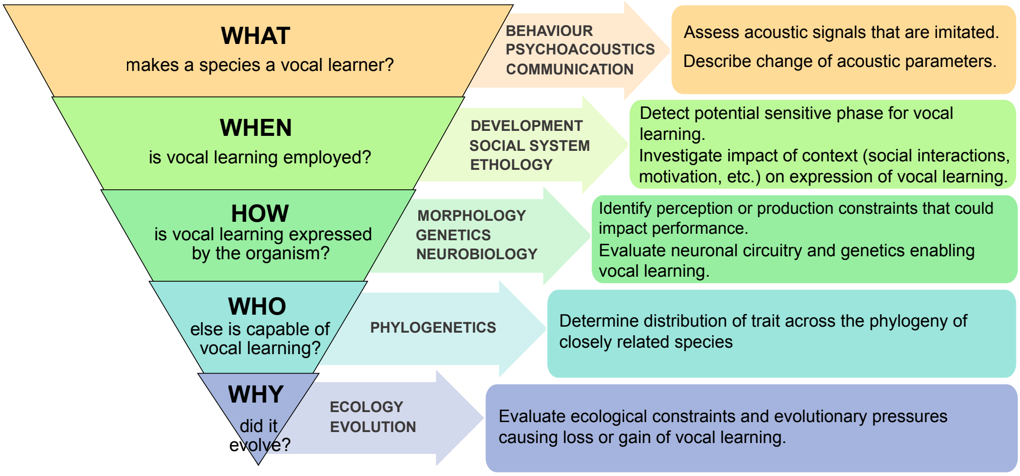
Despite a gradual increase in numbers of recognised vocal learning species, the method of classification of these species has only marginally changed over time. Vocal learning species are still frequently described bimodally as being either vocal learners or non-learners. More valuable however is the idea of a vocal learning continuum — which has previously been proposed as a continuous distribution of species with increasing complexity of vocal learning ability (Petkov and Jarvis, 2012). Although the continuum has been represented as a two dimensional scale, it is probably better thought of as a multidimensional space in which each species has differing abilities and limitations. The vocal learning continuum must take into account capacities and constraints at all levels of organisation; auditory perception, neuronal interface, and vocal-motor production (**Fig. 2.1**). And as noted previously, a large proportion of species are expected to cluster at the low end of the vocal learning continuum (Petkov and Jarvis, 2012). Considering where each species falls on this hypothetical continuum — including non-learning species — will provide much greater insight into shared and divergent characteristics of each animal than a bimodal classification system, and thus greater potential for success in the comparative approach.

A proposed theoretical framework

To extract the most benefit from the comparative approach, a theoretical framework should be established, which enables the structured search for parallels and divergences between species at multiple levels and ultimately allows the full understanding of the communicative system and its evolution. Here, we attempt to develop a framework that could be built upon to facilitate a focused research approach when studying species for their vocal learning ability (**Fig. 2.3**). For each species the following could be considered and compared:

1. WHAT makes a species a vocal learner? Which acoustic signals do they imitate and how do they change their vocalisations? Some animals are capable of copying only species- specific sounds, for example, zebra finches learn a short stereotyped song that they faithfully repeat throughout their lifetime (Slater *et al.*, 1988; Tchernichovski *et al.*, 2001). By comparison, nightingales learn a massive repertoire of sounds that they combine into song in a flexible manner (Todt *et al.*, 1979). Not all forms of vocal learning involve learning from conspecifics as some species, for example, mockingbirds, are capable of mimicking the songs of other animals, and some particularly talented parrots, elephants, and seals can even mimic human speech (with varying degrees of success) (Schusterman, 2008; Pepperberg, 2010; Gammon and Altizer, 2011; Stoeger *et al.*, 2012). Thus it is important to consider what type of signal an animal is capable of learning and to what degree the vocal repertoire can be modified. At one end of the spectrum vocal modifications may involve subtle changes to the structure and spectral content of calls that always retain species-specific characteristics. At the other extreme are animals that can learn a dazzling array of vocalisations with seemingly few restrictions.

2. WHEN is vocal learning employed? Are they able to perform vocal learning during their whole life time or only in specific sensitive phases? Do they require a certain type of context, such as conspecific interaction (e.g. parental influence) or motivation (e.g. presence/absence of social input)? Exploring vocal learning abilities at the appropriate



Assess acoustic signals that are imitated.
Describe change of acoustic parameters.

Detect potential sensitive phase for vocal learning.
Investigate impact of context (social interactions, motivation, etc.) on expression of vocal learning.

Identify perception or production constraints that could impact performance.
Evaluate neuronal circuitry and genetics enabling vocal learning.

Determine distribution of trait across the phylogeny of closely related species

Evaluate ecological constraints and evolutionary pressures causing loss or gain of vocal learning.

Figure 2.3: A proposed theoretical framework: levels at which vocal learning can be studied and associated exemplary questions. The field of vocal learning is of interest for different research areas and is thus approached from different angles. The detailed study of one organism on all these levels (e.g. zebra finches) is important and leads to the in-depth understanding of a confined model system. However, comprehensive cross-species comparisons within each of these levels will advance the understanding of the evolution and transferable requirements for vocal learning as a whole.

ontogenic stage for each individual species is crucial to avoid false-negative classifications. Many animals have been shown to have a flexible learning period early in life, but lose the ability to imitate new sounds later on (open- vs. closed-ended learners) (Brainard and Doupe, 2002). The developmental stage of a potential model species is thus often a sensitive and limiting factor in the study of vocal learning as is it possible to overlook vocal learning if an organism is investigated before it starts producing imitations (e.g. in the memorising phase), or likewise when the sensitive phase for auditory target learning is already over. In many species only one of the sexes actually uses the learned vocalisations, for example, in zebra finches, juvenile male birds learn from an adult male tutor (Tchernichovski *et al.*, 2001; Brainard and Doupe, 2002). Social factors can also determine vocal learning, for example, the superb starling shows no difference in male and female song, but differences in song are rather attributed to social rank (Pilowsky and Rubenstein, 2013). In bats where male–male competition is often strongly expressed and male song is used for territorial defence and mate attraction (Smotherman *et al.*, 2016), vocal learning might also be restricted to males, but this assumption should be rigorously tested in both sexes for bats, and for any other potential vocal learning species. Motivation, influenced by the social environment as well as other factors could also influence the ability of an animal to learn vocalisations. The first reported instance of vocal learning in elephants was hypothesised to have occurred because of deprivation of conspecific company at a particularly sensitive time point during development (Stoeger *et al.*, 2012). Social interactions or a conspecific rich/poor environment might therefore have a strong influence on the expression, and thus visibility, of vocal learning. If the animal is not sufficiently motivated or there is not enough pressure on the animal to learn and modify their vocal output, an animal may appear unable to learn vocalisations, even though the required genetics, morphological, and neuronal structures are present.

3. HOW can vocal learning be expressed by the organisms? Are they subject to physical constraints that impact their auditory perception? How do neural circuitry and gene expression facilitate the required complex processes? How precise is the control of the vocal production apparatus and are there morphological/production constraints? Morphological (e.g. vocal and/or auditory apparatus), neurobiological, and genetic factors shape the biology of an animal, allowing it to perform the complex task of vocal learning. Studies comparing these factors across songbirds and humans have suggested deep-homology in neural circuitry (e.g. cortico-basal ganglia circuits) and genetics (e.g. the *FOXP2* gene) that point to convergent evolutionary mechanisms (Fisher and Scharff, 2009; Enard, 2011; Wheatcroft and Qvarnström, 2015). However, investigation of these features in a broad range of species is needed to determine if these mechanisms are common to all vocal learners. Identifying morphology, gene pathways, and brain structures in one declared vocal learner group is not enough to declare them a prerequisite for vocal learning in general. For example, the notion of 'the structure of a vocal learning brain' is to be taken with caution as it is currently, by necessity, largely based on data from a limited number of closely related bird species (Nelson *et al.*, 1995; Pfenning *et al.*, 2014; Elemans *et al.*, 2015). In order to find structures, circuits, and genes that enable vocal learning, these traits need to be compared at similar resolution across a range of vocal learners, preferably across distinct evolutionary clades. Such studies can also point to reasons why some species are not vocal learners. For example, the morphological differences in the vocal tract of humans and non-human primates have long been posited as a limiting factor in the ability of primates to produce human-like vocalisations. Recent work has shown however that a non-human primate vocal tract is capable of producing speech-like sounds showing that it is instead likely that differences in the neuronal wiring underlying vocal learning and vocal-motor control caused the gap in vocal learning capacity between humans and our closest primate relatives (Fitch *et al.*, 2016; Boë *et al.*, 2017).

4. WHO else is capable of vocal learning? Which of the close relatives of these species shows comparable vocal learning abilities? How is this trait distributed in the phylogeny of the focal species? As can be seen from **figure 2.2**, animals with evidence for vocal learning are spread across the evolutionary tree. However, aside from birds, evidence often comes from one or a few animal species within an order. For example, four species of bats have good evidence as vocal learners, another handful of species are thought to be promising for the study of vocal learning, but to date no bat species has been confirmed as a vocal non-learner (Knörnschild, 2014; Prat *et al.*, 2015; Smotherman *et al.*, 2016). Identifying the distribution of this trait (in its different incarnations as discussed above), and determining which species do not display the trait, will be crucial to pinpointing the biological factors necessary and sufficient for vocal learning and for understanding how this trait evolved.

5. WHY did vocal learning evolve? Which ecological conditions and resulting evolutionary pressures cause a species to gain or lose the capacity for vocal learning? Previously several evolutionary benefits of vocal learning have been proposed (Nowicki and Searcy, 2014). However, which ecological conditions support selection for the generation of a broader call repertoire or create the need for increased information-sharing will have to be investigated across a diverse range of species-specific habitat and living conditions. Taken together, understanding which species have the capacity for vocal learning, how they express the trait at a behavioural level, and the biological factors that underlie this trait will shed light on the evolutionary pressures that caused species to gain or lose this complex and evolutionarily expensive trait. While this framework is by no means exhaustive, if these and related questions are addressed for every species studied for their ability to perform vocal learning, a comparative approach could yield a detailed and informed assignment of different species along the vocal learning continuum. We emphasise that not all of the above-mentioned levels of vocal learning need to be addressed for the comparative approach to work, but they could serve as a research guide. A framework such as this will be key to allow meta-analyses across multiple species, which will be crucial for a clear understanding of the evolution and biological encoding of vocal learning, human speech, and spoken language.

Advantages of the broad comparative approach

With a nuanced approach to the multidimensional continuum of vocal learning abilities and more in-depth understanding of the contributing biological factors will also come the ability to understand the larger forces shaping this trait over evolutionary time. Several traits have previously been identified as frequently co-occurring with, and thus hypothesised to be important for, the evolution of vocal learning (**Fig. 2.1**), but comparative studies are essential to test such hypotheses. With a broader comparative approach to the array of vocal learning abilities in a wide range of species we can begin to answer questions like: Is rhythm perception and rhythmicality an evolutionary prerequisite for, or by-product of, vocal learning (Patel *et al.*, 2009; Schachner *et al.*, 2009; Ravignani *et al.*, 2016; Spierings and Cate, 2016)? How important are the roles of neuronal and motor timing (Prather *et al.*, 2008; Benichov *et al.*, 2015; Roberts *et al.*, 2017)? Is it a coincidence that several established vocal learners use echolocation (Parker *et al.*, 2013)? Is the co-occurrence of vocal learning and turn-taking incidental or required (Chow *et al.*, 2015; Vernes, 2017)? Is vocal learning facilitated by, or even a necessary consequence of, specific social structures in vertebrate groups (Sewall, 2015; Chen *et al.*, 2016; Luo *et al.*, 2017)? And finally what are the factors necessary and sufficient for the emergence of vocal learning? Questions like these cannot be answered by studying only one model system or a small number of phylogenetically closely related species, but must be investigated in many species across multiple different taxa. Therefore, we here reiterate the need for a broader comparative approach applied to the study of vocal learning, in which species are compared depending on their evolutionary history, their ecological and morphological constraints, and their neuronal and genetic underpinnings. Only comparative studies between species with either, shared expression of vocal learning and different underlying mechanisms or shared evolutionary constraints and similar morphological and neurogenetic setup but varying degree of vocal learning, will lead to an in-depth understanding of all factors that are necessary and sufficient for the evolution of vocal learning.

Acknowledgements

This work was funded by a Marie Curie Career Integration Grant (PCIG12-GA-2012-333978), a Max Planck Research Group Award, and a Human Frontiers Science Program (HFSP) Research grant (RGP0058/2016), all awarded to S.C.V. We would like to thank Christina Bergmann for thoughtful comments on the manuscript.

The authors declare no conflict of interest.

References

Agate, R.J.; Scott, B.B.; Haripal, B.; Lois, C. and Nottebohm, F. (2009). Transgenic songbirds offer an opportunity to develop a genetic model for vocal learning. *Proceedings of the National Academy of Sciences of the United States of America* 106, 17963–17967.

Benichov, J.I.; Benezra, S.E.; Vallentin, D.; Globerson, E.; Long, M.A. and Tchernichovski, O. (2015). The forebrain song system mediates predictive call timing in female and male zebra finches. *Current Biology* 26, 1–10.

Boë, L.-J.; Berthommier, F.; Legou, T.; Captier, G.; Kemp, C.; Sawallis, T.R.; Becker, Y.; Rey, A. and Fagot, J. (2017). Evidence of a vocalic proto-system in the baboon (*Papio papio*) suggests pre-hominin speech precursors. *PLoS ONE* 12, e0169321.

Boughman, J.W. (1998). Vocal learning by greater spear-nosed bats. *Proceedings of the Royal Society B* 265, 227–233.

Boughman, J.W. and Moss, C.F. (2003). Social sounds: vocal learning and development of mammal and bird calls. In: A. M. Simmons, R. R. Fay, and A. N. Popper (Eds), *Acoustic Communication - Springer Handbook of Auditory Research*. Springer-Verlag New York, NY, pp. 138–224.

Brainard, M.S. and Doupe, A.J. (2002). What songbirds teach us about learning. *Nature* 417, 351–358.

Chen, Y.; Matheson, L.E. and Sakata, J.T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proceedings of the National Academy of Sciences of the United States of America* 113, 6641–6646.

Chow, C.P.; Mitchell, J.F. and Miller, C.T. (2015). Vocal turn-taking in a non-human primate is learned during ontogeny. *Proceedings of the Royal Society B* 282, 20150069.

Elemans, C.P. (2014). The singer and the song: the neuromechanics of avian sound production. *Current Opinion in Neurobiology* 28, 172–178.

Elemans, C.P.H.; Rasmussen, J.H.; Herbst, C.T.; Düring, D.N.; Zollinger, S.A.; Brumm, H.; Srivastava, K.; Svane, N.; Ding, M.; Larsen, O.N.; Sober, S.J. and Vec, J.G.Š. (2015). Universal mechanisms of sound production and control in birds and mammals. *Nature Communications* 6, 1–13.

Enard, W. (2011). FOXP2 and the role of cortico-basal ganglia circuits of speech and language evolution. *Current Opinion in Neurobiology* 21, 415–424.

Fisher, S.E. and Scharff, C. (2009). FOXP2 as a molecular window into speech and language. *Trends in Genetics* 25, 166–177.

Fitch, W.T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences* 4, 258–267.

Fitch, W.T. (2017). Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review* 24, 3–33.

Fitch, W.T.; de Boer, B.; Mathur, N. and Ghazanfar, A.A. (2016). Monkey vocal tracts are speech-ready. *Science Advances* 2, 1–7.

Gammon, D.E. and Altizer, C.E. (2011). Northern mockingbirds produce syntactical patterns of vocal mimicry that reflect taxonomy of imitated species. *Journal of Field Ornithology* 82, 158–164.

Janik, V.M. (2014). Cetacean vocal learning and communication. *Current Opinion in Neurobiology* 28, 60–65.

Janik, V.M. and Slater, P.J. (2000). The different roles of social learning in vocal communication. *Animal Behaviour* 60, 1–11.

Knörnschild, M. (2014). Vocal production learning in bats. *Current Opinion in Neurobiology* 28, 80–85.

Knörnschild, M.; Nagy, M.; Metz, M.; Mayer, F. and Von Helversen, O. (2012). Learned vocal group signatures in the polygynous bat *Saccopteryx bilineata*. *Animal Behaviour* 84, 761–769.

Knörnschild, M.; Nagy, M.; Metz, M.; Mayer, F. and Von Helversen, O. (2010). Complex vocal imitation during ontogeny in a bat. *Biology Letters* 6, 156–159.

Lee, S.H.; Hsiao, T.Y. and Lee, G.S. (2015). Audio-vocal responses of vocal fundamental frequency and formant during sustained vowel vocalizations in different noises. *Hearing Research* 324, 1–6.

Lipkind, D.; Marcus, G.F.; Bemis, D.K.; Sasahara, K.; Jacoby, N.; Takahasi, M.; Suzuki, K.; Feher, O.; Ravbar, P.; Okanoya, K. and Tchernichovski, O. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* 498, 104–109.

Luo, B.; Huang, X.; Li, Y.; Lu, G.; Zhao, J.; Zhang, K.; Zhao, H.; Liu, Y. and Feng, J. (2017). Social call divergence in bats: a comparative analysis. *Behavioral Ecology* 28, 533–540.

Mitsuya, T.; Munhall, K.G. and Purcell, D.W. (2017). Modulation of auditory-motor learning in response to formant perturbation as a function of delayed auditory feedback. *The Journal of the Acoustical Society of America* 141, 2758–2767.

Nelson, D.A.; Marler, P. and Palleroni, A. (1995). A comparative approach to vocal learning: intraspecific variation in the learning process. *Animal Behaviour* 50, 83–97.

Nowicki, S. and Searcy, W.A. (2014). The evolution of vocal learning. *Current Opinion in Neurobiology* 28, 48–53.

Parker, J.; Tsagkogeorga, G.; Cotton, J. a; Liu, Y.; Provero, P.; Stupka, E. and Rossiter, S.J. (2013). Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* 502, 1–9.

Patel, A.D.; Iversen, J.R.; Bregman, M.R. and Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology* 19, 827–830.

Pepperberg, I.M. (2010). Vocal learning in grey parrots: A brief review of perception, production, and cross-species comparisons. *Brain and Language* 115, 81–91.

Petkov, C.I. and Jarvis, E.D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience* 4, 1–24.

Pfenning, A.R.; Hara, E.; Whitney, O.; Rivas, M.V.; Wang, R.; Roulhac, P.L.; Howard, J.T.; Wirthlin, M.; Lovell, P.V.; Ganapathy, G.; Mouncastle, J.; Moseley, M.A.; Thompson, J.W.; Soderblom, E.J.; Iriki, A.; Kato, M.; Thomas, M.T.P.; Gilbert, P.; Zhang, G.; Bakken, T.; Bongaarts, A.; Bernard, A.; Lein, E.; Mello, C.V.; Hartemink, A.J. and Jarvis, E.D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* 346, 1–13.

Pilowsky, J.A. and Rubenstein, D.R. (2013). Social context and the lack of sexual dimorphism in song in an avian cooperative breeder. *Animal Behaviour* 85, 709–714.

Prat, Y.; Taub, M. and Yovel, Y. (2015). Vocal learning in a social mammal: demonstrated by isolation and playback experiments in bats. *Science Advances* 1, e1500019.

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

Rasilo, H. and Räsänen, O. (2017). An online model for vowel imitation learning. *Speech Communication* 86, 1–23.

Ravignani, A.; Fitch, W.T.; Hanke, F.D.; Heinrich, T.; Hurgitsch, B.; Kotz, S.A.; Scharff, C.; Stoeger, A.S. and Boer, B. de (2016). What pinnipeds have to say about human speech, music, and the evolution of rhythm. *Frontiers in Neuroscience* 10, 1–9.

Roberts, T.F.; Hisey, E.; Tanaka, M.; Kearney, M.G.; Chattree, G.; Yang, C.F.; Shah, N.M. and Mooney, R. (2017). Identification of a motor-to-auditory pathway important for vocal learning. *Nature Neuroscience* 20, 978–986.

Schachner, A.; Brady, T.F.; Pepperberg, I.M. and Hauser, M.D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology* 19, 831–836.

Schusterman, R.J. (2008). Vocal learning in mammals with special emphasis on pinnipeds. In: D.K. Oller and U. Gribel (Eds), *The Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication*. MIT Press, Cambridge, MA, pp. 41–70.

Sewall, K.B. (2015). Social complexity as a driver of communication and cognition. *Integrative and Comparative Biology* 55, 384–395.

Slater, P.J.B.; Eales, L.A. and Clayton, N.S. (1988). Song learning in zebra finches (*Taeniopygia guttata*): progress and prospects. *Advances in the Study of Behavior* 18, 1–34.

Smotherman, M.; Knörnschild, M.; Smarsh, G. and Bohn, K. (2016). The origins and diversity of bat songs. *Journal of Comparative Physiology A* 202, 535–554.

Spierings, M.J. and ten Cate, C. (2016). Zebra finches as a model species to understand the roots of rhythm. *Frontiers in Neuroscience* 10, 1–3.

Stoeger, A.S.; Mietchen, D.; Oh, S.; De Silva, S.; Herbst, C.T.; Kwon, S. and Fitch, W.T. (2012). An Asian elephant imitates human speech. *Current Biology* 22, 2144–2148.

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

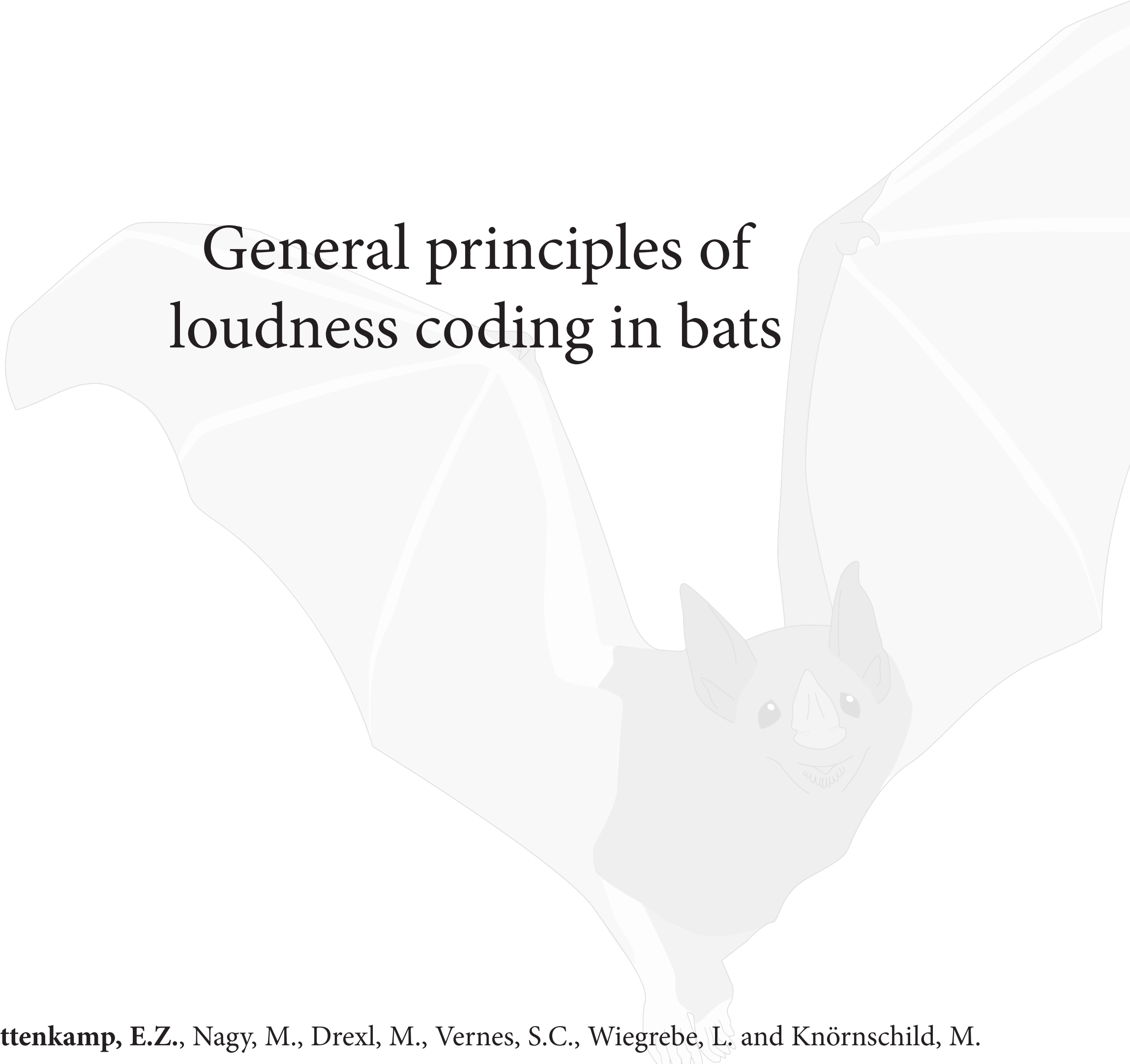
Todt, D.; Hultsch, H. and Heike, D. (1979). Conditions affecting song acquisition in nightingales (*Luscinia megarhynchos* L.). *Ethology* 51, 23–35.

Vernes, S.C. (2017). What bats have to say about speech and language. *Psychonomic Bulletin & Review* 24, 111–117.

Wada, K.; Sakaguchi, H.; Jarvis, E.D. and Hagiwara, M. (2004). Differential expression of glutamate receptors in avian neural pathways for learned vocalization. *Journal of Comparative Neurology* 476, 44–64.

Wheatcroft, D. and Qvarnström, A. (2015). A blueprint for vocal learning: auditory predispositions from brains to genomes. *Biology Letters* 11, 20150155.

Chapter 3



General principles of loudness coding in bats

Lattenkamp, E.Z., Nagy, M., Drexl, M., Vernes, S.C., Wiegrebe, L. and Knörnschild, M. (in prep.). General principles of loudness coding in bats.

Abstract

Our senses are our interface with the environment and different species make use of their sensory systems according to the ecological niche they inhabit. Although sensory systems undergo species-specific adaptations, the general principles are often analogous for closely related taxa. We investigated the peripheral auditory encoding of loudness across frequencies in eleven Neotropical bat species by means of auditory brainstem responses (ABRs). ABRs are minimally invasive recordings of acoustically evoked summary potentials. The frequency-dependent sound levels that create a significant ABR provide an estimate of an individual's audiogram. Furthermore, the growth of the ABR with increasing loudness provides information about loudness encoding in the auditory periphery. We found that the audiograms of the investigated bat species are similar at first sight, but reveal species-, and even sex-specific specialisations. Across all species, we found that the growth of the ABR with increasing loudness is steeper and saturates earlier at lower frequencies mainly corresponding to social call frequencies. At higher ultrasonic frequencies, corresponding to echolocation call frequencies, ABR growth is much shallower and saturation is rarely encountered in the tested loudness range (up to 110 dB SPL). Our data suggest that reliable loudness coding is more important for echolocation, where loudness is a cue for echoacoustic target distance and size. For social calls, absolute auditory sensitivity is equally good, but loudness coding is less flexible. Despite species-specific differences in audiogram shape, this general principle for loudness coding was found in all measured bat species and may extend to other mammals.

Introduction

Our senses are our connection to the world and provide us with the information we need to survive, be it for the localisation or identification of food or predators. Sensory systems are phylogenetically determined and are often based on general principles but can also show species-specific adaptations. The mammalian sensory system for the perception of smell for example works via homologous odorant receptors (Rinaldi, 2007), but the morphology of the organs themselves (i.e. shape of different 'noses') is adjusted for the specific ecological niche of the animals (Zaidi *et al.*, 2017). Which of the senses play the most important role in a species' life is dependent on its lifestyle and habitat. For bats hearing is the dominant sense, enabling not only vocal communication with conspecifics (Griffin, 1980; Wilkinson, 2003; Jones, 2005; Jones and Teeling, 2006; Vernes and Wilkinson, 2019), but also other essential tasks such as navigation and orientation in darkness, foraging and detection of predators.

Acoustic perception in bats has been measured since the 1960s (Grinnell, 1963). However, in comparison to the species-richness of the chiropteran order (more than 1300 extant species), the most principal measure for hearing of bats, the audiogram, has been probed in only a few species (e.g. *Carollia perspicillata* (Koay *et al.*, 2003; Esser and Eiermann, 2008), *Pteronotus parnellii* (Pollak *et al.*, 1979; Henson *et al.*, 1985; Kössl and Vater, 1985b), *Phyllostomus discolor* and *P. hastatus* (Esser and Daucher, 1996; Koay *et al.*, 2002; Wittekindt *et al.*, 2005; Hoffmann *et al.*, 2008; Linnenschmidt and Wiegrebe, 2019), and *Desmodus rotundus* (Schmidt *et al.*, 1991; Gröger and Wiegrebe, 2006; Heffner *et al.*, 2013)). Auditory brainstem recordings (ABRs) have been established for the fast, objective, and minimally invasive assessment of hearing (Jewett, 1970; Jewett *et al.*, 1970; Corwin *et al.*, 1982; Whitlon *et al.*, 2019). However, in the last 60 years, only a few authors reported ABRs from bats (e.g. *Eptesicus fuscus* (Simmons *et al.*, 1990; Burkard and Moss, 1994), *Lasiurus borealis* (Obrist and Wenstrup, 1998), *Myotis lucifugus* (Grinnell, 1963), *Noctilio leporinus* (Wenstrup, 1984), *Phyllostomus discolor* (Linnenschmidt and Wiegrebe, 2019), *Pipistrellus abramus* (Boku *et al.*, 2015; Simmons *et al.*, 2015), *Plecotus townsendii* (Grinnell, 1963), *Rousettus aegyptiacus* (Belknap and Suthers, 1982), and *Tadarida brasiliensis* (Smotherman and Bakshi, 2019)). Species, for which hearing thresholds were investigated showed a close relationship between hearing threshold and the corresponding social call frequency range (Bohn *et al.*, 2006).

Vocal emissions in bats have been studied intensely in the past, with a focus on their echolocation calls (Griffin, 1980; Jones, 2005; Grinnell *et al.*, 2016). Echolocation calls vary in their acoustic parameters dependent on the bats' ecological niche (Jones, 1999; Holderied *et al.*, 2006). In addition to navigation, hunting, and general probing of space (Neuweiler, 1984), they allow individual and species identification (Yovel *et al.*, 2009; Schuchmann and Siemers, 2010a) and can even facilitate social communication (Fenton, 2003; Jones and Siemers, 2011; Knörnschild *et al.*, 2012a). Bat social vocalisations have also been investigated in the past. They are diverse, species-specific, complex, flexible, and vitally important for mother–pup interactions and the social structure of bat communities (Pfalzer and Kusch, 2003; Wilkinson, 2003; Bohn *et al.*, 2007; Knörnschild *et al.*, 2012b; Gillam and Fenton, 2016; Smotherman *et al.*, 2016). These multi-facetted functions highlight the importance of vocalisations in the life-history of bats and the strong evolutionary pressure to accurately perceive these acoustic signals.

Even though echolocation and social calls show clear species-specific characteristics, they function under the same basic principles across species. While social calls are generally used in social interactions occurring in close proximity, echolocation calls need to work over a broad range of distances. We hypothesise that while auditory coding in different bat species is adjusted to the species-specific ecological niche and vocalisations, it also functions based on common principles. In the present study, we measured click and tone pip evoked ABRs in eleven different Neotropical bat species. The magnitudes of the supra-threshold ABRs were then related to the spectral composition of species-specific echolocation and social calls. Common features as well as differences between the species were analysed to highlight general principles of acoustic perception, but also to pinpoint differences according to lifestyle and sexual dimorphisms. We further used this comparison to highlight the usefulness of ABR measurements for the interpretation of ecological factors involved in the lives of bats.

Materials and methods

Animals and experimental approval

ABRs were measured from a total of 86 adult bats from eleven different species (from six families, **Table 3.1**). All bats were caught in the area of Gamboa, Panama, during March and April of 2019. The research was conducted in accordance with the Panamanian government and the regulations of the Smithsonian Tropical Research Institute.

Anaesthetics application

Bats were anaesthetised with a combination drug of Medetomidine, Midazolam, and Fentanyl at least 10 minutes before the experiment (administered amounts per species are listed in **Table 3.2**). The depth of the anaesthesia was quantified by triggering the inter-toe reflex by mechanical stimulation (slight pinching with fine forceps). The drug was injected subcutaneously between the shoulder blades (needle: 0.45 x 12 mm). The reactions of the different bat species to the drug varied strongly, thus in some cases a second dose needed to be injected after 10–15 minutes. The anaesthesia was not antagonised after the measurements were concluded and thus varied in its duration strongly between

Table 3.1: Summary of ABR measurements. Listed is the number of adult bats (N) measured per species, separated by sex.

family (N = 6)	species (N = 11)	males	females	total N per species
Phyllostomidae	*Desmodus rotundus*	3	3	6
	Glossophaga soricina	3	3	6
	Carollia perspicillata	13	0	13
	Phyllostomus hastatus	7	0	7
Emballonuridae	*Saccopteryx bilineata*	7	5	12
	Saccopteryx leptura	1	3	4
	Rhynchonycteris naso	4	1	5
Thyropteridae	*Thyroptera tricolor*	3	7	10
Vespertilionidae	*Myotis nigricans*	8	0	8
Mormoopidae	*Pteronotus parnellii*	8	1	9
Molossidae	*Molossus molossus*	1	5	6

Table 3.2: Administered fentanyl and anaesthesia duration. Listed are weight, amount of injected fentanyl per gram body weight, and duration of the anaesthesia (i.e. time from injection until first reactions to sounds could be noted (i.e. ear movements)) for all measured species. Measurements are given as means and standard deviations. The sample size (N) for the calculation of the mean and standard deviation are given in brackets after the values. Several individuals were used in an additional experiment after the acquisition of the ABRs and their natural anaesthesia durations are not known.

species (N = 11)	bodyweight [gram] (N)	injected fentanyl [µg per g bodyweight] (N)	anaesthesia duration [minutes] (N)
Desmodus rotundus	36.2 ± 1.4 (6)	0.017 ± 0.005 (6)	
Glossophaga soricina	10.1 ± 0.7 (6)	0.009 ± 0.001 (6)	297 ± 34 (3)
Carollia perspicillata	18.6 ± 1.3 (13)	0.011 ± 0.005 (13)	124 ± 90 (12)
Phyllostomus hastatus	118.2 ± 9.2 (7)	0.008 ± 0.002 (7)	574 ± 269 (6)
Saccopteryx bilineata	7.2 ± 0.6 (12)	0.032 ± 0.007 (12)	144 ± 31 (9)
Saccopteryx leptura	5.2 ± 0.7 (4)	0.030 ± 0.008 (4)	69 (1)
Rhynchonycteris naso	3.4 ± 0.4 (5)	0.029 ± 0.016 (5)	
Thyroptera tricolor	4.6 ± 0.7 (10)	0.018 ± 0.005 (10)	362 ± 238 (5)
Myotis nigricans	3.8 ± 0.3 (8)	0.061 ± 0.018 (8)	372 ± 186 (7)
Pteronotus parnellii	22.7 ± 1.0 (9)	0.009 ± 0.001 (9)	280 ± 126 (9)
Molossus molossus	9.8 ± 0.6 (6)	0.031 ± 0.006 (6)	336 ± 77 (6)

the different species (**Table 3.2**). Subsequent to the ABR measurements, some individuals were perfused for a separate experiment. For these individuals, the natural anaesthesia duration is not known (see **Table 3.2**). During anaesthesia, eye cream (Bepanthen®, 5% Dexpanthenol, Bayer AG, Leverkusen, Germany) was applied to prevent the eyes from drying out.

ABR setup

The ABRs were measured in a small sound-attenuating box (PELI 1450 case, peli products, Torrance, CA, USA; inside measurements: 37.1 x 25.8 x 15.2 cm^3, **Fig. 3.1a**). The box was first lined with a fine copper mesh to reduce electrical interferences, and then lined with sound-attenuation foam, reducing reverberations within the box (**Fig. 3.1a**). The bats were positioned frontally facing the loudspeaker (R2004/602000, ScanSpeak, Videbæk, Denmark), which was embedded in the side of the box (**Fig. 3.1b**). The different measured bat species varied in size and ear shape. Thus, the bats were always placed in a way that their outer ear opening was placed at a 4 cm distance from the loudspeaker. The varying sizes of the bats were compensated for with different heights of moveable soft foam pads, so that the heads of the bats were always placed on one level with the horizontal centre of the loudspeaker. The loudspeaker was connected to an amplifier (M032N, Kemo® Electronic, Germany), which was powered via a four-battery pack (NCR18650GA batteries, Li-Ion, 3.6V, 3500 mAh, Panasonic/Sanyo, Ottobrunn, Germany). Both stimulus presentation and ABR recording were done by an audio interface (ADI-2 PRO FS, RME, Haimhausen, Germany), running at a sampling rate of 384 kHz.

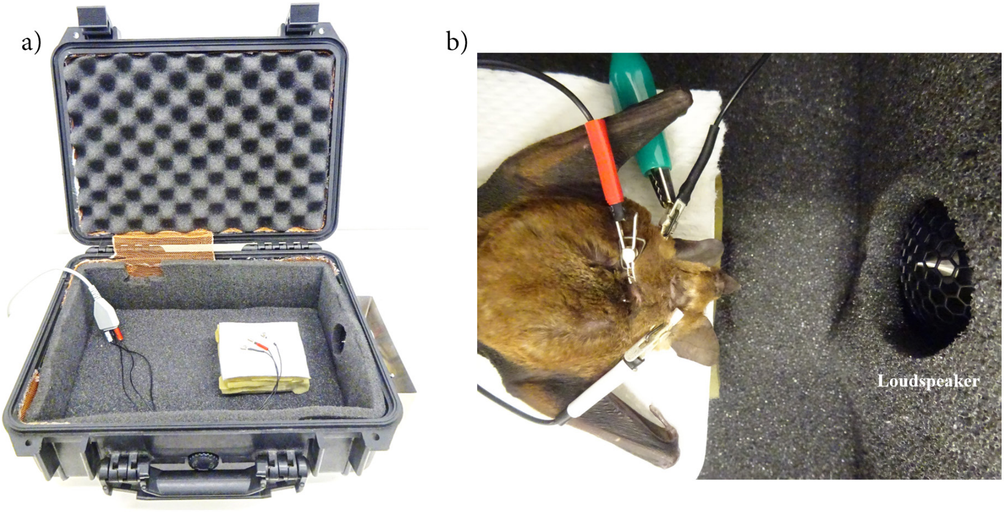


Figure 3.1: ABR setup. a) Picture of the mobile ABR setup in the open configuration. The loudspeaker is embedded in the right side of the box. The setup is lined first with a copper mesh to reduce electrical disturbance (see gap between lid and bottom part of the box) and then covered in sound absorbing foam. The electrode cable is lead into the box via a cut-out gap in the left side of the box. The foam pads are positioned in front of the loudspeaker and covered with a paper towel. b) Positioning of the bats in the setup. The bats were positioned frontally facing the loudspeaker with a distance of 4 cm between the loudspeaker and their outer ear openings. Different heights of moveable soft foam pads were used to compensate for the different sizes of the measured species. The heads of the bats were always placed on one level with the centre of the loudspeaker. The recording electrode (red) was placed at the caudal midline of the head, close to the brainstem. The reference electrode (white) was placed at the dorsal midline of the head between the ears. The ground electrode (black) was placed on the base of the left ear of the animal (or the wing or tail membrane). The ground electrode was connected to the copper lining of the setup via a crocodile clamp (green) in order to ground the animal.

The sound system of the setup was calibrated with a 1/8" measuring microphone (B&K4138 without protective grid, Bruel & Kjaer, Bremen, Germany) connected to a measuring amplifier (B&K Measuring Amplifier Type 2636, Bruel & Kjaer, Bremen, Germany). The acoustic impulse response of the loudspeaker was measured at the bats' head position within the setup using broadband (2–140 kHz) noise. From the measured acoustic impulse response, a compensation impulse response was generated. The used stimuli were convolved with the compensation impulse response in order to present them with a linear phase and a flat frequency response between 3 and 120 kHz.

Stimuli

Recordings of ABRs were done in response to either click or tone pip stimuli presented in the free field. The click stimuli were broadband impulses with a flat power spectrum from 2 to 140 kHz. The tone pips were sinusoids of 2.5 ms duration (Hanning windowed) with different carrier frequencies. The tone pip frequencies were evenly spaced between 5 and 120 kHz in eleven steps of approximately half an octave (550 Cent) on a logarithmic frequency axis. Click and pip stimuli were both presented at sound levels between 0 and 110 dB peak-equivalent sound pressure level (peSPL), with 10-dB increments. All stimuli were presented 256 times with a 44 Hz repetition rate. Every other stimulus was phase-inverted to cancel out electrical stimulus artefacts picked up by the ABR electrodes after averaging in the time-domain. Stimuli were generated at a sampling rate of 384 kHz and a digital word length of 24 bit. Click stimuli were presented first to each tested bat, with the twelve sound levels presented in randomised order. Thereafter, ABRs in response to the tone pip stimuli were measured. For the tone pip stimuli, the twelve sound levels were again randomised, but all eleven frequencies were presented at one sound level first, before the next sound level was chosen. The order of the tone pip carrier frequencies within a sound level was also chosen randomly. A custom written Matlab script (Matlab, R2018b, MathWorks, Natick, NA, USA) was used to generate the stimuli and coordinate their presentation via the above-mentioned audio interface.

ABR recording

ABR measurements return summed electrical responses of both the auditory nerve fibres and the nuclei of the ascending auditory pathway from anaesthetised animals. For the recording of ABRs, two subdermal electrodes (clipped needles, Sterican® brown 0.45 x 12 mm, B. Braun, Melsungen AG, Melsungen, Germany) were placed at the caudal midline of the head, close to the brainstem (recording electrode) and at the dorsal midline of the head between the ears (reference electrode) (**Fig. 3.1b**). The ground electrode was either placed on the base of the left ear of the animal or on the wing or tail membrane if the ears were too small for correct electrode placement. The electrodes were positioned after the fur was cut with scissors to enable their precise placement. The electrodes were connected via alligator clips to a bioamplifier (BMA-200, CWE Inc., USA), which bandpass-filtered the brainstem responses (between 100 Hz and 3 kHz) and amplified the signal by 60 dB. The signal was converted to digital by the above-mentioned audio interface at 384 kHz sampling rate. The ABR signals were downsampled by a factor of 20 before each of the 256 recordings to the same frequency–sound level combination were saved. ABR signals averaged in the time-domain for each combination were displayed to the experimenter for quality monitoring during the recording.

ABR data analysis

The data analysis was done according to previous experiments (Linnenschmidt and Wiegrebe, 2019). The amplitudes of the recorded ABRs were calculated as the root-mean square (RMS) in the time window starting directly after the stimulus presentation and lasting for the duration of the ABR signal (i.e. 1–8 ms after stimulus onset). Bootstrap analyses ($n = 500$; 95% confidence) were performed on the click and tone pip evoked ABR data to statistically verify the presence of an ABR signal (Lv *et al.*, 2007). This procedure tests the statistical likelihood that a recorded signal presents random variation in the data rather than a physiological response. To that end, repeated random resampling (with replacement) of the original data was performed and then it was assessed whether the RMS of the resampled waveform exceeded the original. If 95% of resampled waveforms had a lower RMS than the original waveform, the measurement was considered significant. The lowest sound level evoking a significant ABR signal (at a specific frequency) was conservatively accepted as threshold only when significant ABRs were also obtained for all higher sound levels at the same frequency.

The bootstrap analyses were used to assess the characteristic ABR threshold for each species (mean and standard error of the mean (SEM) were calculated omitting measurements for which the algorithm could not determine the threshold). The click evoked ABR waveforms were also averaged in the time-domain and the mean and SEM for each species is given. Furthermore, the species-specific shape of the ABR threshold in response to tone pips was evaluated by generating the average ABR signal strength for each measured frequency–amplitude combination, which are reported as contour plots. The slope of the growth function at each stimulus frequency was fitted with a sigmoidal curve (using the nlinfit function in Matlab) and the coefficient of determination (R^2) of the fits was assessed. The slope of the fitted functions was extracted and plotted for each species.

Extraction of call parameters

For the comparison between audiograms and spectral parameters of species-specific vocalisations, acoustic parameters of the bats' calls were extracted from our own recordings or from the literature (see supplementary **Table S3.1**). For all species, we reported minimum, maximum, and peak frequencies (i.e. frequency with the highest magnitude in the power spectrum) of echolocation calls and social calls. We selected echolocation calls recorded during search flight for parameter extraction. Whenever possible, recordings from the wild were used to minimise the risk of altered call design due to a confined environment in the laboratory. Echolocation calls in our dataset were either constant frequency with or without an optional frequency modulated part (CF or CF-FM), had a quasi-constant frequency part (QCF) or consisted of downward sweeps with varying degrees of frequency modulation (FM). For CF and QCF calls, we reported minimum, maximum, and peak frequencies of the (quasi-) constant frequency parts of the most prominent harmonic. For FM calls, we reported minimum, maximum, and peak frequencies of the whole call, including all present harmonics. Isolation calls are produced by pups to solicit maternal care (Gould, 1975). We only used isolation calls from non-volant pups and made certain that we did not include echolocation call precursors for parameter extraction. Isolation calls were recorded from hand-held pups in most cases. In all but one species, isolation calls had a lower peak frequency than echolocation calls.

Results

We measured ABRs from a total of 86 adult bats belonging to eleven species from six families (Phyllostomidae, Emballonuridae, Thyropteridae, Vespertilionidae, Mormoopidae, and Molossidae). For four species we measured at least 3 individuals per sex (**Table 3.1**).

ABR thresholds in relation to social vocalisations and echolocation calls

The measured species differed in shape and strength of their click evoked ABRs. The bootstrap thresholds for the detection of a significant click evoked ABR (i.e. stimulus detection onset) were at 60 (± 10) dB peSPL. The tone pip evoked ABR thresholds were generally tub-shaped. Sensitivity to tone pips increased between 5 and approximately 10 kHz (20 kHz for *Rhynchonycteris naso*; **Fig. 3.2c**) and showed a general decrease above 60 kHz (**Fig. 3.2**). The species differed in the slopes and extrema of their audiograms and showed one, two, or three sensitivity peaks. The different iso-response lines of the contour plots indicate the strength of the ABR signal in µV (**Fig. 3.2**). Sensitivity maxima in the mean thresholds and iso-response lines corresponded to echolocation frequency ranges for all species, but one: The echolocation calls of *Thyroptera tricolor* are higher than the frequency range in which we measured ABRs. *Thyroptera tricolor* showed a very pronounced peak in the frequency range of their species-specific isolation calls (**Fig. 3.2i**). Most species also showed a peak in the social call frequency range. Especially for *Saccopteryx bilineata*, peaks in the first contour line correspond perfectly

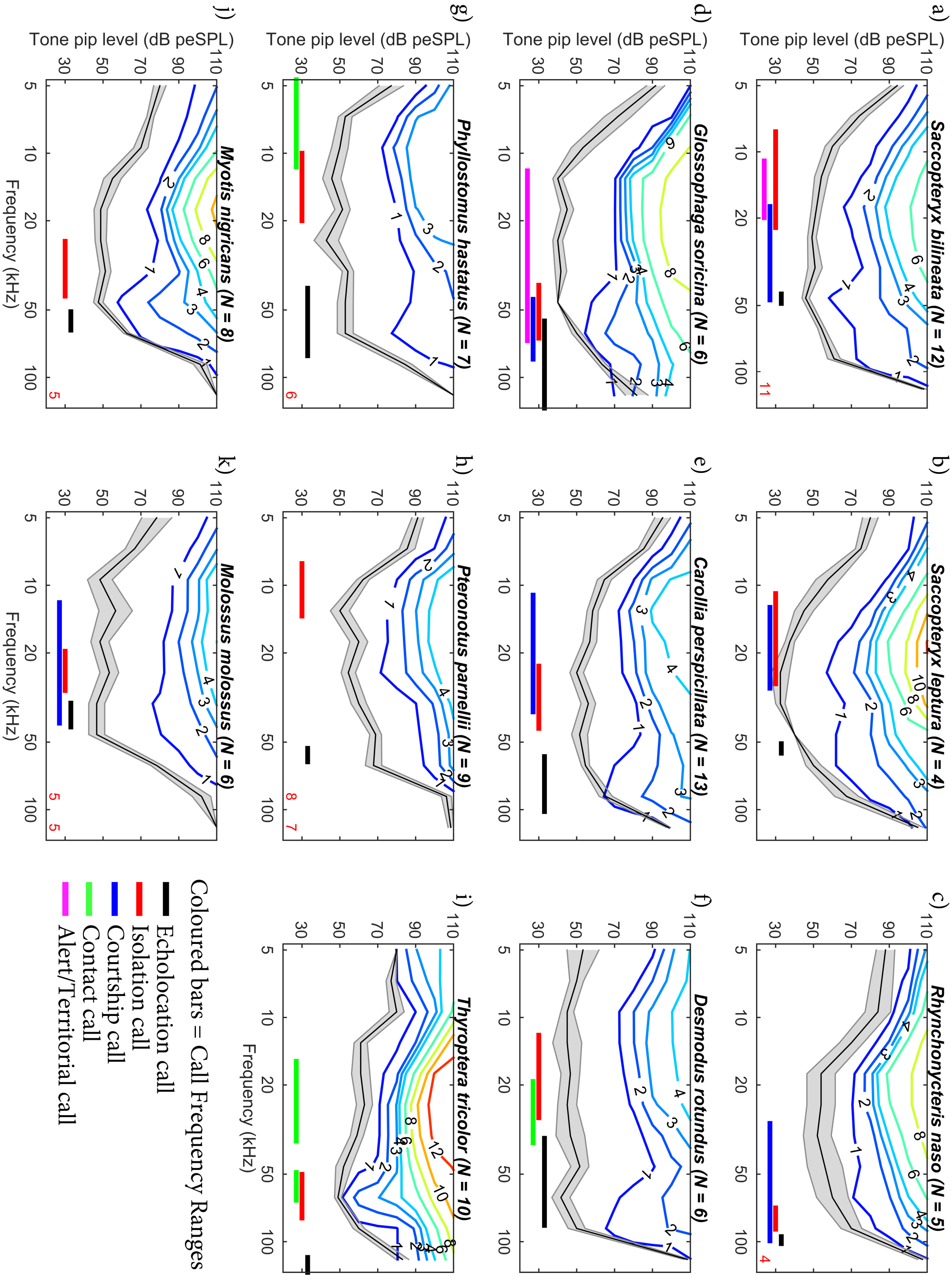


Figure 3.2: Species-specific mean ABR thresholds as calculated via bootstrap analysis. The mean ABR threshold per species is depicted (black line; shading represents SEM). The contour lines represent the strength of the ABR signal (colours and numbers indicate µV response strength). Despite the species-specific shape of the thresholds, a general trend of sensitivity increase between 5 and 15 kHz and decrease between 50 and 80 kHz is noticeable. The number of animals measured per species (N) is given in the plot titles. Vertical bars below the ABR threshold indicate bandwidth of five different call types (determined as described above): echolocation calls (black), isolation calls (red), courtship calls (blue), contact calls (green) and territorial/alert calls (purple) (supplementary **Table S3.1**). Red numbers on the bottom of the panels indicate the number of animals used for the calculation of mean and SEM (if < than total sample size). The numbers are positioned above the frequency for which mean values were determined.

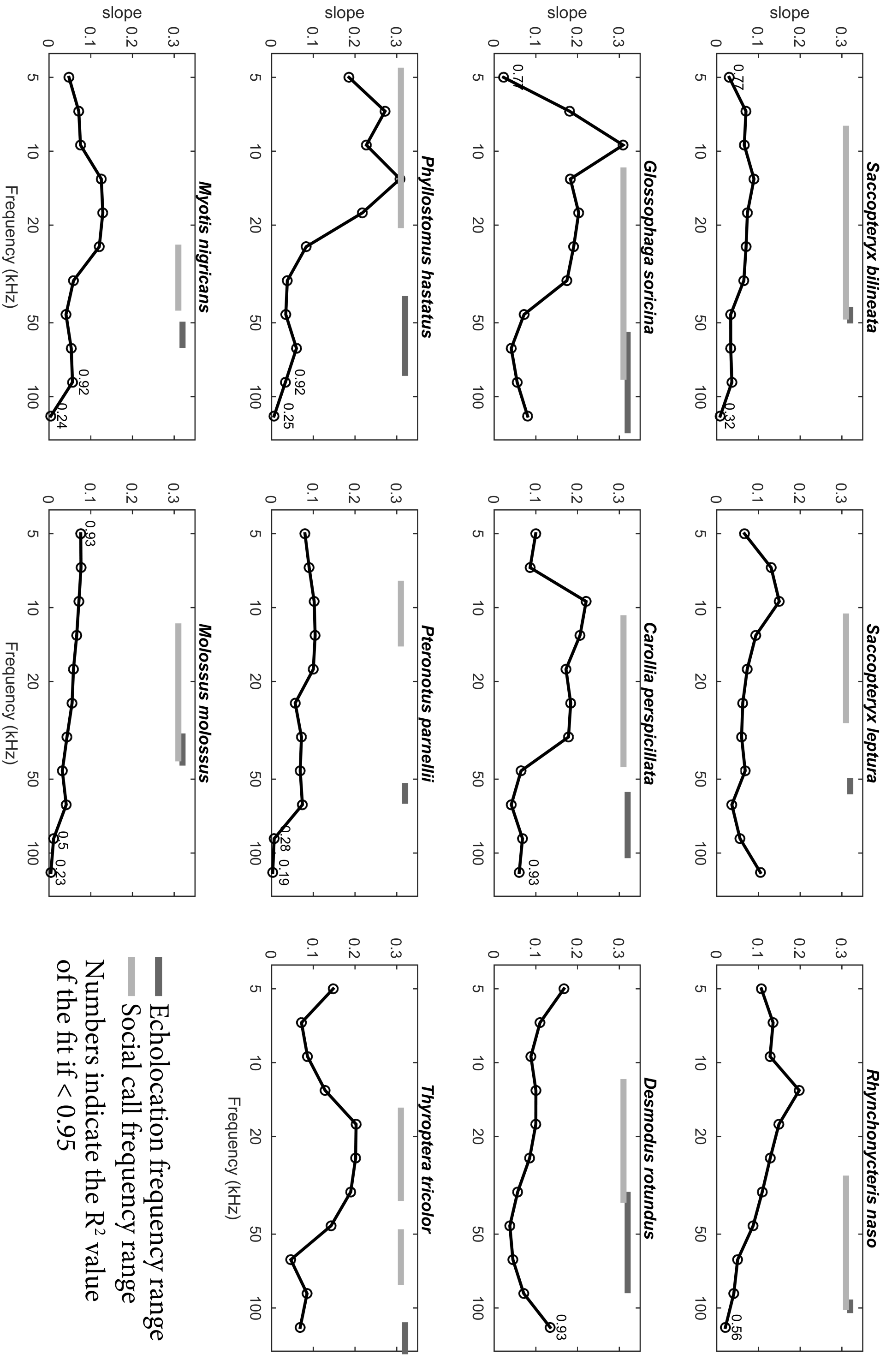


Figure 3.3: Slope of ABR growth functions for each frequency. For each species the growth function per frequency was calculated. The R^2 quality measures of the fits are given above each frequency (if < 0.95). Dark grey bars indicate the echolocation frequency ranges, while the light grey bars indicate the social call frequency ranges.

to the distribution of peak frequencies of echolocation calls (44.3–48.9 kHz) and peak frequencies, where their social calls overlap (17.7–22.1 kHz; **Fig. 3.2a**; supplementary **Table S3.1**).

Dynamic range of hearing in high and low frequency ranges

The distances between the contour lines can be expressed as ABR growth functions for each frequency. The dynamic range of ABRs is the larger, the shallower the slope of the growth function. For the higher frequency ranges corresponding to echolocation calls, the ABR growth functions were generally much shallower than for lower frequency

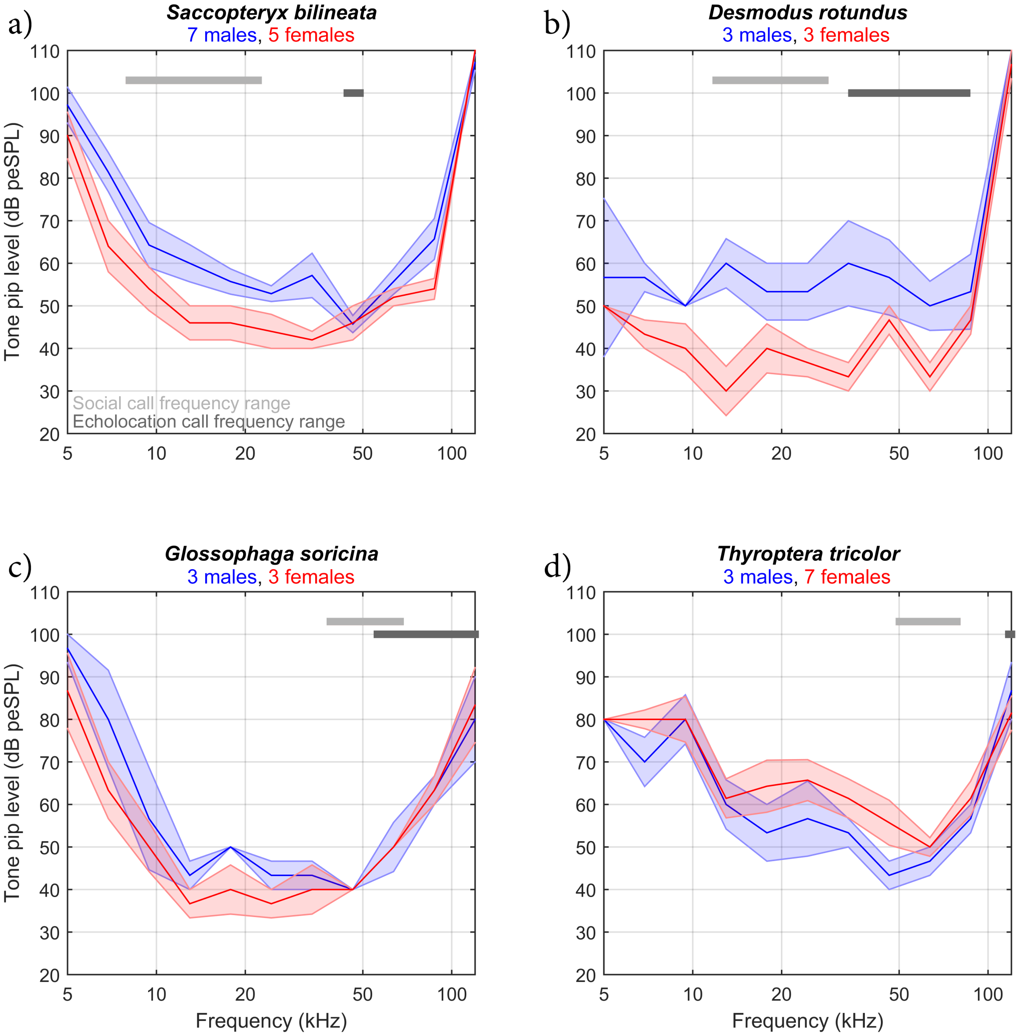


Figure 3.4: Sex-specific ABR thresholds. For four species at least three individuals per sex were measured. Thresholds for females of the species *Saccopteryx bilineata* (a) and *Desmodus rotundus* (b) were significantly lower than the males'. Females of the species *Glossophaga soricina* (c) showed the same trend, but the thresholds of both sexes greatly overlapped. Although the SEM overlapped for males and females of *Thyroptera tricolor* (d) as well, this species showed generally more sensitive thresholds for males than females.

ranges. The steepest growth functions were found between 5 and 20 kHz. The slopes of the ABR growth functions are plotted as a function of frequency in **figure 3.3.** For some species, namely *Saccopteryx leptura*, *Glossophaga soricina*, and *Desmodus rotundus*, the slopes increased again towards higher frequencies (**Fig. 3.3**).

Sexual dimorphism

The audiograms of male and female individuals of four bat species were compared (**Table 3.1**). For two of the species (namely *Saccopteryx bilineata* and *Desmodus rotundus*) the audiograms of male individuals were less sensitive than females of the same species in a frequency range corresponding to pup isolation calls. (i.e. the audiograms did not show an overlap of the SEM in this frequency range) (**Fig. 3.4a,b**). The SEM for males and females greatly overlapped for individuals of the species *Glossophaga soricina* and *Thyroptera tricolor*. Especially in the frequency range of isolation calls, the ABR thresholds of these two species strongly overlap (**Fig. 3.4c,d**).

Discussion

General principles of loudness coding: dynamic ranges of hearing

In the present study, we measured ABRs of 86 bats belonging to eleven species. To date this provides the most comprehensive comparative assessment of the hearing capacity of bats. Our findings show that all bat species have hearing sensitivity peaks in the frequency range of their social vocalisations and echolocation calls (**Fig. 3.2**). These peaks in their hearing curves differ between the species and correspond to their species-specific call types. However, due to the large parameter space probed with the presented ABR measurement regime, we were also able to assess the ABR growth functions for all tested frequencies. Specifically, this allows us to assess the increase in ABR according to the increase in presented stimulus level. Our findings show a general principle of loudness coding: all measured bats showed shallow ABR growth functions in the echolocation call frequency range, while the growth functions were steeper and saturated earlier in the social-call range (**Fig. 3.3**).

Echolocation calls are typically emitted at high sound levels (in some species up to 138 dB SPL (Waters and Jones, 1995; Schuchmann and Siemers, 2010b; Hulgard *et al.*, 2016)), but the returning echoes can be quite faint (Stilz and Schnitzler, 2012). Furthermore, echolocation calls can be dynamically adjusted in sound level, based on habitat, distance, and target strength of the ensonified object (Jakobsen *et al.*, 2013; Luo *et al.*, 2015). Therefore, it is imperative for bats to encode different amplitudes of echolocation calls and their returning echoes. Social calls are emitted and perceived at similarly high sound levels as the calls are generally emitted by conspecifics in close proximity. Hypothetically, the behaviourally relevant information of these calls lies much more in the spectro-temporal modulation of e.g. fundamental frequency and aperiodicity rather than in the call amplitude.

Consistent with our predictions, we found that ABR growth functions in higher frequency ranges (i.e. corresponding to echolocation call frequencies) are shallower than in lower frequency ranges (i.e. corresponding to social call bandwidths). This means that bats have a broader dynamic range for the processing of echolocation call amplitudes than for social call amplitudes. Thus, it is likely that the ABR saturates in response to most communication calls, while it still encodes loudness for echolocation calls and their echoes. This trend is a general feature in all measured species, despite species-specific differences in the spectral content of their calls and is independent of their ecology (e.g. group size, mating system, diet).

Sexual dimorphism in hearing

Evolutionary pressures can exert themselves differently on the sexes and can thus generate sexual dimorphisms in various features (Darwin, 1871). In bats, sexual dimorphisms have been described in e.g. body size, migratory patterns, longevity, and parasitic load (O'Mara *et al.*, 2016; Postawa and Nagy, 2016; Wu *et al.*, 2018; Wilkinson and Adams, 2019). Although sexually dimorphic vocalisations are not unusual in bats (e.g. Jones and Kokurewicz, 1994; Grilliot *et al.*, 2009, 2014), evidence for sexual dimorphism in hearing has not been reported thus far. We here describe significant differences between male and female ABR thresholds for two species (*Saccopteryx bilineata* and *Desmodus rotundus*) in social call frequency ranges, but not in the frequency range of their echolocation calls (**Fig. 3.4a,b**). Such sexually dimorphic audiograms were not measured for two other tested species (specifically *Glossophaga soricina* and *Thyroptera tricolor*; **Fig. 3.4c,d**). This presence or absence of a sexually dimorphic hearing capacity could be an indicator for differences in life history traits, such as parental care. When providing parental care, accurate offspring recognition is crucial to avoid wasting energy on unrelated offspring and thus presents a strong evolutionary pressure on the auditory system. The care-providing sex should thus be especially sensitive in the frequency range of pup vocalisations (Bohn *et al.*, 2006). A correlation between hearing sensitivity and pup isolation call frequency has been shown for thirteen bat species (Bohn *et al.*, 2006), however this meta-analysis did not differentiate hearing by sex.

Desmodus rotundus and *Saccopteryx bilineata*, for which sexually dimorphic hearing was found in the present study (**Fig. 3.4a,b**), represent species in which mainly females care for the offspring, as they live in large female groups (Wilkinson, 1985) or harems (Heckel and Von Helversen, 2003) throughout the year. *Thyroptera tricolor*, for which this dimorphism was not detected (**Fig. 3.4d**), lives in small (i.e. 1–11 individuals, generally

~4 individuals), mixed or sex-separated associations (Vonhof and Fenton, 2004). Parental care could be regularly provided by both sexes in this species. These findings indicate that the social system and life history traits might have a strong effect on the evolution of the auditory system (Bohn *et al.*, 2004). The fourth species investigated here, *Glossophaga soricina,* does not show sexually dimorphic hearing (**Fig. 3.4c**). Although this species lives in colonies generally containing both sexes, they form maternity colonies and thus, paternal care is not provided during early ontogeny (Alvarez *et al.*, 1991). Other evolutionary factors influencing these auditory thresholds are thus also conceivable. As parental care in bats is generally provided by females (Kunz and Hood, 2000; Knörnschild *et al.*, 2013), it would be interesting to test whether sexually dimorphic hearing occurs in most bat species and whether this phenotype is reversed when paternal care is provided to a greater degree. Further research into sexual dimorphism in bat hearing will allow us to connect species ecology and hearing thresholds, thus providing us with a better understanding of the principles of sexual selection influencing this most important sensory system of bats.

ABRs as useful tools: comparison to literature

For some species measured in the present study, audiograms were previously published and serve as a reference point for the presented data. For example, the audiogram of *Desmodus rotundus* was previously measured in two behavioural studies, each testing three individuals (Gröger and Wiegrebe, 2006; Heffner *et al.*, 2013). The comparatively low hearing threshold detected for low frequencies (<10 kHz) shown in both these previous studies is mirrored in the data acquired in the present study. This low threshold is thought to support prey-generated-noise detection in these sanguivorous bats (**Fig. 3.3**; Gröger and Wiegrebe, 2006). The sensitivity peak in their high-frequency hearing varies across studies, but the peak was in all studies in the range of their echolocation calls, which span between 34.6 and 84.6 kHz (Rodríguez-San Pedro and Allendes, 2017). For *Phyllostomus hastatus* a behavioural audiogram (Koay *et al.*, 2002) and a neuronal audiogram (recorded from the inferior colliculus; Esser and Eiermann, 2008) are available. The high frequency peak of all measurements falls within the echolocation call frequency range (40–80 kHz; Bohn *et al.*, 2004). The peak of the ABR threshold and of the first iso-response line fall between the peaks reported in the behavioural and neuronal audiograms. For *Carollia perspicillata,* also one behavioural audiogram (Koay *et al.*, 2003) and one neuronal audiogram (recorded from the inferior colliculus; Sterbing *et al.*, 1994) are available. While the peak of the ABR curve of *Carollia perspicillata* is about 20 kHz lower than reported in the behavioural and neuronal audiograms (at 50 kHz instead of 70 kHz), the first iso-response line perfectly fits the bandwidth of the species-specific echolocation calls (58–102 kHz; Brinkløv *et al.*, 2011; cf. **Fig. 3.3**). *Pteronotus parnellii* is the only bat species measured in this study that makes use of constant frequency (CF) echolocation calls. CF calls have a very small bandwidth, but the echolocation frequencies are individually different and can vary by up to 35 Hz (Keating *et al.*, 1994). The audiograms of *Pteronotus parnellii* show a narrow sensitivity peak, which corresponds to their acoustic fovea (Kössl and Vater, 1985a; b; Kössl, 1994). In this study, we used a predefined parameter space (see methods), which we tested in all bat species. Although we detected a peak in the ABR threshold, which corresponds to the main CF component of *P. parnellii* (cf. **Fig. 3.3**), we most likely did not measure at the exact position of the individual CF bandwidth and thus may have missed the point of most sensitive hearing. Furthermore, the used tone pips may be spectrally too broad to probe the narrow-band peak in their hearing.

Overall, comparing ABR thresholds and shapes of their iso-response lines with previous audiograms, it becomes clear that the threshold shapes are very comparable. While previous audiograms focus on the extraction of an absolute auditory threshold, we here focus on the qualitative hearing sensitivity. This means, the measured ABR thresholds do not match the exact values of previously reported behavioural audiograms, but are less sensitive (up to 60 dB less sensitive than published audiograms, but generally ~20 dB lower (Kössl, 1992; Koay *et al.*, 2002, 2003; Gröger and Wiegrebe, 2006; Esser and Eiermann, 2008; Heffner *et al.*, 2013). The large sensitivity differences may result from the positioning of the loudspeaker at approximately 0° azimuth and elevation. Note that in previous studies, especially those measuring distortion-product otoacoustic emissions, stimulus presentation was focused on one ear, either with a closed sound system in the ear canal, or with a loudspeaker positioned close to the main sensitivity axis of one ear. Typically, this is lateral position, and often significantly deviating from 0° elevation. Despite these differences in absolute sensitivity, ABRs are an immensely useful tool to assess the neural coding of a large and reproducible parameter space of auditory stimuli. In contrast to behaviourally acquired hearing thresholds, ABRs consistently test responses to the same stimuli between all measured individuals. In the present study, the constant, large parameter space used to assess the ABRs of eleven species allowed us to determine not only species-specific hearing curves, but also enabled us to conduct detailed within and cross-species comparisons. Cross-species comparisons were previously challenging to conduct, as the tested frequencies, amplitudes, and setups often varied between measurements of different species. We argue that this consistency in tested parameters is optimal for a cross-species comparative approach. Moreover, this approach also allowed us to analyse frequency-specific growth functions, which, as demonstrated in the present study, can successfully be used to further investigate a species' sensory capacity. Furthermore, the rapid, minimally invasive acquisition of audiograms also supports greater sample sizes and thus enables detailed within species comparisons, which for example allowed the detection of sexually dimorphic hearing in two bat species in the present study.

Conclusions

Due to their ecology, bats are especially dependent on their hearing as it provides them with vital information about their surroundings. We here present the first large-scale comparison of the hearing capacity of eleven bat species. With the help of ABR measurements, we were able to assess hearing in a consistent, comparable way. This large-scale, comparative approach allowed us to investigate species-specific differences in hearing capacity, but moreover we were able to identify species-independent, general principles for the perception of different types of acoustic signals. The observed differences in hearing thresholds align with the species-specific spectral peaks of echolocation and communication calls in most measured bat species. Despite these sensitivity differences between species, we found interspecies principles for the processing of loudness of social and echolocation calls. One the one hand, the loudness of echolocation calls, which can vary drastically and carries vital information for the bat, is dynamically encoded in the auditory periphery (i.e. shallow ABR growth functions). On the other hand, responses to social call loudness, which does not carry essential information for bats, generally saturate quickly (i.e. steep ABR growth functions). We argue that audiograms and ABR growth functions, as acquired in the present study, present a useful tool for the assessment of cross-species hearing capacities, enables the interpretation of behaviours, and provides important information for the future design of acoustic stimuli. Furthermore, our study shows that ABRs can highlight important factors influencing the life of the investigated species (e.g. sexual dimorphic hearing capacity) and thus indicate interesting directions for future research.

Acknowledgements

The authors want to thank the Smithsonian Tropical Research Institute and especially Rachel Page for providing excellent research conditions in Gamboa, Panama. We further want to thank Constance Scharff, Arturo Zychlinsky, Lorenz Knackstedt, Ine Alvarez van Tussenbroek, May Dixon, and Inga Geipel for their help and support during bat capture and measurement. Lastly, we want to thank the workshop of the Ludwig Maximilian University in Munich for their help with building the ABR setup.

Funding

This study was financed by a Heisenberg Fellowship (DFG KN935 3-1) awarded to MK and a Human Frontier Science Program Research Grant (RGP0058/2016) awarded to LW and SCV. MD received funding from a grant of the German Ministry of Education and Research (BMBF) to the German Centre for Vertigo and Balance Disorders (01EO1401). EZL was funded by a short-term stipend of the German Academic Exchange

Program (DAAD) (#57438025) and a travelling fellowship of the Company of Biologists (JEBTF18113).

Authors' contributions

LW, MK, and EZL conceived the experiment. LW, MD, and EZL designed and tested the setup. EZL, MK, and MN collected the data. EZL, MD, MK, MD, and LW analysed the data. EZL wrote the first draft of the manuscript. All authors contributed to the edition and finalising of the manuscript.

The authors declare no conflict of interest.

References

Alvarez, J.; Willig, M.R.; Knox Jones, J. and Webster, D. (1991). *Glossophaga soricina*. *Mammalian Species* 379, 1–7.

Belknap, D.B. and Suthers, R.A. (1982). Brainstem auditory evoked responses to tone bursts in the echolocating bat, *Rousettus*. *Journal of Comparative Physiology A* 146, 283–289.

Bohn, K.M.; Boughman, J.W.; Wilkinson, G.S. and Moss, C.F. (2004). Auditory sensitivity and frequency selectivity in greater spear-nosed bats suggest specializations for acoustic communication. *Journal of Comparative Physiology A* 190, 185–192.

Bohn, K.M.; Moss, C.F. and Wilkinson, G.S. (2006). Correlated evolution between hearing sensitivity and social calls in bats. *Biology Letters* 2, 561–564.

Bohn, K.M.; Wilkinson, G.S. and Moss, C.F. (2007). Discrimination of infant isolation calls by female greater spear-nosed bats, *Phyllostomus hastatus*. *Animal Behaviour* 73, 423–432.

Boku, S.; Riquimaroux, H.; Simmons, A.M. and Simmons, J.A. (2015). Auditory brainstem response of the Japanese house bat (*Pipistrellus abramus*). *The Journal of the Acoustical Society of America* 137, 1063–1068.

Boughman, J.W. (1997). Greater spear-nosed bats give group-distinctive calls. *Behavioral Ecology and Sociobiology* 40, 61–70.

Brinkløv, S.; Jakobsen, L.; Ratcliffe, J.M.; Kalko, E.K. V. and Surlykke, A. (2011). Echolocation call intensity and directionality in flying short-tailed fruit bats, *Carollia perspicillata* (Phyllostomidae). *The Journal of the Acoustical Society of America* 129, 427–435.

Burkard, R. and Moss, C.F. (1994). The brain stem auditory-evoked response in the big brown bat (*Eptesicus fuscus*) to clicks and frequency-modulated sweeps. *The Journal of the Acoustical Society of America* 96, 801–810.

Carter, G.G.; Logsdon, R.; Arnold, B.D.; Menchaca, A. and Medellin, R.A. (2012). Adult vampire bats produce contact calls when isolated: acoustic variation by species, population, colony, and individual. *PLoS ONE* 7, e38791.

Chaverri, G. and Gillam, E.H. (2010). Cooperative signaling behavior of roost location in a leaf-roosting bat. *Communicative & Integrative Biology* 3, 599–601.

Clement, M.J. and Kanwal, J.S. (2012). Simple syllabic calls accompany discrete behavior patterns in captive *Pteronotus parnellii*: an illustration of the motivation-structure hypothesis. *The Scientific World Journal* 2012, 1–15.

Corwin, J.T.; Bullock, T.H. and Schweitzer, J. (1982). The auditory brain stem response in five vertebrate classes. *Electroencephalography and Clinical Neurophysiology* 54, 629–641.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. 1st ed. William Clowes and sons, London.

Engler, S.; Rose, A. and Knörnschild, M. (2017). Isolation call ontogeny in bat pups (*Glossophaga soricina*). *Behaviour* 154, 267–286.

Esser, K.-H. and Daucher, A. (1996). Hearing in the FM-bat *Phyllostomus discolor*: a behavioral audiogram. *Journal of Comparative Physiology A* 178, 779–785.

Esser, K.-H. and Eiermann, A. (2008). Tonotopic organization and parcellation of auditory cortex in the FM-bat *Carollia perspicillata*. *European Journal of Neuroscience* 11, 3669–3682.

Fenton, M.B. (2003). Eavesdropping on the echolocation and social calls of bats. *Mammal Review* 33, 193–204.

Gillam, E. and Fenton, M.B. (2016). Roles of acoustic social communication in the lives of bats. In: M. B. Fenton, A. D. Grinnell, A. N. Popper, and R. R. Fay (Eds), *Bat Bioacoustics*. Springer Handbook of Auditory Research. Springer, New York, NY, pp. 117–139.

Gillam, E.H. and Chaverri, G. (2012). Strong individual signatures and weaker group signatures in contact calls of Spix's disc-winged bat, *Thyroptera tricolor*. *Animal Behaviour* 83, 269–276.

Gould, E. (1975). Neonatal vocalizations in bats of eight genera. *Journal of Mammalogy* 56, 15–29.

Griffin, D.R. (1980). The early history of research on echolocation. In: R. G. Busnel and J. F. Fish (Eds), *Animal Sonar Systems*. Springer US, Boston, MA, pp. 1–8.

Grilliot, M.E.; Burnett, S.C. and Mendonça, M.T. (2009). Sexual dimorphism in big brown bat (*Eptesicus fuscus*) ultrasonic vocalizations is context dependent. *Journal of Mammalogy* 90, 203–209.

Grilliot, M.E.; Burnett, S.C. and Mendonça, M.T. (2014). Sex and season differences in the echolocation pulses of big brown bats (*Eptesicus fuscus*) and their relation to mating activity. *Acta Chiropterologica* 16, 379–386.

Grinnell, A.D. (1963). The neurophysiology of audition in bats: resistance to interference. *Journal of Physiology* 167, 114–127.

Grinnell, A.D.; Gould, E. and Fenton, M.B. (2016). A history of the study of echolocation. In: M. B. Fenton, A. D. Grinnell, A. N. Popper, and R. R. Fay (Eds), *Bat Bioacoustics*. Springer, New York, NY, pp. 1–24.

Gröger, U. and Wiegrebe, L. (2006). Classification of human breathing sounds by common vampire bat, *Desmodus rotundus*. *BMC Biology* 4, 1–8.

Heckel, G. and Von Helversen, O. (2003). Genetic mating system and the significance of harem associations in the bat *Saccopteryx bilineata*. *Molecular Ecology* 12, 219–227.

Heffner, R.S.; Koay, G. and Heffner, H.E. (2013). Hearing in American leaf-nosed bats. IV: the common vampire bat, *Desmodus rotundus*. *Hearing Research* 296, 42–50.

Henson, O.W.; Schuller, G. and Vater, M. (1985). A comparative study of the physiological properties of the inner ear in Doppler shift compensating bats (*Rhinolophus rouxi* and *Pteronotus parnellii*). *Journal of Comparative Physiology A* 157, 587–597.

Hoffmann, S.; Baier, L.; Borina, F.; Schuller, G.; Wiegrebe, L. and Firzlaff, U. (2008). Psychophysical and neurophysiological hearing thresholds in the bat *Phyllostomus discolor*. *Journal of Comparative Physiology A* 194, 39–47.

Holderied, M.W.; Jones, G. and von Helversen, O. (2006). Flight and echolocation behaviour of whiskered bats commuting along a hedgerow: range-dependent sonar signal design, Doppler tolerance and evidence for "acoustic focussing." *The Journal of Experimental Biology* 209, 1816–1826.

Hulgard, K.; Moss, C.F.; Jakobsen, L. and Surlykke, A. (2016). Big brown bats (*Eptesicus fuscus*) emit intense search calls and fly in stereotyped flight paths as they forage in the wild. *The Journal of Experimental Biology* 219, 334–340.

Jakobsen, L.; Brinkløv, S. and Surlykke, A. (2013). Intensity and directionality of bat echolocation signals. *Frontiers in Physiology* 4, 1–9.

Jewett, D.L. (1970). Volume-conducted potentials in response to auditory stimuli as detected by averaging in the cat. *Electroencephalography and Clinical Neurophysiology* 28, 609–618.

Jewett, D.L.; Romano, M.N. and Williston, J.S. (1970). Human auditory evoked potentials: Possible brain stem components detected on the scalp. *Science* 167, 1517–1518.

Jones, G. (1999). Scaling of echolocation call parameters in bats. *The Journal of Experimental Biology* 202, 3359–3367.

Jones, G. (2005). Echolocation. *Current Biology* 15, 484–488.

Jones, G. and Kokurewicz, T. (1994). Sex and age variation in echolocation calls and flight morphology of Daubenton's bats *Myotis daubentonii*. *Mammalia* 58.

Jones, G. and Siemers, B.M. (2011). The communicative potential of bat echolocation pulses. *Journal of Comparative Physiology A* 197, 447–457.

Jones, G. and Teeling, E.C. (2006). The evolution of echolocation in bats. *Trends in Ecology & Evolution* 21, 149–156.

Jung, K.; Kalko, E.K.V. and Von Helversen, O. (2007). Echolocation calls in central American emballonurid bats: signal design and call frequency alternation. *Journal of Zoology* 272, 125–137.

Jung, K.; Molinari, J. and Kalko, E.K.V. (2014). Driving factors for the evolution of species-specific echolocation call design in new world free-tailed bats (Molossidae). *PLoS ONE* 9, e85279.

Keating, A.W.; Henson, O.W.; Henson, M.M.; Lancaster, W.C. and Xie, D.H. (1994). Doppler-shift compensation by the mustached bat: quantitative data. *The Journal of Experimental Biology* 188, 115–129.

Knörnschild, M.; Feifel, M. and Kalko, E.K.V. (2013). Mother–offspring recognition in the bat *Carollia perspicillata*. *Animal Behaviour* 86, 1–8.

Knörnschild, M.; Feifel, M. and Kalko, E.K.V. (2014). Male courtship displays and vocal communication in the polygynous bat *Carollia perspicillata*. *Behaviour* 151, 781–798.

Knörnschild, M.; Jung, K.; Nagy, M.; Metz, M. and Kalko, E. (2012)a). Bat echolocation calls facilitate social communication. *Proceedings of the Royal Society B* 279, 4827–4835.

Knörnschild, M.; Nagy, M.; Metz, M.; Mayer, F. and Von Helversen, O. (2010). Complex vocal imitation during ontogeny in a bat. *Biology Letters* 6, 156–159.

Knörnschild, M.; Nagy, M.; Metz, M.; Mayer, F. and Von Helversen, O. (2012)b). Learned vocal group signatures in the polygynous bat *Saccopteryx bilineata*. *Animal Behaviour* 84, 761–769.

Koay, G.; Bitter, K.S.; Heffner, H.E. and Heffner, R.S. (2002). Hearing in American leaf-nosed bats. I: *Phyllostomus hastatus*. *Hearing research* 171, 96–102.

Koay, G.; Heffner, R.S.; Bitter, K.S. and Heffner, H.E. (2003). Hearing in American leaf-nosed bats. II: *Carollia perspicillata*. *Hearing Research* 178, 27–34.

Kössl, M. (1992). High frequency distortion products from the ears of two bat species, *Megaderma lyra* and *Carollia perspicillata*. *Hearing Research* 60, 156–164.

Kössl, M. (1994). Evidence for a mechanical filter in the cochlea of the constant frequency bats, *Rhinolophus rouxi* and *Pteronotus parnellii*. *Hearing Research* 72, 73–80.

Kössl, M. and Vater, M. (1985)a). Evoked acoustic emissions and cochlear microphonics in the mustache bat, *Pteronotus parnellii*. *Hearing Research* 19, 157–170.

Kössl, M. and Vater, M. (1985)b). The cochlear frequency map of the mustache bat, *Pteronotus parnellii*. *Journal of Comparative Physiology A* 157, 687–697.

Kunz, T.H. and Hood, W.R. (2000). Parental care and postnatal growth in the chiroptera. In: E. G. Crichton and P. H. Krutzsch (Eds), *Reproductive Biology of Bats*. Academic Press, Cambridge, MA, pp. 415–468.

Linnenschmidt, M. and Wiegrebe, L. (2019). Ontogeny of auditory brainstem responses in the bat, *Phyllostomus discolor*. *Hearing Research* 373, 85–95.

Luo, J.; Goerlitz, H.R.; Brumm, H. and Wiegrebe, L. (2015). Linking the sender to the receiver: vocal adjustments by bats to maintain signal detection in noise. *Scientific Reports* 5, 1–11.

Lv, J.; Simpson, D.M. and Bell, S.L. (2007). Objective detection of evoked potentials using a bootstrap technique. *Medical Engineering and Physics* 29, 191–198.

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

O'Mara, M.T.; Bauer, K.; Blank, D.; Baldwin, J.W. and Dechmann, D.K.N. (2016). Common noctule bats are sexually dimorphic in migratory behaviour and body size but not wing shape. *PLoS ONE* 11, 1–13.

Obrist, M.K. and Wenstrup, J.J. (1998). Hearing and hunting in red bats (*Lasiurus borealis*, Vespertilionidae): audiogram and ear properties. *The Journal of Experimental Biology* 201, 143–54.

Pfalzer, G. and Kusch, J. (2003). Structure and variability of bat social calls: implications for specificity and individual recognition. *Journal of Zoology* 261, 21–33.

Pollak, G.; Henson, O.W. and Johnson, R. (1979). Multiple specializations in the peripheral auditory system of the CF-FM bat, *Pteronotus parnellii*. *Journal of Comparative Physiology A* 131, 255–266.

Postawa, T. and Nagy, Z. (2016). Variation of parasitism patterns in bats during hibernation: the effect of host species, resources, health status, and hibernation period. *Parasitology Research* 115, 3767–3778.

Rinaldi, A. (2007). The scent of life. *EMBO Reports* 8, 629–633.

Rodríguez-San Pedro, A. and Allendes, J.L. (2017). Echolocation calls of free-flying common vampire bats *Desmodus rotundus* (Chiroptera: Phyllostomidae) in Chile. *Bioacoustics* 26, 153–160.

Schmidt, U. and Manske, U. (1973). Die Jugendentwicklung der Vampirfledermäuse. *Zeitschrift für Säugetierkunde* 38, 14–33.

Schmidt, U.; Schlegel, P.; Schweizer, H. and Neuweiler, G. (1991). Audition in vampire bats, *Desmodus rotundus*. *Journal of Comparative Physiology A* 168, 45–51.

Schuchmann, M. and Siemers, B.M. (2010)a). Behavioral evidence for community-wide species discrimination from echolocation calls in bats. *The American Naturalist* 176, 72–82.

Schuchmann, M. and Siemers, B.M. (2010)b). Variability in echolocation call intensity in a community of horseshoe bats: a role for resource partitioning or communication? *PLoS ONE* 5, 1–9.

Siemers, B.M.; Kalko, E.K.V. and Schnitzler, H.-U. (2001). Echolocation behavior and signal plasticity in the Neotropical bat *Myotis nigricans* (Schinz, 1821) (Vespertilionidae): a convergent case with European species of *Pipistrellus*? *Behavioral Ecology and Sociobiology* 50, 317–328.

Simmons, A.M.; Boku, S.; Riquimaroux, H. and Simmons, J.A. (2015). Auditory brainstem responses of Japanese house bats (*Pipistrellus abramus*) after exposure to broadband ultrasonic noise. *The Journal of the Acoustical Society of America* 138, 2430–2437.

Simmons, J.A.; Moss, C.F. and Ferragamo, M. (1990). Convergence of temporal and spectral information into acoustic images of complex sonar targets perceived by the echolocating bat, *Eptesicus fuscus*. *Journal of Comparative Physiology A* 166, 449–470.

Simon, R.; Holderied, M.W. and Von Helversen, O. (2006). Size discrimination of hollow hemispheres by echolocation in a nectar feeding bat. *Journal of Experimental Biology* 209, 3599–3609.

Smotherman, M. and Bakshi, K. (2019). Forward masking enhances the auditory brainstem response in the free-tailed bat, *Tadarida brasiliensis*, during a critical time window for sonar reception. *The Journal of the Acoustical Society of America* 145, 19–24.

Smotherman, M.; Knörnschild, M.; Smarsh, G. and Bohn, K. (2016). The origins and diversity of bat songs. *Journal of Comparative Physiology A* 202, 535–554.

Sterbing, S.J.; Schmidt, U. and Rübsamen, R. (1994). The postnatal development of frequency-place code and tuning characteristics in the auditory midbrain of the phyllostomid bat, *Carollia perspicillata*. *Hearing Research* 76, 133–146.

Stilz, W.-P. and Schnitzler, H.-U. (2012). Estimation of the acoustic range of bat echolocation for extended targets. *The Journal of the Acoustical Society of America* 132, 1765–1775.

Thies, W.; Kalko, E.K.V. and Schnitzler, H.U. (1998). The roles of echolocation and olfaction in two Neotropical fruit-eating bats, *Carollia perspicillata* and *C. castanea*, feeding on *Piper*. *Behavioral Ecology and Sociobiology* 42, 397–409.

Vater, M.; Kössl, M.; Foeller, E.; Coro, F.; Mora, E. and Russell, I.J. (2003). Development of echolocation calls in the mustached bat, *Pteronotus parnellii*. *Journal of Neurophysiology* 90, 2274–2290.

Vernes, S.C. and Wilkinson, G.S. (2019). Behaviour, biology, and evolution of vocal learning in bats. *Philosophical Transactions of the Royal Society of London, Series B*, 1–31.

Vonhof, M.J. and Fenton, M.B. (2004). Roost availability and population size of *Thyroptera tricolor*, a leaf-roosting bat, in north-eastern Costa Rica. *Journal of Tropical Ecology* 20, 291–305.

Waters, D.A. and Jones, G. (1995). Echolocation call structure and intensity in five species of insectivorous bats. *The Journal of Experimental Biology* 198, 475–89.

Wenstrup, J.J. (1984). Auditory sensitivity in the fish-catching bat, *Noctilio leporinus*. *Journal of Comparative Physiology A* 155, 91–101.

Whitlon, D.S.; Young, H.; Barna, M.; Depreux, F. and Richter, C.P. (2019). Hearing differences in hartley guinea pig stocks from two breeders. *Hearing Research* 379, 69–78.

Wilkinson, G.S. (1985). The social organization of the common vampire bat. *Behavioral Ecology and Sociobiology* 17, 123–134.

Wilkinson, G.S. (2003). Social and vocal complexity in bats. In: F. B. M. de Waal and P. L. Tyack (Eds), *Animal Social Complexity*. Harvard University Press, London, England, pp. 322–341.

Wilkinson, G.S. and Adams, D.M. (2019). Recurrent evolution of extreme longevity in bats. *Biology Letters* 15.

Wittekindt, A.; Drexl, M. and Kössl, M. (2005). Cochlear sensitivity in the lesser spear-nosed bat, *Phyllostomus discolor*. *Journal of Comparative Physiology A* 191, 31–36.

Wu, H.; Jiang, T.; Huang, X. and Feng, J. (2018). Patterns of sexual size dimorphism in horseshoe bats: testing Rensch's rule and potential causes. *Scientific Reports* 8, 1–13.

Yovel, Y.; Melcon, M.L.; Franz, M.O.; Denzinger, A. and Schnitzler, H.U. (2009). The voice of bats: how greater mouse-eared bats recognize individuals based on their echolocation calls. *PLoS Computational Biology* 5, 1–10.

Zaidi, A.A.; Mattern, B.C.; Claes, P.; McEcoy, B.; Hughes, C. and Shriver, M.D. (2017). Investigating the case of human nose shape and climate adaptation. *PLoS Genetics* 13, 1–31.

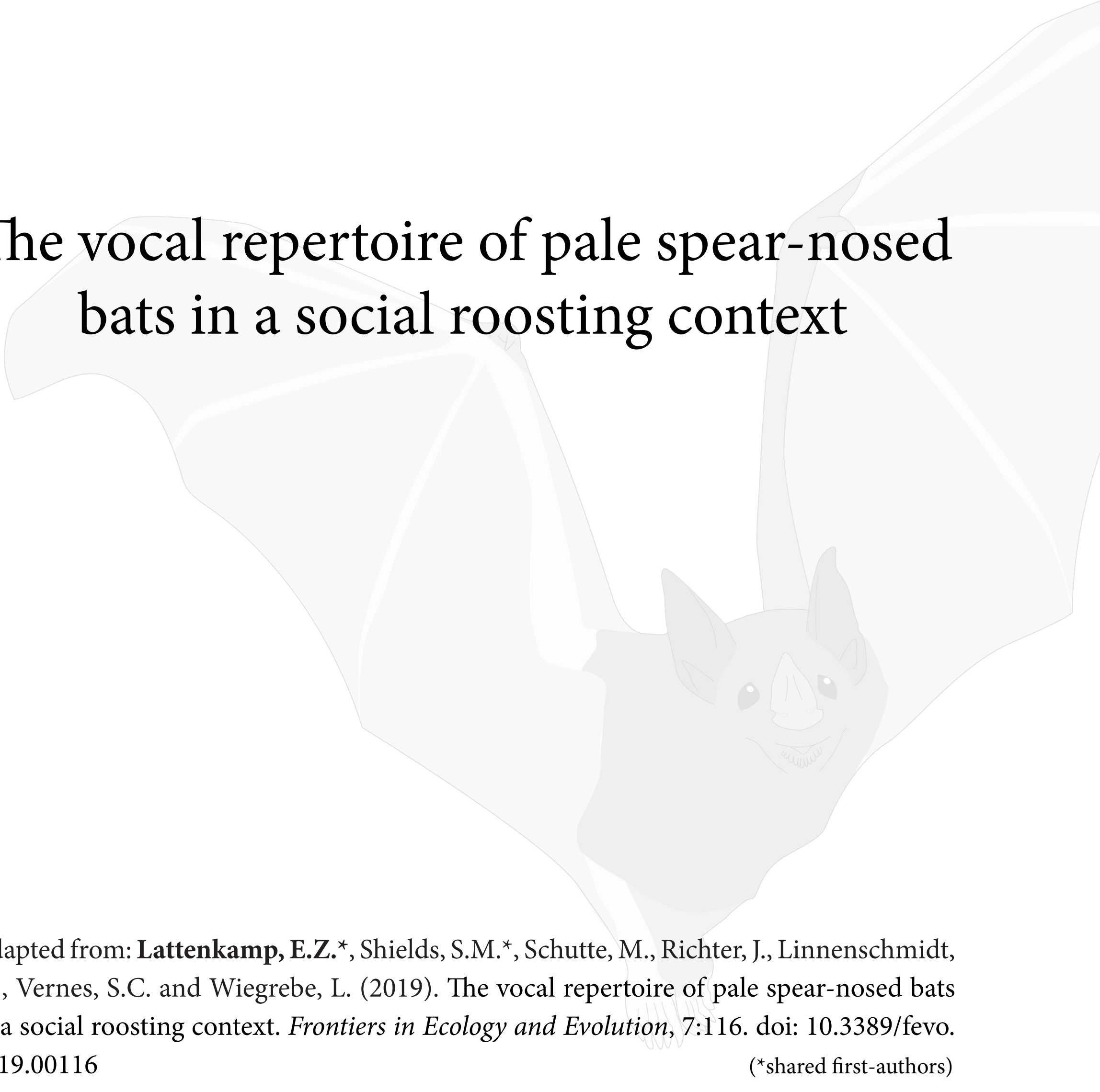
Supplementary Table S3.1: Call parameters (frequency range (min./max.) and peak frequency (pf)) extracted from own data and literature reviews for the phylogenetic comparative analyses. For echolocation calls, we considered either the whole frequency range (FM calls) or the range of the loudest harmonic (usually first; CF calls). Social call parameters were measured for the whole range of calls (bandwidth). For multisyllabic calls, acoustic parameters were assessed on the level of call, not syllable.

species	min. [kHz]	max. [kHz]	pf [kHz]	reference
Echolocation call				
Saccopteryx bilineata	44.3	48.9	45.5-47.2	Knörnschild *et al.*, 2012a
Saccopteryx leptura	50.8	56.0	51.3-54.6	Jung *et al.*, 2007
Rhynchonycteris naso	95.0	102.0	98.2	Jung *et al.*, 2007
Molossus molossus	33.5	42.8	35-41.5	Jung *et al.*, 2014
Myotis nigricans	50.9	61.5	54.2	Siemers *et al.*, 2001
Thyroptera tricolor	117.6	182.4	147.0	unpubl. data I. Geipel: based on 87 calls from 5 bats
Pteronotus parnellii	53.3	61.2	59.8	unpubl. data G. Gessinger: based on 64 calls from 8 bats
Phyllostomus hastatus	40.0	80.0	46.0	Bohn *et al.*, 2004
Desmodus rotundus	34.6	84.6	72.6	Rodríguez-San Pedro and Allendes, 2017
Carollia perspicillata	58.0	102.0	91.0	pf: Brinkløv *et al.*, 2011; min/max: Thies *et al.*, 1998
Glossophaga soricina	56.0	137.0	113.1	pf: unpubl. data G. Gessinger: based on 66 calls from 3 bats; min/max: Simon *et al.*, 2006
Courtship call				
Saccopteryx bilineata	17.7	47.1	30.2	unpubl. data M. Knörnschild
Saccopteryx leptura	12.5	28.7	22.97	unpubl. data M. Knörnschild
Rhynchonycteris naso	29.7	99.1	73.7	unpubl. data M. Knörnschild
Molossus molossus	11.9	41.1	29.6	unpubl. data M. Knörnschild
Carollia perspicillata	11.1	36.6	13.8	Knörnschild *et al.*, 2014
Glossophaga soricina	44.8	82.9	58.8	unpubl. data M. Knörnschild
Isolation calls				
Saccopteryx bilineata	8.1	22.1	11.2	Knörnschild *et al.*, 2012a
Saccopteryx leptura	10.9	27.5	13.6	unpubl. data M. Knörnschild
Rhynchonycteris naso	70.7	88.0	78.9	unpubl. data M. Knörnschild
Molossus molossus	19.7	29.4	23.1	unpubl. data M. Knörnschild
Myotis nigricans	24.7	43.4	27.0	unpubl. data M. Knörnschild
Thyroptera tricolor	50.1	78.4	59.0	unpubl. data G. Chaverri
Pteronotus parnellii	8.0	14.0	10.0	pf: Kanwal unpubl. data [cited in Clement and Kanwal, 2012]; min/max: Vater *et al.*, 2003
Phyllostomus hastatus	10.00	20.00	15.00	pf: Bohn *et al.*, 2004; min/max: Bohn *et al.*, 2007
Desmodus rotundus	12.00	28.00	13.00	pf: Bohn *et al.*, 2006; min/max: Schmidt and Manske, 1973
Carollia perspicillata	22.90	43.40	30.60	Knörnschild *et al.*, 2013

Glossophaga soricina	38.80	66.80	50.20	Engler *et al.*, 2017
Contact call				
Thyroptera tricolor	15.69	35.6	24.44	Chaverri and Gillam, 2010; Gillam and Chaverri, 2012 *inquiry call*
Thyroptera tricolor	49.15	65.34	56.83	Chaverri and Gillam, 2010; Gillam and Chaverri, 2012 *response call*
Phyllostomus hastatus	4.7	11.5	6.7	Boughman, 1997 *screech call*
Desmodus rotundus	19.3	36.2	25.4	Carter *et al.*, 2012
Territorial / alert call				
Saccopteryx bilineata	11	19.91	14.92	Knörnschild *et al.*, 2010 *territorial song*
Glossophaga soricina	24.28	68.51	25.58	unpubl. data M. Knörnschild *alert call type 1*
Glossophaga soricina	11.98	65.87	17.4	unpubl. data M. Knörnschild *alert call type 2*

Chapter 4

The vocal repertoire of pale spear-nosed bats in a social roosting context



Adapted from: **Lattenkamp, E.Z.***, Shields, S.M.*, Schutte, M., Richter, J., Linnenschmidt, M., Vernes, S.C. and Wiegrebe, L. (2019). The vocal repertoire of pale spear-nosed bats in a social roosting context. *Frontiers in Ecology and Evolution*, 7:116. doi: 10.3389/fevo.2019.00116
(*shared first-authors)

Abstract

Commonly known for their ability to echolocate, bats also use a wide variety of social vocalisations to communicate with one another. However, the full vocal repertoires of relatively few bat species have been studied thus far. The present study examined the vocal repertoire of the pale spear-nosed bat, *Phyllostomus discolor*, in a social roosting context. Based on visual examination of spectrograms and subsequent quantitative analysis of syllables, eight distinct syllable classes were defined, and their prevalence in different behavioural contexts was examined. Four more syllable classes were observed in low numbers and are described here as well. These results show that *P. discolor* possesses a rich vocal repertoire, which includes vocalisations comparable to previously reported repertoires of other bat species as well as vocalisations previously undescribed. Our data provide detailed information about the temporal and spectral characteristics of syllables emitted by *P. discolor*, allowing for a better understanding of the communicative system and related behaviours of this species. Furthermore, this vocal repertoire will serve as a basis for future research using *P. discolor* as a model organism for vocal communication and vocal learning and it will allow for comparative studies between bat species.

Introduction

Bats are highly gregarious mammals that have been extensively studied for their ability to echolocate (i.e. gain spatial information from the echoes of prior emitted ultrasonic calls). However, bats also emit social vocalisations to communicate with conspecifics and some bat species have been shown to possess rich vocal repertoires (e.g. Kanwal *et al.*, 1994; Ma *et al.*, 2006; Bohn *et al.*, 2008), supporting intricate social interactions (Wilkinson, 1995, 2003). Current literature on vocal communication in bats illustrates that social vocalisations can be very complex, are highly important for bat sociality, and often vary notably between species. However, research in this field has only been scratching the surface; there is still much to learn about social communication in bats. Relative to the total number of bat species (being the second richest order of mammals with over 1,300 species), very few species have been studied, and even fewer have had their vocal repertoires described. Research on social communication in bats generally focuses on studying a specific subset of vocalisations in a species repertoire—such as neonatal calls (Gould, 1975), calls produced during ontogeny (Knörnschild *et al.*, 2006, 2010a), mother-infant calls (Esser and Schmidt, 1989), male song (Davidson and Wilkinson, 2004)—or more commonly on studying only one particular type of vocalisation—such as distress calls (Russ *et al.*, 2004; Hechavarría *et al.*, 2016) or aggressive calls (Bastian and Schmidt, 2008). Fewer studies have sought to describe the repertoire of a species more comprehensively, defining several types of syllables emitted often in specific behavioural contexts (Behr, 2006; Knörnschild *et al.*, 2010b; Wright *et al.*, 2013). Even fewer have investigated the occurrence of syllable combination and temporal emission patterns (Kanwal *et al.*, 1994; Bohn *et al.*, 2008). These studies have reported a great deal of vocal diversity, ranging from 2 to 22 described vocalisation types per species. The pale spear-nosed bat, *Phyllostomus discolor*, has been in the focus of scientific attention for several years and has been investigated in a variety of psychophysical and neurophysiological studies (Firzlaff *et al.*, 2006; Hoffmann *et al.*, 2008; Heinrich and Wiegrebe, 2013) and, more recently, neurogenetics studies (Rodenas-Cuadrado *et al.*, 2015, 2018). *P. discolor* is a scientifically particularly interesting species as it belongs to the handful of bat species for which evidence of vocal learning (i.e. the ability to produce new or strongly modified vocalisations according to auditory experiences) has been presented (Esser, 1994; Knörnschild, 2014; Lattenkamp *et al.*, 2018 (chapter 5)). Social vocalisations of *P. discolor* are thus especially intriguing as these bats are a valuable system for the study of vocal learning that will help deepen our understanding of this phenomenon (Lattenkamp and Vernes, 2018 (chapter 2)). However, previous studies of social vocalisations in *P. discolor* have mainly focused on mother-infant communication (Esser and Schmidt, 1989; Esser, 1994; Esser and Schubert, 1998; Luo *et al.*, 2017). The current study is the first to assess the vocal communicative repertoire of *P. discolor* in an undisturbed social roosting context, which covers about 80% of their daily activity (La Val, 1970). Pairs and groups of three, four, and six pale spear-nosed bats were repeatedly recorded with a high resolution ultrasonic microphone array under anechoic conditions. Following the methodology of Kanwal and colleagues (1994), vocalisations were initially classified by two independent human raters and the classifications were subsequently statistically verified based on a fixed set of 19 automatically extracted spectral and temporal vocalisation parameters. Eight distinct syllable classes were identified, and four additional, infrequently emitted classes were observed, suggesting that *P. discolor* possesses a diverse vocal repertoire. For the eight distinct syllable classes, the behavioural context at the time of emission was analysed. The combined results present an extensive assessment of the vocal repertoire of the pale spear-nosed bat, *P. discolor*, in a social roosting context.

Materials and methods

Terminology

We follow previous literature in defining syllables as continuous vocal emissions surrounded by periods of silence (Kanwal *et al.*, 1994; Doupe and Kuhl, 1999; Behr and Von Helversen, 2004; Bohn *et al.*, 2008; Gadziola *et al.*, 2012). By this definition, syllables are the smallest, independent acoustic unit of a vocalisation. A call can consist of a single or multiple syllables (Gadziola *et al.*, 2012). For clarity, we specifically focused on studying individual syllables rather than the less objective entity of a call. Syllable classes are used to describe groups of statistically different syllables (cf. Gadziola *et al.*, 2012; Hechavarría *et al.*, 2016), which are assigned depending on the outcome of the classification process described below. We follow the definitions of syllable train and phrase used by Kanwal and colleagues (1994) (cf. 'simple phrase' and 'combination phrase' used by Ma *et al.* (2006)). The term syllable train describes a combination of two or more syllables from the same class, while a phrase describes a combination of syllables from at least two different classes. The silent period between any two syllables in a train or phrase is roughly similar and may be longer than the duration of any one syllable (Kanwal *et al.*, 1994).

Animals

Six adult pale spear-nosed bats were recorded in pairs or groups of three, four, and six. Recordings were done between January and March 2018 for five days per week. The animals recorded in this experiment originated from a breeding colony at Ludwig Maximilian University of Munich, where they were born and housed together throughout their lives. The sex ratio between the bats was equal. One male and one female were approximately 1 year old, while the other bats were between six and nine years old. The bats were provided with a species-specific diet (fruits, supplements, and meal worms) and had *ad libitum* access to water during and outside of the experiment. This experiment was conducted under the principles of laboratory animal care and the regulations of the German Law on Animal Protection. The license to keep and breed *P. discolor* as well as all experimental protocols were approved by the German Regierung von Oberbayern (approval 55.2-1-54-2532-34-2015).

Recording setup

The recording setup was mounted in a sound-insulated chamber (2.24×2.24×1.27 m^3; w×h×d; **Fig. 4.1a**) and consisted of a box containing recording equipment and space for the bats to roost (**Fig. 4.1**). The instrumented box was mounted 1.5 meters above the ground, allowing the bats to fly in and out as they pleased. The ceiling light was only turned on when the experimenter was in the room. Otherwise, the chamber was only dimly illuminated by a small lamp, encouraging the bats to remain in the darker roosting area inside the box. During experimental sessions, the chamber was monitored via an infrared CCD camera (Renkforce CMOS, Conrad Electronic, Hirschau, Germany). Temperature and humidity were monitored from outside the chamber. Vocalisations and behaviours were recorded with both high temporal and spatial resolution via a custom-built acoustic camera. This acoustic camera consisted of a 16-unit ultrasonic microphone array (custom-made on basis of SPU0410LR5H, Knowles Corporation, Itasca, IL, USA) and a high resolution infrared video camera (Point Gray Research Grasshopper3 GS3-U3-41C6NIR; FLIR Integrated Imaging Solutions, Inc., Richmond, BC, Canada) controlled and synchronised via a custom-written Matlab (R2015a, MathWorks, Cambridge, MA, USA) script. By comparing time-of-arrival differences between all microphones of the array, the acoustic camera allows to determine the exact location of a sound source in the recorded video. The camera and microphones were mounted inside of the instrumented box (54×41.5×52 cm^3; w×h×d; **Fig. 4.1**), which was lined with acoustic foam. The bats could enter or exit through a 10 cm wide opening along the bottom of the backside of the box (**Fig. 4.1b**). Two additional doors with latches allowed the experimenter to access the bats and the equipment independently (**Fig. 4.1**). The back wall of the bats' roosting space was lined with mesh for the bats to hang from and crawl on. Two small infrared lights were mounted in the lower corners of the bats' area, illuminating the back wall. An additional infrared light bulb was hung from the mesh mounted on the back wall. This infrared light was used to synchronise the recorded video with the recorded audio. Audio data was recorded via a Horus audio interface (Merging Technologies SA, Puidoux, Switzerland) placed next to the instrumented box in the experimental chamber (**Fig. 4.1a**).

Recording procedure

The six bats were observed in the recording chamber for 47 sessions (either 1.5 or 3 hours long), amounting to a total of 96 h of observation. All 15 possible pair combinations

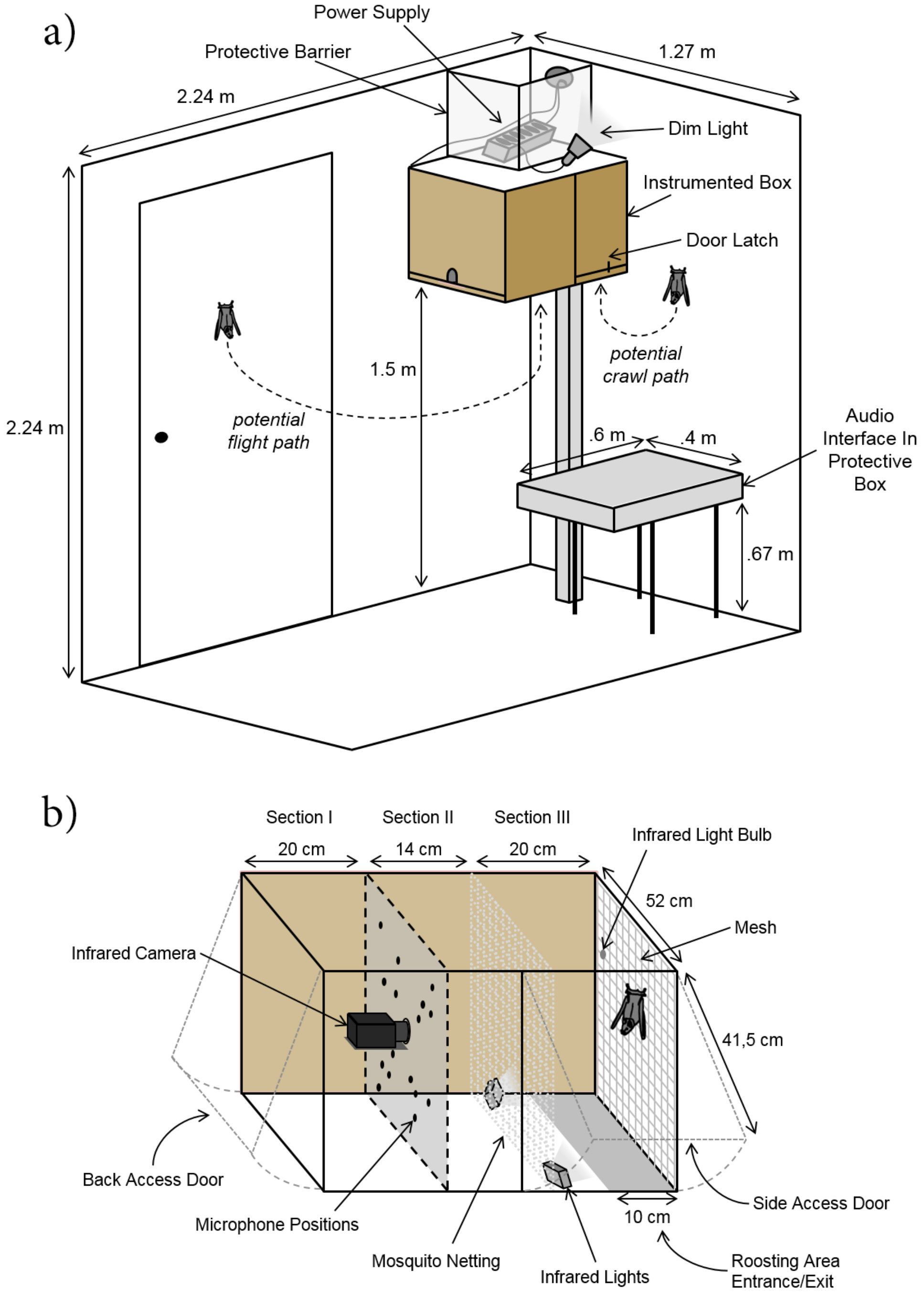


Figure 4.1: Schematic of the setup. a) The sound-insulated chamber containing the instrumented box and the audio interface. The instrumented box was mounted to the wall allowing a free flight path in and out of the box. Dim illumination encouraged the bats to remain in the darker roosting area inside the box. The chamber was monitored via an infrared camera. b) Detailed schematic of the instrumented box containing the acoustic camera. The microphone array and camera faced the freely accessible roosting space. The box was divided into three sections: (I) an area containing the equipment and corresponding cords, (II) a secured space between the bats' roosting area and the panel with the recording equipment, and (III) a section for the bats. The bats' roosting area was illuminated with two infrared lights.

between the six bats were observed for 1.5 h each. On these pair-recording days, the remaining four bats were added into the recording chamber after the first 1.5 h and all six bats were subsequently observed for another 1.5 h. In two additional sessions, first all males and then all females were observed together for 3 h each. Next, all 15 possible combinations of four bats were observed for 3 h as well. During the recording sessions, the bats were monitored in real-time. The recording of audio and visual data was manually triggered by an experimenter from outside the chamber, when social vocalisations were emitted in the chamber. Ultrasonic vocalisations were made audible for the experimenter via real-time heterodyning of two of the 16 microphone channels and presented via headphones. The data acquisition was controlled via a custom-written Matlab script, which saved a 10 s audio ring buffer synchronously for all 16 microphones (sampling rate: 192; microphone gain: 18 dB). The corresponding 10 s long video files were recorded synchronously via StreamPix 6 Single-Camera (NorPix, Inc., Montreal, QC, Canada) (frame rate: 100/s; shutter speed: 9.711ms). The video files were compressed using the Norpix Motion-JPEG Encoder AVI Video Codec.

Acoustic analysis

For the acoustic analysis, we detected and extracted all vocalisations surrounded by silence via a custom-written Matlab script. Syllable detection was based on amplitude peaks identified in the recordings, which were at least 20 dB louder than the background noise and were separated in time from previously detected peaks by at least 5ms. For each identified syllable, the recording from the microphone that picked up the loudest signal was used for analysis. Nineteen acoustic parameters were extracted or calculated for each detected syllable: (1) Syllable duration and (2) maximum syllable amplitude were calculated. To represent the overall frequency content of the syllable, 5 parameters were calculated: (3) spectral centroid frequency (SCF; i.e. weighted mean of the frequencies contained in a syllable), (4) peak frequency (PF; i.e. the frequency with the most energy content), (5) minimum frequency, (6) maximum frequency, and (7) overall syllable bandwidth. The fundamental frequency (f_0) contour of each syllable was detected using the YIN algorithm (de Cheveigné and Kawahara, 2002), and six parameters describing this f_0 contour were then extracted: (8) mean f_0, (9) minimum f_0, (10) maximum f_0, and (11) starting f_0 at the syllable onset. Seven additional parameters describing the f_0 contour were extracted: (12, 13) the coefficients of the best-fitting linear (degree 1) polynomial and (14, 15, 16) quadratic (degree 2) polynomial to the raw contour of the f_0. (17, 18) Furthermore, the root-mean-square errors (RMSE) between the fitted polynomials and the f_0 contours were calculated (19). Lastly, the aperiodicity of syllables was also calculated via the YIN algorithm. It represents how noisy a signal is and functions as a proxy for entropic state of the vocalisation (i.e. an aperiodicity of 0.1 indicates high entropy). The YIN algorithm first assesses the degree of aperiodicity of a recorded call and then tries to assign f_0 to those call segments where aperiodicity is low enough to do so. In the analyses of some quite complex syllables (see below), the f_0 estimate may jump very quickly between quite different values.

Syllable classification

Qualitative categorisation

Following Kanwal and colleagues (1994) and Ma and colleagues (2006), a preliminary classification key consisting of 20 vocalisation classes was generated based on the spectrograms of a subset of recordings and previous literature (Kanwal *et al.*, 1994; Ma *et al.*, 2006). Subsequently, two independent raters visually assessed the spectrograms and waveforms of the extracted syllables based on their duration and frequency information, such as spectral contour, aperiodicity, or suppression of frequencies. The syllables were presented to the raters in four different ways: (1) the waveform of the syllable; (2) the spectrogram of the extracted syllable; (3) the spectrogram of the extracted syllable scaled to a fixed 100ms window; (4) the spectrogram in a 100ms context window, which displayed the recording 50 ms before and after the extracted syllable. This way of displaying the data allowed the raters to determine whether the syllable was extracted well or erroneously. Syllables were either sorted into syllable classes defined in the preliminary classification key, or they were marked as unsuitable for analysis due to low quality (e.g. because of spectral smear, syllable overlap, or incorrect extraction). A few vocalisations were marked as not matching any of the syllable classes present in the preliminary key. These potentially novel syllable classes were later re-examined, and two additional syllable classes were suggested as a result.

Quantitative categorisation

For the quantitative categorisation only high-quality recordings of social syllables that were classified identically by both raters were used. Only classes containing at least 50 detected syllables were analysed. The separability of the classes based on the 19 extracted spectro-temporal parameters was verified and refined based on a 5-fold cross validation procedure (Hastie *et al.*, 2009). The dataset was stratified prior to splitting into folds to avoid empty classes and reduce variance (Forman and Scholz, 2010). In each fold, ~80% of the data for each class were employed to fit a linear discriminant analysis (LDA) classifier (Hastie *et al.*, 2009), and this classifier was used to predict the classes of the remaining 20% of the calls. Each call was used in the test dataset exactly once. A mean confusion matrix was computed from the ground truth labels assigned by the human raters and the labels predicted by the LDA classifier. The confusion matrix was normalised by multiplying each row vector with a constant factor to have row sums of 1. The normalised confusion matrix guided the refinement of the preliminary labels obtained from the qualitative categorisation. As the ultimate goal of the classification process was the development of an automatic classifier, which renders human raters redundant in the future, an algorithmically greedy procedure was used to merge the pair of classes with the highest off-diagonal normalised confusion score. This procedure was done with the input of the human raters, confirming the reasonableness of the merge. The LDA analysis was then rerun on the altered dataset and this algorithm was iterated as long as the human raters agreed that the two candidate classes for merging were non-trivial to separate by their spectrograms. The merging was continued, until a 60% overlap of the human raters and LDA classification was reached.

Behavioural video analysis

We assessed the behavioural context observed during the emission of syllables belonging to the previously established classes. For that reason, an ethogram containing 56 detailed behaviours for *P. discolor* was generated based on personal observations (ML, SMS, EZL). More specifically, the ethogram encompassed 20 behaviours observed in neutral contexts, 18 in prosocial, and 18 in antagonistic behavioural contexts. This ethogram was used by a naïve rater to score the behaviours observed in the video files. The rater was blinded to the emitted syllables contained in the videos. The behavioural scoring was done in the Behavioural Observation Research Interactive Software (BORIS) (Friard and Gamba, 2016), and the behaviour that occurred at the time of syllable emission was extracted.

Results

Within the 96 h of observation 1,434 recordings were made. The automatic syllable finder identified 57,955 vocalisations in these recordings, which were assessed by the two independent raters. The majority of these vocalisations were excluded from the subsequent quantitative analyses for several reasons: 56% (n = 32,551) were excluded, because one or both raters marked them as unsuitable for the classification (due to syllable overlap or low recording quality occurring when vocalisations were emitted outside the instrumented box) or because the two independent raters disagreed on their classification; 2% (n = 1,115) of the recorded sounds were excluded as they presented no vocalisations, but rather scratching noises produced by the bats brachiating on the back wall of the box; and 10% (n = 5,630) of the data were eventually excluded, because not all 19 spectro-temporal syllable parameters could fully be extracted. The remaining 32% (n = 18,658) of the vocalisations represented conservatively selected, high quality syllables classified identically by both independent raters. These syllables were qualitatively and quantitatively assessed as belonging to 13 syllable classes. Of these 13 classes eight were represented by more than 50 syllables and thus evaluated as commonly occurring in this social roosting context (n = 6,162) and four classes were represented by <50 syllables and are thus reported as rarely occurring (n = 81). The largest class (n = 12,416) was comprised of calls with a suppressed f_0 (SF class) and is reported separately below. For the 19 extracted spectro-temporal parameters, the 25th, 50th, and 75th percentiles (i.e. first, second, and third quartiles) are reported below to represent data distribution. These values are presented as follows: Q50 [Q25 Q75]. Additionally, all quartiles for each parameter are listed in supplementary **Table S4.1** for each common syllable class and in supplementary **Table S4.2** for each rare syllable class and the suppressed f_0 class. An example of all commonly occurring syllables is given in **figure 4.2**, while the variation within these classes is illustrated in the supplementary material (supplementary **Fig. S4.1**).

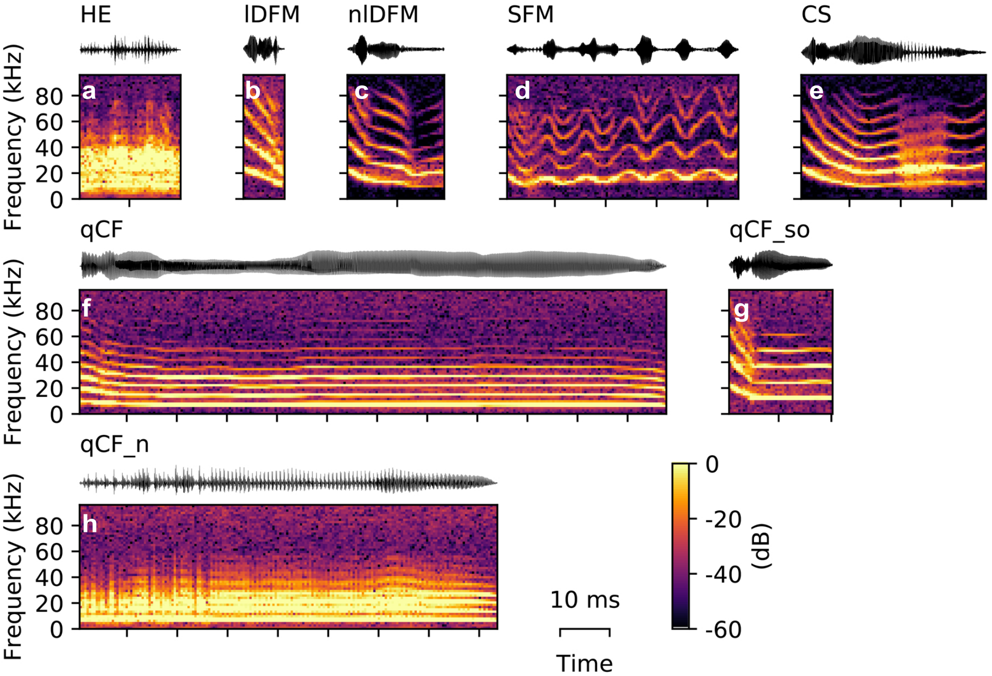


Figure 4.2: Example syllables from the eight commonly occurring classes. From top left to bottom right, one example oscillogram (top) and spectrogram (bottom) of each of the following is displayed: (a) high entropy syllable (HE), (b) linearly downward frequency modulated (lDFM) syllable, (c) non-linearly downward frequency modulated (nlDFM) syllable, (d) sinusoidally frequency modulated (SFM) syllable, (e) composite syllables (CS) with a noisy element within the syllable, (f) long quasi-constant frequency (qCF) syllable, (g) quasi-constant frequency syllable with a steep onset (qCF_so), and (h) noisy quasi-constant frequency (qCF_n) syllable.

Common syllable classes

High entropy (HE) vocalisations

The majority of high quality, commonly emitted social syllables belong to the high entropy (HE) class (n = 3,860; 63% of all syllables in the commonly occurring classes). HE syllables were termed according to their appearance in the spectrogram (i.e. smeared

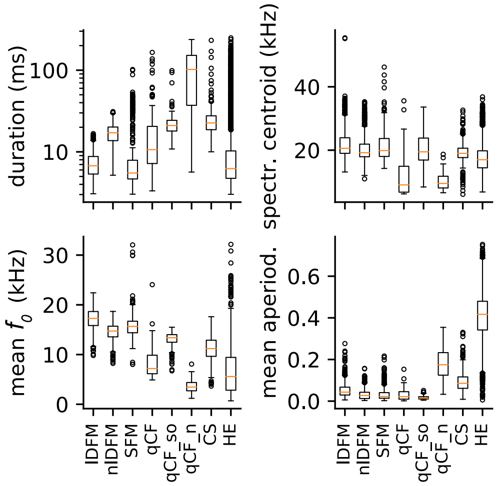
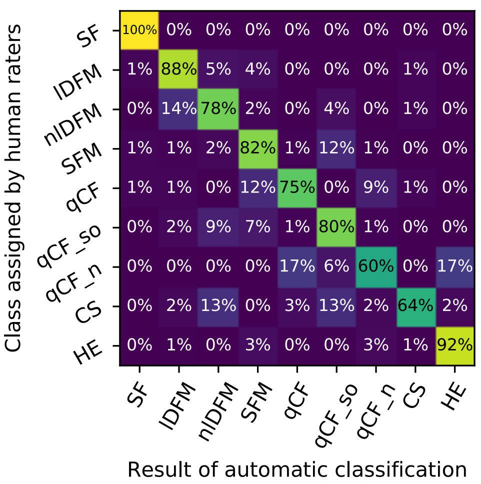


Figure 4.3: Boxplots of four selected spectral and temporal parameters. From top left to bottom right: syllable duration, spectral centroid frequency, mean f_0, and mean aperiodicity. Distributions are shown for the eight commonly occurring syllable classes: linearly downward frequency modulated (lDFM); non-linearly downward frequency modulated (nlDFM), sinusoidally frequency modulated (SFM), quasi-constant frequency (qCF), quasi-constant frequency with a steep onset (qCF_so), noisy quasi-constant frequency (qCF_n), composite syllables (CS), and high entropy syllables (HE).



	SF	lDFM	nlDFM	SFM	qCF	qCF_so	qCF_n	CS	HE
SF	100%	0%	0%	0%	0%	0%	0%	0%	0%
lDFM	1%	88%	5%	4%	0%	0%	0%	1%	0%
nlDFM	0%	14%	78%	2%	0%	4%	0%	1%	0%
SFM	1%	1%	2%	82%	1%	12%	1%	0%	0%
qCF	1%	1%	0%	12%	75%	0%	9%	1%	0%
qCF_so	0%	2%	9%	7%	1%	80%	1%	0%	0%
qCF_n	0%	0%	0%	0%	17%	6%	60%	0%	17%
CS	0%	2%	13%	0%	3%	13%	2%	64%	2%
HE	0%	1%	0%	3%	0%	0%	3%	1%	92%

Figure 4.4: Confusion matrix. Depicted is the distinguishability of the suppressed f_0 (SF) class and the eight commonly occurring syllable classes: linearly downward frequency modulated (lDFM); non-linearly downward frequency modulated (nlDFM), sinusoidally frequency modulated (SFM), quasi-constant frequency (qCF), quasi-constant frequency with a steep onset (qCF_so), noisy quasi-constant frequency (qCF_n), composite syllables (CS), and high entropy syllables (HE). Rows: classes as specified by human raters. Columns: class labels as predicted by an automatic LDA classifier. Rows are normalised to a sum of 100%.

along the frequency axis), and can generally be described as noisy or screechy vocalisations (**Fig. 4.2a**). They can still retain some degree of harmonicity, similar to synthesised tonal noises (iterated rippled noises) (Yost, 1996), and if the residual tonality was strong enough, modulations of f_0 (typically sinusoidal) could be observed (supplementary **Fig. S4.1**). As expected, HE syllables displayed a very high degree of aperiodicity (0.42 [0.34 0.48]; cf. Q50 [Q25 Q75], **Fig. 4.3**). The short average duration of HE syllables (6.24 [4.74 10.27] ms) can be explained by our definition of syllable: The raters observed that long HE calls are often composed of several HE syllables (cf. **Fig. 4.9a**), which were analysed individually, if the call was strongly amplitude modulated and the modulation period longer than 5 ms (cf. 5 ms criterion for syllable separation).

Linearly downward frequency modulated (lDFM) vocalisations

Seven hundred and twenty-seven syllables (12%) are composed of linear downward frequency modulations (lDFM) of f_0 (**Fig. 4.2b**). Linearly DFM syllables are usually relatively short (6.74 [5.35 8.78] ms). They have a steep downward slope (−1.70 [−2.06 −1.41] kHz/ms) and the highest mean f_0 (17.27 [15.83 18.65] kHz; **Fig. 4.3**) of all commonly occurring syllables.

Non-linearly downward frequency modulated (nlDFM) vocalisations

Non-linearly downward frequency modulated (nlDFM) syllables (n = 562; 9%) also sweep downward, but they have a curved shape, or an irregular offset including small constant frequency or upward frequency modulated components (**Fig. 4.2c**). These nlDFM syllables are generally longer than lDFM syllables (17.10 [13.72 20.13] ms; **Fig. 4.3**) and have a lower mean f_0 (14.72 [13.56 15.71] kHz; **Fig. 4.3**). While lDFM and nlDFM syllables have a comparable bandwidth (lDFM: 28.50 [23.25 33.75] kHz; nlDFM: 28.50 [23.25 33.00] kHz), the slope of nlDFM syllables is less steep on average (−0.74 [−1.07 −0.53] kHz/ms).

Sinusoidally frequency modulated (SFM) vocalisations

Also frequently occurring were syllables with a sinusoidal f_0 contour (SFM) (n = 445; 7%). SFM syllables have a stable sinusoidal frequency modulation with small overall variation in modulation depth and modulation frequency, and they generally do not have an onset that notably exceeds the first frequency modulation (**Fig. 4.2d**). However, SFM syllables can also have a steep linear downward sweep onset and a horizontal, ascending, or descending SFM tail (supplementary **Fig. S4.1**). Irregular SFM syllables are also emitted and consist of inconsistent sinusoidal frequency modulations. SFM syllables can vary in both the rate and depth of oscillations. Similar to HE syllables, SFM vocalisations are often strongly amplitude modulated and our definition of syllables thus determines the rather short average durations of the SFM syllables (5.51 [4.66 7.90] ms; **Fig. 4.3**).

Composite (CS) vocalisations

Composite syllables (CS; n = 286; 5%) contain both tonal and noisy elements. Frequently, the syllable begins with a tonal, downward frequency-modulated sweep and then ends with a HE element. One or more HE elements can also occur within syllables (**Fig. 4.2e**). In most cases, a CS is a SFM syllable that is interrupted by one or more HE elements. These syllables had the third highest average aperiodicity (0.09 [0.06 0.12]; **Fig. 4.3**) of the commonly emitted syllables.

Quasi-constant frequency (qCF) vocalisations

Quasi-constant frequency (qCF) syllables (n = 67; 1%) have a near constant f_0 for the duration of the entire syllable (**Fig. 4.2f**). QCF syllables are tonal and have no specific onset, but rather start immediately with the constant frequency element. Overall, syllables in the qCF class tended to have low mean f_0 (7.19 [6.15 9.87] kHz; **Fig. 4.3**).

Quasi-constant frequency vocalisations with a steep onset (qCF_so)

Tonal qCF syllables can also have a steep downward frequency modulated onset (qCF_so; n = 89; 1%; **Fig. 4.2g**). A separate class was created for those qCF_so syllables as they necessarily differ in many parameters from pure qCF syllables, which lack such a clear onset. For example, qCF_so syllables have stronger negative f_0 slopes than the qCF syllables, because of the added onset (qCF_so:−0.40 [−0.51−0.28] kHz/ms; qCF:−0.05 [−0.20 0.01] kHz/ms). For the same reason, the qCF_so syllables are generally longer (qCF_so: 21.03 [17.98 24.33] ms; qCF: 10.64 [7.20 20.52] ms).

Noisy quasi-constant frequency (qCF_n) vocalisations

Noisy quasi-constant syllables (qCF_n) are essentially high entropy versions of the tonal qCF syllables (n = 126; 2%; **Fig. 4.2h**). They also do not start with a frequency modulated onset. Of all syllable classes, qCF_n syllables have the longest average durations (102.20 [37.13 151.90] ms), lowest mean f_0 (3.45 [2.71 4.38] kHz), and lowest spectral centroids (9.51 [8.07 11.78] kHz). They have the second highest average aperiodicity (0.17 [0.12 0.23]; **Fig. 4.3**).

In the quantitative analysis, the LDA classifier performed with an overall accuracy of 87% over the eight classes described above (chance level: 12.5%) (**Fig. 4.4**). The mean overall precision score was 89%, mean overall recall 87%, mean per class precision 67%, and mean per-class recall 76%. **Figure 4.4** reproduces the row-normalised confusion matrix, i.e. each cell shows which percentage of calls of a specific human-rated class is assigned to a specific class label by the automatic classifier. The confusion matrix shows that particularly high recall scores are attained for lDFM and HE calls, which also separate comparatively well univariately (based on mean f_0 and mean aperiodicity, respectively).

Behavioural context of the common syllable classes

For each of the eight commonly occurring syllable classes, 20 videos were scored for the behaviours displayed by the bats during syllable emission. For the lDFM and nlDFM classes only 19 instances could successfully be scored as the behaviour for one instance each was performed outside the field of view of the camera. From the ethogram of 56 detailed behaviours, only 23 behaviours were observed during syllable emission (supplementary **Table S4.3**). Only one single observation was ever made, where a vocalisation was emitted in a neutral behavioural context (**Fig. 4.5a**). More specifically, a single HE syllable was emitted in a context scored as 'brachiating on walls or ceiling'. Other than that, syllables were always emitted either in a prosocial or an antagonistic behavioural context. The behavioural analyses show that the HE syllables are emitted 95% of the time in antagonistic encounters (**Fig. 4.5a**). One exception is the above-mentioned single observation of a HE syllable emitted in a neutral context. All other syllables were, with varying prevalence, emitted in both, prosocial and antagonistic contexts. Syllables from the qCF, SFM, and nlDFM classes were emitted in prosocial behavioural contexts in 75–85% of the scored videos (**Fig. 4.5a**). CS, lDFM, and qCF_so syllables were emitted slightly more often in prosocial than antagonistic contexts (in 55–63% of the videos). Noisy qCF syllables (qCF_n) were emitted in antagonistic

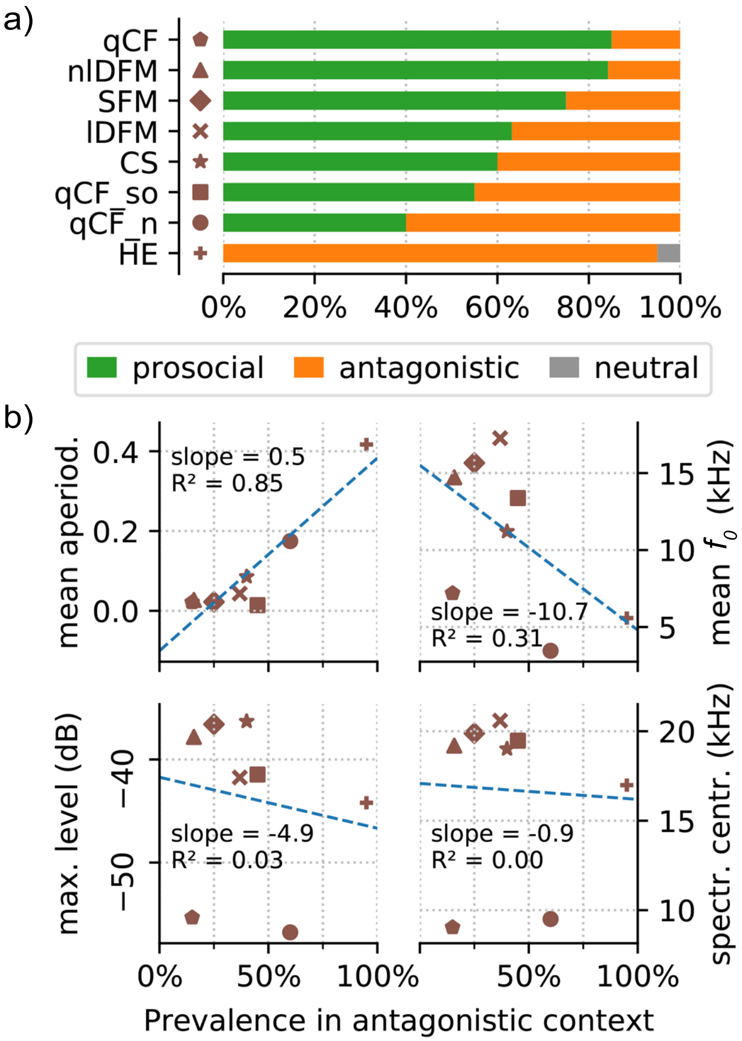


Figure 4.5: Behavioural context of the common syllable classes. (a) The top panel shows the behavioural context scored for the eight commonly occurring syllable classes. (b) The bottom panel shows the correlation of four syllable parameters (top left to bottom right: mean aperiodicity, mean fundamental frequency (f_0), maximum level, and aperiodicity) in dependence of their prevalence in antagonistic encounters. For syllable class abbreviations see **Fig. 4.4**.

behavioural contexts in 40% of the scored videos. Stable correlations were found between some acoustic parameters and the behavioural context in which a syllable was emitted: Specifically, the measured aperiodicity of the syllables was strongly positively correlated with their prevalence in antagonistic encounters (**Fig. 4.5b**). Also syllable f_0 are lower during antagonistic behaviours (**Fig. 4.5b**).

Rare syllable classes

In addition to the commonly occurring syllable classes, several vocalisations were repeatedly, but extremely infrequently emitted. Specifically, out of the total of 18,658 high quality recordings fewer than 50 vocalisations per rare syllable class were recorded. Thus, not enough data are available to include these vocalisations in the statistical analysis. They are described in the following as purely observational and should be considered as rarely emitted, at least in a social roosting context.

Puffs

During the recording sessions, the bats repeatedly emitted air puffs (n = 42), which appeared to result from bats forcefully expelling air through their nostrils. These sounds are not necessarily to be considered sneezing, but are rather short nasal exhalation potentially used to clean the nostrils. The spectrograms of puffs appear to be noisy sound clouds with a sharp onset (**Fig. 4.6a**). As the puffs did not contain a tonal component, the mean aperiodicity and bandwidth of these puffs were the highest of all recorded vocalisations (aperiodicity: 0.43 [0.40 0.47] and bandwidth: 45.75 [42.00 48.75] kHz).

V-shaped vocalisations

Syllables from this class (n = 30) consisted of a downward frequency modulated onset and a subsequent upward sweep, resulting in a characteristic 'V'-shaped frequency contour (**Fig. 4.6b**). Vocalisations in the V-shaped class are comparable to the sinusoidal vocalisations, but always end within the first modulation.

Noisy quasi-constant frequency vocalisations with steep onset (qCF_nso)

The qCF_nso syllables were recorded only five times and were a combination of the qCF_n and the qCF_so syllable classes (**Fig. 4.6c**). They also consist of a steep downward frequency modulated onset followed by a quasi-constant syllable element. However, they were emitted with higher sound pressure levels than qCF_n and higher aperiodicity than qCF_so syllables (supplementary **Tables S4.1**, **S4.2**), resulting in a noisy version of the qCF_so syllable type.

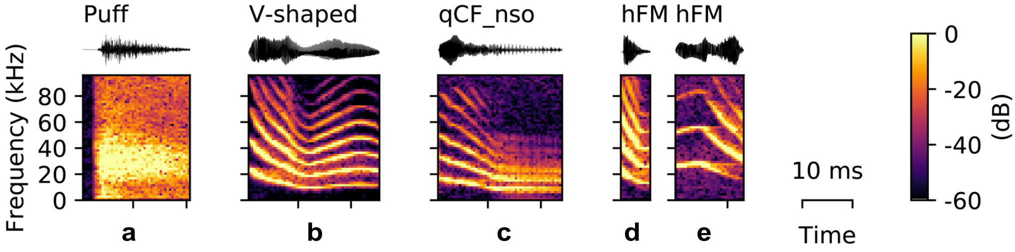


Figure 4.6: Examples of syllables from rarely occurring classes. a) Puff sound, b) V-shaped, c) noisy quasi-constant frequency syllables with a steep onset (qCF_nso), d) upward-hooked, and (e) downward-hooked frequency modulated (hFM) syllables.

Hooked frequency modulated (hFM) vocalisations

Upward- or downward-hooked frequency modulated (hFM) syllables (n = 4) are characterised by the similarity between the shape of the vocalisation displayed in the spectrogram and a hook. These syllables are typically short and can appear in either an upward-hooked (**Fig. 4.6d**) or a downward-hooked (**Fig. 4.6e**) form. These two hFM syllable types were the least abundant (upward-hooked: n = 1; downward-hooked: n = 3). HFM syllables had the highest average spectral centroid aside from syllables with a suppressed f_0 (27.08 [21.71 33.09] kHz). However, comparative results should be taken with care, as the quantitative characteristics of this class are not well-supported due to the small number of syllables detected.

Suppressed fundamental (SF) class

The vast majority of recorded syllables belonged to the suppressed fundamental (SF) class (n = 12,416; 66% of the high quality, uniformly rated syllables). This syllable class can easily be distinguished from all other recorded syllables by its high spectral centroid (**Fig. 4.7**). In fact, the spectral centroid frequency is a parameter showing a clear bimodal distribution of the data, splitting SF syllables and syllables of all other classes (**Fig. 4.7**). Syllables in the SF class have either a fully or partially suppressed f_0, and the dominant harmonic is instead the second or even third harmonic (**Fig. 4.8**). SF syllables typically had short durations (4.07 [3.46 5.04] ms, supplementary **Table S4.2**) and high spectral centroids (43.05 [40.65 46.51] kHz, supplementary **Table S4.2**). Especially the very

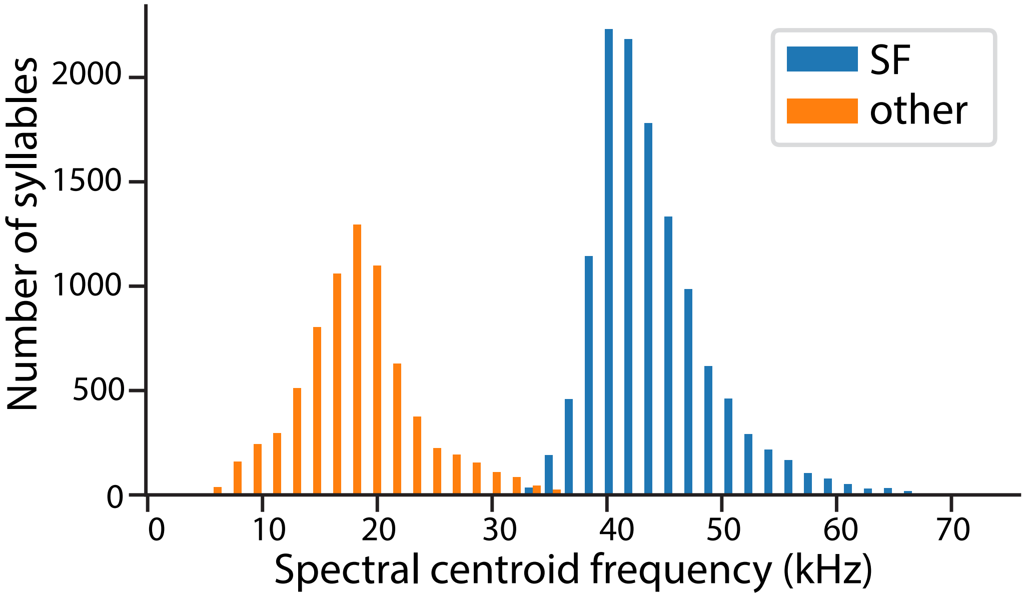


Figure 4.7: Spectral centroid frequencies of all analysed syllables, separated based on whether or not the syllable was classified as belonging to the SF class.

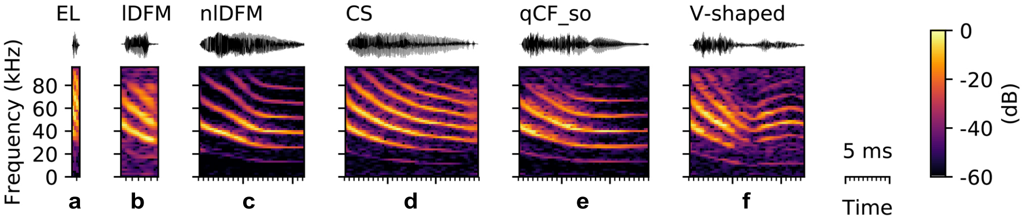


Figure 4.8: Spectrograms of syllables with a suppressed fundamental frequency resembling syllables from other classes. a) species-typical echolocation call (EL), b) linearly downward frequency modulated syllable (lDFM), c) non-linearly downward frequency modulated syllable (nlDFM), d) composite syllable, (e) quasi-constant frequency syllable with a steep onset (qCF_so), and (f) V-shaped syllable.

short durations indicate that this syllable class includes the species-specific echolocation calls, which typically range in duration between 0.3 and 2.5 ms (Rother and Schmidt, 1985; Kwiecinski, 2006; Luo *et al.*, 2015). However, the SF class also included syllables, which structurally resembled syllables from other commonly occurring syllables classes with the only decisive difference that the f_0 was fully or partially suppressed (**Fig. 4.8**). Based on these strong characteristics and the varying shape of the SF syllables, this class can be easily separated from the other classes, but should rather be regarded as a meta-class, containing versions with suppressed f_0 of most other syllable types. The function of these SF calls is currently uncertain and might or might not vary from the normal context of the syllable type with an expressed f_0.

Syllable combinations: trains and phrases

Very few studies have investigated temporal emission patterns of syllables and the existence of consistently occurring syllable combinations (e.g. Kanwal *et al.*, 1994; Bohn *et al.*, 2008; Knörnschild *et al.*, 2014; Smotherman *et al.*, 2016). Previous literature shows, however, that for certain bat species the temporal emission pattern of social vocalisations can be highly complex. *Phyllostomus discolor* also emits combinations of syllables in a standardised order and with constant temporal emission patterns. Temporal relationships between syllables were not analysed in the current work, thus we cannot draw qualitative conclusions about this aspect of the vocalisations. However, during syllable classification we observed several syllable combinations of varying length, complexity, and number of contained syllables (**Fig. 4.9**). Observed syllable trains consist of multiple syllables from the same class repeated with roughly the same temporal distance, whereby the silent interval can be longer than the preceding syllable (**Fig. 4.9b,c**). Syllable trains can be of varying overall length, depending on the number of contained syllables. Phrases consist of syllables from two or more classes, which can be repeated several times (usually in a fixed temporal distance; **Fig. 4.9d–f**). We found eight different types of syllable combination, which were repeatedly recorded over the duration of the experiment. The behavioural purpose of syllable trains and phrases is thus far purely speculative.

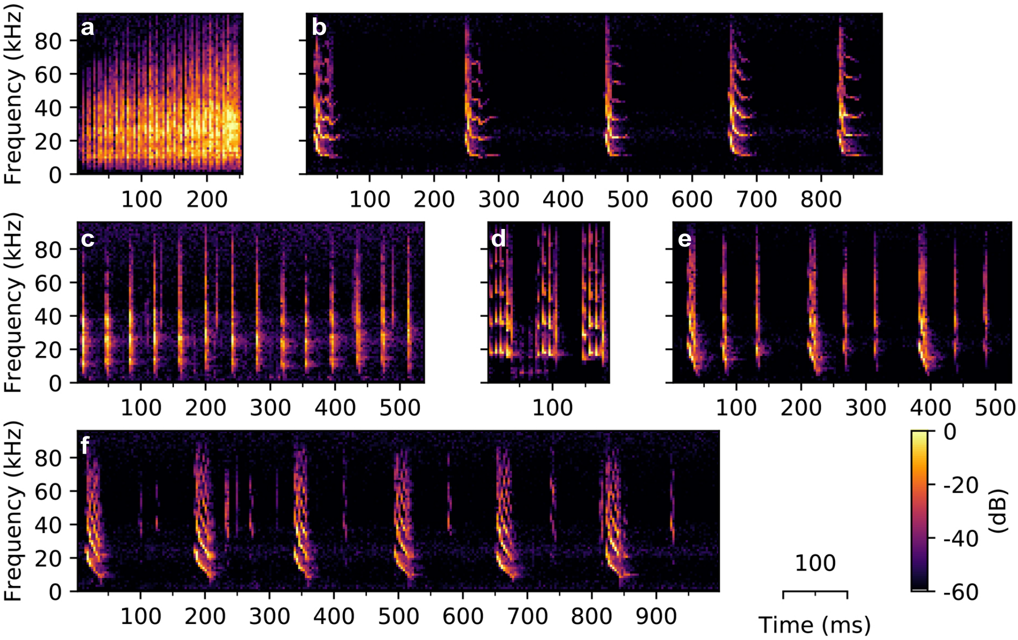


Figure 4.9: Example spectrograms of three syllable trains (a–c) and three phrases (d–f) emitted by *P. discolor*. Syllable trains consist of several repetitions of one syllable type, e.g. (a) high entropy syllables, (b) sinusoidally frequency modulated, or (c) simple linearly frequency modulated syllables. (d–f) Phrases consist of syllables from at least two different classes; phrases are depicted three (d,e) or six (f) times in order to show the temporal relationship between repetitions of the different phrases.

A repetitive emission of phrases might serve to emphasise the transmitted information, but the number of phrase repetitions could also carry information by itself. Though the function and magnitude of syllable trains and phrases in these bats is currently unknown, we want to report our observation of them to encourage further research in this direction.

Discussion

Vocalisations of *Phyllostomus discolor*: known and novel

Here we present an extensive assessment of the vocal repertoire of the pale spear-nosed bat, *P. discolor*. As we recorded vocalisations in a social roosting context, which is the main pastime of *P. discolor* (Kwiecinski, 2006), we are confident that we identified the majority of social vocalisations emitted by this species. From 18,658 high-quality syllable recordings, we were able to define eight distinct classes, uniquely different from each other in their spectro-temporal parameters. We were also able to support the acoustic analysis with a detailed assessment of the behavioural contexts in which these eight syllable classes are generally emitted (supplementary **Table S4.3**; **Fig. 4.5**). Furthermore, we describe four additional call classes, which were only infrequently emitted by the bats and are thus described here, but not analysed on the basis of their spectro-temporal characteristics.

Most syllable classes described in the present study have never before been observed for this species. Especially the quasi-constant frequency modulated (qCF) class and classes containing qCF elements (i.e., qCF_so and qCF_n) have hitherto not been reported for *P. discolor*. From our behavioural observations (supplementary **Table S4.3**) it becomes apparent that all three classes containing syllables with a qCF element are used in very versatile behavioural contexts. This could indicate a loose behavioural association with the syllable structure and one could speculate about a behaviourally more meaningful variation of these syllables in their specific context (e.g. duration of qCF element could indicate special emphasis on a particular meaning). However, such speculations await experimental confirmation. Sinusoidally frequency modulated (SFM) syllables have received considerable attention in previous literature. In *P. discolor*, SFM syllables were found to be used in mother-infant communication (as e.g., maternal directive calls and late forms of infant isolation calls) and can encode individual signatures, and even vocal dialects (Gould, 1975; Esser and Schmidt, 1989; Esser and Lud, 1997; Esser and Schubert, 1998). We can confirm that the majority of the analysed SFM syllables were emitted in the behavioural contexts 'attention seeking' or 'vocal contact', which are both in line with previous observations (supplementary **Table S4.3**). In addition to the usage of SFM syllables in these contexts, we also demonstrated their emission in antagonistic encounters (**Fig. 4.5a**). Emission of one syllable type in a variety of different behavioural contexts suggests complex communicative function or purpose. Thus, our results support previous findings, which advocate syllable subgroups, in which vocalisations with very similar acoustic parameters can be further split up based on associated behaviours (Bohn *et al.*, 2008; Kanwal, 2009). As described above, the syllable classification here presented is based purely on spectrogram shape and the extracted syllable parameters. This allows us to present mathematically distinct syllable classes and validates our first, subjective classification scheme. Nevertheless, the established classes may be further differentiated according to their behavioural contexts. Our behavioural assessments show that syllables from a single class with very similar acoustic characteristics can be used in up to 10 different behavioural contexts (supplementary **Table S4.3**). The establishment of syllable subgroups (i.e. splitting of the presented syllable classes) based on their contextual usage would require extensive, detailed behavioural observations and ideally confirmation via playback experiments. We also want to highlight the possibility that additional syllable classes might be contained in the *P. discolor* repertoire, which were not emitted in the here reported social roosting context.

Comparison to the closely related species *(P. hastatus)*: emerging vocal complexity

The number of distinct syllable classes assessed in this study (eight) is comparable to vocal repertoire descriptions of other bat species, which also found between 2 and 10 syllable types (Nelson, 1964; Gould, 1975; Barclay *et al.*, 1979; Kanwal *et al.*, 1994; Pfalzer and Kusch, 2003; Bohn *et al.*, 2004; Wright *et al.*, 2013; Knörnschild *et al.*, 2014). When comparing the vocal repertoire of *P. discolor* to a closely related species (*P. hastatus*, which lives under essentially identical social and ecological conditions), it is noticeable, that the vocal repertoire of *P. hastatus* is less expansive. In addition to their echolocation calls, only two types of social calls are reported for *P. hastatus*, namely group-specific foraging calls, so-called screech calls, and infant isolation calls (Bohn *et al.*, 2004). The screech calls of *P. hastatus* were shown to be used for the recognition of social group members during foraging, while infant isolation calls help mothers to recognise offspring (Boughman, 1997; Boughman and Wilkinson, 1998; Wilkinson and Boughman, 1998). Vocalisations reported as infant isolation calls are distinctly different between *P. discolor* and *P. hastatus*, with the former using single, clearly sinusoidally frequency modulated calls (Esser and Schmidt, 1989) and the latter typically using a pair of linear or bent frequency modulated calls (Bohn *et al.*, 2007). The broadband, noisy screech calls of *P. hastatus* are similar in their spectral characteristics to the here defined high entropy (HE) syllables (Boughman, 1997). Screech calls are, however, used for the coordination of foraging activities and are not emitted in antagonistic behaviours contexts as observed in this study (supplementary **Table S4.3**). The surprising difference in the size of the vocal repertoires of these closely related species, which are so similar in their ecology and lifestyle, only highlights the value of *P. discolor* as a model species for vocal communication and vocal learning. The vocal repertoires of the other members of the genus (*P. elongatus* and *P. latifolius*) are still unknown. Uncovering the evolutionary background of the emergence of such differences in vocal complexity in closely related species might help us to shed light on the evolution of communicative systems and the capacity for vocal learning in bats.

Similarities to distantly related species: acoustic universals

A number of distantly related bat species were reported to emit high entropy calls during aggressive encounters (Russ *et al.*, 2004; Hechavarría *et al.*, 2016; Prat *et al.*, 2016). It has been hypothesised that aggressive vocalisations tend to always be long, rough, and lower in frequency (Briefer, 2012). We confirmed a strongly positive correlation between the mean syllable class aperiodicity and its prevalence in antagonistic confrontations (**Fig. 4.5b**). We also detected a negative correlation between the mean f_0 of a syllable class and its occurrence during aggressive encounters. Overall, these findings support the idea of shared characteristics of mammalian vocalisations in strongly emotional behavioural contexts and provide further evidence for acoustic universals and potential for interspecies communication (Filippi, 2016; Filippi *et al.*, 2017).

Temporal emission patterns: evidence for higher order vocal constructs

Previous studies suggest that syllable sequences such as trains or phrases can encode combinational meaning or emphasis, thus increasing the available vocal complexity for a given bat species (Behr and Von Helversen, 2004; Bohn *et al.*, 2008; Smotherman *et al.*, 2016; Knörnschild *et al.*, 2017). Sequences of syllables, which present higher order vocal constructs, have been described for a few bat species (review: Smotherman *et al.*, 2016). However, for the family Phyllostomidae, which is a very ecologically diverse and speciose bat family (>140 described species within 56 genera; Wetterer *et al.*, 2000), to date there have been only two published observations of the use of such hetero-syllabic constructs. Specifically, only for Seba's short-tailed bat (*Carollia perspicillata*) and the buffy flower bat (*Erophylla sezekorni*) descriptions of syllable combinations (simple trains and phrases) are available (Murray and Fleming, 2008; Knörnschild *et al.*, 2014). Here we provide further evidence for syntax usage in a phyllostomid bat, which opens this family up for future in-depth research on this topic.

Conclusions

In the framework of this study, 18,658 high-quality social vocalisations of the pale spear-nosed bat, *Phyllostomus discolor*, were recorded under laboratory conditions. From 6,162 of these, it was possible to define eight robust syllable classes, including some vocalisations not previously known to be produced by these bats. Furthermore, we were also able to assess the behavioural contexts in which these syllable classes are generally emitted, and could show that for example high entropy syllables are exclusively emitted in aggressive encounters. We also describe four additional, rarely occurring syllable classes (81 recordings in total). The majority of recorded syllables (n = 12,416) present evidence for a meta-class of vocalisations, i.e. syllables from different classes with the joint characteristic of having a suppressed f_0. Finally, we present tentative evidence for emission of syllable trains and phrases in this Neotropical bat species, highlighting the described complexity of *P. discolor* vocalisations. Together, these results present an extensive assessment of the vocal repertoire of *P. discolor* in a social roosting context and the associated behavioural contexts.

Acknowledgements

The authors want to thank Mirjam Knörnschild for helpful discussions concerning this manuscript and the LMU workshop for their help in constructing the recording setup. We would also like to thank two anonymous reviewers for their helpful and constructive feedback on earlier versions of this manuscript.

Funding

The research was funded by a Human Frontiers Science Program Research Grant (RGP0058/2016) awarded to LW and SCV. SMS was funded by the Fulbright U.S. Student Program.

Author contributions

LW, ML, SCV, and EZL conceived and supervised the study. SMS recorded the data. SMS, EZL, and ML developed the classification key. LW wrote the syllable detection and analysis program. EZL and SMS performed the syllable classification. MS conducted the statistical analyses and data presentation. JR rated the behavioural context. EZL wrote the first draft of the manuscript. All authors contributed to the writing, editing, and revising of the final paper.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary material

The supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo. 2019.00116/full#supplementary-material

References

Barclay, R.M.R.; Fenton, M.B., and Thomas, D.W. (1979). Social behavior of the little brown bat, *Myotis lucifugus*. *Behavioral Ecology and Sociobiology* 6, 137–146

Bastian, A., and Schmidt, S. (2008). Affect cues in vocalizations of the bat, *Megaderma lyra*, during agonistic interactions. *The Journal of the Acoustical Society of America* 124, 598–608.

Behr, O. (2006). The vocal repertoire of the sac-winged bat, *Saccopteryx bilineata*. *Doctoral thesis*, Friedrich-Alexander-Universität Erlangen-Nürnberg.

Behr, O., and Von Helversen, O. (2004). Bat serenades - Complex courtship songs of the sac-winged bat (*Saccopteryx bilineata*). *Behavioral Ecology and Sociobiology* 56, 106–115.

Bohn, K.M.; Boughman, J.W.; Wilkinson, G.S., and Moss, C.F. (2004). Auditory sensitivity and frequency selectivity in greater spear-nosed bats suggest specializations for acoustic communication. *Journal of Comparative Physiology A* 190, 185–192.

Bohn, K.M.; Schmidt-French, B.; Ma, S.T., and Pollak, G.D. (2008). Syllable acoustics, temporal patterns, and call composition vary with behavioral context in Mexican free-tailed bats. *The Journal of the Acoustical Society of America* 124, 1838–1848.

Bohn, K.M.; Wilkinson, G.S., and Moss, C.F. (2007). Discrimination of infant isolation calls by female greater spear-nosed bats, *Phyllostomus hastatus*. *Animal Behaviour* 73, 423–432.

Boughman, J.W. (1997). Greater spear-nosed bats give group-distinctive calls. *Behavioral Ecology and Sociobiology* 40, 61–70.

Boughman, J.W., and Wilkinson, G.S. (1998). Greater spear-nosed bats discriminate group mates by vocalizations. *Animal Behaviour* 55, 1717–1732.

Briefer, E.F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology* 288, 1–20.

Davidson, S.M., and Wilkinson, G.S. (2004). Function of male song in the greater white-lined bat, *Saccopteryx bilineata*. *Animal Behaviour* 67, 883–891.

de Cheveigné, A., and Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America* 111, 1917–1930.

Doupe, A.J., and Kuhl, P.K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Reviews in Neuroscience* 22, 567–631.

Esser, K.H. (1994). Audio-vocal learning in a non-human mammal: the lesser spear-nosed bat *Phyllostomus discolor*. *Neuroreport* 5, 1718–1720

Esser, K.H., and Lud, B. (1997). Discrimination of sinusoidally frequency modulated sound signals mimicking species specific communication calls in the FM bat *Phyllostomus discolor*. *Journal of Comparative Physiology A* 180, 513–522.

Esser, K. H., and Schmidt, U. (1989). Mother-infant communication in the lesser spear-nosed bat *Phyllostomus discolor* (Chiroptera, Phyllostomidae) - evidence for acoustic learning. *Ethology* 82, 156–168.

Esser, K.H., and Schubert, J. (1998). Vocal dialects in the lesser spear-nosed bat *Phyllostomus discolor*. *Naturwissenschaften* 85, 347–349.

Filippi, P. (2016). Emotional and interactional prosody across animal communication systems: a comparative approach to the emergence of language. *Frontiers in Psychology* 7, 1393.

Filippi, P.; Congdon, J.V.; Hoang, J.; Bowling, D.L.; Reber, S.A.; Pašukonis, A.; Hoeschele, M.; Ocklenburg, S.; de Boer, B.; Sturdy, C.B.; Newen, A, and Güntürkün, O. (2017). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals. *Proceedings of the Royal Society B* 284, 1–9.

Firzlaff, U.; Schörnich, S.; Hoffmann, S.; Schuller, G., and Wiegrebe, L. (2006). A neural correlate of stochastic echo imaging. *Journal of Neuroscience* 26, 785–791.

Forman, G., and Scholz, M. (2010). Apples-to-apples in cross-validation studies. *ACM SIGKDD Explorations Newsletter* 12, 49.

Friard, O., and Gamba, M. (2016). BORIS: a free, versatile open-source event logging software for video/audio coding and live observations. *Methods in Ecology and Evolution* 7, 1325–1330.

Gadziola, M.A.; Grimsley, J.M.; Faure, P.A., and Wenstrup, J.J. (2012). Social vocalizations of big brown bats vary with behavioral context. *PLoS ONE* 7, e44550.

Gould, E. (1975). Neonatal vocalizations in bats of eight genera. *Journal of Mammalogy* 56, 15–29.

Hastie, T.; Tibshirani, R., and Friedman, J. (2009). The elements of statistical learning, 2nd Edition. Springer New York, NY.

Hechavarría, J.C.; Beetz, M.J.; Macias, S., and Kössl, M. (2016). Distress vocalization sequences broadcasted by bats carry redundant information. *Journal of Comparative Physiology A* 202, 503–515.

Heinrich, M., and Wiegrebe, L. (2013). Size constancy in bat biosonar? Perceptual interaction of object aperture and distance. *PLoS ONE* 8, e61577.

Hoffmann, S.; Baier, L.; Borina, F.; Schuller, G.; Wiegrebe, L., and Firzlaff, U. (2008). Psychophysical and neurophysiological hearing thresholds in the bat *Phyllostomus discolor*. *Journal of Comparative Physiology A* 194, 39–47.

Kanwal, J.S. (2009). Audiovocal communication in bats. In: L. R. Squire (Eds), *Encyclopedia of Neurosciences*, Academic Press, Oxford, UK, pp. 681–690.

Kanwal, J.S.; Matsumura, S.; Ohlemiller, K., and Suga, N. (1994). Analysis of acoustic elements and syntax in communication sounds emitted by mustached bats. *The Journal of the Acoustical Society of America* 96, 1229–1254.

Knörnschild, M. (2014). Vocal production learning in bats. *Current Opinion in Neurobiology* 28, 80–85.

Knörnschild, M.; Behr, O., and von Helversen, O. (2006). Babbling behavior in the sac-winged bat (*Saccopteryx bilineata*). *Naturwissenschaften* 93, 451–454.

Knörnschild, M.; Blüml, S.; Steidl, P.; Eckenweber, M., and Nagy, M. (2017). Bat songs as acoustic beacons - Male territorial songs attract dispersing females. *Scientific Reports* 7, 1–11.

Knörnschild, M.; Feifel, M., and Kalko, E.K.V. (2014). Male courtship displays and vocal communication in the polygynous bat *Carollia perspicillata*. *Behaviour* 151, 781–798.

Knörnschild, M.; Glöckner, V., and Von Helversen, O. (2010b). The vocal repertoire of two sympatric species of nectar-feeding bats (*Glossophaga soricina* and *G. commissarisi*). *Acta Chiropterologica* 12, 205–215.

Knörnschild, M.; Nagy, M.; Metz, M.; Mayer, F., and Von Helversen, O. (2010a). Complex vocal imitation during ontogeny in a bat. *Biology Letters* 6, 156–159.

Kwiecinski, G.G. (2006). *Phyllostomus discolor*. *Mammalian Species*, 1–11.

La Val, R.K. (1970). Banding patterns and activity periods of some Costa Rican bats. *The Southwestern Naturalist* 15, 1–10.

Lattenkamp, E.Z., and Vernes, S.C. (2018). Vocal learning: a language-relevant trait in need of a broad cross-species approach. *Current Opinion in Behavioral Sciences* 21, 209–215.

Lattenkamp, E.Z.; Vernes, S.C., and Wiegrebe, L. (2018). Volitional control of social vocalisations and vocal usage learning in bats. *Journal of Experimental Biology* 221, jeb180729.

Luo, J.; Goerlitz, H.R.; Brumm, H., and Wiegrebe, L. (2015). Linking the sender to the receiver: vocal adjustments by bats to maintain signal detection in noise. *Scientific Reports* 5, 1–11.

Luo, J.; Lingner, A.; Firzlaff, U., and Wiegrebe, L. (2017). The Lombard effect emerges early in young bats: implications for the development of audio-vocal integration. *Journal of Experimental Biology* 220, 1032–1037.

Ma, J.; Kobayasi, K.; Zhang, S., and Metzner, W. (2006). Vocal communication in adult greater horseshoe bats, *Rhinolophus ferrumequinum*. *Journal of Comparative Physiology A* 192, 535–550.

Murray, K.L., and Fleming, T.H. (2008). Social structure and mating systemof the buffy flower bat, *Erophylla sezekorni* (Chiroptera, Phyllostomidae). *Journal of Mammalogy* 89, 1391–1400.

Nelson, J.E. (1964). Vocal communication in Australian flying foxes (Pteropodidae; Megachiroptera). *Zeitschrift für Tierpsychologie* 21, 857–870.

Pfalzer, G., and Kusch, J.J. (2003). Structure and variability of bat social calls: implications for specificity and individual recognition. *Journal of Zoology* 261, 21–33.

Prat, Y.; Taub, M., and Yovel, Y. (2016). Everyday bat vocalizations contain information about emitter, addressee, context, and behavior. *Scientific Reports* 6, 39419.

Rodenas-Cuadrado, P.; Chen, X. S.; Wiegrebe, L.; Firzlaff, U., and Vernes, S.C. (2015). A novel approach identifies the first transcriptome networks in bats: a new genetic model for vocal communication. *BMC Genomics* 16, 1–18.

Rodenas-Cuadrado, P.M.; Mengede, J.; Baas, L.; Devanna, P.; Schmid, T.A., Yartsev, M.; Firzlaff, U., and Vernes, S.C. (2018). Mapping the distribution of language related genes FoxP1, FoxP2, and CntnaP2 in the brains of vocal learning bat species. *Journal of Comparative Neurology* 526, 1235–1266.

Rother, G., and Schmidt, U. (1985). Die ontogenetische Entwicklung der Vokalisation bei P*hyllostomus discolor* (Chiroptera). *Zeitschrift für Säugetierkunde* 50, 17–26.

Russ, J.M.; Jones, G.; Mackie, I.J., and Racey, P.A. (2004). Interspecific responses to distress calls in bats (Chiroptera: Vespertilionidae): a function for convergence in call design? *Animal Behaviour* 67, 1005–1014.

Smotherman, M.; Knörnschild, M.; Smarsh, G., and Bohn, K. (2016). The origins and diversity of bat songs. *Journal of Comparative Physiology A* 202, 535–554.

Wetterer, A.; Rockman, M.V., and Simmons, N.B. (2000). Phylogeny of phyllostomid bats (Mammalia: Chiroptera): data from diverse morphological systems, sex chromosomes, and restriction sites. *Bulletin of the American Museum of Natural History* 248, 1–200.

Wilkinson, G.S. (1995). Information transfer in bats. *Symposium of the Zoological Society London* 67, 345–360.

Wilkinson, G.S. (2003). Social and vocal complexity in bats. In: F. B. M. deWaal, and P. L. Tyack (Eds), *Animal Social Complexity: Intelligence, Culture and Individualize Societies*, Harvard University Press, Cambridge, MA, pp. 322–341.

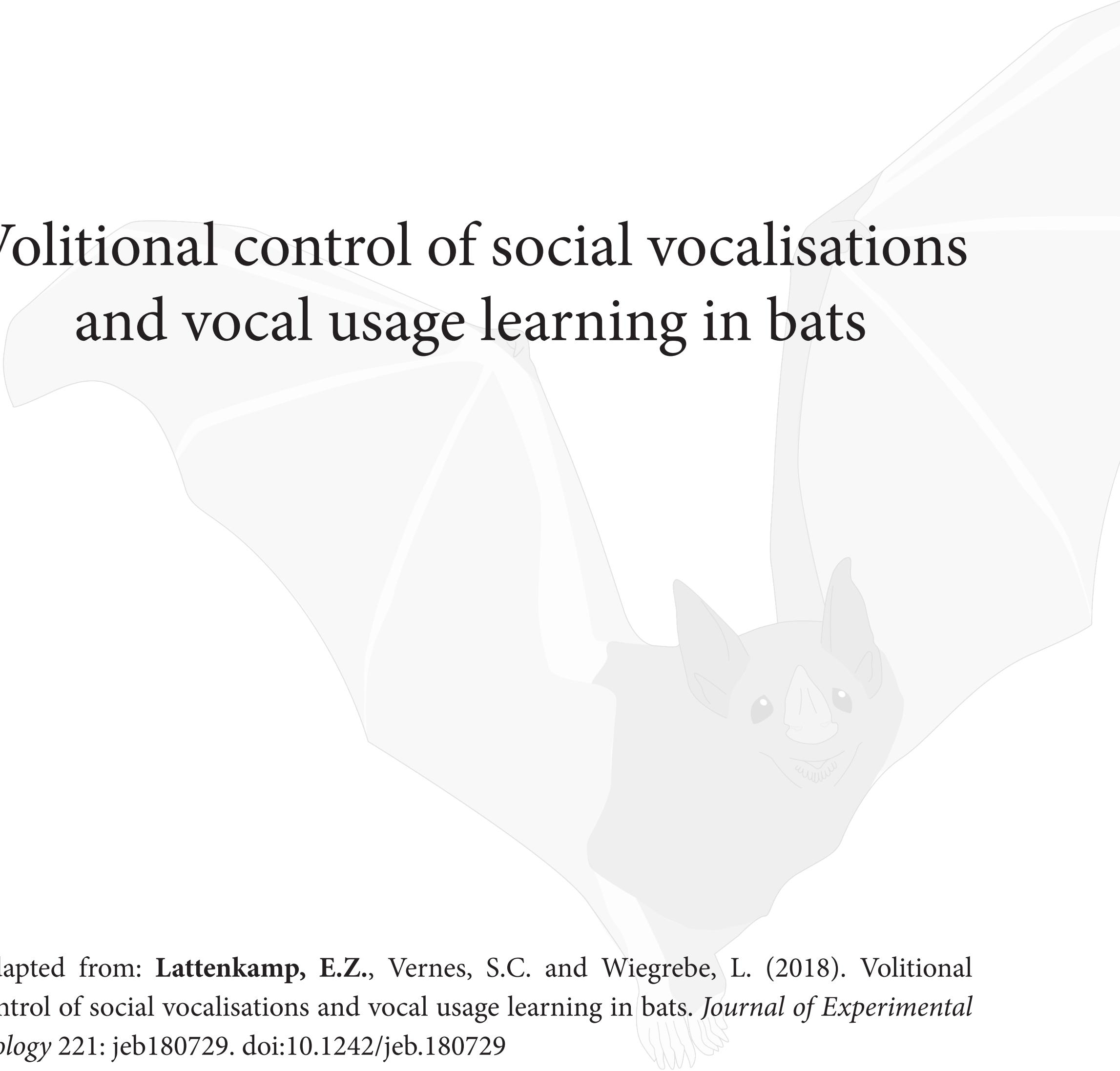
Wilkinson, G.S., and Boughman, J. W. (1998). Social calls coordinate foraging in greater spear-nosed bats. *Animal Behaviour* 55, 337–350.

Wright, G.S.; Chiu, C.; Xian, W.; Wilkinson, G.S., and Moss, C.F. (2013). Social calls of flying big brown bats (*Eptesicus fuscus*). *Frontiers in Physiology* 4, 214.

Yost, W.A. (1996). Pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America* 100, 3329–3335.

Chapter 5

Volitional control of social vocalisations and vocal usage learning in bats



Adapted from: **Lattenkamp, E.Z.**, Vernes, S.C. and Wiegrebe, L. (2018). Volitional control of social vocalisations and vocal usage learning in bats. *Journal of Experimental Biology* 221: jeb180729. doi:10.1242/jeb.180729

Abstract

Bats are gregarious, highly vocal animals that possess a broad repertoire of social vocalisations. For in-depth studies of their vocal behaviours, including vocal flexibility and vocal learning, it is necessary to gather repeatable evidence from controlled laboratory experiments on isolated individuals. However, such studies are rare for one simple reason: eliciting social calls in isolation and under operant control is challenging and has rarely been achieved. To overcome this limitation, we designed an automated setup that allows conditioning of social vocalisations in a new context and tracks spectro-temporal changes in the recorded calls over time. Using this setup, we were able to reliably evoke social calls from temporarily isolated pale spear-nosed bats (*Phyllostomus discolor*). When we adjusted the call criteria that could result in a food reward, bats responded by adjusting temporal and spectral call parameters. This was achieved without the help of an auditory target or social context to direct the bats. Our results demonstrate vocal flexibility and vocal usage learning in bats. Our setup provides a new paradigm that allows the controlled study of the production and learning of social vocalisations in isolated bats, overcoming limitations that have, until now, prevented in-depth studies of these behaviours.

Introduction

Bats are highly vocal animals that possess a rich repertoire of social vocalisations, including sophisticated syllable and song formations (Kanwal *et al.*, 1994; Behr and Von Helversen, 2004; Schwartz *et al.*, 2007; Wright *et al.*, 2013; Smotherman *et al.*, 2016). However, the main body of research on bat vocalisations focuses on their echolocation behaviour (Gillam and Fenton, 2016). This emphasis on echolocation calls is particularly evident when considering controlled laboratory studies of vocalisations, such as psychoacoustic measures to identify vocalisation patterns and perception thresholds in, for example, alternative forced choice experiments (Firzlaff *et al.*, 2006). Even experiments on the Lombard effect (Luo *et al.*, 2015), vocal plasticity (Luo and Wiegrebe, 2016) and vocal learning (Jones and Ransome, 1993) have, to a large extent, been conducted on echolocation calls. This bias has arisen because of the simple and stereotypical structure of echolocation calls and the ease with which they are measured, especially given that they are almost constantly emitted and independent of the social environment.

In contrast, bat social calls have received far less attention. Social calls are strongly associated with their respective behavioural context, contain a great deal of information and show indications of high levels of flexibility (Gillam and Fenton, 2016). Thus far, investigations of social vocalisations in bats have mainly focused on field studies (Boughman and Wilkinson, 1998; Behr and Von Helversen, 2004; Arnold and Wilkinson, 2011; Bohn *et al.*, 2013), recordings in groups (Kanwal *et al.*, 1994; Boughman, 1997, 1998, Bohn *et al.*, 2008, 2009; Knörnschild *et al.*, 2010) or ontogenetic changes in early developmental stages (Esser and Schmidt, 1989; Esser, 1994; Knörnschild *et al.*, 2006; Prat *et al.*, 2015). Social interactions and behavioural context have a strong impact on vocalisations as they influence the state of the emitter. Accounting for such effects is an important requirement in the study of vocal behaviours, but it is often difficult to accomplish. Disentangling changes in vocalisations triggered by social interactions and those initiated by intrinsic factors is challenging, but vital. The same is true for the precise evaluation of vocal changes due to developmental processes or learning. If changes in call parameters are to be attributed to a specific process, such as learning, detailed observations of the vocal behaviour of isolated animals are required (Siemers and Page, 2009). Yet, herein lies the problem: many animals do not spontaneously produce social calls in isolation and isolated individuals tend to fall silent (Carter *et al.*, 2008). As failed attempts of studying vocal behaviour in isolation usually remain unpublished, the literature is biased towards results that do not allow a separation of intrinsic variation of vocal performance and vocal changes triggered by a specific social context (Schusterman, 2008). This problem is well known to bat researchers and, as a consequence, studies on vocal flexibility in isolated adult bats are rare.

In order to use bats as a model species for vocal conditioning and vocal learning experiments, this limitation needs to be overcome. Frequent social vocalisations are required for vocal conditioning experiments as they represent the working point for positive reinforcement (Schusterman, 2008). In some animals, 'easily' elicited social calls, such as food calls induced by presenting food items (Watson *et al.*, 2015), can be brought under volitional control with comparative ease. Bats do not readily emit food calls in isolation, making them a challenging system in which to achieve vocal conditioning. However, once volitional control of vocal output is established, social vocalisations can be studied in detail. Such vocal conditioning experiments provide the basis for the in-depth study of vocal development and vocal learning.

The use of operant conditioning paradigms involving positive reinforcement of the desired (approximate) behaviour has produced positive results in the study of vocalisations in mammalian and avian research (Koda *et al.*, 2007; Manabe and Dooling, 1997; Manabe *et al.*, 1997, 1998; Siemers and Page, 2009). Operant control allows the in-depth investigation of call characteristics and learning behaviour, and the identification of structural constraints on vocalisations (Schusterman, 2008). Based on previous research on vocal control and flexibility in songbirds (Manabe *et al.*, 1997, 2008) and cetaceans (Richards *et al.*, 1984), we developed an automated real-time setup and training regime, within which isolated adult bats were trained to emit social calls. Using this training regime, we aimed to (a) reliably elicit social calls from isolated bats, (b) establish an automated setup, which allows conditioning of social vocalisations in bats, and (c) track spectro-temporal changes of call parameters in response to modifications in the reward schedule. Once trained to emit social calls in isolation, we challenged the isolated bats to adjust temporal and/or spectral parameters of their calls. This was achieved by gradually increasing the lower cut-off frequency, above which the sound level for trigger reward was measured (high-pass criterion). There are a number of possible ways for the bats to adjust their vocalisations in order to overcome the added level of difficulty imposed by the high-pass criterion. To increase the energy content in frequencies above the cut-off, the bats could switch to a different type of call, increase call duration or call level, shift the energy content of the call to higher frequencies or increase the fundamental frequency. The behavioural paradigm did not direct the bats towards any of these options and thus the choice of which strategy to use was left to the individual bats. Via digital analyses, we assessed the recorded calls before and after the activation of the high-pass criterion and, by doing so, were able to demonstrate the bats' changes in temporal and spectral call parameters. The first step of this approach establishes the stimulated but incidental production of social vocalisations. The second step demonstrates induction of vocal modifications through selective positive reinforcement without the use of an auditory target.

Materials and methods

Animals

Four male adult bats of the species *Phyllostomus discolor* Wagner 1843 were used for the experiments. The animals originated from a breeding colony in Department II of the Ludwig Maximilian University (LMU) in Munich (Germany). Training and experiments were conducted at the LMU from July to November 2016, five days per week. The bats were trained in a single session per day with an average length of 3.4 hours (max. 7.5 h). Outside the training sessions, the bats were kept in the colony room together with 25 conspecifics. During training sessions, the bats received food (banana pulp supplemented with infant milk powder, vitamin chalk and honey) as a reward for successful participation in the training. On the two rest days per week, the bats received fruit as well as meal worms (larvae of *Tenebrio molitor*). At all times, the animals had access to water *ad libitum*. This experiment was conducted under the principles of laboratory animal care and the regulations of the German Law on Animal Protection. The licence to keep and breed *P. discolor*, and all experimental protocols were approved by the German Regierung von Oberbayern (approval 55.2-1-54-2532-34-2015).

Experimental setup

The bats were trained in individual boxes (external measurements: 40×48×40 cm^3; w×h×d), which were lined with acoustic foam to reduce sound reflection (**Fig. 5.1c**). All boxes were equipped with one ultrasound microphone (custom made, based on SPU0410LR5H, Knowles Corporation, Itasca, IL, USA) and an infrared surveillance camera (Renkforce CMOS, Conrad Electronic, Hirschau, Germany), which was transmitting a live stream from inside the boxes (**Fig. 5.1b,c**). A self-designed feeding device allowed remote-controlled food reward delivery. The feeding device was a metal box (external measurements: 12×20×8 cm^3; w×h×d) housing one speaker (tweeter XT19NC30-04 Peerless, Tymphany HK Ltd, Sausalito, CA, USA), a flexible PVC tube (13×0.7 cm) for food reward delivery, and a drip tray (**Fig. 5.1d**). To check the bats' usage of the feeder, a photoelectric barrier (EE-SX461-P11 photomicrosensor, Omron Electronics, Langenfeld, Germany) was installed in front of the feeding tube. An orange light emitting diode (LED) next to the feeding tube indicated the feeder's readiness to be activated. The microphone was fixed at 26 cm height on the wall opposite the feeder (horizontal distance between microphone and feeding tube: 18 cm), and connected via a microphone pre-amplifier (OctaMic II, RME, Haimhausen, Germany; level setting: −10 dBV) to a multichannel audio interface (Fireface 800, RME). The loudspeakers were connected to the Fireface via a power amplifier (Harman Kardon AVR 445, Garching, Germany).

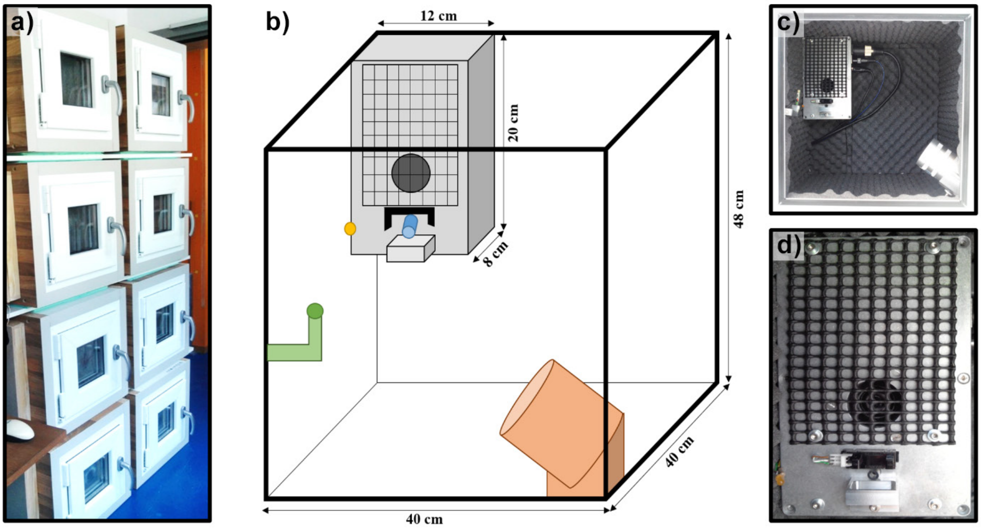


Figure 5.1. Automated behavioural training setup. a) Photograph of eight training boxes. b) Sketch of the inside of a training box equipped with one ultrasound microphone (green), an infrared surveillance camera (orange), an LED (yellow) and a self-designed feeding device (grey), which allowed remote-controlled food reward delivery. The feeding device was a metal box housing one loudspeaker (dark grey), a flexible PVC tube for food reward delivery (blue), and a drip tray. Furthermore, a photoelectric barrier (black) was installed in front of the feeding tube. c) Photograph of the setup inside the training box. d) Close-up picture of the feeding device.

Data acquisition and training

A self-written Matlab (R2007b, v7.5.0.342, MathWorks, Cambridge, MA, USA) script controlled the data acquisition. A ring buffer of 250 ms length recorded the microphone input from all eight boxes simultaneously (sampling rate: 192 kHz). A call level threshold (40 dB sound pressure level, SPL) for feeder activation was integrated over the total buffer size. When a call exceeded this level threshold, the recording was saved and the feeder activated (for 300 ms, which results in around 0.1 ml banana pulp discharge). After each activation, the feeder was disabled for a refractory period of 5 s. Echolocation calls alone did not contain enough energy to exceed the call level threshold.

Training was split into four stages. (1) Initially all bats had a period of two sessions to allow them to become familiar with the setup, the feeding system, and the isolated condition. During this first stage, the feeder was immediately triggered when the bats broke the light barrier located directly above the tube for food delivery (**Fig. 5.1b,d**). Thus, in this stage, no calls needed to be emitted in order to trigger the feeder. (2) In the second stage, the emission of social calls in isolation triggered a food reward. Whenever a vocalisation exceeded the pre-defined sound level of 40 dB SPL integrated over a fixed, 250 ms analysis window (call level threshold), a food reward was delivered and the recording was saved. The isolated bats were stimulated either with playbacks (i.e. random presentation of previously recorded non-aggressive social calls from conspecifics from the same colony) in 20±5 s intervals throughout the training session or constant real-time audio transmission from other boxes. If a bat reliably vocalised in isolation, it was occasionally paired with a second bat in order to demonstrate the expected procedure. Specifically, bat 1 was paired four times each with bat 3 and bat 4; bat 2 was paired once each with bat 3 and bat 4; bat 3 was paired four times with bat 1 and once with bat 2; bat 4 was paired four times with bat 1 and once with bat 2. The pairing lasted 5–60 minutes within a training session and the procedure for food reward delivery was the same as in isolated sessions, i.e. food delivery was triggered when the call level threshold was exceeded by either of the two bats. After 15–22 sessions in the second training stage, all bats reliably produced social calls in isolation. (3) In the third stage, the vocalisations were recorded from the isolated bats, which were not presented with auditory input for at least 16 and up to 25 sessions. All calls recorded on the last three sessions of training stage three, i.e. before the activation of the high-pass criterion, were pooled and comprised the baseline for unstimulated vocalisations in isolation ('pre-criterion' datasets). (4) The fourth stage began with the activation of a spectral high-pass criterion for the feeder trigger. All social calls exceeding the call threshold continued to be saved. However, the feeder was only triggered if the threshold was exceeded in a frequency range above a high-pass cut-off frequency (high-pass criterion). The cut-off frequency was initially set to 25 kHz and then gradually increased to a minimum of 26 kHz and a maximum of 40 kHz. This increase was individually different and dependent on the individual call types (**Fig. 5.2**; supplementary **Table S5.1**). In this training stage, the bats were recorded for up to 3 weeks (9–13 sessions) with the activated high-pass criterion. During the last three recorded sessions of this training stage, i.e. with an active high-pass criterion, all individuals had a constant high-pass cut-off frequency (with the exception of bat 4; see supplementary **Table S5.1**). Calls recorded during these last three sessions were also pooled ('post-criterion' datasets). During the first and second training stages, the bats were assigned random boxes for each session. In the third and fourth stages, the bats were recorded in the same box every training session.

Data analysis

A custom-written Matlab script was used for call analysis. All calls were extracted from 250 ms recordings and individually analysed. Analysed call parameters were call duration, call level, mean fundamental frequency, and spectral centroid (i.e. weighted mean of the frequencies contained in a call). Call levels are given in dB sound pressure level (SPL), measured at the microphone. These measurements do not allow a precise statement of the intensity at the sound source, as the bats were able to move freely in the box. The maximal possible distance between the bat's head and the microphone amounts to 44 cm (corresponding to ~30 dB call level difference). However, the bats usually stayed close to the feeder, at a distance of approximately 10–15 cm to the microphone, which corresponds to a record-level variation of no more than 4 dB. We further assumed constant movement patterns in all datasets and were thus able to compare relative differences in the pre- and post-criterion datasets. The fundamental frequency was tracked using the yin algorithm (de Cheveigné and Kawahara, 2002) in Matlab. Frequent 'harmonic jumps', i.e. falsely tracking the fundamental frequency on the wrong harmonic, were automatically detected and re-calculated to the right frequencies with the help of a custom-written Matlab script. For our analyses, we only used calls with a minimum duration of 5 ms, to conservatively exclude echolocation calls and thus only analyse social calls.

Statistical analysis

For the statistical evaluation of call parameter changes, we pooled data from 3 days before the activation of the high-pass criterion (precriterion; baseline recordings) and from 3 days with the high-pass criterion (post-criterion), with at least 8 sessions of adjustment in between. The pre- and post-criterion datasets contained a total of 6,209 analysed calls. Of these, 377 calls were excluded for being of less than 5 ms length as they are likely to represent echolocation calls rather than communication calls (for exact sample sizes see supplementary **Table S5.1**). For one individual (bat 4), the type of emitted call changed after the activation of the high-pass criterion (see results); thus, we split its post-criterion data into two categories ('short': calls longer than 5 ms, but shorter than 25 ms; 'long': calls of 25 ms or longer). Note that this change in call type led to a greater cut-off frequency for bat 4 than for the other three bats (supplementary **Table S5.1**). Furthermore, the adjustment of the high-pass criterion was more dynamic for bat 4; thus, its pooled post-criterion dataset was recorded with a different cut-off frequency on all three analysed post-criterion days (supplementary **Table S5.1**).

Ultimately, we analysed four pre-criterion and five post-criterion datasets (supplementary **Table S5.1**). All datasets were evaluated separately for each bat. Because of the applied call level recording threshold, not all emitted calls were recorded, which led to a non-normal distribution of the data. The one-sample Kolmogorov–Smirnov test for continuous data confirmed that all our datasets differed significantly from a normal distribution. To examine differences between the data before and after the activation of the high-pass criterion, we thus used the Wilcoxon rank-sum test (also called the Mann–Whitney U-test), which is a non-parametric test for differences in distributions of continuous data. For all datasets, we report the number of analysed calls, median, interquartile range (IQR), and P-values of the Wilcoxon rank-sum test (supplementary **Table S5.2**).

Results

Call types and number of emitted calls

Phyllostomus discolor appears to have a broad repertoire of social calls (see chapter 4). However, after an initial exploration phase, three of the four bats consistently emitted only a single social call in order to trigger the feeder. Bat 4 represented an exception by starting to emit a second call as the training progressed to stage four. The repeatedly emitted social calls were different between the four individuals, i.e. every bat emitted one typical call (or two calls in the case of bat 4 in training stage four) (**Fig. 5.2**). These five different social call types could be clearly distinguished from each other as they differed in duration, fundamental frequency and spectral centroid frequency (**Fig. 5.2**; supplementary **Table S5.2**).

All bats produced broadband, frequency-modulated social calls with fundamental frequencies between 10 and 25 kHz and several harmonics. The loudest component was usually the fundamental frequency, unlike echolocation calls where the third or fourth harmonics are the loudest and the fundamental frequency is strongly suppressed. Moreover, social calls are much longer than echolocation calls, the former being typically between 20 and 80 ms and the latter between 1 and 3 ms. Bats 1, 2, and 3 produced

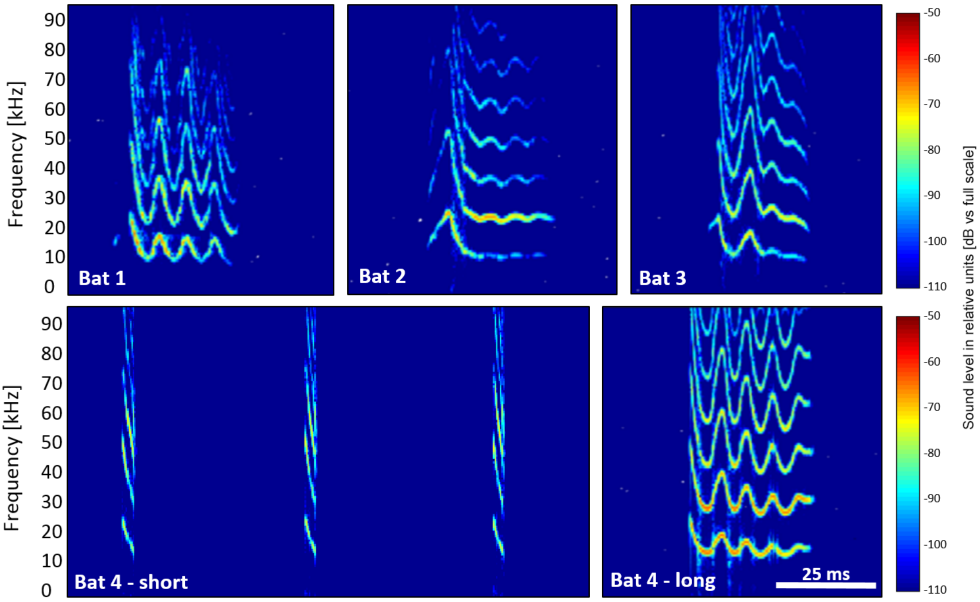


Figure 5.2. Spectrograms of 'typical calls' for bats 1–4. The calls from bats 1, 2 and 3 (top) were recorded from isolated individuals in a sound-attenuated training setup before the activation of the high-pass criterion. These calls were the same in terms of the structure of the fundamental frequency before and after activation of the high-pass criterion. Bat 4 produced a sequence of short calls pre-criterion (lower left panel) but started to also use long calls (lower right panel) in response to activation of the high-pass criterion.

rather long (median 40–54 ms), frequency-modulated calls, while bat 4 initially (i.e. pre-criterion) produced a sequence of shorter calls (median of around 6 ms) and only in training stage four (i.e. post-criterion) emitted longer social calls (median ~60 ms) (**Fig. 5.2**). Nevertheless, these short social calls were still much longer than echolocation calls and dominated by their fundamental frequency. Thus, these short calls of bat 4 were still classified as social calls.

After the activation of the high-pass criterion, three of the four bats continued to emit their typical call. Only bat 4 changed the emitted call type. Before the activation of the high-pass criterion, bat 4 produced sequences of 2–3 short calls, which barely exceeded the call-level threshold. In order to exceed the sound-level threshold after the activation of the high-pass criterion, bat 4 would have needed to emit ≥4 short calls in the 250 ms interval over which the sound level was integrated (see methods). This difficulty may be why bat 4 was the only bat to switch call types (**Fig. 5.2**). The use of an additional call type by bat 4 does not present a gradual change of call parameters and, as such, call parameters were not compared between long and short calls of bat 4 (supplementary **Table S5.2**). The statistical results presented for bat 4 come from a comparison within the short calls of this bat. The long calls of bat 4 circumvent the need to produce calls in a sequence, which led to fewer calls being recorded in the post-criterion phase for this individual (decrease in call emission rate was −0.39 calls min^{-1} post-criterion; supplementary **Table S5.1**). For bats 1, 2, and 3, the number of recorded calls increased dramatically after the activation of the high-pass criterion. The increase of total recorded calls in the post-criterion phase and the consequent increase of the call repetition rate (call rate increased between 0.29 calls min^{-1} (bat 1) and 5.84 calls min^{-1} (bat 2) post-criterion; supplementary **Table S5.1**) is an indicator for the bats' exploratory behaviour in order to meet the challenges of the high-pass criterion.

Changes in call duration and amplitude

The respiratory system is used to control call onset and offset, and thus determines call duration. The respiratory system also determines call amplitude via the control of subglottal pressure. All four individuals showed significant changes in call duration and amplitude after the activation of the high-pass criterion, suggesting volitional control over their respiratory system during vocalisations.

Extensive call prolongation was observed in bat 2, which increased its calls from a median length of 40.3 ms to 53.8 ms (IQR: 8.7 and 10.5 ms, respectively), while keeping all other call parameters approximately constant (**Fig. 5.3**; supplementary **Table S5.2**). Bats 1 and 3 also increased their call duration significantly, but to a lesser extent than bat 2 (**Fig. 5.3**; supplementary **Table S5.2**). After activation of the high-pass criterion, recorded call amplitudes showed statistically significant changes for all individuals (**Fig.**

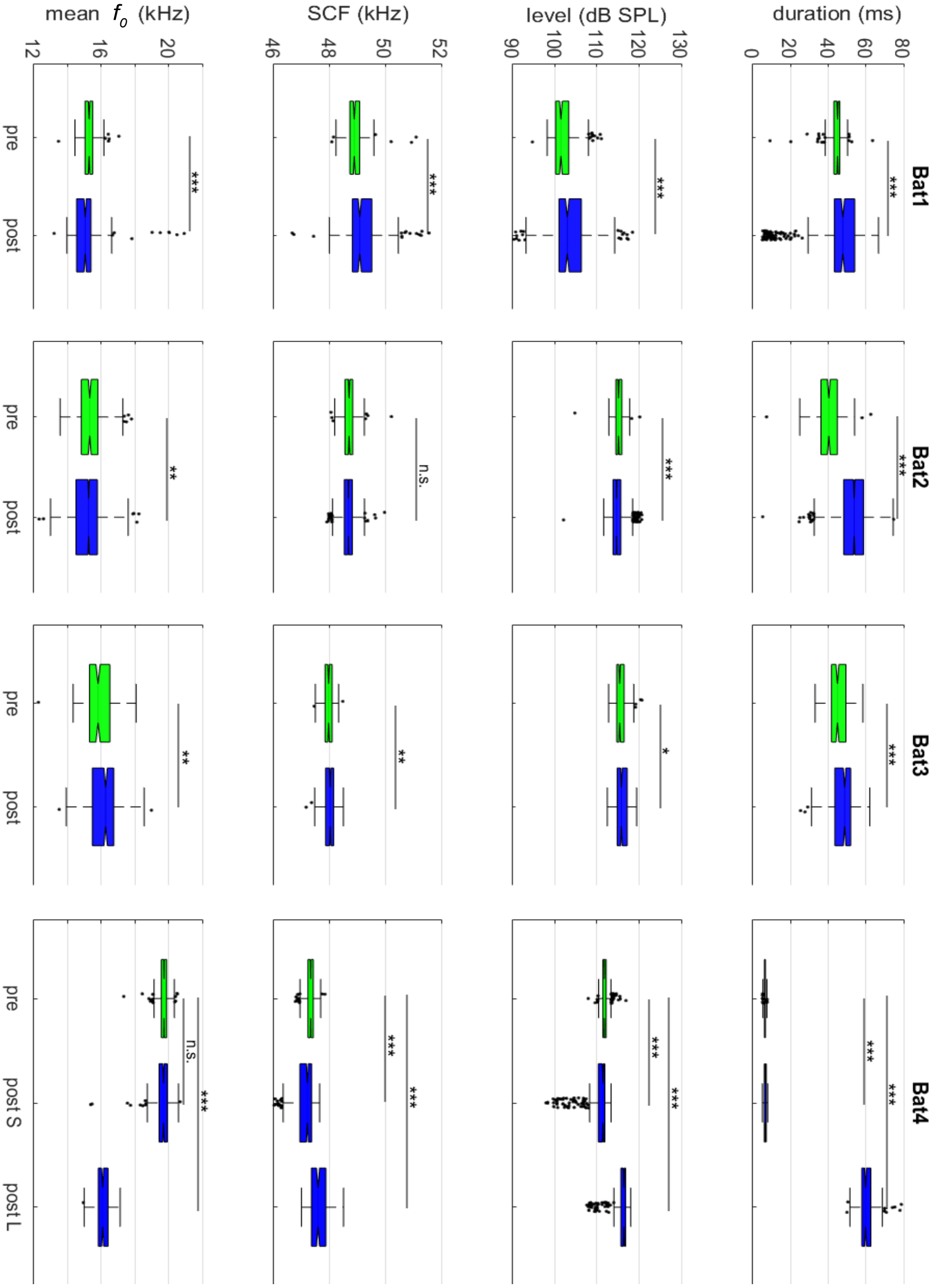


Figure 5.3: Change of temporal and spectral call parameters pre- and post-activation of the high-pass criterion. Changes of call duration, call sound level, spectral centroid frequency (SCF) and mean fundamental frequency (f_0) are shown for bats 1–4 (from left to right). Box plot summary statistics indicate the median line and interquartile ranges for each box. Each bar contains pooled calls from 3 days (pre- and post-criterion). For bat 4, the post-criterion results are split into short (S) and long (L) calls. As these long calls of bat 4 present an additional call type and not a gradual change from the short call type, no statistical comparison between these call types was conducted. Statistical results of the Wilcoxon rank-sum test for the other call types are indicated (n.s., $p \geq 0.05$; *$0.05 > p > 0.01$; **$0.01 \geq p > 0.001$; ***$p \leq 0.001$). For detailed p-values, see supplementary **Table S5.2**.

5.3; supplementary **Table S5.2**). However, these changes in recorded call amplitude are unlikely to be biologically meaningful as recorded call level changes within the same call type were of the order of ≤2 dB for all individuals (**Fig. 5.3**; supplementary **Table S5.2**).

Changes in spectral centroid and mean fundamental frequency

The spectral parameters (i.e. spectral centroid and mean fundamental frequency) were in general much less plastic than the temporal parameters, which is in line with general findings describing the higher level of difficulty associated with modifying the phonatory and filter systems in contrast to the respiratory system (Fitch, 2000; Janik and Slater, 1997). Although, activation of the high-pass criterion did indeed lead to statistically significant changes of the spectral centroid for bats 1, 3, and 4, these changes were of the order of ≤0.2 kHz (see supplementary **Table S5.2** for difference between medians).

Bats 1, 2, and 3 showed a statistically significant change in the mean fundamental frequency of their calls in response to the activation of the high-pass criterion (**Fig. 5.3**). For bats 1 and 2, a significant decrease of the mean fundamental frequency was recorded (change in medians: −0.23 kHz and −0.07 kHz, respectively; supplementary **Table S5.2**). An increase in mean fundamental frequency was detected for bat 3 (increase in median: 0.45 kHz; **Fig. 5.3**; supplementary **Table S5.2**). The mean fundamental frequency of the calls of bat 4 (short calls) decreased by 0.03 kHz in response to activation of the high-pass criterion, which was not a statistically significant change to the baseline calls (p=0.06; supplementary **Table S5.2**).

Discussion

We have established an automated real-time setup, which makes it possible to reliably elicit and record social calls from isolated bats. It allows the conditioning of social vocalisations and the tracking of spectro-temporal changes of call parameters over time. To our knowledge, this constitutes the first report of volitional social call emission, change of spectro-temporal call features and vocal usage learning in isolated bats in a controlled laboratory setup.

Setup and training regime

Our automated setup can be used to achieve conditioning of bat vocalisations over a relatively short time scale. After ≤22 training sessions with acoustic or conspecific stimulation, the four bats reliably produced social calls in isolation to trigger food rewards in the experimental setup. Similar to Manabe and colleagues, who also used a computer-based real-time system (Manabe and Dooling, 1997; Manabe *et al.*, 1995, 1997), we induced differentiation of vocalisations without a target sound. A maximum of just 11 sessions were needed to induce call parameter changes after initial call emission in isolation was established. This makes it a rapid training paradigm and the automated reinforcement makes this a useful method for studying more complex types of vocal learning.

The training regime used in the present study was not based on the stimulation of call emission by signal presentation but rather on the use of reinforcement of incidental vocalisations. Furthermore, no restrictions on the emitted type of call were imposed, which is known to positively support high rates of call emission and improve motivation (Adret, 1993). Studies using operant control paradigms often aim for a directed manipulation of call features (Pierce, 1985); however, this was not our goal. We were interested in the vocal flexibility demonstrated by the bats when trying to overcome the difficulty imposed by the high-pass criterion. This could be achieved by one or a combination of the following call parameter changes: using a different call type from before, increasing call duration or call level, or increasing spectral centroid or fundamental frequency.

Greater change in temporal than spectral call parameters

Temporal call parameters (such as call duration and call repetition rate) and call amplitude are in general considerably more flexible and easier to adjust than frequency characteristics as they are only dependent on the respiratory system (Fitch, 2000; Janik and Slater, 1997). Changes in spectral call parameters are reliant on muscular control over the vocal folds and the exact configuration of the vocal tract (i.e. regulation of the resonance of the produced sound), which require neuromuscular control over the phonatory and filter systems, respectively (Fitch, 2000). These different levels of difficulty for the volitional change in call characteristics are reflected in our findings: while call duration is adjusted strongly and with comparative ease, spectral call features are much more static (**Fig. 5.3**).

Call length was extended as much as 33% (difference between medians, bat 2; supplementary **Table S5.2**) within only a few experimental sessions. At the same time, the spectral call parameters showed little variation: the spectral centroid frequency increased by a maximum of 0.3% (corresponding to 0.13 kHz difference between medians, bat 1; supplementary **Table S5.2**), while the strongest registered change in mean fundamental frequency within one call type was an increase by 2.8% (corresponding to 0.45 kHz difference between medians, bat 3; **Fig. 5.3**; supplementary **Table S5.2**). These very small spectral changes are unlikely to be biologically relevant, or indeed even be perceivable for other bats (Kastein *et al.*, 2013; Krumbholz and Schmidt, 1999, 2001; Preisler and Schmidt, 1998). Although we detected statistically significant changes in recorded call levels over time, the increase did not exceed 2 dB SPL (within one call type; supplementary

Table S5.2). As the bats were freely moving in the boxes, changes in recorded sound level due to a change of distance between the bat and microphone could easily exceed the measured differences.

The presented data do not allow the unequivocal conclusion that *P. discolor* is capable of modifying frequency call parameters volitionally. The observed changes in frequency characteristics of the calls might be too small to be perceivable for the bats. In this case, it is unlikely that the observed spectral changes were due to volitional adjustments of the phonatory and filter systems. Further experiments are needed to investigate their control over the complete vocal system and consequently their vocal learning capacity.

Social call emission for food reward: demonstration of usage learning

The calls emitted by the experimental bats were social calls with a clear harmonic structure and approximately constant length. As only a single call type was repeatedly emitted by each individual, it is not possible to make statements about the general abundance of these calls in the species' call repertoire. We are certain, for several reasons, that we did not observe affective calls associated with a strongly emotional state of the animals (e.g. aggressive, stressed), but rather reliable communication calls. First, the vocalisations are dissociated from a social context, which could lead to an agitated state of the animals, as the recordings were made in isolation. Second, our experimental paradigm and the resulting recordings clearly show the bats' ability to repeatedly produce the same social call over several weeks, independent of other behavioural processes. Third, the call structure is independent of the behavioural context (e.g. a 'food call' emission), as the behavioural context is the same for all individuals, while the emitted calls are dissimilar in structure, spectral features and duration (**Fig. 5.2**).

The volitional emissions of social calls in a new context (e.g. labelling objects: Hage *et al.*, 2013; Richards *et al.*, 1984) is considered contextual usage learning (Janik and Slater, 2000). Bearing in mind the comparable ease of call emission control, it is not surprising that contextual learning (which includes both usage and comprehension learning) can be found much more frequently in the animal kingdom than vocal production learning (Hage *et al.*, 2013; Janik and Slater, 1997; Koda *et al.*, 2007; Seyfarth and Cheney, 2010). We here demonstrated contextual usage learning in *P. discolor* as the bats were able to employ social calls in order to perform a contextually independent task. It is evident that the bats understood the task (i.e. emission of social calls to trigger food rewards) and further changed spectral and/or temporal call parameters when faced with the additional challenge posed by the high-pass criterion. The observed switch between context-independent call types produced by bat 4 (single long calls versus multiple short calls) presents an especially strong case for the demonstration of usage learning in *P. discolor*.

In summary, we succeeded in reliably eliciting social calls from isolated *P. discolor* bats and, by demonstrating their volitional use of these communication calls out of context, captured contextual usage learning in real time. Our results demonstrate usage learning and adjustment of call characteristics without any social feedback. Through positive reinforcement, we were able to connect social calls, which do not have an innate relationship with food, with such food rewards. We thus demonstrate repeatable contextual learning. Moreover, with the help of our automated setup, we were able to track vocal adjustment in response to the implementation of a high-pass criterion without any directing auditory target. We recorded a vocal exploratory phase after changing the threshold for food reward delivery, which resulted in an adjustment of call parameters such as a change in call duration or call type. Exactly which strategy was used differed between individuals, indicating the bats' versatility when faced with the problem of the introduced high-pass criterion. This demonstration of vocal plasticity and usage learning further highlights the value of bats as a model system in the study of vocal learning in mammals (Vernes, 2017). Establishing this behavioural paradigm with bats as a model organism will in future allow the in-depth investigation of the degree of motor control over the vocal system, effects of audio-vocal feedback and consequently vocal learning.

Acknowledgements

The authors want to thank Mirjam Knörnschild, Michael Yartsev, and Uwe Firzlaff for helpful discussion and input leading to the writing of this article. We would also like to thank Lisa Hörtrich for temporarily helping with the setup maintenance in the course of a study project.

Funding

This work was funded by a Human Frontier Science Program (HFSP) Research Grant (RGP0058/2016) awarded to LW and SCV and a Max Planck Research Group Grant awarded to SCV.

Author contributions

Conceptualization: EZL, SCV, LW; Methodology: EZL, LW; Software: LW; Validation: EZL; Formal analysis: EZL; Investigation: EZL; Resources: LW; Data curation: EZL; Writing - original draft: EZL, SCV, LW; Writing - review & editing: EZL, SCV, LW; Visualization: EZL; Supervision: SCV, LW; Project administration: SCV, LW; Funding acquisition: SCV, LW.

The authors declare no competing or financial interests.

Data availability

The raw data and analysis scripts used to prepare the results presented in this article are available from the G-Node Infrastructure repository (Lattenkamp et al., 2018): https://web.gin.g-node.org/LutzW/vocal_usage_learning.

Supplementary material

Supplementary material available online at http://jeb.biologists.org/lookup/doi/10.1242/jeb.180729.supplemental

References

Adret, P. (1993). Vocal learning induced with operant techniques: an overview. *Netherlands Journal of Zoology* 43, 125–142.

Arnold, B.D. and Wilkinson, G.S. (2011). Individual specific contact calls of pallid bats (*Antrozous pallidus*) attract conspecifics at roosting sites. *Behavioral Ecology and Sociobiology* 65, 1581–1593.

Behr, O. and Von Helversen, O. (2004). Bat serenades - Complex courtship songs of the sac-winged bat (*Saccopteryx bilineata*). *Behavioral Ecology and Sociobiology* 56, 106–115.

Bohn, K.M.; Schmidt-French, B.; Ma, S.T. and Pollak, G.D. (2008). Syllable acoustics, temporal patterns, and call composition vary with behavioral context in Mexican free-tailed bats. *The Journal of the Acoustical Society of America* 124, 1838–1848.

Bohn, K.M.; Schmidt-French, B.; Schwartz, C.; Smotherman, M. and Pollak, G.D. (2009). Versatility and stereotypy of free-tailed bat songs. *PLoS ONE* 4, 1–11.

Bohn, K.M.; Smarsh, G.C. and Smotherman, M. (2013). Social context evokes rapid changes in bat song syntax. *Animal Behaviour* 85, 1485–1491.

Boughman, J.W. (1997). Greater spear-nosed bats give group-distinctive calls. *Behavioral Ecology and Sociobiology* 40, 61–70.

Boughman, J.W. and Wilkinson, G.S. (1998). Greater spear-nosed bats discriminate group mates by vocalizations. *Animal Behaviour* 55, 1717–1732.

Carter, G.G.; Skowronski, M.D.; Faure, P.A. and Fenton, B. (2008). Antiphonal calling allows individual discrimination in white-winged vampire bats. *Animal Behaviour* 76, 1343–1355.

de Cheveigné, A. and Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America* 111, 1917–1930.

Esser, K.H. (1994). Audio-vocal learning in a non-human mammal: the lesser spear-nosed bat *Phyllostomus discolor*. *NeuroReport* 5, 1718–1720.

Esser, K.H. and Schmidt, U. (1989). Mother-infant communication in the lesser spear-nosed bat *Phyllostomus discolor* (Chiroptera, Phyllostomidae) - evidence for acoustic learning. *Ethology* 82, 156–168.

Firzlaff, U.; Schörnich, S.; Hoffmann, S.; Schuller, G. and Wiegrebe, L. (2006). A neural correlate of stochastic echo imaging. *Journal of Neuroscience* 26, 785–791.

Fitch, W.T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Science* 4, 258–267.

Gillam, E. and Fenton, M.B. (2016). Roles of acoustic social communication in the lives of bats. In: M.B. Fenton, A.D. Grinnell, A.N. Popper, and R.R. Fay (Eds), *Bat Bioacoustics*. Springer Handbook of Auditory Research. Springer New York, NY, pp. 117–139.

Hage, S.R., Gavrilov, N. and Nieder, A. (2013). Cognitive control of distinct vocalizations in rhesus monkeys. *Journal of Cognitive Neuroscience* 25, 1692–1701.

Janik, V.M. and Slater, P.J.B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior* 26, 59–99.

Janik, V.M. and Slater, P.J.B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour* 60, 1–11.

Jones, G. and Ransome, R.D. (1993). Echolocation calls of bats are influenced by maternal effects and change over a lifetime. *Proceedings of the Royal Society B* 252, 125–128.

Kanwal, J.S.; Matsumura, S.; Ohlemiller, K. and Suga, N. (1994). Analysis of acoustic elements and syntax in communication sounds emitted by mustached bats. *The Journal of the Acoustical Society of America* 96, 1229–1254.

Kastein, H.B., Winter, R., Vinoth Kumar, A.K., Kandula, S. and Schmidt, S. (2013). Perception of individuality in bat vocal communication: discrimination between, or recognition of, interaction partners? *Animal Cognitiono* 16, 945–959.

Knörnschild, M.; Behr, O. and von Helversen, O. (2006). Babbling behavior in the sac-winged bat (*Saccopteryx bilineata*). *Naturwissenschaften* 93, 451–454.

Knörnschild, M.; Glöckner, V. and Helversen, O. Von (2010). The vocal repertoire of two sympatric species of nectar-feeding bats (*Glossophaga soricina* and *G. commissarisi*). *Acta Chiropterologica* 12, 205–215.

Koda, H., Oyakawa, C., Kato, A. and Masataka, N. (2007). Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour* 144, 681–692.

Krumbholz, K. and Schmidt, S. (1999). Perception of complex tones and its analogy to echo spectral analysis in the bat, *Megaderma lyra*. *Journal of the Acoustical Society of America* 105, 898–911.

Krumbholz, K. and Schmidt, S. (2001). Evidence for an analytic perception of multiharmonic sounds in the bat, *Megaderma lyra*, and its possible role for echo spectral analysis. *Journal of the Acoustical Society of America* 109, 1705–1716.

Lattenkamp, E.Z., Vernes, S.C. and Wiegrebe, L. (2018). Usage learning in bats [Dataset]. G-Node. doi: 10.12751/g-node.984192.

Luo, J., Goerlitz, H.R., Brumm, H. and Wiegrebe, L. (2015). Linking the sender to the receiver: vocal adjustments by bats to maintain signal detection in noise. *Scientific Reports* 5, 1–11.

Luo, J. and Wiegrebe, L. (2016). Biomechanical control of vocal plasticity in an echolocating bat. *Journal of Experimental Biology* 219, 878–886.

Manabe, K. and Dooling, R.J. (1997). Control of vocal production in budgerigars (*Melopsittacus undulatus*): selective reinforcement, call differentiation, and stimulus control. *Behavioural Processes* 41, 117–132.

Manabe, K., Kawashima, T. and Staddon, J.E.R. (1995). Differential vocalization in budgerigars: towards an experimental analysis of naming. *Journal of the Experimental Analysis of Behavior* 63, 111–126.

Manabe, K., Staddon, J.E.R. and Cleaveland, J.M. (1997). Control of vocal repertoire by reward in budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology* 111, 50–62.

Manabe, K., Sadr, E.I. and Dooling, R.J. (1998). Control of vocal intensity in budgerigars (*Melopsittacus undulatus*): differential reinforcement of vocal intensity and the Lombard effect. *Journal of the Acoustical Society of America* 103, 1190–1198.

Manabe, K., Dooling, R.J. and Brittan-Powell, E.F. (2008). Vocal learning in budgerigars (*Melopsittacus undulatus*): effects of an acoustic reference on vocal matching. *Journal of the Acoustical Society of America* 123, 1729–1736.

Pierce, J.D. (1985). A review of attempts to condition operantly alloprimate vocalizations. *Primates* 26, 202–213.

Prat, Y.; Taub, M. and Yovel, Y. (2015). Vocal learning in a social mammal: demonstrated by isolation and playback experiments in bats. *Science Advances* 1, e1500019.

Preisler, A. and Schmidt, S. (1998). Spontaneous classification of complex tones at high and ultrasonic frequencies in the bat, *Megaderma lyra*. *Journal of the Acoustical Society of America* 103, 2595–2607.

Richards, D.G., Wolz, J.P. and Herman, L.M. (1984). Vocal mimicry of computer-generated sounds and vocal labeling of objects by a Bottlenosed dolphin, *Tursiops truncatus*. *Journal of Comparative Psychology* 98, 10–28.

Schwartz, C.; Tressler, J.; Keller, H.; Vanzant, M.; Ezell, S. and Smotherman, M. (2007). The tiny difference between foraging and communication buzzes uttered by the Mexican free-tailed bat, *Tadarida brasiliensis*. *Journal of Comparative Physiology A* 193, 853–863.

Schusterman, R.J. (2008). Vocal learning in mammals with special emphasis on pinnipeds. In: D.K. Oller and U. Gribel (Eds), *The Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication*. MIT Press, Cambridge, MA, pp. 41–70.

Seyfarth, R.M. and Cheney, D.L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain and Language* 115, 92–100.

Siemers, B.M. and Page, R.A. (2009). Behavioral studies of bats in captivity: methodology, training, and experimental design. In: T.H. Kunz and S. Parsons (Eds), *Ecological and Behavioural Methods for the Study of Bats*. John Hopkins Press, Baltimore, Maryland, pp. 373–392.

Smotherman, M.; Knörnschild, M.; Smarsh, G. and Bohn, K. (2016). The origins and diversity of bat songs. *Journal of Comparative Physiology A* 202, 535–554.

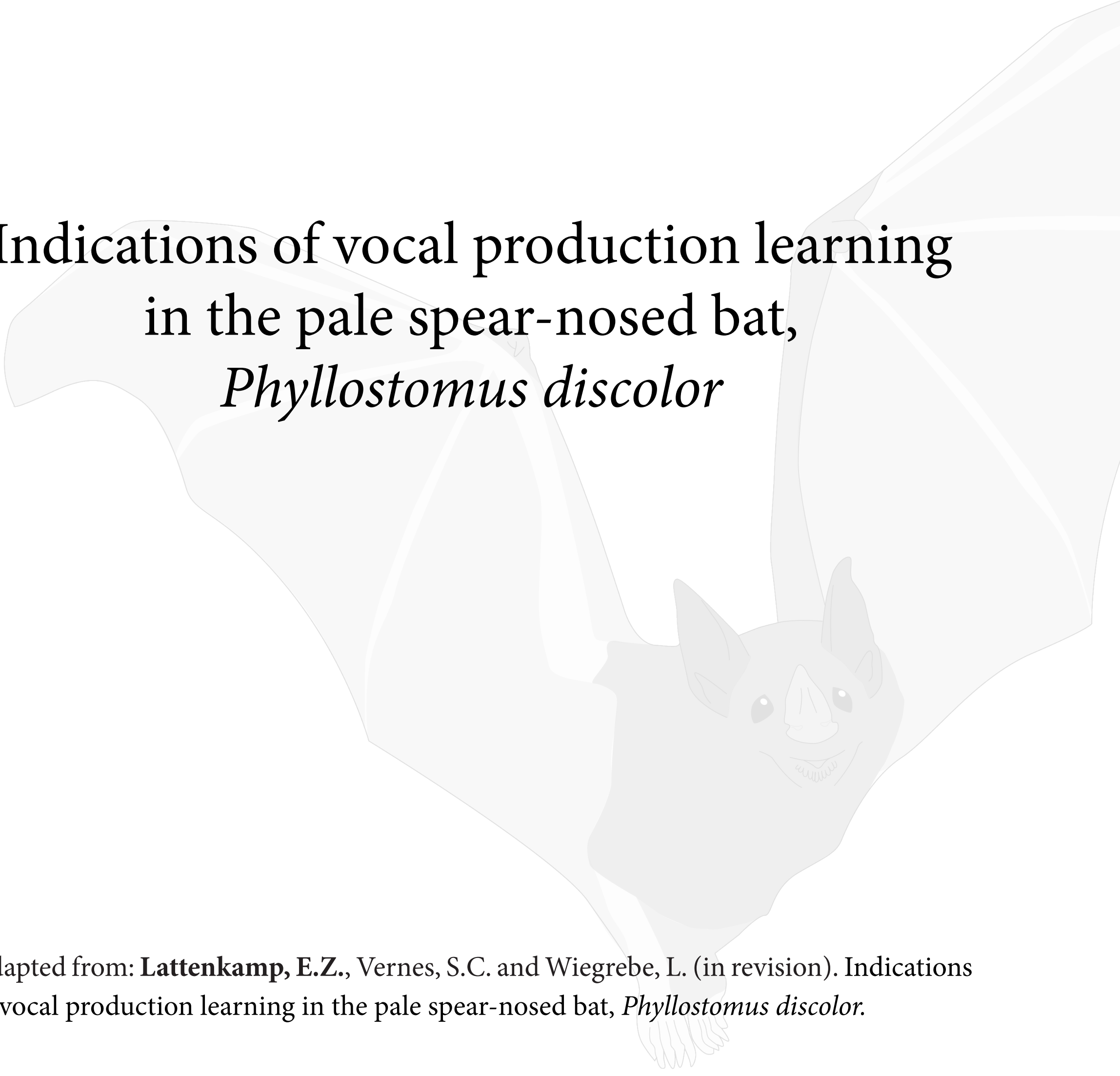
Vernes, S.C. (2017). What bats have to say about speech and language. *Psychonomic Bulletin & Review* 24, 111–117.

Watson, S.K.; Townsend, S.W.; Schel, A.M.; Wilke, C.; Wallace, E.K.; Cheng, L.; West, V. and Slocombe, K.E. (2015). Vocal learning in the functionally referential food grunts of chimpanzees. *Current Biology* 25, 495–499.

Wright, G.S.; Chiu, C.; Xian, W.; Wilkinson, G.S., and Moss, C.F. (2013). Social calls of flying big brown bats (*Eptesicus fuscus*). *Frontiers in Physiology* 4, 214.

Chapter 6

Indications of vocal production learning in the pale spear-nosed bat, *Phyllostomus discolor*



Adapted from: **Lattenkamp, E.Z.**, Vernes, S.C. and Wiegrebe, L. (in revision). Indications of vocal production learning in the pale spear-nosed bat, *Phyllostomus discolor.*

Abstract

Vocal production learning or the ability to modify or acquire vocalisations through the imitation of sounds is a rare trait in the animal kingdom. While humans are exceptional vocal learners, few other mammalian species share the vocal learning trait. Due to their singular ecology and lifestyle, bats are highly specialised for the precise emission and reception of acoustic signals. This specialisation makes them ideal candidates for the study of vocal learning and several bat species have indeed shown indications for vocal learning in the past. Here, we use a sophisticated automated setup and a contingency training paradigm to explore the vocal imitation capacity of isolated pale spear-nosed bats, *Phyllostomus discolor*. We first investigate their ability to lower the pitch of their vocalisations according to an acoustic template, and second, we test their ability to imitate the frequency contour of artificially generated stimuli. We show that these bats are capable of directional change in the fundamental frequency of their calls. This change can be achieved via acoustic templates, but is not necessarily dependent on them. We did not succeed in the demonstration of acoustic convergence towards presented playbacks of artificial templates or even conspecific calls, suggesting that the used setup and training paradigm have their limitations when such a complex imitation task is required. The effectiveness and limitations of the training paradigm are discussed and suggestions for improvements are brought forward.

Introduction

Bats are highly specialised in the use of their auditory system, which allows them not only to orientate in the dark (Neuweiler, 2000), but also to discriminate prey (Anderson and Racey, 1993; Koselj *et al.*, 2011) and surface structures (Grunwald *et al.*, 2004; Firzlaff *et al.*, 2006), and to identify individual conspecifics (Esser and Lud, 1997; Boughman and Wilkinson, 1998; Yovel *et al.*, 2009). This auditory system is used to process acoustic signals with high temporal and spectral resolution. Most research on bat vocal emissions focuses on their echolocation ability, but recently bats have been garnering increased attention due to their capacity for vocal production learning, i.e. the ability to learn new vocalisations or modify known vocalisations (Janik and Slater, 1997; Knörnschild, 2014; Vernes and Wilkinson, 2019). Vocal production learning is a trait shared among few vertebrate species and is crucial for the human capacity for speech learning and spoken language acquisition (Hauser *et al.*, 2002; Fisher and Vernes, 2015).

The mechanisms of vocal production are comparable across mammals and generally involve three physiological systems: the respiratory, phonatory, and filter system (Fitch, 2000). The task of the respiratory system is to regulate the lungs, which determine call duration and amplitude. Air exhalation from the lungs drives the vocal fold oscillations. The vocal folds are located in the larynx, which constitutes the phonatory system and the origin of the produced sound (i.e. the 'sound source'). The speed of the vocal fold oscillations determines the fundamental frequency (f_0) of the vocalisation, while the formants (i.e. the energy rich frequency ranges of the vocal output) are determined by the filter settings of the vocal tract. The filter system regulates the resonance of the vocal tract, meaning that by moving the articulators (e.g. soft pallet, lips, tongue) the attenuation patterns or unimpeded passing of specific frequencies are defined.

Bats change their call intensity and duration (regulated via the respiratory system) depending on the environmental conditions and social context (Smotherman, 2007). A change in call duration and amplitude requires volitional adjustment of air pressure and exhalation duration, and presents a form of contextual learning, the so-called 'usage learning', which is common among several taxa (Janik and Slater, 1997; Koda *et al.*, 2007; Seyfarth and Cheney, 2010; Hage *et al.*, 2013), and which has been shown previously for the pale spear-nosed bat (Lattenkamp *et al.*, 2018, see chapter 5). In contrast, vocal production learning is much more complex as it includes regulation of the phonatory and filter systems, and is accordingly less commonly found in the animal kingdom (Janik and Slater, 2000).

Based on previous research on vocal production learning in songbirds (Manabe *et al.*, 1997, 2008) and cetaceans (Richards *et al.*, 1984), we developed a multi-stage training plan, which was used to train six adult bats via an ultrasonic intercom to adjust their calls according to artificially modified auditory input. This study builds upon our previous research in pale spear-nosed bats, which showed that they have volitional control over their vocalisations, and possess vocal plasticity allowing them to adjust temporal and amplitude parameters of their vocalisations in a context specific task (Lattenkamp *et al.*, 2018, see chapter 5). Initially, we tested the bats' ability to change the f_0 of their vocalisations (experiment I, pitch shift). We assessed the recorded calls before and after the trainings period and thus were able to show a significant change in frequency characteristics of the calls. We then proceeded to a second experiment investigating the bats' ability to structurally imitate the frequency contour of an artificial target sound (experiment II, contour imitation). Both experiments were conducted on isolated bats in order to separate learning effects from the impacts of social interactions or different states of excitement of the animals.

Materials and methods

Experimental animals

For both experiments reported in the study, six adult male pale spear-nosed bats, *Phyllostomus discolor*, were trained. The animals originated from the captive breeding colony at the Biocentre of the Ludwig Maximilian University Munich. The training lasted for max. four hours and was conducted on five days per week. Successful participation during the training sessions resulted in remote-controlled delivery of food rewards (banana pulp supplemented with infant milk powder, vitamin chalk, and honey). The animals always had access to water *ad lib*. When the animals were not participating in the experiment, they were kept with 24 conspecifics in an animal keeping room with a 12h:12h dark:light cycle. This experiment was conducted under the principles of laboratory animal care and the regulations of the German Law on Animal Protection. The license to keep and breed *P. discolor* as well as all experimental protocols were approved by the German Regierung von Oberbayern (approval 55.2-1-54-2532-34-2015). Training and data acquisition for experiment I (pitch shift) lasted from December 2017 until August 2018, while experiment II (contour imitation) was conducted from February until June 2018.

Experimental setup

The bats were trained individually in six separate boxes, which are described in detail elsewhere (Lattenkamp *et al.*, 2018, see chapter 5). Each sound insulated box (outside measurements: 40×48×40 cm^3; w×h×d) was equipped with an ultrasound microphone (custom-made on basis of SPU0410LR5H, Knowles Corporation, Itasca, IL, USA) and an infrared surveillance camera (Renkforce CMOS, Conrad Electronic, Hirschau, Germany), which allowed live monitoring of the bats' behaviour. Food rewards were delivered via a self-designed, remote-controlled feeding device, which also housed a speaker for acoustic stimulus playback (tweeter XT19NC30-04 Peerless, Tymphany HK Ltd., Sausalito, California, USA). A photoelectric barrier (EE-SX461-P11 photomicrosensor, Omron Electronics, Langenfeld, Germany) was installed on the wall adjacent to the feeder, 8 cm below the feeding tube. Interruption of this photoelectric barrier started a run, which was indicated by an orange light emitting diode (LED) that was installed next to the barrier. In each box, the microphone was mounted on the wall opposite to the feeder (horizontal distance: 18 cm). Microphones were connected via a microphone preamplifier (OctaMic II, RME, Haimhausen, Germany; level setting: -10dBV) to a multichannel audio interface (Fireface 800, RME, Haimhausen, Germany) and loudspeakers were connected to the audio interface via a power amplifier (Harman Kardon AVR 445, Garching, Germany).

General data acquisition and training

Data acquisition was controlled via a self-written Matlab script (R2007b, v7.5.0.342, MathWorks, Cambridge, MA, USA). The microphone inputs from all experimental boxes were recorded simultaneously (sampling rate: 192 kHz) onto a ring buffer of 250 ms length. The contents of the ring buffer were saved if a call level threshold of 40 dB sound pressure level (SPL), integrated over the total buffer length, was exceeded. Echolocation calls did not contain enough energy to exceed this call level threshold, requiring the bats to emit social calls in order to trigger the saving of a file. In both experiments, the six bats were trained once per day, five days per week, in a single session with a maximal length of four hours. The training session was terminated prematurely if the bats completed >100 trials successfully. The bats were trained to actively start a run by physically interrupting the photoelectric barrier. After the light barrier was interrupted, the LED lit up for up to five or three seconds (experiment I and II, respectively) indicating the reactive state of the feeder. If, within this light-indicated interval, a vocalisation exceeded the call level threshold, the ring buffer was saved. The reward procedure followed different criteria in experiment I and II and will be described in detail below.

Experiment I: pitch shift

Stimulus generation

During experiment I, the bats were each presented with playbacks of their own 'typical' calls that had been recorded in a previous experiment (see chapter 5). For each bat, six of these 'typical' calls were randomly chosen as templates. The training was split into two phases: phase 1) template calls were presented to the bats in their normal frequency range, as they had been recorded; phase 2) template calls were presented four semitones downward shifted from the frequency range in which they had been recorded.

Specifically, a frequency shift of four semi-tones equals a 24% change of the f_0 of the calls (e.g. a call with a natural f_0 of 15 kHz would be downward pitch-shifted by 3.6 kHz to a new f_0 of 11.4 kHz).

Experimental procedure

When the light barrier was interrupted during the first phase of experiment I, the bats were immediately presented with one out of six randomly chosen playbacks of their own unmodified calls. In this first phase, the food reward was triggered (~ 0.2 ml banana pulp discharge) and the LED switched off as soon as any emitted call exceeded the call level threshold (40 dB) within the 5-sec interval. During the second phase, the bats were instead presented with the downward shifted versions of their own calls when the light barrier was interrupted. In this second phase, a spectral low-pass criterion for the feeder trigger was activated, allowing only those calls to trigger the feeder, which exceeded the call level threshold in a frequency range below the low-pass cut-off frequency (**Fig. 6.1**). Even though not all calls emitted within the reactive 5-sec interval triggered the feeder, all calls were still saved as long as they exceeded the call level threshold.

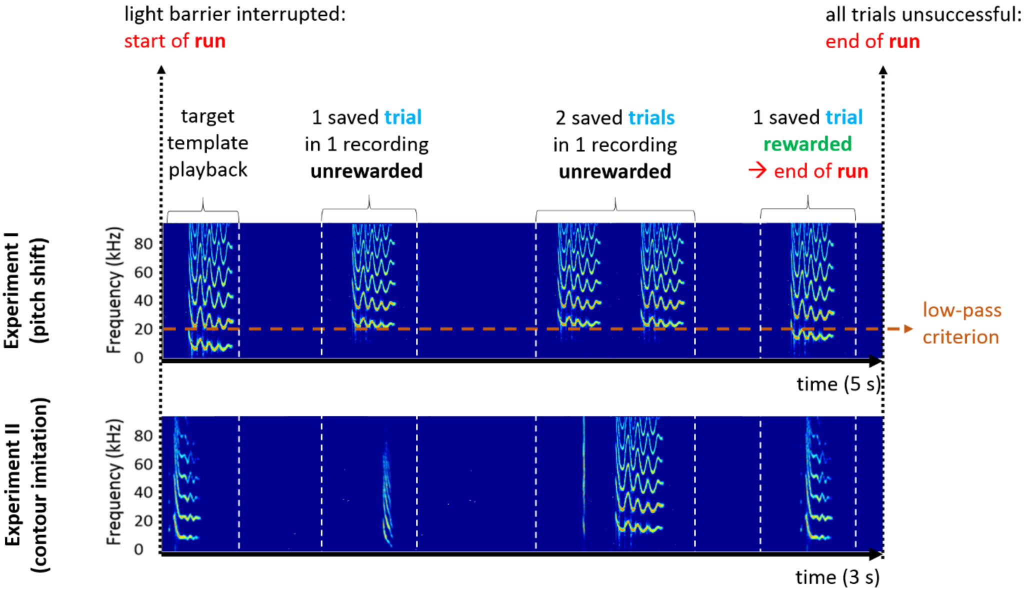


Figure 6.1: Example trials in phase 2 of experiment I (pitch shift) and experiment II (contour imitation). The top panel shows one example run in experiment I. The bats started the run by interrupting the light barrier. They were then presented with a four semi-tones downward pitch-shifted version of their own call (template playback). If calls emitted within 5 seconds exceeded the call level threshold, they were saved in 250 ms long recordings. The f_0 of the saved calls was extracted in real-time. If the extracted f_0 was below the low-pass criterion, the feeder was activated and the run ended immediately. The bottom panel illustrates an example run in experiment II. Also here the bats started the run by interrupting the light barrier. They were then presented with either a control template (as displayed here) or an artificially generated template (cf. **Fig. 6.3**). If calls emitted within three seconds exceeded the call level threshold, they were saved in 250 ms long recordings. The f_0 contour of the saved calls was extracted in real-time and compared to the contour of the playback. If the emitted call contour was similar enough to the playback, the feeder was activated and the run ended immediately.

The low-pass cut-off frequency was initially set to 27, 28, or 30 kHz and was then gradually lowered to frequencies between 13.1 and 15.6 kHz. The recordings were conducted for a minimum of 64 sessions with the low-pass criterion activated. In order to test whether the presentation of the frequency-shifted playbacks or the active low-pass criterion caused the observed change in mean f_0, a subset of four bats was used for ten follow-up training sessions: After the data acquisition for the main experiment ended, the low-pass criterion was deactivated for five days, however, the presented templates were still frequency-shifted ('criterion deactivated' data). After these five days, the presented playback was again switched to the unmodified calls. In this setting, the four bats were again recorded for five days ('unshifted template' data).

Data analysis

A custom-written Matlab script (R2015a, v8.5.0.197613, MathWorks, Cambridge, MA, USA) was used for call analysis. From the saved 250 ms recordings, individual calls were automatically extracted and call duration, level, and mean f_0 was determined. To determine the f_0 of a call the YIN algorithm (de Cheveigné and Kawahara, 2002) was employed in Matlab and falsely detected f_0 jumps were corrected for. From the trace of f_0 over time, the mean f_0 was calculated. Only calls with a minimum duration of 5 ms were considered in the analyses, in order to conservatively exclude echolocation calls from the analyses. A correlation analysis was conducted on all calls of each individual in order to test for correlation between mean f_0 and call duration.

Statistical analysis

In order to determine the individual change of f_0 over time, all calls emitted five days before the start of phase 2 were pooled for each bat ('baseline' data). We compared these baseline data 'after 30 days of training' and 'after 60 days of training'. For these two time points, we again pooled all calls from each individual over five days (i.e. data from days 28-32 and 58-62) (supplementary **Table S6.1**). All data sets were evaluated separately for each bat. With the one-sample Kolmogorov-Smirnov test for continuous data, we confirmed that all our data sets differed significantly from normal distributions. Thus, we used the Wilcoxon rank-sum test (i.e. a non-parametric test for differences in distributions of continuous data) to compare the data sets before and 30 and 60 days after activation of the low-pass criterion and presentation of frequency-shifted templates. For all data sets we report the number of analysed calls, median, interquartile ranges, and *p*-values of the Wilcoxon rank-sum test (supplementary **Table S6.1**).

Experiment II: contour imitation

Stimulus generation

Experiment II was concerned with the imitation of the f_0 contour of an acoustic target. The artificial templates were designed to be similar in length and frequency range to naturally emitted social calls, but should be different in their f_0 contour from any observed species-specific social call. Two artificial templates with a length of 80 ms were generated. While template 1 consisted of a linear upward sweep of the f_0 from 10 to 20 kHz, template 2 also increased in f_0 from 10 to 20 kHz, but in a stepwise manner (i.e. a 15 ms ramp from 10 to 15 kHz was followed by a 50 ms long element of constant 15 kHz, and then another 15 ms ramp to 20 kHz) (**Fig. 6.2**). Furthermore, we used two social calls recorded from conspecifics as control templates (**Fig. 6.2**). These controls are part of the species-typical repertoire and should thus be easily reproducible for all individuals. If the bats understand the imitation task required of them in the paradigm, the reproduction of these control templates should be easily achievable.

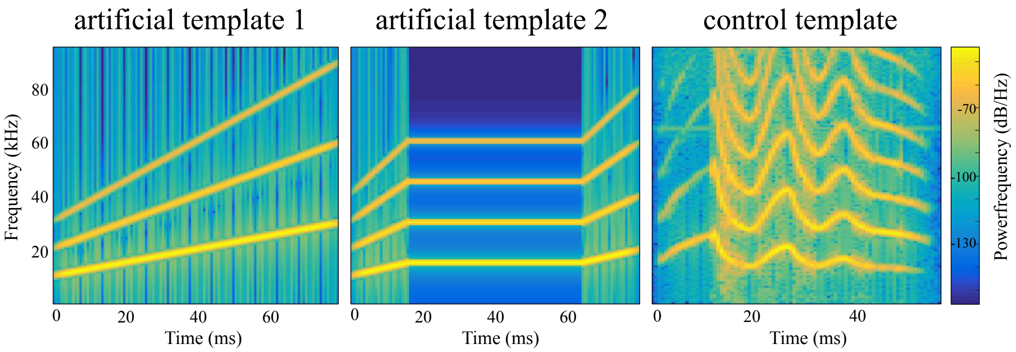


Figure 6.2: Spectrograms of the stimuli used in experiment II (contour imitation). Artificial template I (left panel) consisted of an 80 ms long linear upward sweep from 10 to 20 kHz and was presented to bats 2 and 5. Artificial template 2 (middle panel) also consisted of an 80 ms long playback that increased in a stepwise manner from 10 to 20 kHz (i.e. a 15 ms ramp from 10 to 15 kHz was followed by a 50 ms long element of constant 15 kHz, and then another 15 ms ramp to 20 kHz). Artificial template 2 was presented to bats 3 and 6. The right panel shows one of two control templates, which were recorded from conspecific bats and were presented to bats 1 and 4.

Experimental procedure

For experiment II, the six bats were sorted into three groups of two. Two of these pairs were each presented with one of two artificially generated templates, while the third group was presented with the control templates (i.e. each control bat was played back one of the two conspecific control templates) (**Fig. 6.2**). Artificial template I was used as playback for bats 2 and 5, while bats 3 and 6 were presented with playbacks of artificial template 2. The two control templates were presented to bats 1 and 4. The training for experiment II was also split into two phases: Phase 1) the bats were presented with their individual template (artificial or control) and rewarded whenever they emitted a social vocalisation (exceeding the amplitude threshold (cf. experiment I)) in the 3 s response window. Phase 2) the bats were only rewarded if the emitted vocalisation resembled the presented template in f_0 contour. In both phases, the bats needed to interrupt the light barrier in order to start a run and to be presented with their specific template. Afterwards they had a 3 s time window, during which they had the possibility to trigger a food reward

(~ 0.2 ml banana pulp discharge) (**Fig. 6.1**). Trials were saved whenever a call exceeded the amplitude threshold (cf. experiment I). If a trial was saved, the YIN algorithm (de Cheveigné and Kawahara, 2002) was employed in Matlab to extract the f_0 contour of the emitted call in real time. The f_0 contour was then compared to the contour of the template call and the root-mean-square (RMS) distance between the frequency contours of template and recorded call was calculated. For the calculation of the RMS distance, the length of the artificial templates was adjusted to the length of the recorded call. This was not the case for the two control templates. Thus, in case the extracted f_0 contour of a call was much longer than the contour of the control template, it resulted in a high RMS distance value. A high RMS distance in general indicated a large discrepancy between the template f_0 contour and the f_0 contour of the recorded call. In phase 1 of this experiment, the bats were rewarded independent of the RMS distance threshold (data recorded in this phase was the 'baseline data'). In phase 2, we used a continuously decreasing threshold as reward criterion. This way, the bats were motivated to adjust the contour of their emitted calls to the previously presented template.

For each bat an individual RMS distance starting value was calculated at the beginning of phase 2 (**Table 6.1**). This starting value was yielded by calculating the RMS distance of the baseline data calls to the specific template presented to each bat and then using half of this distance as starting value. This way, half of the calls emitted by the bat should by chance still trigger a food reward. The starting values of the RMS distance threshold were different between the bats as the typical calls of the bats were different from the individual templates. Following the contingency training paradigm, we then decreased this threshold over time depending on the performance of the different individuals (**Table 6.1**). In particular, artificial template 1 had the largest initial RMS distance to the calls of bats 2 and 5.

Table 6.1: The applied RMS distance threshold for reward delivery. The mean RMS distance between the calls emitted by each bat before training onset and the presented template was calculated and used as starting point for training. With the mean RMS distance as threshold, 50 % of the emitted calls of each bat should still trigger a food reward as they will fall below the threshold by chance. The threshold was then lowered in the course of 57 training days depending on the individual training success of the six bats. The difference (of RMS distances) indicates the absolute lowering of the threshold over the course of the training.

	Threshold (RMS distance) at training onset	Threshold (RMS distance) at the end of training	difference	template
Bat 1	5310	4600	-710	control
Bat 2	9410	5500	-3910	art. 1
Bat 3	5110	2300	-2810	art. 2
Bat 4	4680	2900	-1780	control
Bat 5	9260	7400	-1860	art. 1
Bat 6	4080	3600	-480	art. 2

art. 1 = artificial template 1; art. 2 = artificial template 2 (see **Fig. 6.2**)

Data analysis

After 57 days of data acquisition, the RMS distance between the individual playbacks and all calls emitted by the bats were calculated for each training day. The abundance of a certain RMS distance was plotted as heat map to visualise the progress of the bats (i.e. the convergence of the frequency contour of the emitted calls towards the presented template call, **Fig. 6.6**). Furthermore, the recordings were visually assessed and example spectrograms were selected from every seventh recording session to further illustrate the change in spectral shape of the emitted calls (supplementary **Fig. S6.3**).

Results

Experiment I (pitch shift)

In the course of the experiment, a total of 28,452 calls was recorded within the different data sets, which each encompassed recordings from five days (**Fig. 6.3**, supplementary **Tables S6.1**-**3**). Two bats (bats 1 and 3) were not tested beyond 62 days as they finished data collection significantly later than the other four bats. The number of emitted calls increased for all bats within 30 days of training and even further after 60 days of training. Subsequent to the deactivation of the low-pass filter, the number of emitted calls decreased again for all four tested bats (bats 1 and 3 were not recorded after training day 62). This demonstrates the increased level of difficulty the different training phases presented to the bats, as more calls were required to receive sufficient food rewards.

Changes in call characteristics in the course of the training

All six bats significantly lowered the mean f_0 of their calls after 30 days of training with pitch-shifted template calls ($p < 0.01$, supplemenatry **Table S6.1**). Five of them further decreased their f_0 within the next 30 training days, while one bat started to increase its call f_0 again (**Fig. 6.4b**). However, after 60 days of training the mean f_0 of the calls of all six bats were significantly lower ($p < 0.01$, supplementary **Table S6.1**) than in the 'baseline' data (**Fig. 6.4b**). The maximal decrease in f_0 after 60 days of training was achieved by bat 2, which lowered its mean f_0 by 990 Hz (supplementary **Table S6.1**).

All bats also significantly prolonged their calls in response to the training regime. The strongest increase in median call duration was noted for bat 5, which increased median call duration by as much as 12.6 ms within 60 days of training (this corresponds to a call prolongation of 33%; supplementary **Table S6.2**). Over the same time span, other bats increased their calls by only 2 or 3.5 ms (bats 1 and 4, respectively; supplementary **Table S6.2**). Mean f_0 and call duration were negatively correlated for most bats (bats 2-6), but slightly positively correlated for bat 1 (supplementary **Fig. S6.2**).

Moreover, a significant increase in median call levels was detected for four of the six bats within 60 days of training (supplementary **Table S6.3**). Bat 2 increased its call levels only insignificantly by 1 dB, while bat 5 decreased its call levels by the same amount (supplementary **Table S6.3**). The magnitude of all level changes was generally small (1–5 dB; corresponding to < 5% change) and was in the range of possible recording variability. As the bats could move freely in the box, the bats' physical distance to the microphones and their head positions were variable and could account for the observed call level change. Although the change in call levels was negligible, the fact that median call levels increased for most bats indicates that they generally respond with call level increase to the pressure the training paradigm exerts over them. The same behaviour has been observed in a previous experiment (see chapter 5).

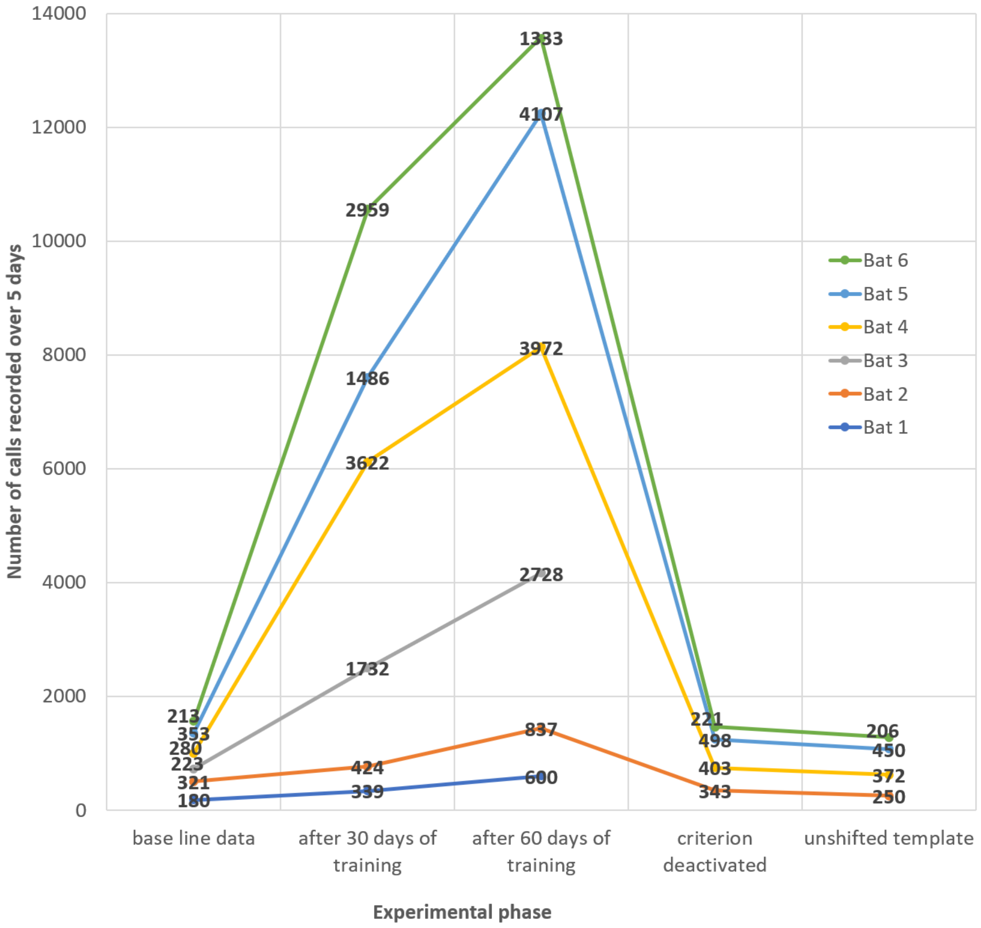


Figure 6.3: Number of calls recorded within 5 days for each experimental phase. In the course of the different experimental phases, a total of 28,452 calls was recorded and analysed. Calls recorded between the different 5-day bins were stored, but not used in this comparison. The number of calls emitted increased with training duration and thus level of difficulty (i.e. the low-pass filter criterion was continuously adjusted downwards depending on the individual participation). Four bats (bats 2, 4, 5, and 6) were recorded as well after the low-pass criterion was deactivated and subsequently when the template was presented again in its normal frequency range (i.e. unshifted). The number of emitted calls drastically decreased after the low-pass filter was deactivated and even further when the template was no longer pitch-shifted. For bats 1 and 3 data was only acquired until training day 62.

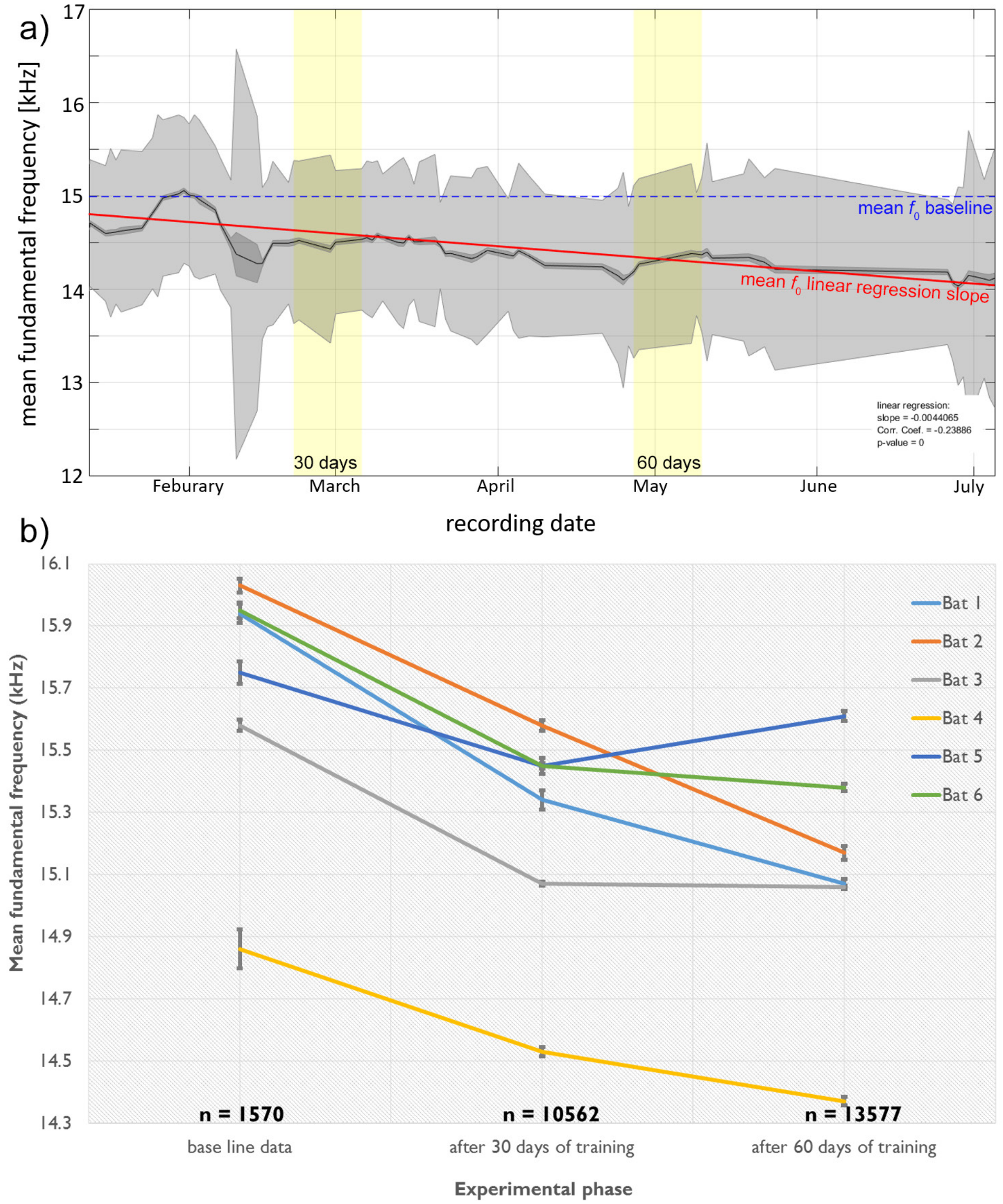


Figure 6.4: Change of mean fundamental frequency over the course of the training with a pitch-shifted template and an activated low-pass criterion. a) Illustrated is the exemplary individual trajectory of the decrease in f_0 of bat 4. The mean f_0 of the baseline dataset (~15 kHz) is indicated with a dotted blue line, while the actual mean f_0 is shown with a solid black line. Dark grey shading shows the standard error of the mean (SEM), while light grey shading shows the standard derivation. The linear regression slope of the mean f_0 is indicated with a solid red line. Yellow shading indicates the time points '30 days of training' and '60 days of training', which encompasses the data analysed in the present study. b) This panel shows the decrease in mean f_0 of all six bats over the course of the training. The different bats are represented by different colours and grey error bars indicate the SEM. The overall number of calls recorded and analysed in each experimental phase (n) is given at the bottom of the graph. All six bats significantly reduced their mean f_0 within 30 days of training. Although one bat increased its f_0 between training days 30 and 60, the decrease in mean f_0 between the baseline data and the calls after 60 days of training is significant (p-value < 0.002, supplementary **Table S6.1**).

Effect of template presentation and low-pass criterion activation

The successful initiation of pitch shift in all animals demonstrates their capacity to directionally modify the spectral parameters of their vocalisations. However, in order to test if this directed change was induced by the presentation of the pitch-shifted playbacks or if the applied low-pass filter criterion alone would be sufficient to elicit this change, an additional test was conducted with four bats subsequently to the data acquisition period (**Fig. 6.5**, supplementary **Fig. S6.1**). For this additional test, we first deactivated the low-pass filter criterion (the template was still pitch-shifted) for a duration of five days and only then presented the bats with an unshifted playback of their own calls. If the bats pitch-shifted their vocalisations according to the template and independently from the reward schedule, we expected to see no change or even a further decrease in the f_0 of their calls once the low-pass criterion was deactivated, but an increase in f_0 when unshifted templates were presented. This expectation was fulfilled for bat 4, which continued to significantly decrease the f_0 of its calls even after the low-pass filter criterion was deactivated ($p < 0.01$, bat 4) and increased its f_0 drastically when unshifted templates were presented ($p < 0.01$, bat 4, **Fig. 6.5**, supplementary **Table S6.1**). Another bat did not show a significant change in pitch after the low-pass criterion was deactivated ($p = 0.78$, bat 5), but significantly increased the f_0 of its calls once unshifted templates were presented again ($p < 0.01$, bat 5, **Fig. 6.5**, supplementary **Table S6.1**). Two out of four bats

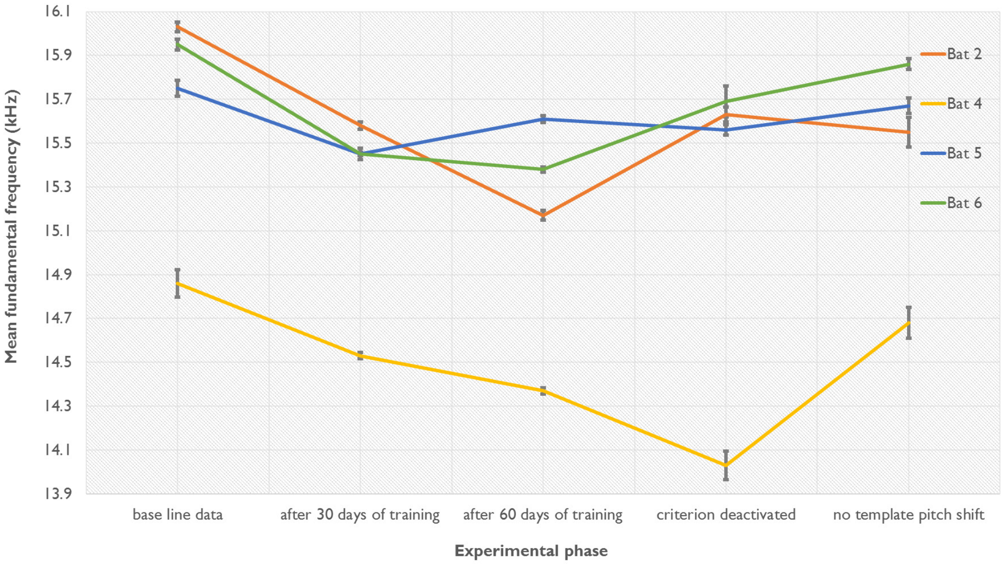


Figure 6.5: Change of mean fundamental frequency within the first 60 days of training and with a deactivated low-pass criterion and an unshifted template. After the low-pass criterion was deactivated, bat 4 continued to decrease the fundamental frequency of its calls, while bat 5 did not show a significant change in the f_0 of its calls (p-value < 0.01, supplementary **Table S6.1**). Bats 2 and 6 significantly increased the f_0 of their calls (p-value < 0.01, supplementary **Table S6.1**), although the presented playback of their own calls were still pitch-shifted. Without an activated low-pass criterion and a playback call in the normal frequency-range (unshifted) bats 2 and 6 do not significantly change their f_0 (p-value > 0.2, supplementary **Table S6.1**), while bats 4 and 5 further increase the fundamental frequency of their calls (p-value < 0.01, supplementary **Table S6.1**).

increased the f_0 of their calls significantly after the low-pass criterion was deactivated ($p < 0.01$, bats 2 and 6). Neither of them significantly changed the f_0 after the unshifted templates were played back ($p = 0.78$ (bat 2) and $p = 0.21$ (bat 6), **Fig. 6.5**, supplementary **Table S6.1**).

Experiment II (contour imitation)

Data acquisition for experiment II was conducted over 57 days of training. Not all bats were trained on every single training day, as training was dependent on the bats' participation and in individual cases bats needed to be excluded from training for certain periods due to weight loss. The RMS distance threshold that was applied as reward criterion was lowered over time in dependence of the success rate of the bats (**Table 6.1**). No change in RMS distance between the emitted calls and the presented template was noted in the course of the training (**Fig. 6.6**). This holds true for the four bats presented with either of the two artificial templates (**Fig. 6.6b,c**) and the control group

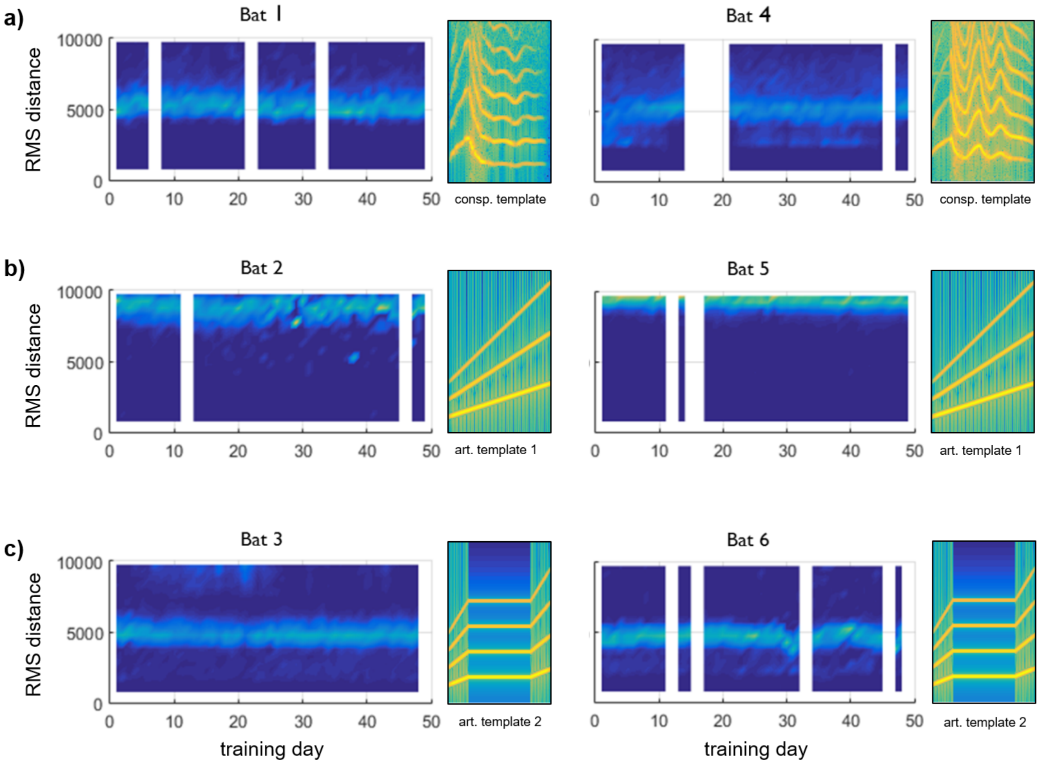


Figure 6.6: Change of RMS distance between the presented template call (right) and the calls emitted by the bats. The number of calls with a certain RMS distance to the template is indicated via the colour axis. A light colour indicates a high percentage of calls emitted with a certain RMS distance, while dark blue indicates that no calls with this RMS distance to the template were emitted. a) Bats 1 and 4 were presented with a conspecific control template. b) Bats 2 and 5 were presented with artificial template 1. c) Bats 3 and 6 were presented with artificial template 2. For more detail on the templates, see **figure 6.2**. The white bars indicate necessary breaks in the training due to weight loss of the animals. Note that there is no change in the RMS distance for bats 1 and 4, which were presented with conspecific calls as templates.

with conspecific templates (**Fig. 6.6a**). Visual inspection of the recorded calls for all bats supported these findings (supplementary **Fig. S6.3**).

Discussion

Changing the spectral parameters of a vocalisation requires the involvement of at least one of the three sound production systems. Either the phonatory system is employed in order to change f_0 or the filter system is adjusted to change the energy distribution (i.e. formants) of the vocalisation, or both systems are involved. Using a setup and training regime modified from the song bird literature (Manabe *et al.*, 2008), we here demonstrate that the pale spear-nosed bat, *Phyllostomus discolor*, can perform a directed shift of the f_0 of their social calls. While the setup and training paradigm were well suited to condition isolated bats to perform relatively simple modification of their vocalisations, they also show clear limitations when complex imitation is concerned. Specifically, it was not possible to elicit copying of different call types (conspecific or artificial templates), and thus this paradigm could not be used to assess the ability to modify frequency contours according to an acoustic template.

Lowering the f_0 according to an auditory template (experiment I)

All six bats decreased the f_0 of their calls significantly after 30 days of training following playback of a pitch-shifted version of their own calls and an activated low-pass criterion for food reward delivery (**Fig. 6.4b**). The maximal observed decrease in mean f_0 (990 Hz or 6.2% (difference of medians), bat 2, **Fig. 6.5**) was a smaller shift than the pitch shift applied to the presented template call (3.6 kHz or 24%). Nevertheless, the bats were lowering the f_0 of their calls actively and in a directional manner. This is corroborated by previous findings, which demonstrated that *P. discolor* is capable of perceiving a frequency change as little as 1% (Esser and Kiefer, 1996). At a f_0 of approximately 15 kHz, a 1% change would amount to a lowering of the f_0 by 150 Hz, which was exceeded by even the smallest observed change (260 Hz (difference in medians), bat 5, **Fig. 6.4b**, supplementary **Table S6.1**). In order to demonstrate vocal production learning, this observed change should be due to auditory experiences of the bats (e.g. adjustment towards an auditory template).

To test the driving force behind the decrease in f_0, we first deactivated the low-pass criterion (‘criterion deactivated’ data) and only afterwards presented an unshifted template (‘unshifted template’ data). If the bats had been listening to the template in order to adjust their f_0, we would have expected an increase in f_0 only once the unshifted templates were presented. If the bats had been lowering their pitch in response to the low-pass criterion, we would have expected an immediate upward shift after the criterion was deactivated. The results show that different bats used either of these strategies in the training paradigm. In other words, the driving force behind the lowering of the f_0 was different for the different animals. While bat 4 responded to the pitch-shifted acoustic template, bats 2 and 6 started to increase the f_0 of their calls while the presented templates were still downwards pitch-shifted, but the low-pass criterion was deactivated (**Fig. 6.5**). We thus demonstrate the flexibility of *P. discolor* to lower the f_0 of their calls and report strong evidence for vocal production learning in this species, supported by the results found for bat 4.

The observed shift in mean f_0 is an interesting finding, as it demonstrates the vocal plasticity of *P. discolor* in the spectral domain. We also report the change of other call parameters (namely call duration and level) in response to the paradigm. The call parameters mean f_0 and call duration were often negatively correlated, but this correlation was not imperative as the vocalisations of bat 1 showed the opposite trend. It would be interesting to also investigate if this correlation changes over the course of the training. A trajectory of this correlation would also enable us to gain deeper insights into the call changes affected by the training paradigm and the influence of the acoustic template on these changes.

The increased emission of calls during the training period (**Fig. 6.3**) and the decrease in success rate for feeder activation (supplementary **Fig. S6.1**) indicate that the task was getting progressively harder for the animals. As we calculated the mean f_0 over all emitted calls (i.e. all calls that exceeded the amplitude threshold and not only those that triggered a food reward), the observed change in mean f_0 might result from a shift in usage of calls with a lower f_0. In order to investigate the shift within the calls, it would be instrumental to separate the data sets into successful and unsuccessful trials (cf. **Fig. 6.1**). This has not yet been attempted as we aimed to investigate the overall change in response to the training paradigm. However, the observed effect size has likely been underestimated due to the pooling and averaging of all data recorded for each individual.

Modification of the frequency contour according to an auditory template (experiment II)

In the course of experiment II (contour imitation), we were not able to demonstrate a convergence between vocalisations emitted by the bats and the acoustic templates presented to them. This was true for all stimuli, artificial call templates as well as conspecific calls, which are expected to be within their natural call repertoire. The contingency training paradigm we employed increased the pressure on the bats gradually in the course of the progressing training. However, the bats continued to emit their baseline calls independent of the pressure exerted over them via the RMS distance reward threshold (supplementary **Fig. S6.3**). With the setup and paradigm used in experiment II, we failed to demonstrate an adjustment of the emitted calls to any of the presented acoustic stimuli, including conspecific vocalisations. There are several possible explanations for the bats' failure to successfully participate in experiment II. The task itself was very advanced and the bats might not have comprehended the action demanded from them in this training paradigm. The exclusively male and adult status of the animals might have contributed to their inflexibility. Furthermore, all six bats were used in both experiments, and an effect of overtraining from experiment I might have contributed to their lack of adjustment. A repetition of experiment II with naïve bats could result in more flexible behaviour and increased participation in the experiment. We want to caution the reader not to conclude that the bats are incapable of imitating acoustic templates. Because the control animals also did not converge their calls with conspecific templates, we conclude that the bats were not paying attention to the task, rather than displaying their inability to reproduce the template. We thus want to stress the point that the training paradigm was most likely too advanced or unsuitable for the bats. One possible change in the training paradigm could be a continuous change in the presented template simultaneously with the lowering of the applied thresholds. A gradual transformation of the template from a baseline call towards the artificial template call could be facilitating the comprehension of the required task and prevent habituation to the presented stimuli.

Conclusion

In this study, we have demonstrated that adult pale spear-nosed bats have some vocal learning abilities, as they have control over the pitch of their vocalisations and change their calls in a directional manner according to a pitch-shifted template. Specifically, we trained six bats to reduce the f_0 of their calls in response to a contingency training paradigm, in the course of which they were presented with a pitch-shifted version of their own calls. This paradigm shows that auditory experience is not essential for this change to occur, but can be sufficient to maintain or even further this effect. While the contingency training paradigm and basic setup used in the presented experiments was successfully used in songbirds in previous studies, the application for acoustic imitation training in bats has its limitations. Structural convergence of vocalisations towards artificially generated playback sounds could not be demonstrated. It is conceivable that the physical presence of a tutor or conspecifics would have positive effects on the learning behaviour of these bats, as social context has been shown to be an important factor facilitating vocal production learning in other species.

Acknowledgements

The authors want to thank the LMU workshop for help with constructing the setup used in this study and Dr. Mark D. Scherz for helpful discussion and feedback on the manuscript.

Funding

The research was funded by a Human Frontiers Science Program Research Grant (RGP0058/2016) awarded to LW and SCV. SCV is funded by a Max Planck Research Group Award from the Max Planck Gesellschaft.

Author contributions

LW, SCV, and EZL perceived and designed the study. LW and EZL developed the setup and analysis program. EZL acquired the data wrote the first draft of the manuscript. All authors critically revised and agreed on the final version of the manuscript.

The authors declare no conflict of interest.

References

Anderson, M.E. and Racey, P.A. (1993). Discrimination between fluttering and non-fluttering moths by brown long-eared bats, *Plecotus auritus*. *Animal Behaviour* 46, 1151–1155.

Boughman, J.W. and Wilkinson, G.S. (1998). Greater spear-nosed bats discriminate group mates by vocalizations. *Animal Behaviour* 55, 1717–1732.

de Cheveigné, A. and Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America* 111, 1917–1930.

Esser, K.H. and Kiefer, R. (1996). Detection of frequency modulation in the FM-bat *Phyllostomus discolor*. *Journal of Comparative Physiology A* 178, 787–796.

Esser, K.H. and Lud, B. (1997). Discrimination of sinusoidally frequency modulated sound signals mimicking species specific communication calls in the FM bat *Phyllostomus discolor*. *Journal of Comparative Physiology A* 180, 513–522.

Firzlaff, U.; Schörnich, S.; Hoffmann, S.; Schuller, G. and Wiegrebe, L. (2006). A neural correlate of stochastic echo imaging. *Journal of Neuroscience* 26, 785–791.

Fisher, S.E. and Vernes, S.C. (2015). Genetics and the language sciences. *Annual Review of Linguistics* 1, 289–310.

Fitch, W.T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences* 4, 258–267.

Grunwald, J.-E.; Schörnich, S. and Wiegrebe, L. (2004). Classification of natural textures in echolocation. *PNAS* 101, 5670–5674.

Hage, S.R.; Gavrilov, N. and Nieder, A. (2013). Cognitive control of distinct vocalizations in rhesus monkeys. *Journal of Cognitive Neuroscience* 25, 1692–1701.

Hauser, M.D.; Chomsky, N. and Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579.

Janik, V.M. and Slater, P.J.B. (1997). Vocal learning in mammals. *Advances in the Study of Behaviour* 26, 59–99.

Janik, V.M. and Slater, P.J.B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour* 60, 1–11.

Knörnschild, M. (2014). Vocal production learning in bats. *Current Opinion in Neurobiology* 28, 80–85.

Koda, H.; Oyakawa, C.; Kato, A. and Masataka, N. (2007). Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour* 144, 681–692.

Koselj, K.; Schnitzler, H.-U. and Siemers, B.M. (2011). Horseshoe bats make adaptive prey-selection decisions, informed by echo cues. *Proceedings of the Royal Society B* 278, 3034–3041.

Lattenkamp, E.Z.; Vernes, S.C. and Wiegrebe, L. (2018). Volitional control of social vocalisations and vocal usage learning in bats. *The Journal of Experimental Biology* 221, jeb180729.

Manabe, K.; Dooling, R.J. and Brittan-Powell, E.F. (2008). Vocal learning in budgerigars (*Melopsittacus undulatus*): effects of an acoustic reference on vocal matching. *The Journal of the Acoustical Society of America* 123, 1729–1736.

Manabe, K.; Staddon, J.E.R. and Cleaveland, J.M. (1997). Control of vocal repertoire by reward in budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology* 111, 50–62.

Neuweiler, G. (2000). *The biology of bats*. Oxford University Press, Oxford, UK.

Richards, D.G.; Wolz, J.P. and Herman, L.M. (1984). Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus*. *Journal of Comparative Physiology A* 98, 10–28.

Seyfarth, R.M. and Cheney, D.L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain and Language* 115, 92–100.

Smotherman, M.S. (2007). Sensory feedback control of mammalian vocalizations. *Behavioural Brain Research* 182, 315–326.

Vernes, S.C. and Wilkinson, G.S. (2019). Behaviour, biology, and evolution of vocal learning in bats. *Philosophical Transactions of the Royal Society of London, Series B*, 1–31.

Yovel, Y.; Melcon, M.L.; Franz, M.O.; Denzinger, A. and Schnitzler, H.U. (2009). The voice of bats: how greater mouse-eared bats recognize individuals based on their echolocation calls. *PLoS Computational Biology* 5, 1–10.

Supplementary material

Table S6.1: Change in median fundamental frequency of the calls emitted by bat 1-6 in experiment I (pitch shift). The table shows the comparison of the different data sets (i.e. before and 30 and 60 days after activation of the low-pass criterion and presentation of frequency-shifted templates) acquired during experiment I. The difference between the data sets was tested using the Wilcoxon rank-sum test. For all data sets we report the number of analysed calls, median, interquartile ranges, and *p*-values.

	N	median (kHz)	IQR (kHz)	N	median (kHz)	IQR (kHz)	diff. medians (Hz)	diff. medians (%)	*p* -value
Bat 1									
base line vs. **after 30 days of training**	180	15.87	0.47	339	15.37	0.51	-500	-3.2	***
base line vs. **after 60 days of training**	180	15.87	0.47	600	15.05	0.50	-820	-5.2	***
after 30 days of training vs. **after 60 days of training**	339	15.37	0.51	600	15.05	0.50	-320	-2.1	***
Bat 2									
base line vs. **after 30 days of training**	321	16.09	0.51	424	15.62	0.46	-470	-2.9	***
base line vs. **after 60 days of training**	321	16.09	0.51	837	15.10	0.75	-990	-6.2	***
after 30 days of training vs. **after 60 days of training**	424	15.62	0.46	837	15.10	0.75	-520	-3.3	***
after 60 days of training vs. **criterion deactivated**	837	15.10	0.75	343	15.54	0.83	440	2.9	***
criterion deactivated vs. **no template pitch-shift**	343	15.54	0.83	250	15.53	1.02	-10	-0.1	0.7763
Bat 3									
base line vs. **after 30 days of training**	223	15.59	0.34	1732	15.07	0.38	-520	-3.3	***
base line vs. **after 60 days of training**	223	15.59	0.34	2728	15.05	0.34	-540	-3.5	***
after 30 days of training vs. **after 60 days of training**	1732	15.07	0.38	2728	15.05	0.34	-20	-0.1	0.0387
Bat 4									
base line vs. **after 30 days of training**	280	14.97	1.23	3622	14.38	0.96	-590	-3.9	***
base line vs. **after 60 days of training**	280	14.97	1.23	3972	14.22	0.75	-750	-5.0	***
after 30 days of training vs. **after 60 days of training**	3622	14.38	0.96	3972	14.22	0.75	-160	-1.1	***
after 60 days of training vs. **criterion deactivated**	3972	14.22	0.75	403	14.14	1.32	-80	-0.6	***
criterion deactivated vs. **no template pitch-shift**	403	14.14	1.32	372	14.70	1.72	560	4.0	***
Bat 5									
base line vs. **after 30 days of training**	353	15.69	0.60	1486	15.43	1.35	-260	-1.7	***
base line vs. **after 60 days of training**	353	15.69	0.60	4107	15.54	1.29	-150	-1.0	0.0015
after 30 days of training vs. **after 60 days of training**	1486	15.43	1.35	4107	15.54	1.29	110	0.7	***
after 60 days of training vs. **criterion deactivated**	4107	15.54	1.29	498	15.56	0.68	20	0.1	0.7749
criterion deactivated vs. **no template pitch-shift**	498	15.56	0.68	450	15.71	0.61	150	1.0	***
Bat 6									
base line vs. **after 30 days of training**	213	15.95	0.41	2959	15.42	0.47	-530	-3.3	***
base line vs. **after 60 days of training**	213	15.95	0.41	1333	15.38	0.44	-570	-3.6	***
after 30 days of training vs. **after 60 days of training**	2959	15.42	0.47	1333	15.38	0.44	-40	-0.3	***
after 60 days of training vs. **criterion deactivated**	1333	15.38	0.44	211	15.82	0.52	440	2.9	***
criterion deactivated vs. **no template pitch-shift**	211	15.82	0.52	206	15.83	0.24	10	0.1	0.2128

*** = *p*-value < 0.0001; N = number of analysed calls ; IQR = interquartile range in kHz; diff. medians = difference between the medians of the two compared data sets (in kHz and %); 'baseline' = data from 5 days before the activation of the low-pass criterion and the pitch-shifted template; 'after 30 days of training' = data from training days 28–32; 'after 60 days of training' = data from training days 58–62; 'criterion deactivated' = data from recording days 1–5 after the low-pass criterion was deactivated, but the template was still pitch-shifted; 'unshifted template' = data from recording days 6-10 after the low-pass criterion was deactivated, the template was not pitch-shifted anymore.

Table S6.2: Change in median call duration emitted by bat 1-6 in experiment I (pitch shift). The table shows the comparison of the different data sets (i.e. before and 30 and 60 days after activation of the low-pass criterion and presentation of frequency-shifted templates) acquired during experiment I. The difference between the data sets was tested using the Wilcoxon rank-sum test. For all data sets we report the number of analysed calls, median, interquartile ranges, and *p*-values.

	N	median (ms)	IQR (ms)	N	median (ms)	IQR (ms)	diff. medians (ms)	diff. medians (%)	*p*-value
Bat 1									
base line vs. **after 30 days of training**	180	50.6	4.7	339	51.6	5.4	1.0	2.0	0.0046
base line vs. **after 60 days of training**	180	50.6	4.7	600	52.9	4.7	2.3	4.5	***
after 30 days of training vs. **after 60 days of training**	339	51.6	5.4	600	52.9	4.7	1.3	2.5	***
Bat 2									
base line vs. **after 30 days of training**	321	47.8	3.4	424	49.0	4.2	1.2	2.5	***
base line vs. **after 60 days of training**	321	47.8	3.4	837	54.3	8.5	6.5	13.6	***
after 30 days of training vs. **after 60 days of training**	424	49.0	4.2	837	54.3	8.5	5.3	10.8	***
after 60 days of training vs. **criterion deactivated**	837	54.3	8.5	343	55.4	8.6	1.1	2.0	0.8016
criterion deactivated vs. **no template pitch-shift**	343	55.4	8.6	250	53.7	9.5	-1.7	-3.1	0.0335
Bat 3									
base line vs. **after 30 days of training**	223	58.6	5.1	1732	61.3	5.3	2.7	4.6	***
base line vs. **after 60 days of training**	223	58.6	5.1	2728	70.6	6.3	12.0	20.5	***
after 30 days of training vs. **after 60 days of training**	1732	61.3	5.3	2728	70.6	6.3	9.3	15.2	***
Bat 4									
base line vs. **after 30 days of training**	280	53.9	9.8	3622	56.1	9.2	2.2	4.1	0.0002
base line vs. **after 60 days of training**	280	53.9	9.8	3972	57.4	7.6	3.5	6.5	***
after 30 days of training vs. **after 60 days of training**	3622	56.1	9.2	3972	57.4	7.6	1.3	2.3	***
after 60 days of training vs. **criterion deactivated**	3972	57.4	7.6	403	45.3	11.2	-12.1	-21.1	***
criterion deactivated vs. **no template pitch-shift**	403	45.3	11.2	372	45.4	10.9	0.1	0.2	0.8215
Bat 5									
base line vs. **after 30 days of training**	353	37.8	6.1	1486	54.1	12.1	16.3	43.1	***
base line vs. **after 60 days of training**	353	37.8	6.1	4107	50.4	11.1	12.6	33.3	***
after 30 days of training vs. **after 60 days of training**	1486	54.1	12.1	4107	50.4	11.1	-3.7	-6.8	***
after 60 days of training vs. **criterion deactivated**	4107	50.4	11.1	498	43.2	7.0	-7.2	-14.3	***
criterion deactivated vs. **no template pitch-shift**	498	43.2	7.0	450	42.0	6.2	-1.2	-2.8	0.0034
Bat 6									
base line vs. **after 30 days of training**	213	62.7	6.4	2959	72.1	9.2	9.4	15.0	***
base line vs. **after 60 days of training**	213	62.7	6.4	1333	70.6	8.2	7.9	12.6	***
after 30 days of training vs. **after 60 days of training**	2959	72.1	9.2	1333	70.6	8.2	-1.5	-2.1	***
after 60 days of training vs. **criterion deactivated**	1333	70.6	8.2	211	63.4	5.7	-7.2	-10.2	***
criterion deactivated vs. **no template pitch-shift**	211	63.4	5.7	206	59.3	3.0	-4.1	-6.5	***

*** = *p*-value < 0.0001; N = number of analysed calls ; IQR = interquartile range in ms; diff. medians = difference between the medians of the two compared data sets (in ms and %); 'baseline' = data from 5 days before the activation of the low-pass criterion and the pitch-shifted template; 'after 30 days of training' = data from training days 28–32; 'after 60 days of training' = data from training days 58–62; 'criterion deactivated' = data from recording days 1–5 after the low-pass criterion was deactivated, but the template was still pitch-shifted; 'unshifted template' = data from recording days 6–10 after the low-pass criterion was deactivated, the template was not pitch-shifted anymore.

Table S6.3: Change in median call level emitted by bat 1-6 in experiment I (pitch shift). The table shows the comparison of the different data sets (i.e. before and 30 and 60 days after activation of the low-pass criterion and presentation of frequency-shifted templates) acquired during experiment I. The difference between the data sets was tested using the Wilcoxon rank-sum test. For all data sets we report the number of analysed calls, median, interquartile ranges, and *p*-values.

	N	median (dB)	IQR (dB)	N	median (dB)	IQR (dB)	diff. medians (dB)	diff. medians (%)	*p* -value
Bat 1									
base line vs. **after 30 days of training**	180	106	3	339	107	3	1	0.9	0.0032
base line vs. **after 60 days of training**	180	106	3	600	111	3	5	4.7	***
after 30 days of training vs. **after 60 days of training**	339	107	3	600	111	3	4	3.7	***
Bat 2									
base line vs. **after 30 days of training**	321	105	3	424	105	3	0	0.0	0.0041
base line vs. **after 60 days of training**	321	105	3	837	106	4	1	1.0	0.05
after 30 days of training vs. **after 60 days of training**	424	105	3	837	106	4	1	1.0	***
after 60 days of training vs. **criterion deactivated**	837	106	4	343	105	3	-1	-0.9	***
criterion deactivated vs. **no template pitch-shift**	343	105	3	250	105	3	0	0.0	0.5657
Bat 3									
base line vs. **after 30 days of training**	223	111	4	1732	112	2	1	0.9	***
base line vs. **after 60 days of training**	223	111	4	2728	112	3	1	0.9	***
after 30 days of training vs. **after 60 days of training**	1732	112	2	2728	112	3	0	0.0	***
Bat 4									
base line vs. **after 30 days of training**	280	108	3	3622	109	4	1	0.9	***
base line vs. **after 60 days of training**	280	108	3	3972	111	6	3	2.8	***
after 30 days of training vs. **after 60 days of training**	3622	109	4	3972	111	6	2	1.8	***
after 60 days of training vs. **criterion deactivated**	3972	111	6	403	107	3	-4	-3.6	***
criterion deactivated vs. **no template pitch-shift**	403	107	3	372	108	4	1	0.9	0.0004
Bat 5									
base line vs. **after 30 days of training**	353	108	4	1486	107	3	-1	-0.9	***
base line vs. **after 60 days of training**	353	108	4	4107	107	3	-1	-0.9	***
after 30 days of training vs. **after 60 days of training**	1486	107	3	4107	107	3	0	0.0	0.9637
after 60 days of training vs. **criterion deactivated**	4107	107	3	498	107	3	0	0.0	***
criterion deactivated vs. **no template pitch-shift**	498	107	3	450	108	4	1	0.9	***
Bat 6									
base line vs. **after 30 days of training**	213	113	3	2959	113	2	0	0.0	***
base line vs. **after 60 days of training**	213	113	3	1333	113	3	0	0.0	***
after 30 days of training vs. **after 60 days of training**	2959	113	2	1333	113	3	0	0.0	***
after 60 days of training vs. **criterion deactivated**	1333	113	3	211	111	3	-2	-1.8	***
criterion deactivated vs. **no template pitch-shift**	211	111	3	206	112	2	1	0.9	0.0001

*** = *p*-value < 0.0001; N = number of analysed calls ; IQR = interquartile range in dB; diff. medians = difference between the medians of the two compared data sets (in dB and %); ‘baseline’ = data from 5 days before the activation of the low-pass criterion and the pitch-shifted template; ‘after 30 days of training’ = data from training days 28–32; ‘after 60 days of training’ = data from training days 58–62; ‘criterion deactivated’ = data from recording days 1–5 after the low-pass criterion was deactivated, but the template was still pitch-shifted; ‘unshifted template’ = data from recording days 6–10 after the low-pass criterion was deactivated, the template was not pitch-shifted anymore.

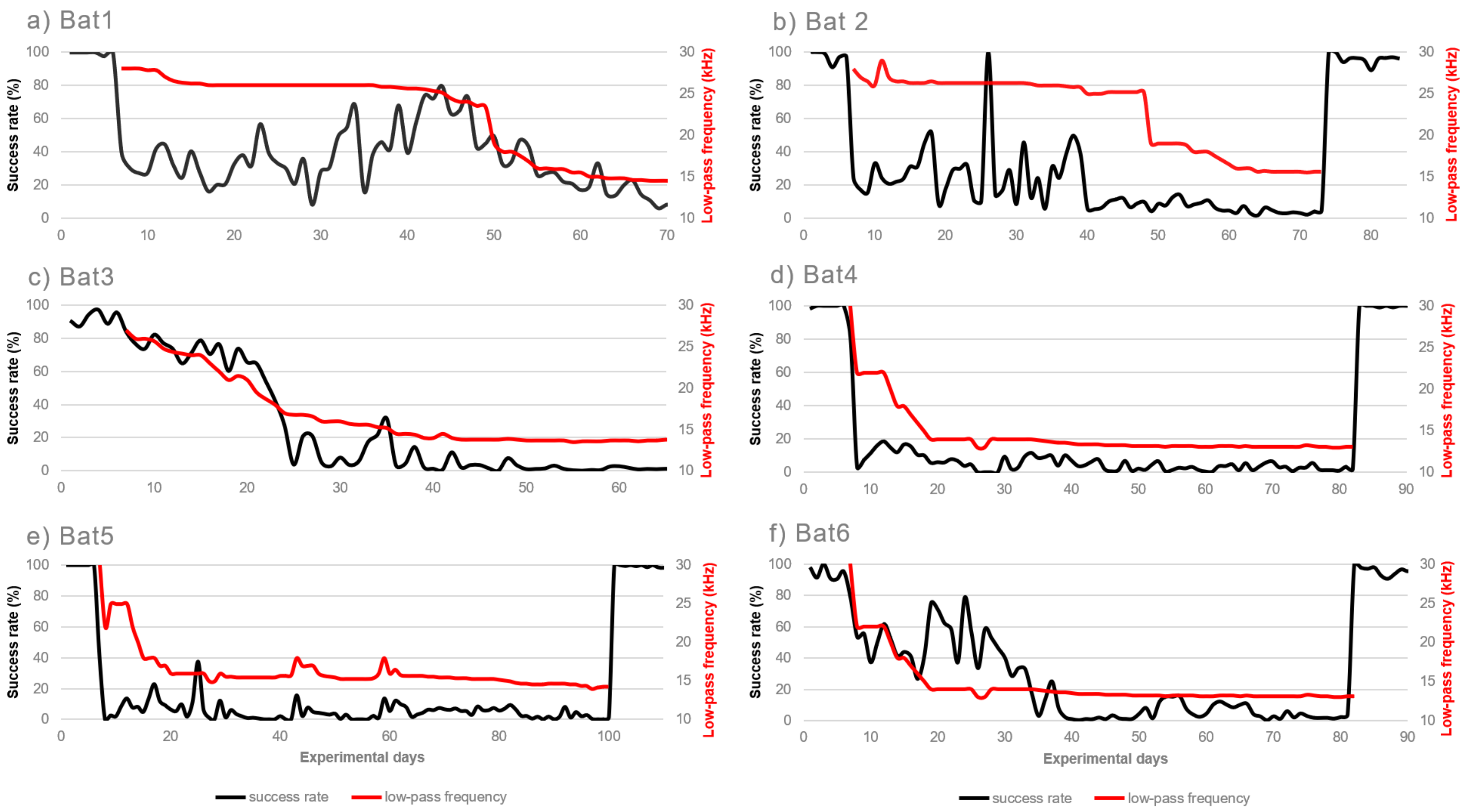


Figure S6.1: The success rate in dependence of the applied low-pass filter criterion in experiment I (pitch shift). The success rate (black line, left y-axis) represents the percentage of food rewards triggered depending on the total number of calls emitted during the respective training session. The low-pass filter criterion (red line, right y-axis) was activated after the first five days of data acquisition (baseline data) and subsequently adjusted according to the performance of the individual bats. In such a contingency training paradigm, the aim is to keep the success rate low as this indicates continuous pressure exerted over the animals. In this specific case, this means that there is constant pressure to further decrease the fundamental frequency of their vocalisations. For bats 2,4,5,6 the low-pass filter criterion was deactivated, after the data acquisition was finalised and additional recordings were made with the deactivated criterion and subsequently with a non-pitch-shifted template. The success rate immediately goes back to ca. 100% when the criterion is deactivated, as every emitted vocalisation that exceeds the level threshold is rewarded. Data acquisition for the different animals went on for different time spans, as the participation in the training varied.

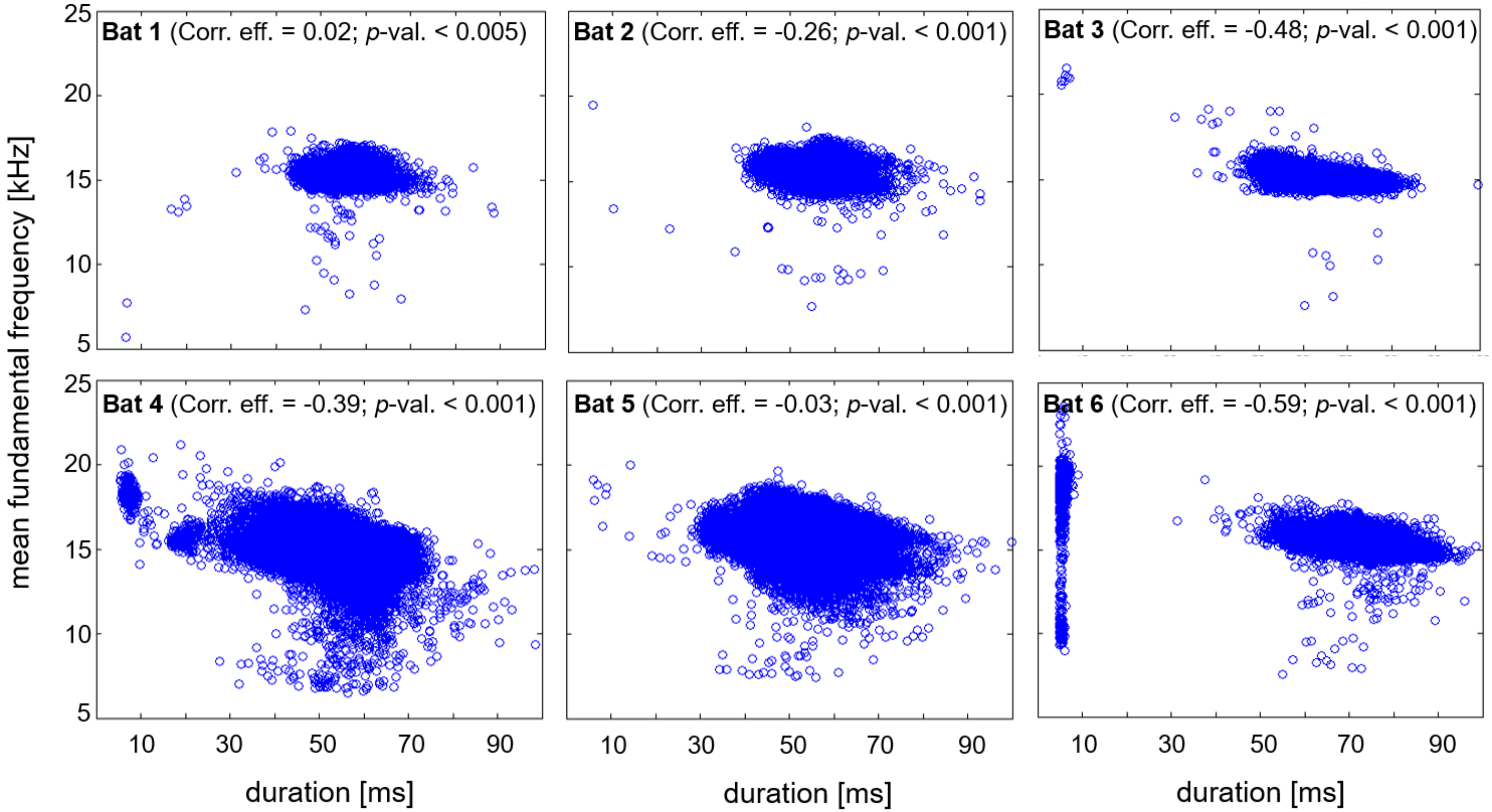


Figure S6.2: Results of the correlation analysis of the mean fundamental frequency and the call duration. Illustrated is a scatterplot of the pooled calls of each individual. The correlation analysis demonstrates that fundamental frequency can be negatively correlated with call duration (bats 2-6), but also that this correlation is not imperative (bat 1). Bat 6 used two call types (see also chapter 5), one of which was shorter and was more variable in its detected mean fundamental frequency. The negative correlation between mean fundamental frequency and call duration of bat 6 was probably overestimated due to the use of two call types, however a negative correlation can be conceived even within the longer calls of this individual.

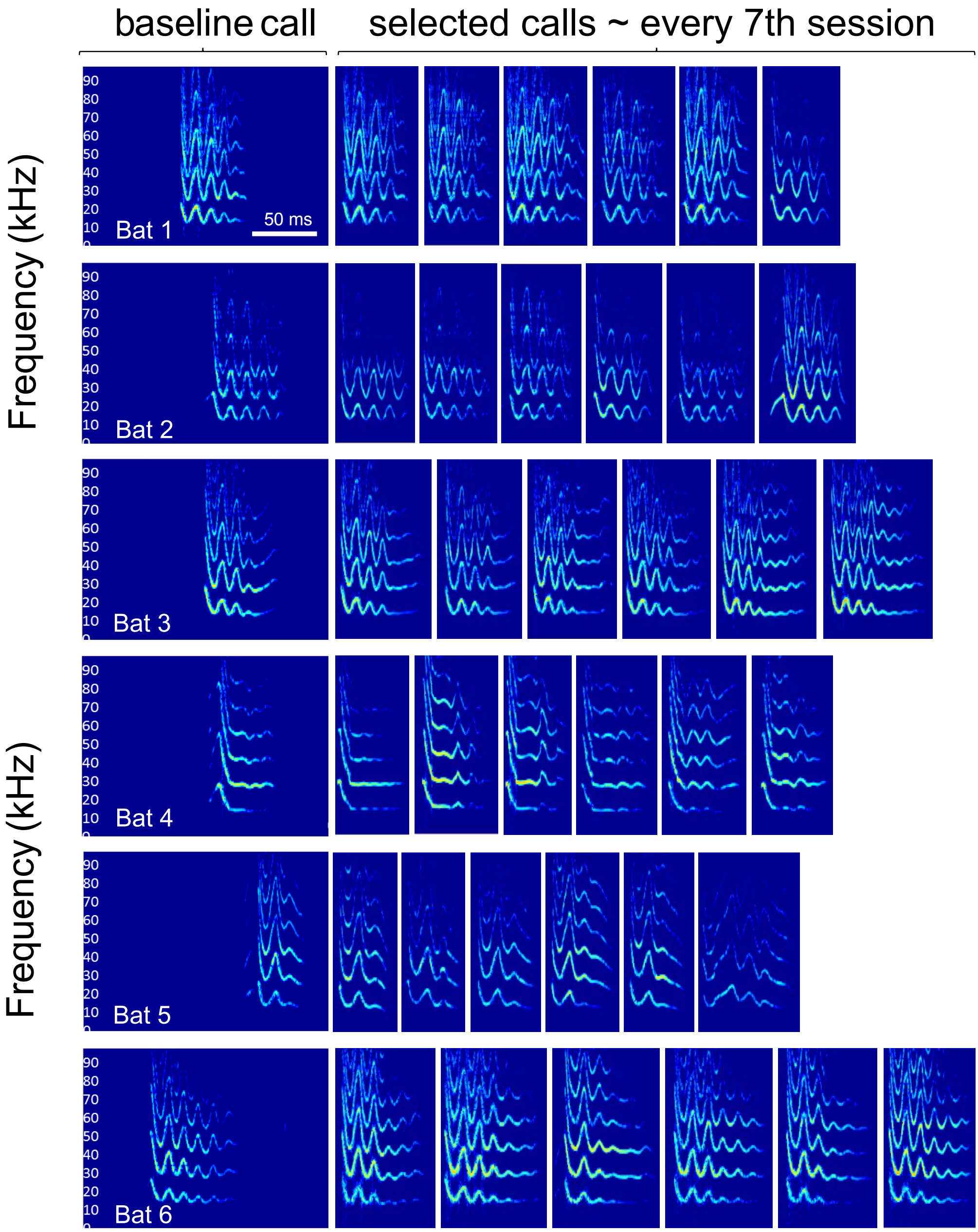


Figure S6.3: Spectrograms of example calls emitted by each bat in the course of training in experiment II (contour imitation). The leftmost spectrogram always indicates the 'typical call' for each bat, recorded in the baseline data set before the presentation of artificial or control templates. The following six spectrograms are manually selected calls from ca. every 7th recording session. The spectrograms show the unchanged, stereotypical shape of each individual's call.

Chapter 7

Discussion

General discussion

Vocal learning is a complex behavioural trait identified in only a few vertebrate species. Understanding the foundation of the vocal learning capacity in other study systems will allow insights in the evolutionary and functional basis in the human capacity for speech and spoken language acquisition. In this thesis, I have discussed the need for a common, unifying framework for the cross-species study of vocal learning (chapter 2) and then went on to establish the perceptional basis for the investigation of vocal learning in bats (chapter 3). Finally, I addressed three topics concerning the vocal learning capacity of the pale spear-nosed bat, *Phyllostomus discolor* (cf. section 1.7). (I) I showed how acoustically and behaviourally complex the vocal emissions of pale spear-nosed bats are by describing their vocal repertoire and the behavioural contexts in which the different identified syllable classes are emitted (chapter 4). (II) I showed that *P. discolor* vocalisations can be brought under operant control and demonstrated their capacity for vocal usage learning (chapter 5). (III) I demonstrated that *P. discolor* have control over call amplitude and duration (chapter 5), and moreover showed their capacity for directional change of their fundamental frequency (or ‘pitch’) according to an auditory target (chapter 6).

In the following, I will briefly summarise the findings presented in the previous chapters (section 7.1) before I discuss these findings in the context of the literature and current definitions in the field (section 7.2). I will discuss limitations of the presented work (section 7.3) and provide an outlook for future research that will address gaps in our current understanding of vocal learning (section 7.4).

7.1 Summary of the findings presented in this thesis

7.1.1 The importance of cross-species approaches to vocal learning (chapter 2).

In order to support cross-species approaches to the study of vocal learning, I introduced a theoretical framework that is aimed at strategically combining findings from various species. The proposed framework is intended as a guideline for the design of future vocal learning studies through the identification of gaps in the existing knowledge. Furthermore, the framework should function as a scaffold that aids researchers to place their findings into context and thus to facilitate the organisation of findings within the field. In the past, research on vocal learning has been diverse not only in the animal models investigated, but also in the studied developmental stage, evolutionary approaches, behavioural assessments, and acoustic analyses. Due to the heterogeneous nature of the experimental studies conducted in the field of vocal learning, this framework is especially important to combine and focus future research efforts. Integrating results from many species by applying this framework will allow the determination of common factors in the evolution of vocal learning across different species and thus highlight shared evolutionary pressures acting to support the emergence of this trait.

7.1.2 Investigation of loudness encoding in bats (chapter 3).

Knowledge about the sound perception of a species allows us insights in the way these organisms hear the world and is an important prerequisite for the meaningful design of acoustic stimuli in experimental studies. In chapter 3, I presented a large-scale comparative study, in which I described the audiograms and characteristics of loudness encoding for eleven Neotropical bat species. I found that despite species-specific differences in audiogram shape, a general principle for loudness coding was present in all measured bat species, which may extend to other mammals.

7.1.3 Vocal and behavioural complexity of the pale spear-nosed bat (chapter 4).

Establishing a vocal repertoire is extremely valuable for the assessment of vocal learning, as the comparison between the normally emitted vocalisations (baseline) and newly acquired vocalisations (learned) is often required for clear demonstration of vocal learning. Knowledge about the natural range of a species' vocal production is thus an important criterion for the study of vocal learning. In chapter 4, I assessed the vocal repertoire of the focal species of this dissertation, the pale spear-nosed bat. I described eight classes of commonly occurring syllables and four further syllable classes which were emitted less frequently. Moreover, I showed that most of these syllables were emitted in distinct behavioural contexts.

7.1.4 Volitional control of social vocalisations and usage learning (chapter 5).

Having established the perception- (chapter 3) and production- (chapter 4) related groundwork for vocal learning in pale spear-nosed bats, I subsequently focused on the demonstration of volitional control over their vocal emissions (chapter 5). For this demonstration of vocal control and usage learning I set up a behavioural paradigm, in which isolated bats used social vocalisations in order to elicit food rewards. Vocal conditioning of isolated individuals of such a gregarious species was challenging, but I succeeded in eliciting social vocalisations in isolation. This demonstration of usage learning, e.g. volitionally emitting a social vocalisation to achieve a context-independent goal, such as food reward delivery, was the first application of such a paradigm in a bat model. Furthermore, I also showed that these bats can modify their calls in response to challenges imposed on them by the paradigm (chapter 5). The introduction of a spectral high-pass filter challenged the bats to modify their vocalisations and demonstrated their capacity to change their social vocalisations volitionally. This study provided evidence for the control over call duration and amplitude and indications for volitional spectral changes.

7.1.5 Indications for vocal production learning in the pale spear-nosed bat (chapter 6).

The modification of spectral call parameters is widely considered to be more challenging than the modification of temporal parameters or call amplitude, as it requires the modification of the phonatory and/or filter system. In chapter 6, I focused on the demonstration of directional change of spectral call parameters. I used the behavioural paradigm established in chapter 5 in order to show the ability of *P. discolor* to change the fundamental frequency, or pitch, of their vocalisations in a directional manner in response to an auditory target and a contingency training paradigm (chapter 6). Within 30 days of training, the bats decreased the fundamental frequency of their calls significantly according to the presented auditory target. By deactivating first the reward criterion and then the pitch-shift of the target, I was able to demonstrate that at least one of the trained bats modified its vocalisations by following the acoustic target rather than the reward criterion.

7.2 State of the art: vocal learning with particular reference to bats

My research efforts were aimed at answering one main question: Is the pale spear-nosed bat a vocal learning species? In order to answer this question, it is essential to know which characteristics identify a given species as a vocal learner. These characteristics should be extractable from the definitions of vocal learning, which are expected to clearly distinguish between vocal flexibility and vocal production learning. However, as the experimental approaches and the species showing the vocal learning capacity are very diverse, finding a common, overarching definition is difficult. Despite the release of several recent reviews (Petkov and Jarvis, 2012; Janik, 2014; Knörnschild, 2014; Tyack, 2016; Vernes and Wilkinson, 2019) the line between mammalian vocal learning and vocal flexibility is still blurry, to say the least. In order to definitively answer whether *Phyllostomus discolor* is capable of vocal production learning, I will first highlight differences and similarities between the most prominent definitions of vocal learning and then evaluate previous findings in different bat species in the light of these definitions. Only after taking a critical look at the current literature will it be possible to assess if the pale spear-nosed bat is indeed capable of vocal production learning based on the findings presented in this thesis.

7.2.1 Key definitions of Vocal Production Learning

The most comprehensive and frequently referred to definitions for mammalian vocal learning are found in reviews from Janik and Slater (1997, 2000), Boughman and Moss (2003), Petkov and Jarvis (2012), and Tyack (2016). Janik and Slater (1997, 2000) clearly distinguish between contextual learning, which includes sound comprehension and vocal usage learning, and vocal production learning (VPL). Their definition of sound comprehension learning comprises reacting in a specific way to an external acoustic signal (e.g. a dog learning to behave in a certain way in response to a spoken command). The capacity for vocal usage learning contains conditional call control and also a change of context in which a certain call is produced. These two forms of contextual learning are distinctly separated from VPL, which is the learned generation of new sounds. More specifically, according to Janik and Slater (1997) VPL is the modification of vocal signals 'as a result of experiences with those of other individuals'. Here 'other individuals' generally refers to conspecifics, but heterospecifics or electronic tutors are also conceivable.

For Boughman and Moss (2003) vocal learning is intrinsically connected to the 'modification of acoustic signals produced in a social context'. They thus approach the topic of vocal learning under the strict prerequisite of social interactions and exclude vocal learning studies in which the social component was missing. They distinguish between two types of VPL: learned acquisition and social modification (Boughman and Moss, 2003). Learned acquisition refers to the addition of new vocalisations to an individual's repertoire, while social modification entails the alteration of existing calls in response to social interactions. Through learned acquisition individuals should thus be able to broaden their repertoire, while social modifications will generally lead to increased acoustic similarity between the social partners.

While the definitions by Janik and Slater (1997, 2000) and Boughman and Moss (2003) focus on the mechanisms leading to vocal learning and present a rather categorical concept (i.e. either you are a vocal learner or not), the VPL definitions of Petkov and Jarvis (2012) and Tyack (2016) aim at a more qualitative assessment of the phenotype. Tyack (2016) 'discriminates limited vocal learning, which uses auditory input to fine-tune acoustic features of an inherited auditory template, from complex vocal learning, in which novel sounds are learned by matching a learned auditory target'. Petkov and Jarvis (2012) published a similar description of the vocal learning phenotype, which they separate into discrete steps (i.e. vocal non-learners, limited, moderate, complex, and high vocal learners). Both definitions agree that the majority of vocal learning species are considered limited vocal learners, and only a few taxa show complex (Tyack, 2016) or even high vocal learning (Petkov and Jarvis, 2012).

The differences between these definitions seem only slight, however, they can lead to quite differing conclusions about the assessment of VPL. One conceivable example is the convergence of the call repertoire of two merging animal populations. Repertoire convergence could be caused by the acquisition of novel calls or modulation of call characteristics in one or both populations. This case would be considered a demonstration of VPL according to all definitions, with Petkov and Jarvis (2012) and Tyack (2016) specifying it as complex vocal learning. However, the convergence could also be due to changes in call type usage caused by the change in social affiliations between the animals. This latter case would be considered VPL by Boughman and Moss (2003), but not by Janik and Slater (1997). Furthermore, it would be considered limited vocal learning by Tyack (2012), but would be assessed as moderate vocal learning by Petkov and Jarvis (2012). This example demonstrates the importance for detailed reporting of the experimental findings and the applied definition, when claiming the occurrence of VPL.

7.2.2 The difficult theoretical assessment of VPL: the role of novelty

Distinguishing between different forms of contextual and vocal learning can be challenging and is dependent on the definition used. Thus, the conclusions about the demonstration of VPL drawn from different studies can vary strongly. As definitions might provide inconclusive guidelines, experimental studies often focus on the demonstration of especially strong evidence. Such a commonly accepted, convincing argument for the definite occurrence of VPL is the demonstration of the acquisition of a novel call (Janik and Slater, 1997; Boughman and Moss, 2003; Petkov and Jarvis, 2012; Tyack, 2016). However, herein lays the problem: proving that a newly recorded call was not pre-existing in an individual's repertoire is a difficult undertaking. This difficulty is rooted in the common prevailing lack of exhaustive species-specific vocal repertoires. The acquisition of a complete vocal repertoire requires recordings of all behavioural contexts, life history events, social interactions, and developmental stages from several individuals of both sexes. As wild animals are often inaccessible and difficult to track over long distances or time spans, complete vocal repertoires are often near impossible to acquire. However, without characterising the vocal repertoire it is challenging to demonstrate that a novel recorded call did not already exist in the animal's repertoire. Currently, there are several theoretical approaches available to estimate the repertoire size of a given animal species, however, single novel calls could always present a missed outlier in the estimation process rather than a newly acquired vocalisation (Kershenbaum *et al.*, 2015).

To make things even more challenging, Janik and Slater (1997) mention several ways in which new call characteristics can be introduced into a species vocal repertoire that may not be due to learning. Examples are the accidental recording of a newcomer to the group with a different call repertoire, changes in environmental conditions that can have an effect on call parameters (for example temperature and humidity), or maturation processes. Recording animals under stable laboratory conditions, as done in chapters 4-6 of this thesis, helps to exclude most of these uncertainty factors. However, a laboratory setting cannot exclude all conceivable aspects, which might give rise to novel vocalisations independent of VPL.

One interesting potential cause of novelty mentioned by Janik and Slater (1997) is the invention of calls. The invention of new calls can be considered vocal flexibility, contextual learning, or VPL depending on the context. If, for example, invention included random sound generation, it would be considered a sign of vocal flexibility, because no auditory experience is needed in order to induce these changes. If new sequences were generated out of pre-existing calls, this type of invention would instead mostly be considered contextual learning. Emitting previously known calls in a new sequential order would be considered VPL by Janik and Slater (2000) only under one condition: if the new sequence as a whole is the meaningful unit and thus the complete sequence would have to be learned anew, i.e. if the novel sequence presented a new signal with an adaptive function. Lastly, if invention is used in order to generate a new call that avoids spectral or temporal overlap with other sounds, it would always be considered VPL. Although VPL through avoidance also occurs in the absence of an acoustic target, auditory feedback is in this case still necessary for a directed change of the vocalisation parameters.

This difficulty in evaluating vocal novelty highlights the importance of clarifying not only which precise definition of VPL is applied, but also the significance of a detailed reporting culture. The acquisition of reference data, such as call repertoires (see chapter 4) or baseline calls (see chapters 5-6), is a prerequisite for the comparison against newly arisen changes in vocal parameters. Only the comparison between observed learning effects and natural variation can indicate which type of learning has been demonstrated and if the learning effect is biologically relevant with regard to the focal species. Furthermore, extensive background knowledge about the social behaviours and ecology of a species are beneficial, in order to assess the context in which the vocalisations are emitted or being changed.

7.2.3 VPL evidence from bats

Bats have been singled out as one of the few and, due to their comparatively easy maintenance and small size, arguably one of the most promising mammalian groups for the study of VPL (Knörnschild, 2014; Vernes, 2017; Vernes and Wilkinson, 2019). However, evidence for VPL in different bat species has been assessed based on a variety of different experimental approaches and studied call parameters (**Table 7.1**). The experimental approaches to the study of VPL in bats include observational field studies, behavioural conditioning, isolation and playback studies, transfer, and acoustic avoidance studies. The tested and reported parameters comprise among others the number of frequency modulation cycles per call, call duration, call repetition rate, fundamental frequency, and bandwidth of the fundamental frequency, Euclidean distance, entropy, and spectral centroid of the vocalisations (**Table 7.1**). Generally, only a selection of the tested parameters is reported and rarely do two studies report the same call features. These discrepancies impede the assessment of VPL in bats as findings from different studies are often not comparable. Especially in the light of the growing number of experimental studies reporting VPL in bats, the need for comparability is becoming more and more pressing (see chapter 2). In the following, I will focus in detail on four bat species, namely *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus hastatus*, and *Phyllostomus discolor* (sections 7.2.4–7.2.7), which have shown indications of VPL. These four species were selected based on the relative comparability of reported acoustic parameters (for more detail on other bat species with indications of VPL see section 1.5). From these studies, I will isolate a set of acoustic parameters, which have previously been used for the demonstration of VPL in bats. More importantly, I will extract the exact values for the observed change in these studies. This extraction of explicit evidence is essential to get an impression of the absolute call parameter changes observed in bats due to VPL. By directly comparing the results of my research to this set of extracted parameters (**Table 7.1**), I will be able to put my findings into the context of current bat VPL literature. This comparison will allow me to discuss the VPL capacity of the pale spear-nosed bat and its placement on the vocal learning continuum.

7.2.4 Evidence from *Rhinolophus ferrumequinum*

Jones and Ransome (1993) showed in a long-term study that pups of the greater horseshoe bat, *Rhinolophus ferrumequinum*, adapt the resting frequency of the constant element of their echolocation calls to the frequency their mothers use. While the structure of the calls stays the same and seems to be inherited, the resting frequencies differ between siblings born in different years. This was robustly demonstrated, as the resting frequency of the mothers' echolocation calls changes over their lifetime and the resting frequency of their pups show accordingly consistent differences. However, although the differences in resting frequency were significant, they were only of the order of 300 Hz. This absolute change in frequency is surprisingly small, considering that 200 Hz has been shown to be the normal seasonal fluctuation in the resting frequency in these animals (Long and Schnitzler, 1975). Even though this study is regularly cited as evidence for vocal learning in bats and demonstrates the impact of external acoustic feedback on vocalisations, even Jones and Ransome (1993) caution that the transmission of echolocation call frequency from mother to infant could have both, genetic and learned components, and suggest further experimental evidence to solidify the evidence for vocal learning in *R. ferrumequinum*. Findings about social call VPL have not yet been published for this species.

7.2.5 Evidence from *Rousettus aegyptiacus*

The Egyptian fruit bat, *Rousettus aegyptiacus*, was first indicated as a potential vocal learning species by Prat *et al.* (2015), who showed the influence of isolation and playback experiments on the vocal development of this species. Prat and colleagues (2015) raised two groups of five pups each in absence of adult animals. The experimental group was acoustically isolated, while the control group was presented with playbacks of adult conspecifics. Pups from both groups showed a decreasing trajectory of the average fundamental frequency of their vocalisations over time. However, after 210 days the vocalisations of the experimental group showed a difference of 500 Hz to the fundamental frequency of the control group and of adult vocalisations (Prat *et al.*, 2015). Within these 210 days the acoustic differences between pups and adult bats (measured as Euclidean distance in an acoustic feature space) also decreased. Both, experimental and control pups, started from an acoustic difference of 3.0–3.2 to the adult vocalisations. While the control animals reduced the difference to adult vocalisation to 0 within 110 days, the isolated pups still showed a difference of ca. 0.2 after 210 days (Prat *et al.*, 2015). They also found that the spectral entropy (measure of noisiness/tonality of a vocalisation) of the isolation group calls was lower (0.64) than the control group (0.65) (Prat *et al.*, 2015). These results indicate not only that isolation results in slower vocal development towards an adult repertoire, but also that juvenile calls converge toward the adult baseline with or without adult acoustic input. This experiment rather demonstrated delayed vocal development, than classical vocal learning, but it did show the absence of a fixed learning phase and a continuous vocally plastic phase for this species.

In a follow-up experiment (Prat *et al.*, 2017), vocal dialects were elicited in three groups of pups (n = 5, 5, and 4). The pups were raised together with their mothers and, for the duration of one year, exposed additionally to playbacks of conspecific, agonistic calls. The playbacks were assembled to mainly consist of calls with either a high (average fundamental of 1,871 Hz), a low (average fundamental frequency of 303 Hz), or a 'normal' (average fundamental frequency of 564 Hz) fundamental frequency distribution. Although the forming of vocal dialects was reported, as all groups showed a significantly higher fraction of calls with the target fundamental frequency in contrast to the other test groups, the average fundamental frequency of all groups was ca. 600 Hz, suggesting an innate preference (Prat *et al.*, 2017). Recently, further evidence for continuous vocal plasticity in this species was published. For the duration of two weeks, Genzel and colleagues (2019) played noise of a specific spectral content to four groups of *R. aegyptiacus* (n = 6 or 4) and found that changes in 11 vocalisation characteristics occurred, which were persistent even after the noise exposure ended. The strongest change was observed in the spectral centroid of the calls. The maximal observed change was an increase of the average spectral centroid of group 3 (n = 4; 6.65 ± 1.91 kHz) from the baseline calls to after the noise exposure (Genzel *et al.*, 2019).

7.2.6 Evidence from *Phyllostomus hastatus*

Evidence for vocal learning in the greater spear-nosed bat, *Phyllostomus hastatus*, comes from the analysis of group-distinctive screech calls. Boughman (1997) showed that 28 *P. hastatus* bats from three different social groups used distinctly different screech calls between groups, but that these were indistinguishable within groups. As the social groups of *P. hastatus* are known to generally consist of unrelated females (McCracken and Bradbury, 1977, 1981), genetic inheritance can be excluded as a driver for group-specific similarity of these screech calls. Boughman (1997) investigated twelve call parameters of which nine significantly contributed to group discrimination. However, pulse duration (or 'syllable duration', as defined in chapter 4) was found to have the strongest effect and resulted in group discrimination by itself. These group-specific screech calls are identifiable by *P. hastatus*, meaning that they can discriminate between own-group and alien-group playbacks (Boughman and Wilkinson, 1998). These group-specific differences were hypothesised to be caused by call convergence between unrelated females in a group (Boughman, 1997) in order to enable group coherence during e.g. foraging (Boughman and Wilkinson, 1998).

This acoustic convergence hypothesis was tested in a follow-up transfer experiment. Boughman (1998) reciprocally transferred five females from one captive colony room to another. In total ten transfer bats and thirteen resident females were recorded before and after the transfer. Boughman (1998) demonstrated that frequency and temporal parameters of the new groups (both transfer and resident animals) were adjusted so that after five months the groups were again distinctly different from each other, but calls within groups were practically indistinguishable. Notably, while the resident females change their vocalisations drastically in the course of the experiment, the transfer bats only change their vocal emissions minimally. For example, resident group 1 shows a decrease of ~1.5 kHz in call bandwidth (pre vs. after 5 months), while transfer group 1 shows an increase of only ~0.15 kHz over the same time span. Furthermore, while resident group 1 shows a 250 ms decrease in call repetition rate (pre vs. after 5 months), transfer group 1 increases its call repetition rate by only 20 ms (Boughman, 1998). This strong change in resident call characteristics in contrast to the small change of the transfer group is surprising, especially considering Boughman's hypotheses that (i) vocal convergence is used to conserve group differences and that (ii) these group-specific differences should be difficult to copy to avoid cheating in the signalling of group affiliation (Boughman, 1998). Although this experiment shows clear indications for vocal convergence and controls for age-differences, it would strongly profit from the inclusion of control groups, and thus the reporting of naturally occurring call parameter shifts in undisturbed roosts.

7.2.7 Evidence from *Phyllostomus discolor*

One of the earliest studies on vocal learning in bats was published by Esser and Schmidt (1989), who showed that isolation calls of *Phyllostomus discolor* pups become progressively more similar to maternal directive calls over time. *P. discolor* pups reportedly emit four types of isolation calls of which only one was examined in this study as the others cease to be emitted by week two after birth. The authors investigated the development within the first 47 postnatal days, during which the bats were kept together in a group consisting of eight mothers and nine pups. The similarity between isolation and directive calls was measured on the basis of 20 calls per animal, which were selected based on unreported criteria. The number of frequency modulation (FM) cycles per call and the temporal distance between the first and last FM peak (modulation duration) were used as similarity metrics. After 47 days, calls emitted by five out of eight pups showed corresponding distribution patterns in the number of the FM cycles of their calls (as assessed by χ^2 test) (Esser, Schmidt, 1989). Calls from four out of eight pups showed matching patterns to the temporal call structure of their mothers (as assessed by χ^2 test) (Esser and Schmidt, 1989). Summarising, Esser and Schmidt (1989) demonstrated that the number of FM peaks of *P. discolor* pups generally increase during ontogeny until they match their mothers' maternal directive call. Furthermore, they showed that the modulation duration of half of the tested pups increased until it matched their mothers'. No absolute numbers were reported for the variation between number of FM cycles or modulation duration between pup and maternal vocalisations.

In order to support these initial findings, Esser (1994) later used an isolation experimental approach to show vocal adjustment of *P. discolor* pups towards an acoustic playback. For this, ten pups were hand-raised in isolation. Five of these were presented daily with a playback of one maternal directive call, while the other five were kept in total acoustic isolation. Esser (1994) found that pups from the playback group developed calls that were more similar to the playback than the calls of a control group. The similarity metrics used in this study were number of FM cycles, bandwidth of the fundamental frequency, and modulation duration. The experiment closely followed the vocal development of these ten pups for the first 50 postnatal days and additionally included one observation at day 100. Evidence for vocal learning was reported mainly by comparing the development between experimental day 50 to 100, although no further acoustic stimulation occurred after experimental day 50. The mean number of FM peaks in the playback group increased from 6 to 8.5, while the control group only increased from 5 to 6 FM cycles. The mean maximal frequency of the control group decreased from 19 to 18 kHz, while the playback group decreased from 18 to 15 kHz. A smaller decrease was reported for the mean minimal frequency, which decreased for the control group from 14.5 to 13 kHz and for the playback group from 12.5–12 kHz. The mean modulation duration is only given for day 50 of the experiment and shows a broader spread for the

species	study type	parameter	quantity	direction of change	reference
R. ferrumequinum	Observational	resting frequency of echolocation calls	± 0.3 kHz	adjustment towards maternal call (seasonal variation: 0.2 kHz)	Jones & Ransome, 1993
R. aegyptiacus	Isolation/playback	fundamental frequency	+ 0.5 kHz	higher than adult vocalisation at day 210	Prat *et al.*, 2015
R. aegyptiacus	Isolation/playback	Euclidean distance	0.2	difference to adult vocalisation after 210 days (initial difference (week 1): ~3.1)	Prat *et al.*, 2015
R. aegyptiacus	Isolation/playback	entropy	0.64	difference to adult vocalisation after 210 days (control group: 0.65)	Prat *et al.*, 2015
R. aegyptiacus	Acoustic avoidance	spectral centroid	+ 6.65 kHz	variation from baseline after noise exposure	Genzel *et al.* , 2019
P. hastatus	Observational	bandwidth	± 1.25 kHz	difference between social groups	
P. hastatus	Observational	pulse duration	± 75 ms	difference between social groups (strongest effect out of 9 significant parameters)	Boughman, 1997
P. hastatus	Group transfer	bandwidth	- 1.5 kHz	decrease compared to baseline (change occurs in resident group)	Boughman, 1998
P. discolor	Observational	number of FM cycles	± 2 cycles	imitation of maternal call	Esser & Schmidt, 1989
P. discolor	Observational	call duration	± 30 ms	imitation of maternal call	Esser & Schmidt, 1989
P. discolor	Isolation/playback	mean bandwidth of fundamental frequency	2.5 kHz	decrease compared to baseline (playback group)	Esser, 1994
P. discolor	Isolation/playback	number of FM cycles	+ 2.5	increase as adjustment towards playback (control group: 1 FM cycle increase)	Esser, 1994
P. discolor	Operant conditioning	call duration	+ 13.5 ms	increase compared to baseline (in absence of playback)	chapter 5
P. discolor	Operant conditioning	spectral centroid	+ 0.2 kHz	increase compared to baseline (in absence of playback)	chapter 5
P. discolor	Isolation/playback	fundamental frequency	- 0.99 kHz	decrease; adjustment towards playback	chapter 6

Table 7.1: Comparison between the extracted detailed parameter changes reported as evidence for VPL in bats. This table lists the type of experiment and extracted acoustic parameters which have previously been used to demonstrate vocal learning in the specific bat species. The same reported parameters are highlighted in the same colour. Some of the studies referred to measured more than the here listed parameters, however, the most significant or important parameters are named here.

control group (22–45 ms) than for the playback group (20–32 ms) (Esser, 1994). Although isolation/playback experiments have the potential to provide convincing evidence for VPL and are useful tools to exclude normal maturation effects, the results should be evaluated carefully. Janik and Slater (1997) already caution to accept this specific study as evidence for vocal learning: 'One problem with isolation experiments is the general lack of stimulation. Bats might simply need auditory input of some non-specific sort to develop normal calling behaviour. An experiment with two groups hearing different calls would clarify whether learning is involved' (Janik and Slater, 1997).

7.2.8 Placing this thesis in the context of previous research

With my research on the vocal learning capacity of *Phyllostomus discolor*, I aimed to fulfil three criteria outlined by Janik and Slater (2000) for the conclusive demonstration of VPL: (1) the imitation of a playback should be demonstrated on at least two types of playbacks to be conclusive (a non-stimulated control group is not sufficient, but additionally recommended); (2) the auditory target should be atypical for the species; (3) an alteration of specific, single call characteristics should be shown. None of the above listed studies reporting indications for vocal learning in bats produced this kind of evidence. In the experiments conducted and presented in chapters 5 and 6, I considered these issues, and my experimental design specifically addressed all three of these points. While the alteration of specific call parameters within a particular call was already shown in chapter 5, in chapter 6 I focused on the demonstration of imitating an acoustic target (criteria 1–3). Using an approach that combines the ideas used in the studies of Esser (1994) and Prat *et al.* (2017), I first trained bats to imitate playbacks of their own calls, that were shifted only in their frequency content and thus addressing the third criterion stated by Janik and Slater (2000). I was able to show significant directional change of the fundamental frequency of their calls of as much as 990 Hz within 60 days of training (chapter 6, experiment I). In experiment II of chapter 6, two artificial, species-atypical acoustic targets were used to train two bats each, and a third control group was trained with a conspecific target playback (addressing criteria 1 and 2). Although the results of this last experiment were inconclusive, as the training failed due to limitations of the setup and training regime (addressed in section 7.3), the results obtained in these studies contribute substantially to our current understanding of vocal learning in *Phyllostomus discolor*. Moreover, my findings demonstrated that *P. discolor* has volitional control over call parameters, which they can generally modify independently from each other (chapters 5 and 6). Call duration increased by about 13.5 ms within a call type and call repetition rates increased by as much as 5.84 calls per minute. The maximal change in spectral centroid was on the order of 200 Hz and the strongest change in mean fundamental frequency was an increase of 450 Hz. Importantly, these changes were observed in response to a training paradigm which operated in the absence of an acoustic target (chapter 5).

Comparing the VPL evidence for the four bat species discussed above (**Table 7.1**), highlights again how diverse the experimental approaches and measured parameters in VPL studies are (even within taxa). *Phyllostomus discolor* is one of the most in-depth studied vocal learning bat species. Knowledge about their acoustic perception and production, their vocal plasticity and their capacity for vocal learning has been and is still accumulating. In the light of past and present findings, the vocal learning capacity of this species is indisputable. However, where exactly this species falls on the vocal learning continuum is a persistent question. Positioning *P. discolor* along the vocal learning gradient is not only dependent on the capacity of this specific species, but also on the capacity of any species it is contrasted with. This direct comparison is hindered by the selective measuring and reporting of specific acoustic parameters. A simple example for this can be seen in the assessment of level of control over call duration in bats. While pulse duration in *Phyllostomus hastatus* was shown to naturally vary between populations by as much as 150 ms (observational study, Boughman, 1997), *P. discolor* volitionally changes the mean call duration by 13.5 ms (operant conditioning paradigm, chapter 5). However, even though *P. discolor* pups were shown to adjust their vocalisations towards their maternal directive calls, a variation of up to 30 ms between pup and mother calls was shown (observational study, Esser and Schmidt, 1989). While call duration is an easily modifiable and measurable call parameter, the reported vocal control over this parameter varies widely even within bats. This example highlights the need for comprehensive and systematic testing of VPL in order to eventually be able to sort different species along the vocal learning continuum described in chapter 2.

7.3 Limitations of the current experimental design

The setup and general training regime used in chapters 5 and 6 are the central experimental pieces of this thesis. The setup was adopted from successful studies with songbirds (Manabe and Dooling, 1997; Manabe *et al.*, 2008) and modified for the use with bats. An essential requirement for the success of this setup was the achievement of vocal conditioning of isolated bats, which had not previously been achieved. Specifically, adult bats had not been trained to use non-aggressive social vocalisations in order to trigger a food reward in the past. However, the level of vocal conditioning achieved with this setup was limited in comparison to similar studies in songbirds or non-human primates, as only the elicited emission but not the withholding of vocalisations was shown (e.g. Pierce, 1985; Adret, 1993; Stansbury and Janik, 2019). Although signal-triggered or withheld vocalisations are stronger demonstrations of control over vocalisations, this was an unnecessary level of control for the studies presented here (chapters 5 and 6). In light of the observed level of vocal control demonstrated by the bats, I hypothesise that the bats would indeed be able to show also this higher level of control, but the training period would probably be substantially prolonged.

Another difference between the presented training paradigm and those previously used is the duration the animals remain within the setup. Bird studies investigating the effect of acoustic targets on vocal development frequently keep the animals in the experimental setup for the full duration of the experiment; often several months (e.g. Tchernichovski *et al.*, 2004). In the vocal conditioning experiments presented here, the bats were trained for 3-4 hours per day and spent the rest of the day in company of their conspecifics in the colony rooms (chapters 5 and 6). Longer training durations as well as total social deprivation might very well result in stronger effect sizes and the effect sizes reported in chapters 5 and 6 might thus be an underestimation of their full potential. However, *P. discolor* is a very gregarious species and social isolation has a substantial negative effect on the health, motivation, and general condition of such social animals (Hall, 1998). Therefore, although my results might indeed underestimate the possible effect sizes, considering the well-being of the animals, a full-time training regime is not advisable in this species. A possible option to still achieve stronger effect sizes would be a general prolongation of the training or continuous training for weeks or months in groups, as has been done in *Rousettus aegyptiacus* (Prat *et al.*, 2015, 2017; Genzel *et al.*, 2019). It is not uncommon for behavioural studies to demand training periods of several months to years, but a prolongation of the presented experiments was not possible due to time restrictions within the framework of a doctoral thesis. The presented results should be seen as indications for achievable vocal control and effect sizes in *P. discolor*, but future long-term studies might result in even stronger evidence.

A third consideration, concerning the limitations of the setup/training is the effectiveness of acoustic targets vs. physical tutors. Behavioural training aimed at vocal learning has reportedly been more successful and effective when a physical tutor was present (Chen *et al.*, 2016). Songbird studies have shown that visual stimulation (video of a tutor bird), was more successful than training with an acoustic target alone, but the results were still inferior to the presence of a physical tutor (Benichov *et al.*, 2015; Halfwerk *et al.*, 2019). Using visual stimulation alongside acoustic stimulation is difficult when studying *P. discolor* as these bats are less visually oriented than songbirds. Using screens for visual stimulation is not effective with these bats and difficult to employ as screens often emit ultrasonic noises (Genkin *et al.*, 2018), which might disturb the bats. The most promising training method, using a live conspecific tutor, has two major disadvantages: (i) the behavioural state of the experimental animal will be dependent on the social interaction (e.g. a competitive male tutor might have a different effect than a receptive female tutor) and (ii) the experimental control over the tutored vocalisations is reduced. As vocal learning in the pale spear-nosed bat has not been established by a large body of research, I focused on an extremely controlled paradigm, in which I had absolute control over the acoustic stimulation the bats received during the training sessions. Furthermore, this setup and training regime allowed the assessment of individual training trajectories, which are difficult to investigate if two or more individuals are recorded together. Moreover, the experimental design used in chapters 5 and 6 of this thesis allowed me to disregard changes in behavioural states of the animals due to social interactions.

Aside from concerns about the setup and training paradigm, there are also limitations concerning the choice of experimental animals. All bats used in the presented studies were adults. The trained bats in chapters 5 and 6 were exclusively adult males. When designing my experiments, it was always an option to deviate from the used experimental animals and instead conduct my research with younger and/or female individuals. As the experiments showed positive results, the need to change experimental animals did not arise. Nevertheless, it might be that the vocal learning trait is more strongly expressed in younger bats, as they might have a stronger need to, for example, express their affiliation to a social group, which has been suggested as a potential driver for vocal learning (Nottebohm and Liu, 2010; Nowicki and Searcy, 2014). Furthermore, it is conceivable that vocal learning has a sex-specific expression, as the need for vocal learning might be dependent on sexual selection pressures (Jarvis, 2006; Nowicki and Searcy, 2014). Therefore, it would be valuable to repeat these experiments with a broader demographic sampling, including younger and female bats. This would enable us to get a more general understanding of the vocal learning ability in these bats and test for potential age- and sex-specific differences.

Not only sex and age of the experimental animals might have an impact on their vocal learning performance. As touched upon in the discussion of chapter 6, previous training experience might also affect the outcome of the study. The same animals were trained and recorded in chapters 5 and 6. Moreover, three of them had been trained previously in other psychophysical experiments. Although it is important to consider the effect of previous experiences of the animals used in behavioural studies, these previous experiments were concerned with the recording of echolocation calls and did not require any vocal training or vocal control. However, training the same animals in two subsequent vocal learning paradigms could potentially lead to an effect of overtraining, thus reducing their capacity to respond plastically to a new task (as discussed in chapter 6). The alternative, using naïve experimental animals, generally presents a trade-off: preventing effects from overtraining on the one hand, but possibly adding months of initial training and habituation time on the other.

Concluding, the setup, training regime, and choice of experimental animals used in chapters 5 and 6 of this thesis were optimal to provide evidence for vocal control, vocal usage learning, and even vocal production learning in *P. discolor* under controlled laboratory conditions. A continuation of these experiments with younger, female, and naïve experimental animals, prolonged experimental durations, possibly stricter training paradigms, and the use of a conspecific or even heterospecific tutor could indeed lead to even stronger effect sizes and further evidence bridging the gaps still persisting in our knowledge about vocal learning in the pale spear-nosed bat.

7.4 Conclusions and outlook

The prevailing motivation for animal vocal learning research is that its efforts will ultimately help us to understand the evolution, biological basis, and functional background of the human capacity for speech. The research collected in this thesis has helped to strengthen the foundation for research on the vocal learning capacity of bats in general and in particular of the pale spear-nosed bat, *Phyllostomus discolor*. To lay the foundation for behavioural trainings, I have investigated bat hearing and shown general principles of sound perception in a variety of bat species (chapter 3). Knowledge about their audiograms gives us a better understanding of their signal perception and enables us to design species-specific acoustic targets for the behavioural investigation of vocal learning. Furthermore, I presented the vocal repertoire of the pale spear-nosed bat (chapter 4). Understanding the vocal complexity of investigated bat species allows the detection of novel vocalisations and vocalisation modifications during behavioural assessment of vocal learning in this species. I further studied the vocal plasticity of these bats by first demonstrating their capacity for volitional call control and usage learning (chapter 5) and subsequently their ability to change the pitch of their vocalisations in a directional manner (chapter 6). The research presented here thus supports *P. discolor* as a model system for the study of vocal learning and presents a solid foundation for future in-depth investigation of their vocal learning capacity.

Aside from the direct evidence for vocal learning in *P. discolor*, the training paradigms established in chapters 5 and 6 will be useful for future vocal learning research in bats. Future work will build on these paradigms to understand the boundaries of vocal learning behaviour in bats. They can also be used as a basis for investigating the neuronal or genetic underpinnings of vocal learning behaviours. For example, these paradigms would allow the monitoring of neuronal circuit activation during vocal learning, or the behavioural consequences of genetic manipulations, such as knockdown of language-relevant genes. As such, the research on bat vocal learning behaviour presented in this thesis enables follow-up studies, not only for increasing our understanding of vocal learning behaviour, but also its underlying biology. In this way, it brings us one step closer to understanding vocal learning and the human capacity for speech and language.

References

Adret, P. (1993). Vocal learning induced with operant techniques: an overview. *Netherlands Journal of Zoology* 43, 125–142.

Benichov, J.I.; Benezra, S.E.; Vallentin, D.; Globerson, E.; Long, M.A. and Tchernichovski, O. (2015). The forebrain song system mediates predictive call timing in female and male zebra finches. *Current Biology* 26, 1–10.

Boughman, J.W. (1998). Vocal learning by greater spear-nosed bats. *Proceedings of The Royal Society B* 265, 227–233.

Boughman, J.W. and Moss, C.F. (2003). Social sounds: vocal learning and development of mammal and bird calls. In: A. M. Simmons, R. R. Fay, and A. N. Popper (Eds), *Acoustic Communication - Springer Handbook of Auditory Research*. Springer-Verlag New York, Inc., New York, pp. 138–224.

Boughman, J.W. and Wilkinson, G.S. (1998). Greater spear-nosed bats discriminate group mates by vocalizations. *Animal Behaviour* 55, 1717–1732.

Boughman, J.W. (1997). Greater spear-nosed bats give group-distinctive calls. *Behavioral Ecology and Sociobiology* 40, 61–70.

Chen, Y.; Matheson, L.E. and Sakata, J.T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proceedings of the National Academy of Sciences of the United States of America* 113, 6641–6646.

Esser, K.H. (1994). Audio-vocal learning in a non-human mammal: the lesser spear-nosed bat *Phyllostomus discolor*. *NeuroReport* 5, 1718–1720.

Esser, K.H. and Schmidt, U. (1989). Mother-infant communication in the lesser spear-nosed bat *Phyllostomus discolor* (Chiroptera, Phyllostomidae) - evidence for acoustic learning. *Ethology* 82, 156–168.

Fitch, W.T.; Neubauer, J. and Herzel, H. (2002). Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour* 63, 407–418.

Genkin, D.; Pattani, M.; Schuster, R. and Tromer, E. (2018). Synesthesia: detecting screen content via remote acoustic side channels. In: *IEEE Symposium on Security & Privacy.* , pp. 1–31.

Genome 10K Community of Scientists (2009). Genome 10K: A Proposal to Obtain Whole-Genome Sequence for 10 000 Vertebrate Species. *Journal of Heredity* 100, 659–674.

Genzel, D.; Desai, J.; Paras, E. and Yartsev, M.M. (2019). Long-term and persistent vocal plasticity in adult bats. *Nature Communications* 10, 3372.

Halfwerk, W.; Varkevisser, J.; Simon, R.; Mendoza, E.; Scharff, C. and Riebel, K. (2019). Toward testing for multimodal perception of mating signals. *Frontiers in Ecology and Evolution* 7, 2013–2019.

Hall, F.S. (1998). Social deprivation of neonatal, adolescent, and adult rats has distinct neurochemical and behavioral consequences. *Critical Review in Neurobiology* 12, 129–162.

Janik, V.M. (2014). Cetacean vocal learning and communication. *Current Opinion in Neurobiology* 28, 60–65.

Janik, V.M. and Slater, P.J.B. (1997). Vocal learning in mammals. *Advances in the Study of Behaviour* 26, 59–99.

Janik, V.M. and Slater, P.J.B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour* 60, 1–11.

Jarvis, E.D. (2006). Selection for and against vocal learning in birds and mammals. *Ornithological Science* 5, 5–14.

Jones, G. and Ransome, R.D. (1993). Echolocation calls of bats are influenced by maternal effects and change over a lifetime. *Proceedings of the Royal Society B* 252, 125–128.

Kershenbaum, A.; Freeberg, T.M. and Gammon, D.E. (2015). Estimating vocal repertoire size is like collecting coupons: a theoretical framework with heterogeneity in signal abundance. *Journal of Theoretical Biology* 373, 1–11.

Knörnschild, M. (2014). Vocal production learning in bats. *Current Opinion in Neurobiology* 28, 80–85.

Long, G.R. and Schnitzler, H.-U. (1975). Behavioural audiograms from the bat, *Rhinolophus ferrumequinum*. *Journal of Comparative Physiology A* 100, 211–219.

Manabe, K. and Dooling, R.J. (1997). Control of vocal production in budgerigars (*Melopsittacus undulatus*): selective reinforcement, call differentiation, and stimulus control. *Behavioural Processes* 41, 117–132.

Manabe, K.; Dooling, R.J. and Brittan-Powell, E.F. (2008). Vocal learning in budgerigars (*Melopsittacus undulatus*): effects of an acoustic reference on vocal matching. *The Journal of the Acoustical Society of America* 123, 1729–1736.

McCracken, G.F. and Bradbury, J.W. (1977). Paternity and genetic heterogeneity in the polygynous bat, *Phyllostomus hastatus*. *Science* 198, 303–306.

McCracken, G.F. and Bradbury, J.W. (1981). Social organization and kinship in the polygynous bat *Phyllostomus hastatus*. *Behavioral Ecology and Sociobiology* 8, 11–34.

Nottebohm, F. and Liu, W.C. (2010). The origins of vocal learning: new sounds, new circuits, new cells. *Brain and Language* 115, 3–17.

Nowicki, S. and Searcy, W.A. (2014). The evolution of vocal learning. *Current Opinion in Neurobiology* 28, 48–53.

Petkov, C.I. and Jarvis, E.D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience* 4, 1–24.

Pierce, J.D. (1985). A review of attempts to condition operantly alloprimate vocalizations. *Primates* 26, 202–213.

Prat, Y. (2019). Animals have no language, and humans are animals too. *Perspectives on Psychological Science*, 174569161985840.

Prat, Y.; Azoulay, L.; Dor, R. and Yovel, Y. (2017). Crowd vocal learning induces vocal dialects in bats: playback of conspecifics shapes fundamental frequency usage by pups. *PLoS Biology*, 1–14.

Prat, Y.; Taub, M. and Yovel, Y. (2015). Vocal learning in a social mammal: demonstrated by isolation and playback experiments in bats. *Science Advances* 1, e1500019.

Stansbury, A.L. and Janik, V.M. (2019). Formant modification through vocal production learning in gray seals. *Current Biology*, 1–6.

Tchernichovski, O.; Lints, T.J.; Derégnaucourt, S.; Cimenser, A. and Mitra, P.P. (2004). Studying the song development process: rationale and methods. *Annals of the New York Academy of Sciences* 1016, 348–363.

Teeling, E.C.; Vernes, S.C.; Dávalos, L.M.; Ray, D.A.; Gilbert, M.T.P. and Myers, E. (2018). Bat biology, genomes, and the Bat1K project: to generate chromosome-level genomes for all living bat species. *Annual Review of Animal Biosciences* 6, 23–46.

Tyack, P.L. (2016). Vocal learning and auditory-vocal feedback. In: R.A. Suthers, W.T. Fitch, R.R. Fay, and A.N. Popper (Eds), *Vertebrate Sound Production and Acoustic Communication*. Springer Handbook of Auditory Research 53, pp. 261–296.

Vernes, S.C. (2017). What bats have to say about speech and language. *Psychonomic Bulletin & Review* 24, 111–117.

Vernes, S.C. and Wilkinson, G.S. (2019). Behaviour, biology, and evolution of vocal learning in bats. *Philosophical Transactions of the Royal Society of London, Series B*, 1–31.

Appendix

English Summary

Nederlandse Samenvatting

Research Data Management

Acknowledgements

Curriculum Vitae

Publication List

MPI Series

English Summary

The complexity of human spoken language in structure and expression is unmatched in the animal kingdom. However, several animal species possess the capacity for vocal production learning (in the following simply referred to as 'vocal learning'), which is a requirement for the acquisition and production of human speech. Vocal learning is defined as the ability to imitate new sounds or modify existing signals from a pre-existing repertoire according to auditory input. Even though this may seem simple enough, this is a complex trait that requires the functional interplay of precise auditory perception, neuronal integration of the signal (including template memorisation and matching), and accurate neuronal motor control of the vocal production systems (i.e. lungs, larynx, and vocal tract). Invasive experiments, such as *in vivo* gene modification or electrophysiology, are crucial to gain an in-depth understanding of the evolution and the biological basis of human speech. As invasive experiments are impossible in humans, animal models are of great importance for the experimental approach to answering questions about the origin and functional basis of human language and speech acquisition. Vocal learning as a behavioural trait can be found in different levels of expression within the animal kingdom and should be regarded as a continuum.

Several vertebrate species have shown indications for vocal learning in the past. These include three groups of birds (songbirds, hummingbirds, and parrots) and several mammalian taxa (cetaceans, pinnipeds, elephants, and bats). In the past decades, songbirds, which occupy a position toward the upper end of the vocal learning continuum, have been studied extensively and this research has yielded invaluable insights in the physiological prerequisites, neurobiology, and genetics required for vocal learning. While research on avian model organisms dominates the field, mammalian vocal learning by comparison has largely been neglected in the past. Comparing the available sample sizes, ease of study, and capabilities different mammalian vocal learners provide, bats present themselves as a useful candidate model system and several bat species have shown indications of vocal learning in the past. However, controlled, repeatable experiments with isolated adult bats have thus far been lacking.

With my thesis, I aim to expand the existing knowledge about bat vocal learning, with a specific focus on the vocal learning capacity of the pale spear-nosed bat, *Phyllostomus discolor*. I start my thesis by introducing a theoretical framework that is aimed at strategically combining findings from various species to support cross-species approaches to the study of vocal learning (**chapter 2**). Incorporating previously suggested definitions, I propose a multi-layered study of related fields of research, which in combination provide an in-depth understanding of a species' vocal learning capacity and its placement along

the vocal learning continuum. The proposed framework is intended as a guideline for the design of future vocal learning studies through the identification of gaps in existing knowledge. Furthermore, the framework should function as a scaffold that aids researchers to place their findings into context and thus to facilitate the organisation of findings within the field. Integrating results from many species by applying this framework will allow the determination of common factors in the evolution of vocal learning across different species and thus highlight shared evolutionary pressures acting to support the emergence of this trait. I then go on to describe the perceptional basis for the investigation of vocal learning in bats (**chapter 3**). Knowledge about the principles of sound perception in different species affords us insights in the way these organisms hear the world. It is an important prerequisite for understanding the ability to react to or imitate acoustic signals and thus the meaningful design of acoustic stimuli in experimental studies. If the correct perception of a target sound is hindered (for example due to low temporal or spectral resolution of the animals' hearing), the attempted imitation of this acoustic target will always seem flawed. Being informed about the acoustic perceptional abilities of a species is thus important for the evaluation of its imitation and learning success. I here present a large-scale comparative study, in which I describe the audiograms and characteristics of loudness encoding for eleven Neotropical bat species. I found that despite species-specific differences in audiogram shape, a general principle for loudness coding was present in all measured bat species, which may extend to other mammals.

In the main part of my thesis, I address three topics specifically concerning the vocal learning capacity of *P. discolor*. First, I demonstrate the acoustical and behavioural complexity of the vocal emissions of *P. discolor* by describing their vocal repertoire and the behavioural contexts in which the identified vocalisation classes are emitted (**chapter 4**). Establishing a vocal repertoire is extremely valuable for the assessment of vocal learning, as the comparison between the normally emitted vocalisations (baseline) and newly acquired vocalisations (learned) is often required for clear demonstration of vocal learning. Knowledge about the natural range of a species' vocal production is thus an important criterion for the study of vocal learning. Second, I show that *P. discolor* vocalisations can be brought under operant control and demonstrate their capacity for vocal usage learning in the absence of any social feedback, meaning that they can use specific vocalisations in new behavioural contexts (**chapter 5**). For this demonstration of vocal control and usage learning, I set up a behavioural paradigm, in which isolated bats used social vocalisations in order to elicit food rewards. This is the first demonstration of such a behavioural paradigm in a bat model. Furthermore, I also show that these bats can modify their calls in response to challenges imposed on them by the paradigm, in consequence providing evidence for their control over call duration and amplitude and indications for volitional spectral changes. Third, I focus on the demonstration of directional change of spectral call parameters according to an auditory target (**chapter 6**). The modification of spectral call parameters is widely considered to be more challenging than the modification of temporal parameters or call amplitude, as it requires the modification of the phonatory and/or filter system. By modifying the setup and training regime previously established, I was able to show that the bats decreased the fundamental frequency of their calls significantly according to the presented auditory target. This automated setup and conditioning training paradigm developed in my thesis allows for the spectro-temporal tracking of call parameters and their changes over time. This approach enables the detailed study of vocal parameters and behaviours on an individual level. Future work will build on this setup and paradigm to better understand the boundaries of vocal learning behaviour in bats. For example, this setup and paradigm would allow the monitoring of neuronal circuit activation during vocal learning, or the behavioural consequences of genetic manipulations, such as knockdown of language-relevant genes. Such follow-up work will further deepen our understanding of the vocal learning trait and its underlying biology.

In this thesis, I highlight the value of bats as a mammalian model system for the study of vocal learning and provide experimental evidence for the vocal learning capacity of the pale spear-nosed bat. My research presents a solid foundation for future in-depth investigation of bat vocal learning and will ultimately bring us one step closer to understanding the human capacity for speech and language acquisition.

Parts adapted from: **Lattenkamp, E.Z.**, Vernes, S.C. and Wiegrebe, L. (2018). Mammalian models for the study of vocal learning: a new paradigm in bats. In: C. Cuskley, M. Flaherty, H. Little, L. McCrohon, A. Ravignani, and T. Verhoef (Eds). *The Evolution of Language: Proceedings of the 12th International Conference.* pp. 235–237. doi:10.12775/3991-1.056

Nederlandse Samenvatting

De complexiteit van door de mens gesproken taal is ongeëvenaard in het dierenrijk. Verscheidene diersoorten bezitten echter het vermogen om nieuwe geluiden te leren produceren. Deze eigenschap is een vereiste voor het verwerven en produceren van menselijke spraak en wordt ook wel aangeduid als vocaleproductieleren of vocaal leren. Vocaal leren wordt gedefinieerd als het vermogen om nieuwe geluiden of bestaande signalen van een reeds bestaand repertoire te wijzigen op basis van auditieve input. Hoewel dit eenvoudig mag lijken, is dit een complex systeem dat het samenspel vereist van precieze auditieve perceptie, neuronale integratie van het signaal (inclusief het memoriseren en matchen van een template) en nauwkeurige neuronale motorische controle van het vocale productiesysteem (d.w.z. longen, strottenhoofd en stembanden). Invasieve experimenten, zoals *in vivo* genmodificatie of elektrofysiologie, zijn cruciaal voor een diepgaand inzicht in de evolutie en de biologische basis van menselijke spraak. De experimentele benadering om vragen te beantwoorden over de oorsprong en functionele basis van menselijke taal en spraakverwerving. Vocaal leren als een gedragskenmerk komt in verschillende mate tot expressie binnen het dierenrijk en moet derhalve worden beschouwd als een continuüm.

Verschillende gewervelde soorten hebben in het verleden aanwijzingen gegeven dat zij in staat zijn tot vocaal leren. Deze omvatten drie groepen vogels (zangvogels, kolibries en papegaaien) en meerdere taxa van zoogdieren (walvisachtigen, pinnipeds, olifanten en vleermuizen). In de afgelopen decennia zijn zangvogels, die een positie innemen aan de bovenkant van het continuüm voor vocaal leren, uitgebreid bestudeerd en geëvalueerd met betrekking tot de fysiologische vereisten, neurobiologie en genetica die nodig zijn voor vocaal leren. Terwijl onderzoek naar ornithologische modelorganismen het veld domineert, is vocaal leren bij zoogdieren in het verleden grotendeels verwaarloosd. De beschikbare steekproefgrootte, bestudeerbaarheid en capaciteiten van verschillende soorten in overweging genomen, bieden vleermuizen een geschikte kandidaat als modelsysteem. Meerdere vleermuissoorten hebben tekenen van vocaal leren laten zien. Gecontroleerde, herhaalbare experimenten met geïsoleerde volwassen vleermuizen ontbreken echter tot dusver.

Met mijn proefschrift wil ik de bestaande kennis over het vocaal leren van vleermuizen uitbreiden, met een specifieke focus op het vocale leervermogen van de bonte lansneusvleermuis, *Phyllostomus discolor*. Ik begin mijn scriptie door een theoretisch kader te introduceren met als doel de bevindingen van meerdere diersoorten op strategische wijze te controleren om de studie van vocaal leren over meerdere diersoorten te ondersteunen (**hoofdstuk 2**). Gebruikmakend van eerder voorgestelde definities, stel ik een meerlagige bestudering voor van verwante onderzoeksgebieden, die samen een diepgaand inzicht verschaffen in het vocale leervermogen van een soort en de plaatsing ervan in het vocale leercontinuüm. Het voorgestelde kader is bedoeld als leidraad voor het ontwerp van toekomstige studies van vocaal leren doordat het gaten in onze huidige kennis aanwijst. Bovendien moet het raamwerk als een kapstok fungeren die onderzoekers in staat stelt om hun bevindingen in context te plaatsen en die daardoor helpt om bevindingen in het onderzoeksveld te organiseren. Door resultaten van vele soorten volgens dit raamwerk te integreren zal het mogelijk zijn om gemeenschappelijke factoren in de evolutie van vocaal leren aan te wijzen tussen verschillende diersoorten en zo de evolutionaire krachten die het ontstaan van deze eigenschap ondersteunen zichtbaar te maken. Hierna ga ik verder om de perceptieve basis te beschrijven voor de studie van vocaal leren in vleermuizen (**hoofdstuk 3**). Kennis over de principes van geluidsbeleving in verschillende soorten geeft ons inzicht in de wijze waarop deze organismen de wereld horen. Dit is een belangrijke voorwaarde om te begrijpen op welke akoestische signalen dieren kunnen reageren en om dus tot een zinvol ontwerp van akoestische stimuli in experimentele studies te kunnen komen. Als de correcte waarneming van een voorbeeldgeluid wordt verhinderd (bijvoorbeeld door lage temporele of spectrale resolutie van het hoorvermogen van een dier), dan zal iedere imitatiepoging voor de waarnemer verkeerd klinken.. Informatie over de akoestische waarnemingsmogelijkheden van een soort is dus belangrijk om zijn imitatie en leersucces te kunnen evalueren. Ik presenteer hier een grootschalig vergelijkend onderzoek, waarin ik de audiogrammen en de eigenschappen voor de codering van luidheid beschrijf in elf neotropische vleermuissoorten. Ik heb ontdekt dat ondanks soortspecifieke verschillen in de vorm van het audiogram, alle soorten een algemeen principe vertonen voor de codering van luidheid dat mogelijk zou kunnen generaliseren naar andere zoogdieren.

In het hoofddeel van mijn proefschrift bespreek ik drie onderwerpen met betrekking tot het vocale leervermogen van *P. discolor*. Ten eerste laat ik de akoestische en ethologische complexiteit zien van de vocale emissies van *P. discolor* door zijn vocale repertoire en de gedragscontexten waarin de geïdentificeerde vocalisatieklassen worden uitgezonden te beschrijven (**hoofdstuk 4**). Het opstellen van een vocaal repertoire is uiterst waardevol voor de beoordeling van vocaal leren, omdat de vergelijking tussen de normaal uitgezonden vocalisaties (referentie) en nieuw verworven vocalisaties (geleerd) vaak vereist is voor het aantonen van vocaal leren. Kennis over het natuurlijke bereik van de vocale productie van een soort is dus een belangrijk criterium voor de studie van vocaal leren. Ten tweede laat ik zien dat vocalisaties van *P. discolor* onder operante controle kunnen worden gebracht en toon ik hun vermogen tot vocaal gebruiksleren in de afwezigheid van enige sociale feedback, wat betekent dat ze specifieke vocalisaties kunnen gebruiken in nieuwe gedragscontexten (**hoofdstuk 5**). Voor deze demonstratie van vocale controle en gebruiksleren heb ik een gedragsexperiment opgezet waarin geïsoleerde vleermuizen sociale vocalisaties moeten gebruiken om voedselbeloningen te verkrijgen. Dit is de eerste keer dat een dergelijk paradigma wordt toepast op een vleermuismodel. Hiermee laat ik zien dat deze vleermuizen hun roep kunnen veranderen in reactie op eisen die het experiment stelt aan hun vocalisaties voor een beloning. De resultaten bewijzen dat ze controle hebben over de duur en amplitude van hun vocalisatie en geven een indicatie van vrijwillige aanpassingen van hun vocalisatiespectrum. Ten derde ga ik in op directionele verandering van spectrale vocalisatieparameters naar een gehoord voorbeeld (**hoofdstuk 6**). De modificatie van spectrale vocalisatieparameters wordt vaak beschouwd als een uitdagendere taak dan de modificatie van temporele parameters of vocalisatieamplitude, aangezien het vleermuizen vereist hun fonetische en / of filtersysteem aan te passen. Door de opzet en het trainingregime uit het vorige hoofdstuk aan te passen was ik in staat om te laten zien dat de vleermuizen de fundamentele frequentie van hun oproepen verlaagden volgens het gehoorde voorbeeld. Deze geautomatiseerde opzet en het conditionerende trainingsparadigma die ik heb ontwikkeld in mijn proefschrift maken het mogelijk om spectro-temporele vocalisatieparameters te meten inclusief hun veranderingen over tijd. Deze methode maakt een gedetailleerde studie van vocale parameters en gedrag op individueel niveau mogelijk. Toekomstig werk zal voortbouwen op deze opstelling en dit paradigma om de grenzen van vocaal leergedrag in vleermuizen beter te begrijpen. Deze opstelling en dit paradigma zouden het bijvoorbeeld mogelijk maken om neuronale circuitactivering tijdens vocaal leren te monitoren, of om de effecten op gedrag van genetische manipulaties te onderzoeken, zoals knockdown van taalrelevante genen. Dergelijk vervolgonderzoek zal ons begrip van vocaal leren en de onderliggende biologie ervan verder verdiepen.

In dit proefschrift onderstreep ik de waarde van vleermuizen als een zoogdiermodel voor het bestuderen van vocaal leren en geef ik experimenteel bewijs voor het vocale leervermogen van de bonte lansneusvleermuis. Mijn onderzoek vormt een solide basis voor toekomstig diepgaand onderzoek naar vocaal leren door vleermuizen en zal ons een stap dichter brengen bij het begrijpen van de menselijke capaciteit voor spraak- en taalverwerving.

Aangepast van: **Lattenkamp, E.Z.**, Vernes, S.C. and Wiegrebe, L. (2018). Mammalian models for the study of vocal learning: a new paradigm in bats. In: C. Cuskley, M. Flaherty, H. Little, L. McCrohon, A. Ravignani, and T. Verhoef (Eds). *The Evolution of Language: Proceedings of the 12th International Conference.* pp. 235–237. doi:10.12775/3991-1.056

Research Data Management

This thesis research has been carried out under the institute research data management policy of the Max Planck Institute for Psycholinguistics (Nijmegen, The Netherlands). This policy can be found here: https://www.mpi.nl/page/research-data-management (date of last access: 23rd January 2020).

Data sets and code were produced for chapters 3-6. For chapters 1, 2, and 7 no data or code was produced. Chapter 3 was published under the DOI 10.3389/fevo.2019.00116. The data sets and codes produced in this project are permanently stored at the lrz data repository of the Ludwig Maximilian University Munich (Germany) and can be accessed upon request from Dr. Sven Schörnich. The data set and codes for chapter 4 are currently deposited with a collaborator (Prof. Mirjam Knörnschild) at the Humboldt University to Berlin (Germany). The corresponding manuscript is in preparation and the data set and codes will be made publicly available in the publishing process. The data set and codes for chapters 5 and 6 are accessible via the online G-Node Infrastructure repository: Chapter 5: DOI 10.12751/g-node.984192 (https://doid.gin.g-node.org/10.12751/g-node.984192). Chapter 6: DOI 10.12751/g-node.d9dd71 (https://doi.org/10.12751/g-node.d9dd71)

Acknowledgements

First and foremost, I want to thank my supervisors **Sonja C. Vernes** and **Lutz Wiegrebe** for four years of close supervision and all the time they invested in me. Despite your relatively different supervision styles the two of you make a really good team and I learned a lot from you. Especially in the last year, I was very grateful for your support and motivation. Thank you for making my work and this PhD possible! I also want to thank my external advisor **Mirjam Knörnschild** for her emotional and scientific support. Your feedback and opinions helped me a lot and made me rethink important decisions. You guided my steps in a personal, supportive way, which I very much appreciated. For providing me with the infrastructure and opportunity to conduct my PhD under such excellent circumstances, I want to thank my promotor **Simon Fisher**, director of the Language and Genetics Department of the MPI for Psycholinguistics, and **Benedikt Grothe**, chair of the Division of Neurobiology of the LMU. I also want to thank the Company of Biologists, the German Academic Exchange Service, and the Radboud University for funding some of my conference participations and my research stay in Panama. So many experiences would not have been had without this support. I want to thank my thesis reading committee, **Stephen C. Levinson**, **Carel ten Cate**, and **Susanne Hoffmann** for their time and consideration.

I want to thank my colleagues from both of my working groups. The **members of the NVC group**, who shared their interesting research with me every week in our lab meetings. Especially, I want to thank **Paolo, Ine, Ksenia, Kai,** and **Midas**. Paolo, who seems to stand like a rock in the rough waves academia presents and helps everyone who passes his way. Ine, with whom I spent some exciting weeks in Panama. I am so glad you never got annoyed at me for talking every night in my sleep. Ine and Ksenia took on the job of being my paranymphs, for which I am eternally grateful. Midas and Kai, you were my smart, kind, and funny anchors in Nijmegen and my visits would have been so much bleaker without you. I was lucky enough to have two working groups and my every-day working group at the LMU also deserves my thanks. **Kathi** and **Meike**, two clever, strong, fabulous women, who helped me find my way through the university jungle from the first minute on. **Leonie** I want to thank for a lot of things: for her inexhaustible energy to organise and help with practically everything, for her mental support, and for introducing me to **Theater Torschlusspanik**. **Michael**, thank you for being an endless source of tidbits of interesting information. I never know what I will know after meeting you. Thank you for being the good soul of the group and I hope we will spend some time cooking together in the future. Especially in the last months of my PhD, **Marga** was my pillar. You were there in the hard times and the good ones. Your laughter and our karaoke saved my days and my moods. Without you, these four years would have been nearly unbearable. Thank you from the bottom of my heart, my favourite Koala-fan. My thanks also go to **Lisa, Rebecca, Stephanie, Jassica, Meli,** and **Eva**, students I supervised during the last four years, who gave me a chance to share what I had learned. Supervising you was great fun and it meant a lot to me to see you succeed in your different paths. Although she was not my student but Leonie's, I want to thank **Kristin**, for her friendship, her trust throughout the years, and her ability to turn just about everything into a funny story. I want to thank the lunch group, my fellow PhD students. You sharing the same, but still individually different struggles provided me with urgently needed camaraderie and distraction. I want to thank especially **Alex, Delwen, Ela, Eli, Sara P., Dardo, Alisha, Simon**, and **Gregory**. A special thank goes to **Diana**, who ignores my constant forgetfulness about responding to messages and still hasn't given up on me. Thank you also for urging me to get involved with the **CNS social hour** people, who made the time at the LMU so much more fun. For introducing me to the **Boogiebären**, my second home in Munich, I want to thank **Sara H**. Several people from the Neurobiology department of the LMU also deserve special thanks: **Steffen**, for his interest in me and my work and showing me that academia can be personal and supportive; **Andrea** and **Misku**, for a never-ending supply of coffee and four open ears over lunch; **Sven**, for his sarcasm that helped to set the PhD problems into perspective; and **Pepe**, who regrettably decided to work with… birds. A few people facilitating my work in the last years deserve my thanks as well: **Sabrina**, the best bat caretaker in the world, **Florian** and **Stefan** from the LMU workshop for helping with anything setup-related and always being good for a laugh, and **Evelyn** and **Simone**, for being the most helpful secretaries one could hope for.

Some people crossed my path in academia that very much improved the whole experience. Thank you, **Holger** and **Markus,** for being good examples that sticking with your principles about good scientific practice is possible against all external pressures. Thank you for showing me that science can be good without selling your soul and also thank you for showing me that choosing this path is extremely difficult. I also want to thank my former flatmates, **Daniel** and **Theresa**. You were my gateway drug into the bat world, my colleagues, and now my close friends. Thank you for being there and being you and being so special. I hope we never lose touch.

Ich danke meinen Freunden, die ich in Berlin zurückgelassen habe, die aber immer für mich da sind, wenn ich sie brauche. Wenn meine Gedanken nach Berlin streifen, weiß ich, dass es immer weitergeht im Leben. Ich danke euch allen **Petra, Zara, Kim, Carol, Julia, Felix** und natürlich meiner liebsten **Meister-Familie**!

Gertraud und **Anna**, ihr beide habt einen erheblichen Anteil daran, dass ich den PhD nicht abgebrochen habe, obwohl ich oft genug darüber nachgedacht habe. Ihr habt mir äußere Umstände gewährt, die so entspannend, wohltuend und sicher waren, dass ich auch die anstrengendsten Zeiten überstehen konnte. Vielen Dank für alles was ihr mir ermöglicht habt und ich freue mich sehr auf die bevorstehende intensive Zeit mit euch. Was die Entspannung angeht, bin ich natürlich auch **Amelia** sehr dankbar.

Liebster **Mark**, ich kann gar nicht in Worte fassen, wie sehr ich dir für deine Hilfe, deine Zeit und deine Liebe danke. Ich weiß, ich hätte das alles irgendwie auch ohne dich geschafft, aber du machst jeden Tag zu einer Freude, jede Aufgabe machbar, und wenn ich denke ich kann nicht mehr, bist du da, um mich vom Gegenteil zu überzeugen. Ich danke dir von ganzem Herzen für jedes der vergangenen Jahre und freue mich auf jede Minute unserer gemeinsamen Zukunft. Mein Dank schließt auch die ganze Scherz Familie ein. Danke **Sandy** und **Heidi**, dass ihr mich aufgenommen habt und auch eure hervorragenden, liebevollen **Mays** mit mir teilt.

Inge und **Andi**, ich danke euch. Nicht im Besonderen für die letzten vier Jahre, sondern eher für die letzten 32. Aber im Grunde verdanke ich euch beiden alles.

Curriculum Vitae

Ella Zoe Lattenkamp was born in 1988 in Berlin, Germany. In 2013, she obtained her Bachelor's degree in Biology from the Humboldt University to Berlin (Germany). She received dual Master's degrees in Evolutionary Biology from the University of Uppsala (Sweden) and the Ludwig Maximilian University Munich (Germany) in 2015. In 2016, she started her PhD as a collaboration of the Neurogenetics of Vocal Communication group (Max Planck Institute for Psycholinguistics) and the Division of Neurobiology (Ludwig Maximilian University Munich).

Publication List

11.) **Lattenkamp, E.Z.**; Linnenschmidt, M., Mardus, E., Vernes, S.C., Wiegrebe, L. and Schutte, M. (accepted). Impact of auditory feedback on bat vocal development. *The Evolution of Language: Proceedings of the 13th International Conference*.

10.) **Lattenkamp, E.Z.***; Shields, S.M.*; Schutte, M.; Richter, J.; Linnenschmidt, M.; Vernes, S.C. and Wiegrebe, L. (2019). The vocal repertoire of pale spear-nosed bats in a social roosting context. *Frontiers in Ecology and Evolution* 7:16. doi:10.1016/j.cobeha.2018.04.007

9.) **Lattenkamp, E.Z.** and Vernes, S.C. (2018). Vocal learning: a language-relevant trait in need of a broad cross-species approach. *Current Opinion in Behavioral Sciences* 21:209–215. doi:10.1016/j.cobeha.2018.04.007

8.) **Lattenkamp, E.Z.**; Vernes, S.C. and Wiegrebe, L. (2018). Volitional control of social vocalisations and vocal usage learning in bats. *Journal of Experimental Biology* 211: jeb180729. doi:10.1242/jeb.180729

7.) **Lattenkamp, E.Z.**; Kaiser, S.; Kaučič, R.; Großmann, M.; Koselj, K. and Görlitz, H.R. (2018). Environmental acoustic cues guide the biosonar attention of a highly specialised echolocator. *Journal of Experimental Biology* 221: jeb165696. doi:10.1242/jeb.165696

6.) **Lattenkamp, E.Z.**; Vernes, S.C. and Wiegrebe, L. (2018). Mammalian models for the study of vocal learning: a new paradigm in bats. In Cuskley, C., Flaherty, M., Little, H., McCrohon, L., Ravignani, A. & Verhoef, T. (Eds.): *The Evolution of Language: Proceedings of the 12th International Conference*: 235–237. doi:10.12775/3991-1.056

5.) Teeling, E.C.; Vernes, S.C.; Dávalos, L.M.; Ray, D.A.; Gilbert, M.T.P.; Myers, E.; **Bat1K Consortium** (2018). Bat biology, genomes and the Bat1K proect: to generate chromosome-level genomes for all living bat species. *Annual Review of Animal Biosciences* 6: 23–46. doi:10.1146/annurev-animal-022516-022811

4.) Cortázar-Chinarro, M.; **Lattenkamp, E.Z.**; Luquet, E.; Meyer Lucht, Y.; Laurila, A. and Höglund, J. (2017). Drift, selection, or migration? Processes affecting genetic differentiation and variation along a latitudinal gradient in an amphibian. *BMC Evolutionary Biology* 17(189):1–14. doi:10.1186/s12862-017-1022-z

3.) **Lattenkamp, E.Z.**; Mandák, M. and Scherz, M.D. (2016). The advertisement call of *Stumpffia be* Köhler, Vences, D'Cruze & Glaw, 2010. *Zootaxa* 4205(5): 483–485. doi:10.11646/zootaxa.4205.5.7

2.) Niemela, P.T.; **Lattenkamp, E.Z.** and Dingemanse, N.J. (2015). Personality-related survival and sampling bias in wild cricket nymphs. *Behavioral Ecology* 26(3): 936–946. doi:10.1093/beheco/arv036

1.) Ludwig, A.; Vernesi, C.; Lieckfeldt, D.; **Lattenkamp, E.Z.**; Wiethölter, A. and Lutz, W. (2012). Origin and patterns of genetic diversity of German fallow deer as inferred from mitochondrial DNA. *European Journal of Wildlife Research* 58(2): 495–501. doi:10.1007/s10344-011-0571-5

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*
17. Moving eyes and naming objects. *Femke van der Meulen*

18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*

19. Morphology in speech comprehension. *Kerstin Mauth*

20. Morphological families in the mental lexicon. *Nivja H. de Jong*

21. Fixed expressions and the production of idioms. *Simone A. Sprenger*

22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*

23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*

24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*

25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*

26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*

27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*

28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*

29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*

30. A grammar of Jalonke argument structure. *Friederike Lüpke*

31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*

32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*

33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*

34. Phonetic and lexical processing in a second language. *Mirjam Broersma*

35. Retrieving semantic and syntactic word properties. *Oliver Müller*

36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*

37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*

38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*

39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*

40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*

41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*

42. The acquisition of auditory categories. *Martijn Goudbeek*

43. Affix reduction in spoken Dutch. *Mark Pluymaekers*

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*

45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*

46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*

47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*

48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*

49. Phoneme inventories and patterns of speech sound perception. *Anita Wagner*

50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*

51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*

52. Prosodic structure in speech production and perception. *Claudia Kuzla*

53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*

54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*

55. Processing the fine temporal structure of spoken words. *Eva Reinisch*

56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*

57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*

58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*

59. Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*

60. Processing casual speech in native and non-native language. *Annelie Tuinman*

61. *Split intransitivity in Rotokas, a Papuan language of Bougainville. Stuart Robinson*

62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*

63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*

64. The meaning and use of ideophones in Siwu. *Mark Dingemanse*

65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*

66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*

67. The relevance of early word recognition: Insights from the infant brain. *Caroline Junge*

68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias J. Sjerps*

69. Structuring language. Contributions to the neurocognition of syntax. *Katrien R. Segaert*

70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. *Birgit Knudsen*

71. Gaze behavior in face-to-face interaction. *Federico Rossano*

72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space. *Conny de Vos*

73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. *Attila Andics*

74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*

75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*

76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*

77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*

78. Contributions of executive control to individual differences in word production. *Zeshu Shao*

79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*

80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*

81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*

82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*

83. The many ways listeners adapt to reductions in casual speech. *Katja Poellmann*

84. Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners. *Giuseppina Turco*

85. Morphological processing in younger and older people: Evidence for flexible dual-route access. *Jana Reifegerste*

86. Semantic and syntactic constraints on the production of subject-verb agreement. *Alma Veenstra*

87. The acquisition of morphophonological alternations across languages. *Helen Buckler*

88. The evolutionary dynamics of motion event encoding. *Annemarie Verkerk*

89. Rediscovering a forgotten language. *Jiyoun Choi*

90. The road to native listening: Language-general perception, language-specific input. *Sho Tsuji*

91. Infants'understanding of communication as participants and observers. *Gudmundur Bjarki Thorgrímsson*

92. Information structure in Avatime. *Saskia van Putten*

93. Switch reference in Whitesands. *Jeremy Hammond*

94. Machine learning for gesture recognition from videos. *Binyam Gebrekidan Gebre*

95. Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish. *Beyza Sümer*

96. An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music. *Salomi Savvatia Asaridou*

97. lncrementality and Flexibility in Sentence Production. *Maartje van de Velde*

98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. *Edwin van Leeuwen*

99. The request system in Italian interaction. *Giovanni Rossi*

100. Timing turns in conversation: A temporal preparation account. *Lilla Magyari*

101. Assessing birth language memory in young adoptees. *Wencui Zhou*

102. A social and neurobiological approach to pointing in speech and gesture. *David Peeters*

103. Investigating the genetic basis of reading and language skills. *Alessandro Gialluisi*

104. Conversation Electrified: The Electrophysiology of Spoken Speech Act Recognition. *Rósa Signý Gisladottir*

105. Modelling Multimodal Language Processing. *Alastair Smith*

106. Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing. *Florian Hintz*

107. Situational variation in non-native communication. *Huib Kouwenhoven*

108. Sustained attention in language production. *Suzanne Jongman*

109. Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntatic, and orthographic effects. *Malte Viebahn*

110. Nativeness, dominance, and the flexibility of listening to spoken language. *Laurence Bruggeman*

111. Semantic specificity of perception verbs in Maniq. *Ewelina Wnuk*

112. On the identification of FOXP2 gene enhancers and their role in brain development. *Martin Becker*

113. Events in language and thought: The case of serial verb constructions in Avatime. *Rebecca Defina*

114. Deciphering common and rare genetic effects on reading ability. *Amaia Carrión Castillo*

115. Music and language comprehension in the brain. *Richard Kunert*

116. Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language. *Nietzsche H.L. Lam*

117. The biology of variation in anatomical brain asymmetries. *Tulio Guadalupe*

118. Language processing in a conversation context. *Lotte Schoot*

119. Achieving mutual understanding in Argentine Sign Language. *Elizabeth Manrique*

120. Talking Sense: the behavioural and neural correlates of sound symbolism. *Gwilym Lockwood*

121. Getting under your skin: The role of perspective and simulation of experience in narrative comprehension. *Franziska Hartung*

122. Sensorimotor experience in speech perception. *Will Schuerman*

123. Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond. *Ashley Lewis*

124. Influences on the magnitude of syntactic priming. *Evelien Heyselaar*

125. Lapse organization in interaction. *Elliott Hoey*

126. The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech. *Sophie Brand*

127. The neighbors will tell you what to expect: Effects of aging and predictability on language processing. *Cornelia Moers*

128. The role of voice and word order in incremental sentence processing. *Sebastian Sauppe*

129. Learning from the (un)expected: Age and individual differences in statistical learning and perceptual learning in speech. *Thordis Neger*

130. Mental representations of Dutch regular morphologically complex neologisms. *Laura de Vaan*

131. Speech production, perception, and input of simultaneous bilingual preschoolers: Evidence from voice onset time. *Antje Stoehr*

132. A holistic approach to understanding pre-history. *Vishnupriya Kolipakam*

133. Characterization of transcription factors in monogenic disorders of speech and language. *Sara Busquets Estruch*

134. Indirect request comprehension in different contexts. *Johanne Tromp*

135. Envisioning Language - An Exploration of Perceptual Processes in Language Comprehension. *Markus Ostarek*

136. Listening for the WHAT and the HOW: Older adults' processing of semantic and affective information in speech. *Juliane Kirsch*

137. Let the agents do the talking: on the influence of vocal tract anatomy on speech during ontogeny and glossogeny. *Rick Janssen*

138. Age and hearing loss effects on speech processing. *Xaver Koch*

139. Vocabulary knowledge and learning: Individual differences in adult native speakers. *Nina Mainz*

140. The face in face-to-face communication: Signals of understanding and non-understanding. *Paul Hömke*

141. Person reference and interaction in Umpila/Kuuku Ya'u narrative. *Clair Hill*

142. Beyond the language given: The neurobiological infrastructure for pragmatic inferencing. *Jana Bašnáková*

143. From Kawapanan to Shawi: Topics in language variation and change. *Luis Miguel Rojas-Berscia*

144. On the oscillatory dynamics underlying speech-gesture integration in clear and adverse listening conditions. *Linda Drijvers*

145. Understanding temporal overlap between production and comprehension. *Amie Fairs*

146. The role of exemplars in speech comprehension. *Annika Nijveld*

147. A network of interacting proteins disrupted in language-related disorders. *Elliot Sollis*

148. Fast speech can sound slow: Effects of contextual speech rate on word recognition. *Merel Maslowski*

149. Reason-giving in everyday activities. *Julija Baranova*

150. Speech planning in dialogue - Psycholinguistic studies of the timing of turn taking. *Mathias Barthel*

151. The role of neural feedback in language unification: How awareness affects combinatorial processing. *Valeria Mongelli*

152. Exploring social biases in language processing. *Sara Iacozza*

153. Vocal learning in the pale spear-nosed bat, *Phyllostomus discolor*. *Ella Lattenkamp*