Isochronous rhythmic organization of learned animal vocalizations

Kumulative Dissertation zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
im Fach Biologie

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

von
Philipp Norton

Berlin im Dezember 2018

Die vorliegende Arbeit wurde von Juni 2015 bis Dezember 2018 am Institut für Biologie in der Arbeitsgruppe Verhaltensbiologie unter Leitung von Prof. Constance Scharff angefertigt.

1. Gutachterin:	Prof. Constance Scharff, Ph.D.
2. Gutachterin:	PD Dr. Henrike Hultsch
Disputation am:	12.02.2019

For Donald B. Norton

“Outside of a dog, a book is man's best friend. Inside of a dog it's too dark to read.”

Groucho Marx

Danksagung

Vielen Dank an alle, die mich über die letzten dreieinhalb Jahre unterstützt haben, ob durch Beratung, anregende Gespräche und Diskussionen, seelischen Beistand, oder Süßigkeiten. Vor allem Süßigkeiten. Ihr wisst, ich bin kein Mann ausschweifender Worte. Wisst aber auch, dass ich ohne eure Hilfe hoffnungslos verloren gewesen wäre, und dass ihr mein Leben im Laufe der letzten Jahre nachhaltig zum Guten geprägt habt:

Constance Scharff, die nicht nur Betreuerin meiner Arbeit und Chefin war, sondern eine wunderbare Mentorin in jedem Aspekt des wissenschaftlichen (und manchen des sonstigen) Lebens.
Henrike Hultsch für die gemeinsamen Lehren auf Augenhöhe und die schönen Feierabende.
Robert Ullrich für anregende, teilweise weltanschauungserweiternde Gespräche bei wohlgestapeltem Mensaessen, und für die Arschtritte.
Meinen Kollaborateuren in dieser Arbeit: Andrea Ravignani, Mirjam Knörnschild, Lara Burchardt und Ezequiel Mendoza.
Meinen Büromitbewohnern, die mir den Alltag versüßt haben: Jennifer Kosubek-Langer, Adriana Schatton, Daniela Vallentin, Jonathan Benichov, Stefan Wilczek, Susanne Seltmann, Anna Proß, Linda Bistere, und allen anderen aus der AG für die schöne Zeit.
Johanna und Vanessa Norton, die immer für mich da sind (und ich für sie).
Ivo und Fine für ein Zuhause wie es Loris gebührt.
Meiner Ca.

Publikationsliste

Zum Zeitpunkt der Dissertationseinreichung waren zwei der vier Einzelarbeiten veröffentlicht, eine zur Veröffentlichung angenommen und eine in Vorbereitung.

Publikation A

Ravignani, A.*, and **Norton, P.*** (2017). Measuring rhythmic complexity: A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations. *Journal of Language Evolution* 2, 4–19. https://doi.org/10.1093/jole/lzx002.

* The authors contributed equally to this study.

Publikation B

Norton, P., and Scharff, C. (2016). 'Bird Song Metronomics': Isochronous organization of zebra finch song rhythm. *Frontiers in Neuroscience* 10:309. https://doi.org/10.3389/fnins.2016.00309.

Publikation C

Burchardt, L., **Norton, P.**, Behr, O., Scharff, C., and Knörnschild, M. (accepted for publication). General isochronous rhythm in echolocation calls and social vocalizations of the bat Saccopteryx bilineata. *Royal Society Open Science*.

Publikation D

Mendoza, E., **Norton, P.**, Barschke, P., and Scharff, C. (in preparation). Effects on song learning differ after lentivirally mediated knockdown of FoxP1, FoxP2 or FoxP4 in Area X of zebra finches. *Journal of Neuroscience*.

Allgemeine Kurzfassung der Ergebnisse

Der evolutionäre Pfad der uns zu Musik geführt hat, wie wir sie heute kennen, ist schwer nachzuverfolgen. Artübergreifend vergleichende Forschung hilft uns das biologische Substrat zu ergründen, das es dem Menschen ermöglichte, dieses eigenartige Verhalten zu entwickeln. Rhythmus, die zeitliche Organisation von Ereignissen, ist ein zentraler Bestandteil der Struktur jeglicher Form von Musik. Musikalischer Rhythmus ruft oft die Wahrnehmung eines isochronen Takts, oder Pulses, hervor. Gelernte Vokalisationen nicht-menschlicher Tiere, wie Vogelgesang und die Gesänge bestimmter Fledermausarten, zeigen auffällige Parallelen zu Vokaler Musik (d.h. menschlicher Gesang). Diese Dissertation untersucht solche Vokalisationen auf das Vorkommen einer isochron rhythmischen Struktur, die es einem zuhörenden Artgenossen erlauben könnte einen solchen Takt wahrzunehmen. Zu diesem Zweck entwickelte ich eine Methode (genannt ‚generate-and-test'; GAT) um einen isochronen Puls aus einer zeitlichen Sequenz von Ereignissen, z.B. Notenanfängen, zu extrahieren. Diese Methode wird verglichen mit einer Reihe von existierenden analytischen Techniken zur Analyse unterschiedlicher Aspekte von Rhythmen in Vokalisationen, Bewegungen und anderen Verhaltensweisen die sich zeitlich entwickeln. Die Eignung der verschiedenen Methoden, um bestimmte Fragestellungen zu beantworten wird anhand einer Reihe von Beispielen veranschaulicht. Die Anwendung des GAT Ansatzes auf verschiedene Vokalisationstypen der großen Sackflügelfledermaus (*Saccopteryx bilineata*) brachte eine gemeinsame zeitliche Regelmäßigkeit zum Vorschein, die einen auf eine interessante Beziehung zwischen physiologisch determiniertem Rhythmus und dem Rhythmus von erlernten sozialen Lauten hindeuten könnte. In den Gesängen von Zebrafinken (*Taeniopygia guttata*) entdeckten wir eine hierarchisch isochrone Struktur die an die metrische Struktur vieler Musikarten erinnert. Wir berichten dann von dem Effekt genetischer Manipulationen auf den Gesangslernerfolg von Zebrafinken. Die Expression von FoxP2, einem Gen, das im Spracherwerb und im Gesangserwerb bei Singvögeln involviert ist, sowie von zwei verwandten Genen, FoxP1 und FoxP4, wurde in juvenilen Vögeln während ihrer Gesangslernphase experimentell reduziert. Neben anderen Effekten produzierten die adulten Vögel Gesänge mit einer beeinträchtigten isochronen Struktur. Überraschenderweise zeigten Kontrollvögel, deren FoxP Expression nicht reduziert wurde, einen ähnlichen Effekt in diesem Zusammenhang. Ich diskutiere dieses Ergebnis angesichts aktueller Kenntnisse über neuronale Mechanismen und Verhaltensprozesse im Bezug auf Gesangslernen und –produktion.

Table of Contents

List of Figures

General Introduction

Publication A

Publication B

Publication C

Publication D

List of Tables

Introduction

Publication A

Publication B

General Abstract

The evolutionary path that led to music as we know it today is difficult to trace. Cross-species comparative research can help us uncover the biological substrates that enabled humans to develop this peculiar behavior. Rhythm, the organization of events in time, is a central component in the structure of all forms of music. Oftentimes musical rhythm gives rise to a perceptionally isochronous beat, or pulse. Learned vocalizations of non-human animals, such as birdsong and the songs of certain bat species, show striking parallels to vocal music (i.e. human song). This thesis investigates these vocalizations for the presence of an isochronous rhythmic structure that could allow a conspecific listener to perceive such a beat. To this end, I have developed a generate-and-test (GAT) method to extract an isochronous pulse from a temporal sequence of events, such as the onsets of notes. This method is compared to a variety of existing analytic techniques for analyzing different aspects of rhythms in vocalizations, movements and other behaviors developing over time. The suitability of the different methods for addressing particular questions is illustrated through various examples. The application of the GAT approach to different types of vocalizations of the greater sac-winged bat (*Saccopteryx bilineata*) revealed a common temporal regularity that might point towards an interesting relationship between physiologically determined rhythm and the rhythm of learned social vocalizations. In the songs of zebra finches (*Taeniopygia guttata*) we discovered a hierarchical isochronous structure that is reminiscent of the metrical structure of many types of music. We then report the effect of genetic manipulations on the song learning success of zebra finches. The expression of FoxP2, a gene involved in speech acquisition and birdsong learning, as well as of two related genes, FoxP1 and FoxP4, was experimentally reduced in juvenile birds during their learning period. Among other effects, the adult birds produced song with an impaired isochronous structure. Surprisingly, control animals whose FoxP levels were not reduced, showed a similar effect in this regard. I discuss possible interpretations of this result in the light of current knowledge about neural mechanisms and behavioral processes of song learning and production.

General Introduction

Music has often been called the “universal language”. This term is not without its problems. It seems obvious, however, that music is intimately linked to the human experience. It plays an important role in social life across cultures (Trehub et al. 2015). Already as newborns our auditory system is tuned to musical features such as pitch and rhythm similarly to the way it is in adults (Stefanics et al., 2009; Winkler et al. 2009). These and other observations suggest that it is somehow grounded in our biology. If so, how did this peculiar behavior, whose adaptive function is far from obvious, evolve?

Uncovering the evolutionary roots of music is severely hampered by missing archeological evidence, as behaviors such as vocal music do not fossilize (Honing et al., 2015). As musicologist Gary Tomlinson put it: “Following such developments may seem to pose intractable problems, even imponderable ones; for what kind of evidence can be brought to bear on the case? The question is pressing for musicking [music-making] since, whatever else it is, it is an evanescent act or set of acts that fades as it sounds. Its product does not have the staying power of mammoths painted on cave walls or the heft of carved ‘Venus’ figurines” (Tomlinson, 2015, p. 12). The earliest direct and undisputed archeological evidence for human musicality is that of musical instruments that were found in different parts of Eurasia, in the form of flutes carved out of animal bones (Buisson, 1990; Conard et al., 2009; Hahn and Münzel, 1995; Kunej and Turk, 2001). Carbon dating of these artefacts proves the existence of a widespread musical tradition that is at least around 40.000 years old, shortly after the first humans settled in Europe (Higham et al., 2012). The relative sophistication of the instruments suggests that they were likely predated by simpler instruments, many of which were made of biodegradable material and might never be recovered (Trehub et al., 2015). Even the earliest tools built specifically for making music are thought to have drawn upon an already established repertoire of musical behavior unaided by tools, i.e. vocal music (song) and percussive use of the body, e.g. clapping (Honing et al., 2015; Morley, 2014). Song is thought to be a universal in human music (Brown and Jordania, 2011; Nettl, 2001; Trehub, 2001) and vocalization – the production of sound with the vocal organs – a promising candidate for an evolutionary progenitor to music (Fitch, 2006; Mehr and Krasnow, 2017).

Cross-species comparative research can offer insights into the evolution of music

In the face of this scarcity of empirical evidence, comparative studies can greatly aid the endeavor of piecing together the evolutionary path that led to our present music faculty. Investigating which musical behaviors are shared among all human musical cultures (‘musical universals') can help disentangle whether they are more likely shaped by innate cognitive dispositions or developed through cultural evolution. This approach hinges on the availability of multiple independent data points. In the case of music the opportunities to gather these are rapidly diminishing, as western music – through globalization – leaves its mark even on the most remote cultures (Huron, 2008).

Many authors have stressed the utility of species comparative animal research in this very endeavor (e.g. Carterette and Kendall, 1999; Fitch, 2006; Honing, 2018; Hulse and Page, 1988). Here it is important to distinguish between *musicality*, "a natural, spontaneously developing set of traits based on and constrained by our cognitive abilities and their underlying biology" and *music*, "a social and cultural construct based on that very musicality" (Honing, 2018). In this sense musicality is what allowed humans to develop music. Both the traits that make up musicality, as well as the biological foundations that enable those traits, can be broken up into their constituent components. One can then ask which of these components we share with other animal species. Musical universals across human cultures can help uncover which features of music are fundamentally rooted in musicality. The discovery and study of components of musicality shared with non-human species can in turn inform our understanding of the evolution of musicality. Rhythm is a central component of musical structure and its perception and production fundamental to musicality. At its most basic conception, it is simply the organization of events in time. As such, it governs the wealth of animal behavior, from locomotion to foraging to communication. It thus constitutes a potent domain for inquiry into universals of musicality.

Isochronous rhythmic structure in human music

One of the countless definitions for rhythm was formulated by Patel (2008, p. 96), specifically for application in music and speech: "Rhythm [is] the systematic patterning of sound in terms of timing, accent, and grouping". When rhythmic patterns are regularly repeated, they give rise to periodicity. In terms of grouping, for example, a melodic phrase of four different notes might be repeated several times, so that each of those notes periodically reoccurs, interspersed with the remaining three. Like such phrasal patterns, also temporal patterns can be periodic. Consider the first seven notes of the children's song 'Mary had a little lamb' (**Figure 1A**). The timing of the onsets of all seven notes is *isochronous*, i.e. the onsets are equally spaced in time. In other words, all notes have the same inter-onset interval (IOI). In this simple case of temporal periodicity, the temporal pattern is repeated with each note. If you were to clap your hands to this tune, your claps might coincide with every second note ('Mary had a little lamb'). This is then the tempo of the beat, "a perceptually isochronous pulse to which one can synchronize with periodic movements such as taps or footfalls" (Patel, 2008, p. 97). In this example, there are notes that occur between single beats (i.e. single events in the beat), namely every second note. The beat period is thus twice as long in duration as the IOI. In melodies with more complex patterns, some single beats might not coincide with salient acoustic events (e.g. note onsets), instead occurring in the space between successive events (**Figure 1B**). Thus, the beat is not necessarily instantiated in the acoustic signal itself, although it can be emphasized by an instrument, as the bass drum in electronic dance music often does. Instead it is a cognitive construct of the listener that is implicit in the acoustic structure of music (Arom, 1991, p. 230; Fitch, 2013).

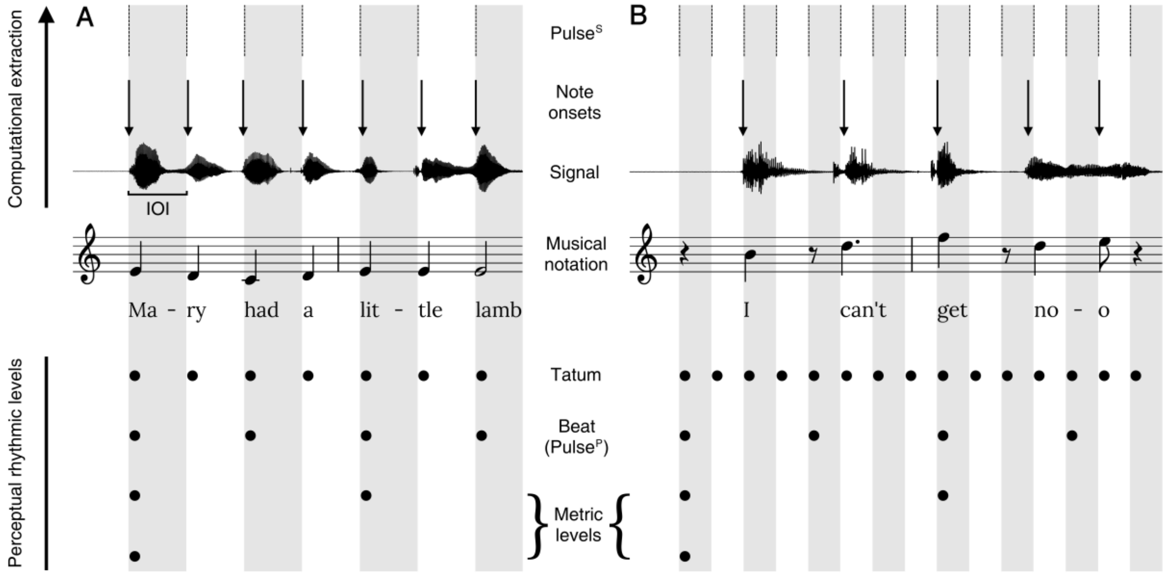


Figure 1 – Isochronous rhythmic subdivisions of two example song sections. The examples are the first two bars of vocalization of the nursery rhyme 'Mary Had a Little Lamb' (A) and the Rolling Stones song '(I Can't Get No) Satisfaction' (B). The middle row in each panel shows the Western musical notation of the vocal melody and lyrics. The musical symbols here are spaced according to their durational values. The dots below represent possible perceptual rhythmic subdivisions and the terms used for those in this thesis (center column). Tatum: the inferred "time division which most highly coincides with all note onsets" (Bilmes, 1993). Beat or perceived pulse (pulse^P): "a perceptually isochronous pulse to which one can synchronize with periodic movements such as taps or footfalls" (Patel, 2008, p. 97). Note that listeners might perceive the beat on a different level, e.g. at twice the tempo. Metric levels: hierarchic pattern of more strongly accentuated events of the beat pattern. Above the musical notation is a visual representation of the audio signal (oscillogram) of the sung melody. The arrows indicate the time points of the note onsets. Also indicated in (A) is the first inter-onset interval (IOI). The extracted note onsets can be used to computationally fit a signal-derived pulse (pulse^S). Note that the tatum and the pulse^S coincide, while the tempo of the pulse^P often is a small integer fraction of that of the tatum – ½ in (A) and ¼ in (B). Different listeners might perceive the beat on different levels, e.g. at twice or half the tempo. Further note that the strong accent on the beginning of (B) is carried in the song by non-vocal instruments not depicted here. De-emphasizing the highest metrical level is rare in music, but can be used as a stylistic device, e.g. the 'one drop' rhythm in Jamaican popular music (Oliver, 2013, p. 244).

The rhythmic pattern is often further organized into a recurrent hierarchical pattern of weakly and strongly accentuated beats, the meter (Fitch, 2013; Longuet-Higgins, 1976). In the example, the first and third beat, coincident with the first and fifth note, is percieved as more strongly accentuated than the others (‘**Ma**ry had a **lit**tle lamb’). The metric hierarchy can have multiple levels, such that the first beat, for example, could be percieved as having an even stronger accent than the third (**Figure 1**).

As the term beat often implies the presence of a metrical structure, I will call the bare periodic percept without any assumptions of heterogeneous accentuation the (isochronous) pulse. I make a further distinction between this subjectively *percieved pulse* (abbreviated as pulse^{P}) and a *signal-derived pulse* (pulse^{S}), that is computationally extracted from an acoustic signal or pattern (extraction methods are exemplified in **Publication A**, sections 2.6 and 2.7). The events in such a pattern (e.g. note onsets) can themselves be quasi-isochronous, as in **Figure 1A**. If they are not (**Figure 1B**), the events can still be temporally organized in a regularity that gives rise to an *isochronous rhythmic structure*. That structure can be revealed through the pulse^{S} and the robustness of its isochronicity evaluated through the temporal fit between pulse^{S} and events. Given sufficient goodness of fit, this pulse^{S} indicates a regularity required for a listener to percieve a pulse^{P}. The pulse^{S} corresponds temporally to the tatum, the inferred “time division which most highly coincides with all note onsets” (inspired by a portmanteau of 'temporal *atom*' and named after jazz pianist Art Tatum; Bilmes, 1993). The temporal level of the pulse^{P}, however, depends on the listener. People show a preference for pulses^{P} with a period of around 500–700 ms, although they can attend to a much wider range (Parncutt, 1994; van Noorden and Moelants, 1999). This tempo also exhibits a degree of inter-cultural variability (Drake and El Heni, 2003) that may partly depend on knowledge of the specific music structure (Toiviainen and Eerola, 2003).

Note that not all music has an isochronous beat. Notable exceptions include the music played on the Chinese Guqin (or Ch’in) lute. The traditional notation for the Guqin contains no temporal markings for individual notes (Patel, 2008, p. 97-98). However, a recent study on a diverse global set of 304 music recordings found an isochronous beat to be one of six ‘statistical universals’ of human music, a feature that was present in the majority of songs sampled from each of nine geographical regions spanning the globe (Savage et al., 2015).

This thesis was motivated by the question of whether an equivalent to this central characteristic of human music can be found in the behaviors of non-human animals. For my specific research pertaining to the evolution of vocal music, some animal behaviors are more suited than others. A variety of animal species produce highly rhythmic communication signals through different mechanisms and modalities. Examples can be found in the visual domain, such as the claw waving displays of fiddler crabs (Kahn et al., 2014) and the bioluminescent flashes in fireflies (Buck, 1938; Moiseff and Copeland, 2010). The rhythmic chirps of bush crickets (Greenfield and Roizen, 1993; Sismondo, 1990) are acoustic signals, but are produced by mechanical stridulation of the wings, rather than through vocalizations. Various species of frogs produce vocalizations with high temporal regularity (reviewed by Greenfield, 2005). The rhythms underlying these vocalizations, however, are thought to be controlled by vocal pattern generators in the brainstem, and to be more homologous to largely innate human vocalizations like laughter and crying (Bass et al., 2008; Hage, 2018; Yamaguchi et al., 2017). Rhythmic drumming behavior as exhibited, for example, by palm cockatoos (Heinsohn et al., 2017), woodpeckers (Dodenhoff et al., 2001), and chimpanzees (Babiszewska et al., 2015) is non-vocal, but might be highly informative in the context of the evolution of human percussive music, especially considering that it is present in great apes, our closest living relatives (Fitch, 2015; Ravignani et al., 2017).

Vocal learning

A fundamental prerequisite for the development of vocal music is our capacity for vocal production learning. It is defined as the process of modifying one's vocalizations as a result of experience with those of other individuals (Janik and Slater, 1997, 2000). Production learning is distinct from contextual learning, i.e. learning the context in which an existing vocalization is used. The latter comprises vocal comprehension learning – the extraction of a novel meaning from the use of a vocalization by another individual – and vocal usage learning – the learned production of an existing vocalization in a new context (Janik and Slater, 2000). Contextual learning is relatively wide-spread in vertebrates (Schusterman, 2008). Dogs, for example, can learn to sit in response to hearing the word 'sit' (comprehension learning) and can be trained to bark in response to a specific signal (usage learning; Salzinger and Waller, 1962). They cannot, however, learn to produce the word 'sit' themselves, instead being restricted to their innate vocalizations. Vocal production learning (henceforth only referred to as 'vocal learning') appears to be a comparatively rare trait in the animal kingdom. Apart from humans, it has been clearly documented so far only in songbirds, hummingbirds and parrots as well as several species of bats, some marine mammals, and elephants (reviewed by Petkov and Jarvis, 2012). Note that only few species have been systematically tested for this capacity, and in many more it may still await discovery. The faculty of vocal learning has traditionally (at least implicitly) been discussed as a dichotomy, separating animals into vocal learners (those listed above) and non-learners. Findings that some species generally considered 'non-learners' show some degree of flexibility in their otherwise innate vocalizations have led to the formulation of the continuum hypothesis for vocal learning (Arriaga and Jarvis, 2013; Petkov and Jarvis, 2012). Species on one end of this continuum can subtly modify their vocalizations using auditory feedback, but are able to develop their species-typical vocal repertoire without an external model ('limited vocal learning'; Scharff et al., in press). On the other end of the spectrum are those that learn most of their vocalizations by imitative learning from an external model ('extensive vocal learning'; ibid.).

Birdsong as a genuine model for vocal learning

Among the most accomplished vocal imitators, next to humans, are many species of songbirds and parrots. Vocal learning has been studied more extensively in songbirds than in parrots (reviewed by Catchpole and Slater, 2008; c.f. Pepperberg, 2010). Much of that research in the last couple of decades has been motivated by striking parallels between the processes through which we learn to speak, and birds learn to sing, as well as in the underlying neural and genetic systems (Bolhuis et al., 2010; Doupe and Kuhl, 1999; Marler, 1970; Prather et al., 2017). Both children and juvenile songbirds learn their vocalizations by imitating adult conspecifics. Early in their development, songbirds start to produce relatively unstructured sounds called 'subsong', akin to the babbling phase of babies. Like children, they then gradually modify those sounds through imitative learning to increasingly resemble the memorized sounds of their vocal models (e.g. adults). This learning process

(i) depends on external auditory input; birds that grow up in complete acoustic isolation barely progress past subsong and end up with a highly impoverished song as adults (Fromkin et al., 1974; Thorpe, 1958).

(ii) is shaped by innate dispositions; when given the choice between multiple different vocal models, they preferentially learn from their conspecifics (Wheatcroft and Qvarnström, 2015). An experiment by Fehér et al. (2009) aptly demonstrated that the neural substrate of zebra finches, a vocal learning songbird species, already carries relatively specific dispositions for particular features of their species' song (Bolhuis et al., 2010): They established a colony consisting exclusively of isolated juvenils that never heard a 'normal' adult song, which then tutored their offspring with their subsong. Over several generations the songs in the colony increasingly resembled wild-type zebra finch song.

(iii) depends on auditory feedback; in modifying their vocalizations toward their memorized template, individuals have to compare the two in order to correct the perceived mismatch error (Brainard and Doupe, 2000; Möttönen and Watkins, 2009).

The production of learned vocalizations in both humans and songbirds rely on comparable specialized forebrain regions. These include vocal motor pathways that evolved independently, likely out of a motor pathway that existed in the last common ancestor of birds and mammals (Chakraborty and Jarvis, 2015; Jarvis, 2004), accompanied by convergent transcriptional specializations (Pfenning et al., 2014).

While these parallels have mostly been studied in the context of comparative language research, many apply more broadly to vocal learning in general and thus to the basis of vocal music. In some aspects of form and function birdsong is more reminiscent of music. Birdsong, like human song, is repeated again and again, both in practice and performance contexts (in zebra finches the two distinct modes are called 'undirected' and 'directed' song). Many bird songs sound very musical to the human ear, which has prompted several composers to incorporate them into their music, or to emulate birdsong in their arrangements (Baptista and Keister, 2005; Taylor, 2014). The very word bird<u>song</u> reflects this perception. Despite differences in absolute pitch and duration, songs from a wide range of songbird species tend to exhibit similar descending or arched melodic contours as human songs (Savage et al., 2017). These might come out of basic energetic and motor constraints (Tierney et al., 2011), and/or shared perceptual preferences. The rhythmic structure of birdsong has received surprisingly little attention so far. Two related methods have been developed to visualize and explore the overall developmental dynamics of birdsong rhythm (Saar and Mitra, 2008; Sasahara et al., 2015), but have found little application so far.

Some recent work has examined the balance between repetition and novelty in the structure of bird songs. This well-studied balance is highly abundant in music (e.g. Hargreaves, 1984; Leach and Fitch, 1995; Sallavanti et al., 2015). David Huron (2006) has posited that one of the main drivers of the emotive power of music is the interplay between fulfillment and violation of expectations in the form of successful anticipations and surprises. For this interplay to occur, there has to be a 'sweet spot' (or rather range) between simplicity and complexity: an extremely repetitive song without any form of variation might quickly loose one's interest. On the other hand, a constantly changing song devoid of any recurring patterns might strain the cognitive capacities of the listener and might not even allow for expectations to form. A study

(Janney et al., 2016) examined the temporal reoccurrence of motifs in up to hundreds of phrases consecutively sung by pied butcherbirds (*Cracticus nigrogularis*), a species noted for its virtuosity (Taylor, 2008; Taylor and Lestel, 2011). The authors found that individuals with a large repertoire tended to maximize the regularity in the repetition of motifs, while those with smaller repertoires showed reduced regularity. One of several alternative explanations they offer for this negative correlation of temporal diversity with repertoire size is somewhat related to Huron's theory. It assumes that potential mates prefer accurate reproduction of song motifs, which has been shown to be the case in several bird species (Catchpole and Slater, 2008; Riebel, 2009). To evaluate the accuracy of a motif's performance, i.e. compare multiple renditions, they must memorize it between subsequent presentations. A bird with a large repertoire should avoid exceeding the memory capacities of its avian listener by minimizing the temporal distance between renditions and thus increasing temporal regularity. For a bird with a small repertoire this is less of a concern, and it could instead reduce the chances of the listener habituating to its performance by 'mixing it up', i.e. increasing the temporal diversity in its song. An isochronous pulse could potentially serve as a temporal scaffolding for song and thus as a strong driver for anticipations in the time-domain of rhythm.

The parallels in vocal learning between humans and songbirds has motivated a wealth of research into the behavioral, neurobiological and genetic mechanisms of birdsong over the last decades. The overall ethical acceptance of invasive studies in non-human animals has enabled the attainment of a finer grained understanding of basic biological mechanisms. Electrophysiological and optogenetic experiments, genetic manipulation and measurement of behavior-dependent gene expression provide a window into the proximate 'how' of vocal learning that can inform our inquiry into the neural and genetic substrate for speech and vocal music. The zebra finch (*Taeniopygia guttata*) has been established as the main model species in this line of research (Lattenkamp and Vernes, 2018), particularly in the domain of neuroscience (Griffith and Buchanan, 2010). The high resolution of knowledge about song and its development in the zebra finch, as well as the fact that its song is short and highly stereotyped makes it a good first candidate for investigations into isochronous rhythmic structure in learned vocalizations of non-human animals.

The zebra finch

Zebra finches live in large flocks where they have to navigate a complex social environment. Much of their communication takes place in the acoustic modality. Both male and female zebra finches are very vocal, uttering a variety of unlearned calls in different social situations

(Zann, 1996). Only the males produce a learned song. Each adult male sings an individual repeated motif of roughly 1s duration (**Figure 2**). A motif is composed of about 3–9 bioacoustically distinct notes (also called syllables or elements; range from this thesis), which are separated by silent gaps of short but deep inhalations ('minibreaths'; Wild et al., 1998) . Song notes consist of one or more sub-note elements that correspond to neuromuscular gestures (henceforth simply called 'gestures'; Amador et al., 2013). These are characterized by discontinuities in motor control parameters in the vocal organ, like membrane tension and air pressure, and often result in sudden frequency shifts (**Figure 2**, top). Animals mostly sing in bouts of several repetitions of the motif. These bouts are typically preceded by a variable number of usually identical elements called introductory notes. The sequential order of song notes is generally very stereotyped, and many birds sing only a single motif variant (e.g. abcd abcd abcd). Others have several motif variants and note insertions and deletions are common occurrences (Helekar et al., 2000).

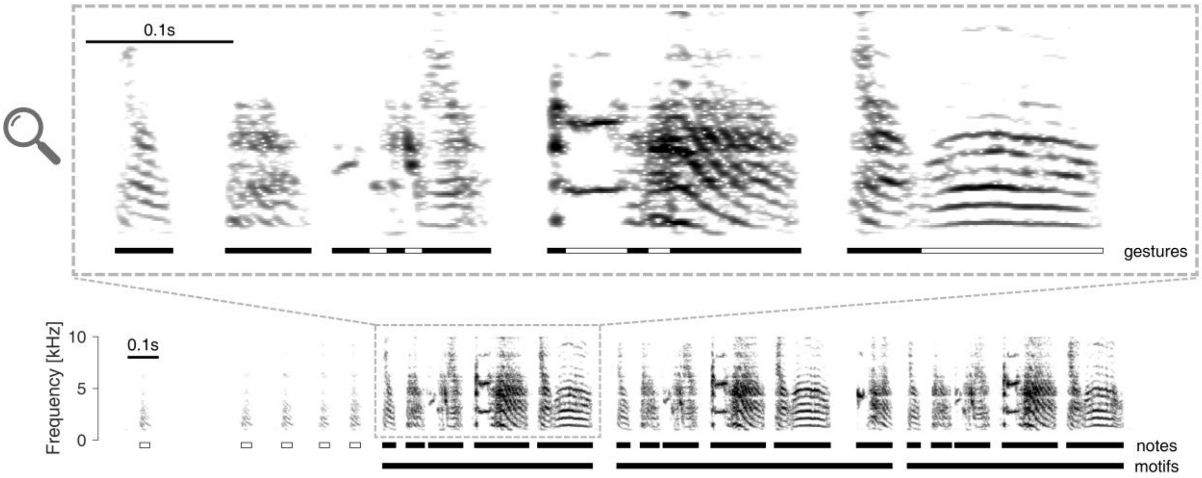


Figure 2 – Units of zebra finch song. A sonogram of a typical zebra finch song bout (bottom) and a magnification of the first motif of this bout (top). The song bout begins with a series of short introduction notes (white bars; bottom), followed by three motifs (black bars). The first and third motif consist of five song notes (black bars above the motif bars), separated by silent inhalation gaps. The second motif is a variant with a sixth note added at the end. The magnification of the first motif highlights the sub-note elements called gestures (alternating black and white bars).

As so-called closed-ended (or age limited) vocal learners, zebra finch juveniles go through a finite learning period, after which their song remains largely unchanged. This distinguishes them from open-ended learners like canaries (Nottebohm et al., 1986; Nottebohm and Nottebohm, 1978), European starlings (Mountjoy and Lemon, 1995) and nightingales (Kiefer et al., 2006), whose song plasticity seasonally reopens. The song learning period of a male zebra finch consists of two overlapping phases: The sensory and the sensorimotor phase. During the sensory phase, which lasts from around 25 to 65 days after hatching ('post-hatch day', PHD), juvenile birds memorize the song heard from adult male tutors, preferentially their social father's song (Immelmann, 1969; Mann and Slater, 1995; Roper and Zann, 2006; Zann, 1990). Beginning around the same time, the birds start to produce noisy and relatively unstructured vocalizations called subsong, analogous to infants babbling (Doupe and Kuhl, 1999; Marler, 1970). In the sensorimotor phase, from around 35 PHD onward, these vocalizations are gradually modified, increasingly resembling the memorized tutor song (Derégnaucourt, 2011). At the time of sexual maturation, around 90 PHD, vocal plasticity closes and the song crystallizes (Immelmann, 1969).

Two specialized neural pathways in the songbird brain are involved in song learning and production. Together they form what is commonly called the 'song system' (**Figure 3**). Both pathways consist of anatomically discrete brain regions (nuclei) connected by projection neurons and both originate at the pallial nucleus HVC (historically an abbreviation, now used as a proper name; Reiner et al., 2004). Although birds do not possess the layered neocortex found in the mammalian brain, the avian pallium is argued by many to be analog, if not homolog to the mammalian neocortex (e.g. Jarvis et al., 2005).

The vocal motor pathway (or song motor pathway, SMP) is essential for song production. Lesions in any part of this pathway lead to either elimination or severe disruption of song (Nottebohm et al., 1976; Simpson and Vicario, 1990). Neurons project from HVC to the robust nucleus of the arcopallium (RA) and from there to the tracheosyringeal portion of the hypoglossus (nXIIts), which innervates the muscles of the syrinx, the bird's vocal organ (Wild, 1997).

The second, the anterior forebrain pathway (AFP), forms a closed cortico – basal ganglia – thalamic loop. It plays a crucial role in sensorimotor learning and adult plasticity of song (Kao and Brainard, 2006; Scharff and Nottebohm, 1991; Thompson et al., 2011), and contributes to recognition of conspecific song (Brenowitz, 1991; Scharff et al., 1998). It is involved in generating variability across song renditions, thereby facilitating the learning process through motor exploration and subsequent performance evaluation (Aronov et al., 2008; Kojima et al., 2018; Ölveczky et al., 2005). This pathway passes through Area X, the nucleus dorsolateralis anterior pars medialis (DLM) and the lateral magnocellular nucleus of the anterior nidopallium (LMAN). From there it converges with the SMP at the RA.

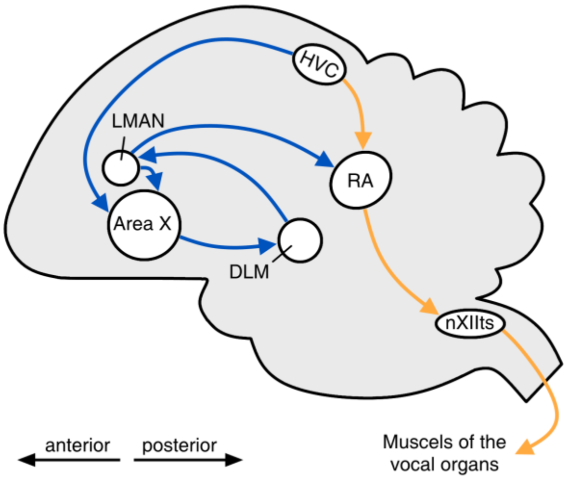


Figure 3 – Schematic overview of the song system pathways in the songbird brain. The song motor pathway (SMP, orange arrows, right) includes the nuclei HVC (used as a proper name), the robust nucleus of the arcopallium (RA), and the tracheosyringeal portion of the hypoglossus (nXIIts). The anterior forebrain pathway (AFP, dark blue arrows, left) passes through HVC, Area X, the nucleus dorsolateralis anterior pars medialis (DLM), the lateral magnocellular nucleus of the anterior nidopallium (LMAN) and ends at the RA. Image based on Bolhuis et al. (2010).

FoxP2: A genetic building block for vocal learning

Another intriguing parallel between songbirds and humans in terms of vocal learning is one on the genetic level. In 1990 an extended family was discovered of which some members showed a severe speech disorder called 'developmental verbal dyspraxia' (DVD) or 'childhood apraxia of speech' (Hurst et al., 1990). Unlike earlier described cases of DVD, the disorder was autosomal dominantly inherited in this family. It was found that the affected members carried a heterozygous point-mutation in the gene coding for FOXP2 (forkhead box protein P2; Lai et al., 2001). It belongs to a large family of transcription factors – proteins that bind to the DNA to repress or enhance the expression of other genes. Patients with DVD caused by intragenic mutations of FOXP2 typically exhibit difficulties in sequencing speech sound into syllables, words and sentences and in planning or programming of oral movements (Morgan et al., 2016). Symptoms also commonly include receptive and expressive language deficits, both semantic and syntactic. Non-verbal IQ and general fine motor skills are not – or in some cases only mildly – affected (Morgan and Webster, 2018). To my knowledge there are no reports on rhythm production and perception abilities in patients with FoxP2-related DVD to date.

The genes encoding these proteins are highly conserved across vertebrates (Scharff and Petri, 2011; Schatton & Scharff, 2016). In songbirds, FoxP2 expression is regulated developmentally and in relation to singing activity in Area X (Haesler et al., 2004; Teramitsu et al., 2010). Studies in which Foxp2 was experimentally downregulated in Area X of juvenile zebra finches, resulted in impairment of proper song learning (Haesler et al., 2007). These birds copied fewer notes from their tutors, those that were copied were copied less accurately and the sequential order of notes was jumbled. In addition to inaccurate copying of spectral features, the duration of the copied notes was significantly more different from their tutors than was the case in control birds. In the adult song of these birds, the variability of note durations was also significantly increased. FoxP2 thus appears to be involved in both learning and production of accurate timing in song.

Besides FoxP2, two other members of the FoxP family of transcription factors are expressed in the brain: FoxP1 and FoxP4 (Lu et al., 2002; Teufel et al., 2003). These have been demonstrated to form homo- and heterodimers in cultures of human (Sin et al., 2015) and mouse neural tissue (Li et al., 2004) and in the zebra finch brain (Mendoza and Scharff, 2017). Dimerization seems to be essential for the transcriptional function of FoxP2 (Li et al., 2004). FoxP2, as well as FoxP1 and FoxP4 may therefore play a role in the development of zebra finch song rhythm.

Echolocating bats as mammalian vocal learners

Despite the apparently high convergence of the neural mechanisms underlying vocal learning in songbirds and humans, the divergent evolutionary paths since their last common ancestor led to marked differences, e.g. in brain morphology. A species comparative approach in the search for the evolutionary path to isochronous rhythmic structure in music could be benefited greatly by including species that are phylogenetically closer to humans.

One group of mammals in which the capacity for vocal learning evolved apparently independently from humans, are echolocating bats. Compared to the breadth of research on vocal learning in songbirds, studies on bat vocal learning are relatively scarce (Knörnschild, 2014). Pups of the greater sac-winged bat (*Saccopteryx bilineata*) has been shown to learn territorial songs from adult males through vocal imitation (Knornschild et al., 2010). The vocalizations produced early in their learning phase show similarities to human babbling and the subsong in songbirds (Knörnschild et al., 2006). Within the phylogenic group of bats FoxP2 exhibits a high level of sequence diversity, an intriguing deviant to the gene's otherwise marked conservation (Li et al., 2007). The significance of this diversity for vocal learning is yet unclear, but it offers new opportunities for research into the molecular mechanisms of FoxP2 (e.g. Chen et al., 2013).

Thesis outline

This thesis begins with a review of different statistical methods and visualization tools for the analysis of temporal structure in vocalizations, movements and other behaviors developing over time (**Publication A**: Ravignani & Norton, 2017). This article provides an overview of the various methods available for both the quantification of rhythmic complexity in single vocalizations and the comparison of rhythmic structure between multiple vocalizations (or other behaviors). Most of the presented methods are demonstrated by application to a set of computer-generated temporal sequences that differ in their rhythmic structure. Thereby we describe the appropriateness of the tests to particular hypotheses and provide possible interpretations of the exemplary results. The article incudes the code that was used to carry out the exemplary analyses and create the figures as supplementary material, including a detailed explanation and instructions on its usage. Due to the broad applicability of these analytic tools in diverse disciplines and the importance of cross-species comparative research, the paper also intends to provide a common point of reference for language researchers, musicologists and behavioral biologists, hopefully facilitating interdisciplinary exchange and fostering comparability of results.

One of the methods presented in Publication A is the generate-and-test approach (GAT) I developed as a tool to investigate zebra finch song for an underlying isochronous regularity. **Publication B** describes this method and reports the results of its application to the songs of adult zebra finches (Norton & Scharff, 2016). The GAT algorithm finds a signal-derived isochronous pulse (pulse^S) that fits best to a series of temporal events. Here we chose note onsets as the events to which the pulses^S were aligned. In each of the 15 birds whose songs were analyzed, the frequencies (i.e. tempo) of the best fitting pulses^S clustered closely around a dominant frequency between 10 and 60Hz (25–45Hz for most birds), that differed from individual to individual. These pulses fit significantly better than pulses aligned to artificial 'songs' with randomized note and gap durations but identical sequential structure. This result indicates that the analyzed songs have a strong isochronous temporal structure in the timing of notes, and thus the potential for a listener to perceive a pulse akin to the beat in many types of music. Interestingly, this regularity extends to the sub-note level: the transitions between gestures within complex notes coincided with the pulse significantly more often than expected by chance. This implies a hierarchical temporal structure, on the lowest level of which lies the gesture, that is reminiscent of the metrical rhythm of many types of music.

In **Publication C** the GAT method was applied to three different vocalization types produced by the greater sac-winged bat. Echolocation calls, one of the three types, are emitted by the bat during flight. Echolocation calls during search flights are known to be coupled to wing beat frequency. The other two types –learned male territorial songs and innate pup isolation calls – are produced while the bats are stationary and therefore not coupled to wingbeat. The frequencies of the best fitting pulse^S mostly clustered in a range of 6–20Hz for all animals. Interestingly, the range for the non-coupled vocalization types was in a similar range to the wingbeat-coupled echolocation sequences. This led us to speculate that attentional tuning to the rhythms of conspecifics' echolocation calls might have an influence on the rhythm of social vocalizations.

Publication D reports the effects of experimentally reduced levels of FoxP1, 2 and 4 in Area X of juvenile male zebra finches on their song learning success. Four groups of animals received lentivirus-mediated downregulation targeting either FoxP1, FoxP2 or FoxP4, or with a non-targeting control construct. All non-control treatment groups showed impaired song learning and an impoverished song as adults, compared to the control group. Despite overlap in the parameter-space of the affected song features, discriminant analysis revealed that the treatment groups diverged from each other in linear combinations of these features. This suggests that developmental manipulations of FoxP1, FoxP2 and FoxP4 in Area X differentially impact adult song. One of the strongest discriminating factor was related to the variability in note timing. Surprisingly the songs of the control group birds had a markedly reduced isochronous rhythmic structure, comparable to those of the knockdown groups. The **general discussion** offers some possible interpretations of these results in light of recent research on neural and behavioral mechanisms of song learning and production.

Table 1 – Different terms used for important concepts in the four publications.

Introduction	**Publication A**	**Publication B**	**Publication C**	**Publication D**	**Description/Definition**
pulse^S	pulse	pulse^S	rhythm^S	pulse	Isochronous sequence of timepoints that best aligns to a given temporal pattern
note	note	note	element	song element	Unit of animal vocalizations that is separated by silent inhalation intervals (gaps)
gap	-	gap	gap	gap	Silent inhalation interval

References

Amador, A., Perl, Y. S., Mindlin, G. B., and Margoliash, D. (2013). Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature* 495, 59–64. doi:10.1038/nature11967.

Arom, S. (1991). *African Polyphony and Polyrhythm.* Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511518317.

Aronov, D., Andalman, A. S., and Fee, M. S. (2008). A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science* 320, 630–634. doi:10.1126/science.1155140.

Arriaga, G., and Jarvis, E. D. (2013). Mouse vocal communication system: Are ultrasounds learned or innate? *Brain Lang.* 124, 96–116. doi:10.1016/j.bandl.2012.10.002.

Babiszewska, M., Schel, A. M., Wilke, C., and Slocombe, K. E. (2015). Social, contextual, and individual factors affecting the occurrence and acoustic structure of drumming bouts in wild chimpanzees (Pan troglodytes). *Am. J. Phys. Anthropol.* 156, 125–134. doi:10.1002/ajpa.22634.

Baptista, L. F., and Keister, R. A. (2005). Why birdsong is sometimes like music. *Perspect. Biol. Med.* 48, 426–443. doi:10.1353/Pbm.2005.0066.

Bass, A. H., Gilland, E. H., and Baker, R. (2008). Evolutionary origins for social vocalization in a vertebrate hindbrain-spinal compartment. *Science* 321, 417–421. doi:10.1126/science.1157632.

Bilmes, J. (1993). Techniques to foster drum machine expressivity. In *Proc. Int. Comp. Music Conf.*, 276–283.

Bolhuis, J. J., Okanoya, K., and Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* 11, 747–759. doi:10.1038/nrn2931.

Brainard, M. S., and Doupe, A. J. (2000). Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* 404, 762–766. doi:10.1038/35008083.

Brenowitz, E. (1991). Altered perception of species-specific song by female birds after lesions of a forebrain nucleus. *Science* 251, 303–305. doi:10.1126/science.1987645.

Brown, S., and Jordania, J. (2011). Universals in the world's musics. *Psychol. Music* 41, 229–248. doi:10.1177/0305735611425896.

Buck, J. B. (1938). Synchronous rhythmic flashing of fireflies. *Q. Rev. Biol.* 13, 301–314. doi:10.1086/516403.

Buisson, D. (1990). Les flûtes paléolithiques d'Isturitz (Pyrénées-Atlantiques). *Bull. la Société préhistorique française* 87, 420–433.

Carterette, E. C., and Kendall, R. A. (1999). "Comparative music perception and cognition," in

The Psychology of Music, ed. D. Deutsch (San Diego, CA: Academic Press), 725–791.

Catchpole, C., and Slater, P. (2008). *Bird Song: Biological Themes and Variations*. 2nd edition. Cambridge, UK: Cambridge University Press doi:10.1017/CBO9781107415324.004.

Chakraborty, M., and Jarvis, E. D. (2015). Brain evolution by brain pathway duplication. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370, 20150056. doi:10.1098/rstb.2015.0056.

Chen, Q., Zhu, T., Jones, G., Zhang, J., and Sun, Y. (2013). First knockdown gene expression in bat (Hipposideros armiger) brain mediated by lentivirus. *Mol. Biotechnol.* 54, 564–571. doi:10.1007/s12033-012-9596-6.

Conard, N. J., Malina, M., and Münzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature* 460, 737–740. doi:10.1038/nature08169.

Derégnaucourt, S. (2011). Birdsong learning in the laboratory, with especial reference to the song of the Zebra Finch (Taeniopygia guttata). *Interact. Stud.* 12, 324–350. doi:10.1075/is.12.2.07der.

Dodenhoff, D. J. D., Stark, R. D. R., and Johnson, E. V (2001). Do woodpecker drums encode information for species recognition? *Condor* 103, 143–150. doi:10.1650/0010-5422(2001)103[0143:DWDEIF]2.0.CO;2.

Doupe, A. J., and Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annu. Rev. Neurosci.* 22, 567–631. doi:10.1146/annurev.neuro.22.1.567.

Drake, C., and El Heni, J. Ben (2003). Synchronizing with music: Intercultural differences. *Ann. N. Y. Acad. Sci.* 999, 429–437. doi:10.1196/annals.1284.053.

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. doi:10.1038/nature07994.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition* 100, 173–215. doi:10.1016/j.cognition.2005.11.009.

Fitch, W. T. (2013). Rhythmic cognition in humans and animals: Distinguishing meter and pulse perception. *Front. Syst. Neurosci.* 7, 68. doi:10.3389/fnsys.2013.00068.

Fitch, W. T. (2015). Four principles of bio-musicology. *Philos. Trans. R. Soc. B Biol. Sci.* 370, 20140091. doi:10.1098/rstb.2014.0091.

Fromkin, V., Krashen, S., Curtiss, S., Rigler, D., and Rigler, M. (1974). The development of language in genie: A case of language acquisition beyond the “critical period.” *Brain Lang.* 1, 81–107. doi:10.1016/0093-934X(74)90027-3.

Greenfield, M. D. (2005). “Mechanisms and Evolution of Communal Sexual Displays in Arthropods and Anurans,” in *Advances in the Study of Behavior* 35, 1–62. doi:10.1016/S0065-3454(05)35001-7.

Greenfield, M. D., and Roizen, I. (1993). Katydid synchronous chorusing is an evolutionarily stable outcome of female choice. *Nature* 364, 618–620. doi:10.1038/364618a0.

Griffith, S. C., and Buchanan, K. L. (2010). The Zebra Finch: The ultimate Australian supermodel. *Emu - Austral Ornithol.* 110, v–xii. doi:10.1071/MUv110n3_ED.

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 Biol.* 5, 2885–2897. doi:10.1371/journal.pbio.0050321.

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E. E., Lints, T., Jarvis, E. D., et al. (2004). FoxP2 expression in avian vocal learners and non-learners. *J. Neurosci.* 24, 3164–3175. doi:10.1523/JNEUROSCI.4369-03.2004.

Hage, S. R. (2018). Dual neural network model of speech and language evolution: New insights on flexibility of vocal production systems and involvement of frontal cortex. *Curr. Opin. Behav. Sci.* 21, 80–87. doi:10.1016/j.cobeha.2018.02.010.

Hahn, J., and Münzel, S. (1995). Knochenflöten aus dem Aurignacien des Geißenklösterle bei Blaubeuren, Alb-Donau-Kreis. *Fundberichte aus Baden-Württemb.* 20, 1–12.

Hargreaves, D. J. (1984). The effects of repetition on liking for music. *J. Res. Music Educ.* 32, 35–47. doi:10.2307/3345279.

Heinsohn, R., Zdenek, C. N., Cunningham, R. B., Endler, J. A., and Langmore, N. E. (2017). Tool-assisted rhythmic drumming in palm cockatoos shares key elements of human instrumental music. *Sci. Adv.* 3, e1602399. doi:10.1126/sciadv.1602399.

Helekar, S., Marsh, S., Viswanath, N., and Rosenfield, D. (2000). Acoustic pattern variations in the female-directed birdsongs of a colony of laboratory-bred zebra finches. *Behav. Processes* 49, 99–110.

Higham, T., Basell, L., Jacobi, R., Wood, R., Ramsey, C. B., and Conard, N. J. (2012). Testing models for the beginnings of the Aurignacian and the advent of figurative art and music: The radiocarbon chronology of Geißenklösterle. *J. Hum. Evol.* 62, 664–676. doi:10.1016/j.jhevol.2012.03.003.

Honing, H. (2018). On the biological basis of musicality. *Ann. N. Y. Acad. Sci.* 1423, 51–56. doi:10.1111/nyas.13638.

Honing, H., ten Cate, C., Peretz, I., and Trehub, S. E. (2015). Without it no music: Cognition, biology and evolution of musicality. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370, 20140088. doi:10.1098/rstb.2014.0088.

Hulse, S. H., and Page, S. C. (1988). Toward a comparative psychology of music perception. *Music Percept. An Interdiscip. J.* 5, 427–452. doi:10.2307/40285409.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA:

The MIT Press.

Huron, D. (2008). Science & music: lost in music. *Nature* 453, 456–457. doi:10.1038/453456a.

Hurst, J. A., Baraitser, M., Auger, E., Graham, F., and Norell, S. (1990). An extended family with a dominantly inherited speech disorder. *Dev. Med. Child Neurol.* 32, 352–355. doi:10.1111/j.1469-8749.1990.tb16948.x.

Immelmann, K. (1969). “Song development in the zebra finch and other estrildid finches,” in *Bird Vocalizations*, ed. R. A. Hinde (Cambridge, UK: Cambridge University Press), 64–74.

Janik, V. M., and Slater, P. J. B. (1997). Vocal Learning in Mammals. *Adv. Study Behav.* 26, 59–99. doi:10.1016/S0065-3454(08)60377-0.

Janik, V., and Slater, P. (2000). The different roles of social learning in vocal communication. *Anim. Behav.* 60, 1–11. doi:10.1006/anbe.2000.1410.

Janney, E., Taylor, H., Scharff, C., Rothenberg, D., Parra, L. C., and Tchernichovski, O. (2016). Temporal regularity increases with repertoire complexity in the Australian pied butcherbird’s song. *R. Soc. Open Sci.* 3. doi:10.1098/rsos.160357.

Jarvis, E. (2004). Learned birdsong and the neurobiology of human language. *Ann.* N. Y. *Acad. Sci.* 1016, 749–777.

Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., et al. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nat. Rev. Neurosci.* 6, 151–159. doi:10.1038/nrn1606.

Kahn, A. T., Holman, L., and Backwell, P. R. Y. (2014). Female preferences for timing in a fiddler crab with synchronous courtship waving displays. *Anim. Behav.* 98, 35–39. doi:10.1016/j.anbehav.2014.09.028.

Kao, M. H., and Brainard, M. S. (2006). Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *J. Neurophysiol.* 96, 1441–1455. doi:10.1152/jn.01138.2005.

Kiefer, S., Spiess, A., Kipper, S., Mundry, R., Sommer, C., Hultsch, H., et al. (2006). First-year common nightingales (Luscinia megarhynchos) have smaller song-type repertoire sizes than older males. *Ethology* 112, 1217–1224. doi:10.1111/j.1439-0310.2006.01283.x.

Knörnschild, M. (2014). Vocal production learning in bats. *Curr. Opin. Neurobiol.* 28, 80–85. doi:10.1016/j.conb.2014.06.014.

Knörnschild, M., Behr, O., and Von Helversen, O. (2006). Babbling behavior in the sac-winged bat (Saccopteryx bilineata). *Naturwissenschaften* 93, 451–454. doi:10.1007/s00114-006-0127-9.

Knörnschild, M., Nagy, M., Metz, M., Mayer, F., and von Helversen, O. (2010). Complex vocal imitation during ontogeny in a bat. *Biol. Lett.* 6, 156–159. doi:10.1098/rsbl.2009.0685.

Kojima, S., Kao, M. H., Doupe, A. J., and Brainard, M. S. (2018). The avian basal ganglia are a source of rapid behavioral variation that enables vocal motor exploration. *J. Neurosci.* 38, 9635–9647. doi:10.1523/JNEUROSCI.2915-17.2018.

Kunej, D., and Turk, I. (2001). “New perspectives on the beginnings of music: Archeological and musicological analysis of a middle Paleolithic bone ‘flute,’” in *The Origins of Music*, eds. N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 235–268.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., and Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413, 519–523. doi:10.1038/35097076.

Lattenkamp, E. Z., and Vernes, S. C. (2018). Vocal learning: a language-relevant trait in need of a broad cross-species approach. *Curr. Opin. Behav. Sci.* 21, 209–215. doi:10.1016/j.cobeha.2018.04.007.

Leach, J., and Fitch, J. (1995). Nature, music, and algorithmic composition. *Comput. Music J.* 19, 23. doi:10.2307/3680598.

Li, G., Wang, J., Rossiter, S. J., Jones, G., and Zhang, S. (2007). Accelerated FoxP2 evolution in echolocating bats. *PLoS One* 2, e900. doi:10.1371/journal.pone.0000900.

Li, S., Weidenfeld, J., and Morrisey, E. E. (2004). Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. *Society* 24, 809–822. doi:10.1128/MCB.24.2.809.

Longuet-Higgins, H. C. (1976). Perception of melodies. *Nature* 263, 646–653. doi:10.1038/263646a0.

Lu, M. M., Li, S., Yang, H., and Morrisey, E. E. (2002). Foxp4: a novel member of the Foxp subfamily of winged-helix genes co-expressed with Foxp1 and Foxp2 in pulmonary and gut tissues. *Gene Expr. Patterns* 2, 223–228. doi:10.1016/S1567-133X(02)00058-3.

Mann, N. I., and Slater, P. J. B. (1995). Song tutor choice by zebra finches in aviaries. *Anim. Behav.* 49, 811–820. doi:10.1016/0003-3472(95)80212-6.

Marler, P. (1970). Birdsong and speech development: Could there be parallels? *Am. Sci.* 58, 669–673.

Mehr, S. A., and Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. *Evol. Hum. Behav.* 38, 674–684. doi:10.1016/j.evolhumbehav.2016.12.005.

Mendoza, E., and Scharff, C. (2017). Protein-protein interaction among the FoxP family members and their regulation of two target genes, VLDLR and CNTNAP2 in the zebra finch song system. *Front. Mol. Neurosci.* 10, 1–15. doi:10.3389/fnmol.2017.00112.

Moiseff, A., and Copeland, J. (2010). Firefly synchrony: A behavioral strategy to minimize visual clutter. *Science* 329, 181. doi:10.1126/science.1190421.

Morgan, A., Fisher, S. E., Scheffer, I., and Hildebrand, M. (2016). “FOXP2-Related Speech and Language Disorders,” in *GeneReviews® [Internet]*, eds. M. P. Adam, H. H. Ardinger, and R. A. Pagon, et al. (Seattle, WA: University of Washington).

Morgan, A. T., and Webster, R. (2018). Aetiology of childhood apraxia of speech: A clinical practice update for paediatricians. *J. Paediatr. Child Health* 54, 1090–1095. doi:10.1111/jpc.14150.

Morley, I. (2014). A multi-disciplinary approach to the origins of music: Perspectives from anthropology, archaeology, cognition and behaviour. *J. Anthropol. Sci.* 92, 147–177. doi:10.4436/JASS.92008.

Möttönen, R., and Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* 29, 9819–9825. doi:10.1523/JNEUROSCI.6018-08.2009.

Mountjoy, D. J., and Lemon, R. E. (1995). Extended song learning in wild European starlings. *Anim. Behav.* 49, 357–366. doi:10.1006/anbe.1995.0048.

Nettl, B. (2001). “An ethnomusicologist contemplates universals in musical sound and musical culture,” in *The Origins of Music*, eds. N. K. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 463–472.

Norton, P., and Scharff, C. (2016). “Bird song metronomics”: Isochronous organization of zebra finch song rhythm. *Front. Neurosci.* 10, 309. doi:10.3389/fnins.2016.00309.

Nottebohm, F., and Nottebohm, M. E. (1978). Relationship between song repertoire and age in the canary, Serinus canarius. *Z. Tierpsychol.* 46, 298–305. doi:10.1111/j.1439-0310.1978.tb01451.x.

Nottebohm, F., Nottebohm, M. E., and Crane, L. (1986). Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei. *Behav. Neural Biol.* 46, 445–471. doi:10.1016/S0163-1047(86)90485-1.

Nottebohm, F., Stokes, T. M., and Leonard, C. M. (1976). Central control of song in the canary, Serinus canarius. *J. Comp. Neurol.* 165, 457–486. doi:10.1002/cne.901650405.

Oliver, R. (2013). “Groove as familiarity with time,” in *Music and Familiarity: Listening, Musicology and Performance*, eds. E. King and H. M. Prior (London, UK: Routledge), 239–252.

Ölveczky, B. P., Andalman, A. S., and Fee, M. S. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol.* 3, e153. doi:10.1371/journal.pbio.0030153.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Percept.* 11, 409–464. doi:10.2307/40285633.

Patel, A. D. (2008). *Music, Language, and the Brain*. New York, NY: Oxford University Press.

Pepperberg, I. M. (2010). Vocal learning in Grey parrots: A brief review of perception, production, and cross-species comparisons. *Brain Lang.* 115, 81–91. doi:10.1016/j.bandl.2009.11.002.

Petkov, C. I., and Jarvis, E. D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* 4, 12. doi:10.3389/fnevo.2012.00012.

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V, Wang, R., Roulhac, P. L., et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* 346, 1256846. doi:10.1126/science.1256846.

Prather, J. F., Okanoya, K., and Bolhuis, J. J. (2017). Brains for birds and babies: Neural parallels between birdsong and speech acquisition. *Neurosci. Biobehav. Rev.* 81, 225–237. doi:10.1016/j.neubiorev.2016.12.035.

Ravignani, A., Honing, H., and Kotz, S. A. (2017). Editorial: The evolution of rhythm cognition: Timing in music and speech. *Front. Hum. Neurosci.* 11, 303. doi:10.3389/fnhum.2017.00303.

Ravignani, A., and Norton, P. (2017). Measuring rhythmic complexity: A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations. *J. Lang. Evol.* 2, 4–19. doi:10.1093/jole/lzx002.

Reiner, A., Perkel, D. J., Bruce, L. L., Butler, A. B., Csillag, A., Kuenzel, W., et al. (2004). Revised nomenclature for avian telencephalon and some related brainstem nuclei. *J. Comp. Neurol.* 473, 377–414. doi:10.1002/cne.20118.

Riebel, K. (2009). Song and female mate choice in zebra finches: A review. *Adv. Study Behav.* 40, 197–238. doi:10.1016/S0065-3454(09)40006-8.

Roper, A., and Zann, R. (2006). The onset of song learning and song tutor selection in fledgling zebra finches. *Ethology* 112, 458–470. doi:10.1111/j.1439-0310.2005.01169.x.

Saar, S., and Mitra, P. P. (2008). A technique for characterizing the development of rhythms in bird song. *PLoS One* 3, e1461. doi:10.1371/journal.pone.0001461.

Sallavanti, M. I., Szilagyi, V. E., and Crawley, E. J. (2015). The role of complexity in music uses. *Psychol. Music* 44, 757–768. doi:10.1177/0305735615591843.

Salzinger, K., and Waller, M. B. (1962). The operant control of vocalization in the dog1. *J. Exp. Anal. Behav.* 5, 383–389. doi:10.1901/jeab.1962.5-383.

Sasahara, K., Tchernichovski, O., Takahasi, M., Suzuki, K., and Okanoya, K. (2015). A rhythm landscape approach to the developmental dynamics of birdsong. *J. R. Soc. Interface* 12, 20150802. doi:10.1098/rsif.2015.0802.

Savage, P. E., Brown, S., Sakai, E., and Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci.* 112, 8987–8992. doi:10.1073/pnas.1414495112.

Savage, P. E., Tierney, A. T., and Patel, A. D. (2017). Global music recordings support the motor constraint hypothesis for human and avian song contour. *Music Percept. An Interdiscip. J.* 34, 327–334. doi:10.1525/mp.2017.34.3.327.

Scharff, C., Knörnschild, M., and Jarvis, E. D. (in press). "Vocal learning and spoken language: Insights from animal models with an emphasis on genetic contributions," in *Language in Interaction: The Human Language Faculty from Genes to Behavior*, ed. P. Hagoort (Cambridge, MA: MIT Press).

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. *J. Neurosci.* 11, 2896–2913.

Scharff, C., Nottebohm, F., and Cynx, J. (1998). Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway. *J. Neurobiol.* 36, 81–90. doi:10.1002/(SICI)1097-4695(199807)36:1<81::AID-NEU7>3.0.CO;2-6.

Scharff, C., and Petri, J. (2011). Evo-devo, deep homology and FoxP2: Implications for the evolution of speech and language. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 366, 2124–2140. doi:10.1098/rstb.2011.0001.

Schatton, A., and Scharff, C. (2016). Next stop: Language. The 'FOXP2' gene's journey through time. *Mètode Rev. difusió la Investig.* 7, 25–33. doi:10.7203/metode.7.7248.

Schusterman, R. J. (2008). "Vocal learning in mammals with special emphasis on pinnipeds," in *Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication*, eds. D. K. Oller and U. Griebel (Cambridge, MA: MIT Press), 41–70. doi:10.7551/mitpress/9780262151214.003.0003.

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

Sin, C., Li, H., and Crawford, D. A. (2015). Transcriptional regulation by FOXP1, FOXP2, and FOXP4 dimerization. *J. Mol. Neurosci.* 55, 437–448. doi:10.1007/s12031-014-0359-7.

Sismondo, E. (1990). Synchronous, alternating, and phase-locked stridulation by a tropical katydid. *Science* 249, 55–58. doi:10.1126/science.249.4964.55.

Stefanics, G., Háden, G. P., Sziller, I., Balázs, L., Beke, A., and Winkler, I. (2009). Newborn infants process pitch intervals. *Clin. Neurophysiol.* 120, 304–308. doi:10.1016/j.clinph.2008.11.020.

Taylor, H. (2008). Decoding the song of the pied butcherbird: An initial survey. *Rev. Transcult.*

Música Transcult. Music Rev. 12, 24–25.

Taylor, H. (2014). Whose bird is it? Messiaen's transcriptions of Australian songbirds. *Twentieth-Century Music* 11, 63–100. doi:10.1017/S1478572213000194.

Taylor, H., and Lestel, D. (2011). The Australian pied butcherbird and the natureculture continuum. *J. Interdiscip. Music Stud.* 5, 57–83.

Teramitsu, I., Poopatanapong, A., Torrisi, S., and White, S. A. (2010). Striatal FoxP2 is actively regulated during songbird sensorimotor learning. *PLoS One* 5. doi:10.1371/journal.pone.0008548.

Teufel, A., Wong, E. A., Mukhopadhyay, M., Malik, N., and Westphal, H. (2003). FoxP4, a novel forkhead transcription factor. *Biochim. Biophys. Acta - Gene Struct. Expr.* 1627, 147–152. doi:10.1016/S0167-4781(03)00074-5.

Thompson, J. A., Basista, M. J., Wu, W., Bertram, R., and Johnson, F. (2011). Dual pre-motor contribution to songbird syllable variation. *J. Neurosci.* 31, 322–330. doi:10.1523/JNEUROSCI.5967-09.2011.

Thorpe, W. H. (1958). The learning of song patterns by birds, with especial reference to the song of the chaffinch Fringilla coelebs. *Ibis.* 100, 535–570. doi:10.1111/j.1474-919X.1958.tb07960.x.

Tierney, A., Russo, F. a, and Patel, A. D. (2011). The motor origins of human and avian song structure. *Proc. Natl. Acad. Sci. U. S. A.* 108, 15510–15515. doi:10.1073/pnas.1103882108.

Toiviainen, P., and Eerola, T. (2003). Where is the beat? Comparison of Finnish and South-African listeners. *Proc. 5th Trienn. ESCOM Conf.* 5, 6–9.

Tomlinson, G. (2015). *A million years of music: The emergence of human modernity*. Cambridge, MA: MIT Press.

Trehub, S. (2001). "Human processing predispositions and musical universals," in *The Origins of Music*, eds. N. K. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 427–448.

Trehub, S. E., Becker, J., and Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370, 20140096. doi:10.1098/rstb.2014.0096.

van Noorden, L., and Moelants, D. (1999). Resonance in the perception of musical pulse. *J. New Music Res.* 28, 43–66. doi:10.1076/jnmr.28.1.43.3122.

Wheatcroft, D., and Qvarnström, A. (2015). A blueprint for vocal learning: Auditory predispositions from brains to genomes. *Biol. Lett.* 11; 20150155. doi:10.1098/rsbl.2015.0155.

Wild, J. M. (1997). Neural pathways for the control of birdsong production. *J. Neurobiol.* 33, 653–670. doi:10.1002/(SICI)1097-4695(19971105)33:5<653::AID-NEU11>3.0.CO;2-A.

Wild, J. M., Goller, F., and Suthers, R. a (1998). Inspiratory muscle activity during bird song. *J. Neurobiol.* 36, 441–453.

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., and Honing, H. (2009). Newborn infants detect the beat in music. *Proc. Natl. Acad. Sci. U. S. A.* 106, 2468–2471. doi:10.1073/pnas.0809035106.

Yamaguchi, A., Cavin Barnes, J., and Appleby, T. (2017). Rhythm generation, coordination, and initiation in the vocal pathways of male African clawed frogs. *J. Neurophysiol.* 117, 178–194. doi:10.1152/jn.00628.2016.

Zann, R. (1990). Song and call learning in wild zebra finches in south-east Australia. *Anim. Behav.* 40, 811–828. doi:10.1016/S0003-3472(05)80982-0.

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

Publication A: Measuring rhythmic complexity

Ravignani, A.*, and **Norton, P.*** (2017). Measuring rhythmic complexity: A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations. *Journal of Language Evolution* 2, 4–19. doi:10.1093/jole/lzx002.

* The authors contributed equally to this study.

Measuring rhythmic complexity: A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations

Andrea Ravignani[1,2,*] & Philipp Norton[3,*]

[1]Artificial Intelligence Lab, Vrije Universiteit Brussel
[2]Language and Cognition Department, Max Planck Institute for Psycholinguistics
[3]Department of Animal Behaviour, Freie Universität Berlin
* The authors contributed equally to this study

Abstract

Research on the evolution of human speech and phonology benefits from the comparative approach: Structural, spectral, and temporal features can be extracted and compared across species in an attempt to reconstruct the evolutionary history of human speech. Here we focus on analytical tools to measure and compare temporal structure in human speech and animal vocalizations. We introduce the reader to a range of statistical methods usable, on the one hand, to quantify rhythmic complexity in single vocalizations, and on the other hand, to compare rhythmic structure between multiple vocalizations. These methods include: time series analysis, distributional measures, variability metrics, Fourier transform, auto- and cross-correlation, phase portraits, and circular statistics. Using computer-generated data, we apply a range of techniques, walking the reader through the necessary software and its functions. We describe which techniques are most appropriate to test particular hypotheses on rhythmic structure, and provide possible interpretations of the tests. These techniques can be equally well applied to find rhythmic structure in gesture, movement and any other behaviour developing over time, when the research focus lies on their temporal structure. This introduction to quantitative techniques for rhythm and timing analysis will hopefully spur additional comparative research, and will produce comparable results across all disciplines working on the evolution of speech, ultimately advancing the field.

1. Introduction

Research on the evolution of speech has greatly benefitted from the comparative approach (Fitch, 2000; Yip, 2006; Rauschecker & Scott, 2009; Fedurek & Slocombe, 2011; Scharff & Petri, 2011; de Boer, 2012; Rauschecker, 2012; Ghazanfar, 2013; Lameira, Maddieson & Zuberbühler, 2014; Lameira et al., 2015; Fehér, 2016; Filippi, 2016; Gustison & Bergman, under review). In fact, relating human speech, and animal vocal production learning can inform the evolution of speech both by comparing the signals produced across species, and by inverse inference, comparing the neurocognitive machinery used to generate the signals (Bolhuis, Okanoya, & Scharff, 2010; Fitch, 2014; Bowling & Fitch, 2015; Collier et al., 2014).

Three main avenues seem particularly relevant to the comparative study of the evolution of speech (**Figure 1**).

Positional regularities: how the building blocks of speech or language, each taken holistically, are organized and related to each other (Fitch, 2014; Kershenbaum et al., 2014; ten Cate, 2016).

Spectral characteristics: how different frequencies and their intensities relate to one another, giving rise to tone in speech, liaison, vowel quality, harmonicity, etc. (Fitch, 2000; Yip, 2006; Spierings & ten Cate; 2016).

Temporal structure: how the speech/vocal signal is organized and develops over time (Ramus, Nespor, & Mehler, 1999; Goswami, & Leong, 2013; Bekius et al., 2016; Benichov et al., 2016; Hannon et al., 2016; Jadoul et al., 2016). Timing, together with some spectral characteristics, contributes to rhythm (see **Table 1** for a definition of rhythm and other major concepts presented in this paper).

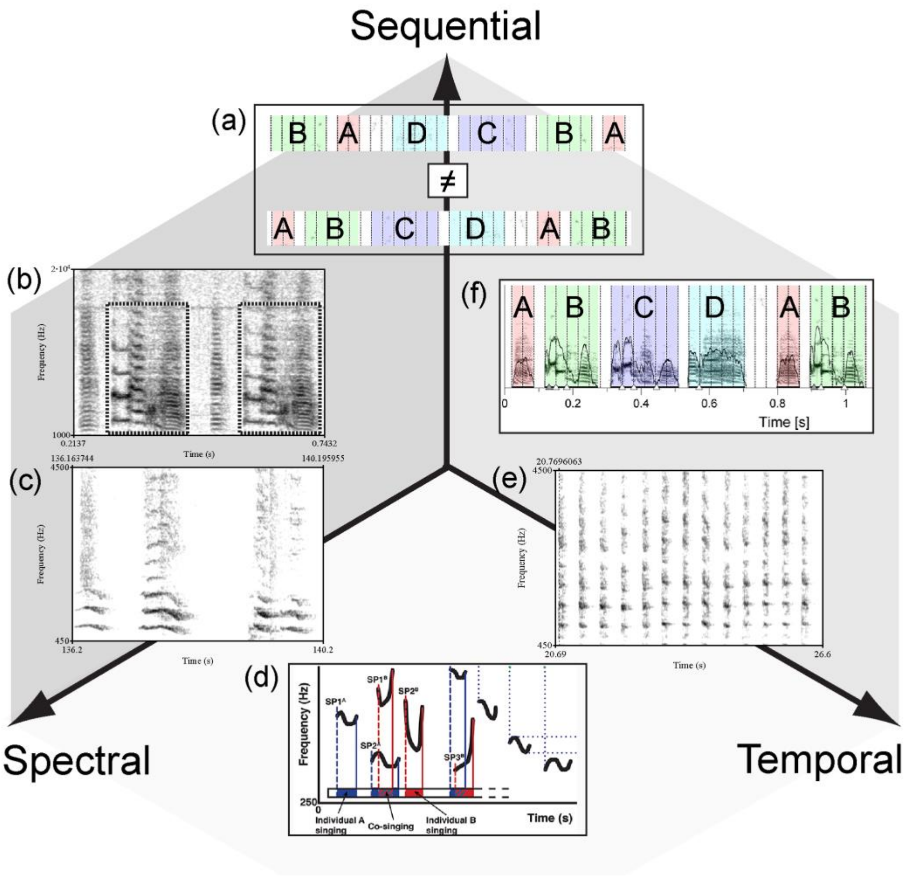


Figure 1 – Three approaches to the analysis of animal vocalizations, including human speech, laying emphasis on sequential, spectral or temporal features of the signal. The three-dimensional space presents a common multicomponent approach to speech and animal vocalizations, showing that these components constitute complementary dimensions, rather than a mutually exclusive classification. Depending on the species and scientific question, one dimension might be particularly relevant. Research focus on structural, spectral and temporal information is exemplified along the three main axes, with many cases of vocalizations falling between two categories. Counter-clockwise, the figure shows structures where a possible research emphasis is laid on: (a) *Sequential:* an abstract representation of a series of song notes, and its mirror permutation, containing the same elements and having the same duration but different sequential properties; (b) *Spectro-sequential:* spectrogram of a chunk of zebra finch song (*Taeniopygia guttata*), where complex spectral information is combined into a sequential repetitive structure (dashed boxes); (c) *Spectral:* spectrogram of a harbour seal pup call (*Phoca vitulina*), showing harmonic features, with less emphasis on structural/positional regularities; (d) *Spectro-temporal:* schematic representation of a spectrogram showing a singing lemur vocal duet (*Indri indri*), where spectral and temporal information might be interacting (Gamba et al., 2016); (e) *Temporal:* spectrogram of a California sea lion's bark (*Zalophus californianus*), showing a clear rhythmic, isochronous pattern: barks contain little spectral information and variability, but are produced with remarkable regularity, like a metronome; (f) *Tempo-sequential:* notes composing a zebra finch song, which in turn possesses an underlying rhythmic structure (Norton and Scharff, 2016). Notice that these examples concern the research focus, rather than the nature of the signal: for instance, panel (e) is taken as a prototype of temporal signal, but it also has a clear spectral structure. This paper focuses on useful methods to analyse cases (d), (e) and (f), that is, when emphasis is put on analysing the temporal structure of vocalizations. Panel (d) reproduced and modified from Gamba et al., 2016 and panels (a) and (f) from Norton and Scharff, 2016, both published open access under the CC BY 4.0. Other panels generated using Praat (Boersma & Weeninck, 2013).

Table 1 – Definition of crucial concepts in comparative rhythm research

Name (abbreviation)	**Definition**
Rhythm	Structure over time; durations marked by (acoustic) events (McAuley, 2010)
Isochronous	A series of events repeating at a constant rate
Synchronous	A series of events individually occurring at the same time as events from another series
Time series	Sequence of events occurring over time, sampled at regular time intervals
Inter-onset interval (IOI)	Time elapsed between the beginning of one event (i.e. onset) and the beginning of the next event
Meter	Hierarchical organization of timed events based on spectral properties (stress, e.g. loudness alternation, pitch variation)
Syllable-timed	Language in which all syllables (both accented and unaccented) are roughly isochronous
Stress-timed	Language with an isochronous occurrence of stressed syllables
Mora-timed	Language in which all moras (syllables and some combinations thereof) are isochronous
Pulse	Isochronous grid most suitable to a given temporal pattern.
Iambic	Metrical form alternating a weak (unstressed) syllable with a strong (stressed) syllable
Trochaic	Metrical form alternating a strong syllable with a weak syllable

Rhythmic properties of human speech (Ramus, et al., 1999; Grabe & Low, 2002; Tilsen, 2009; Lehiste, 1977) and animal vocal communication (Saar & Mitra, 2008; Norton & Scharff, 2016; Ravignani et al., 2016) are often investigated. However, lack of an interdisciplinary statistical approach to rhythmicity of vocalizations across species hinders comparative and comparable research.

Here we introduce the reader to a broad range of quantitative methods useful for rhythm research within and across species. The paper is structured as follows. A first distinction between methods concerns measuring either rhythmic *complexity of one pattern*, that is, how regular and predictable a sequence of events is (section 2), or *rhythmic relationships* between multiple patterns, that is, similarities in the temporal structure of two or more sequences of events (section 3).

In section 2 we characterize inter-onset intervals, used to measure temporal information, and show how rhythmicity can be quantified using: distributional measures, such as histograms (2.2), Kolmogorov-Smirnov D (2.3) and normalized pairwise variability index (2.4); autocorrelation (2.5); beat-finding algorithms, such as Fourier transform (2.6) and Pulse generate-and-test (2.7); time series analysis (2.10). We also present two powerful techniques to visualize regularities in rhythmic patterns: phase portraits (2.8) and recurrence plots (2.9).

Section 3 extends section 2 by focusing on measures of temporal similarity between two patterns, namely: cross-correlation (subsection 3.1); multidimensional time series analysis, testing how one pattern can be linearly estimated (3.2) or causally predicted (3.3) from the other; statistical analysis of circular data, including visualization with rose plots (3.4).

Section 4 describes some research areas where the techniques we present may be particularly useful. Section 5 provides some overall conclusions. Section 6 describes how to access the data and scripts used in this paper. To get a better understanding of the methods presented in this article, we encourage the reader to use the provided code (Supplementary Material) to experiment with different patterns. New versions of the code will be uploaded at http://userpage.fu-berlin.de/phno/mrc/.

2. Within-pattern analytical techniques: Quantifying rhythmicity

2.1. Overview

The temporal structure of vocalizations can be analyzed on the raw audio signal, or more commonly on time series extracted from the signal, like the amplitude envelope. Often however, one is interested in the timing of certain discrete events, e.g. syllable onsets, amplitude peaks, or any other behaviour developing over time. Some of the methods described here apply to patterns represented by their inter-onset intervals (IOIs), while others – like autocorrelation and Fourier transform – are more readily applied to a time series, i.e. a sequence of values sampled at equally spaced points in time (see Supplementary Material for details).

To demonstrate the methods described in this section, we generated 5 sequences that differ in their rhythmic structure, each consisting of 24 events (**Figure 2**, left column). To simplify notation, we demonstrate techniques for events with IOIs in the range of 1–6 seconds (s), though all presented techniques can be used at any time scale, depending on the observed behavior.

2.2. IOI distribution: Which intervals occur in the pattern, and how often?

A histogram of all IOIs in a sequence provides a first visualization of its rhythmic structure (**Figure 2**, right column). A normal distribution suggests a roughly isochronous pattern, where a lower spread indicates a more isochronous pattern (**Figure 2.b**). The existence of a few distinct IOI categories may appear as a multimodal distribution (**Figure 2.c**). Uniformly distributed IOIs might hint at a lack of such categories or a rhythmic structure in general (**Figure 2.a**). However, when many durational categories exist, a uniform distribution might also appear, concealing higher order structure (**Figure 2.d**). A particular limitation of histograms and statistical measures of durational distributions is their lack of sensitivity to structure: for instance, any permutation of the order of IOIs will change the structure while leaving distributional measures unaltered.

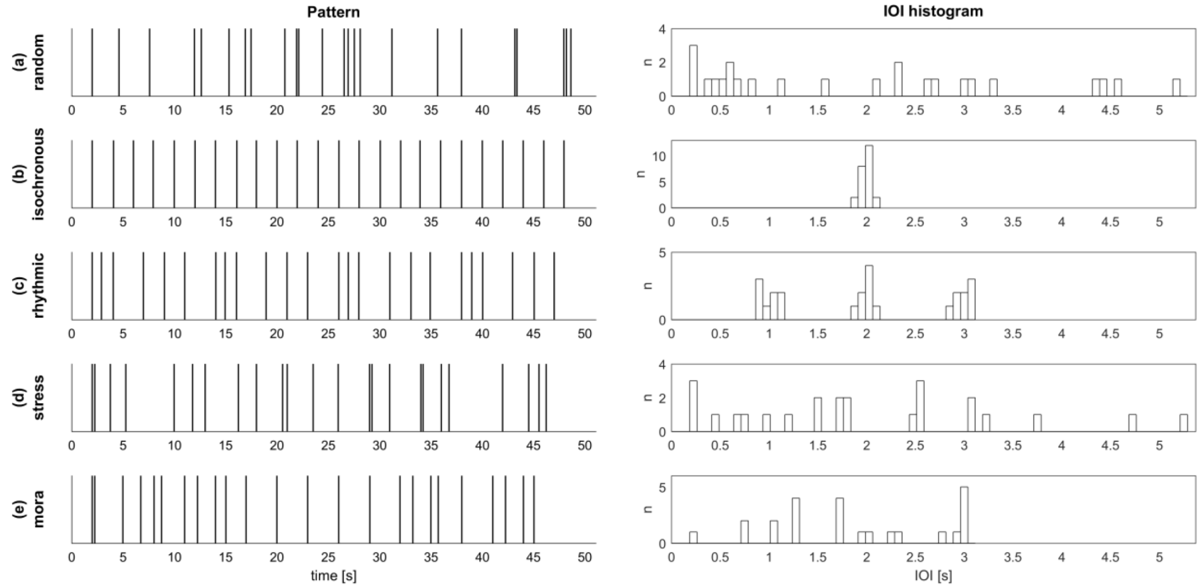


Figure 2 – Five different example patterns that differ in their rhythmic structure (a–e, left column), as well as histograms of their inter-onset intervals (IOI, right column). (a) *Random sequence*: The time points of events of the first sequence are pseudorandomly drawn from a uniform distribution. (b) *Isochronous sequence*: The second sequence is isochronous, i.e. IOIs of successive events are equal for all events (in this case 2s). This sequence with added noise is taken to loosely represent the hypothesized syllable-timed structure of some languages (Grabe & Low, 2002). (c) *Rhythmic sequence*: The third pattern contains 4 repetitions of a sub-pattern consisting of 6 events (IOIs: 1s, 1s, 3s, 2s, 2s, 3s). (d–e) The final two sequences are loosely based on concepts of timing division in language: stress timing and mora timing. (d) The *stress sequence* consists of clusters of 4 events each. The duration of all clusters is 8s and the pattern of IOIs within a cluster are different each time (e.g. [0.5,1.5,3.75,2.25], [2,1,3.5,1.5], …). (e) The *mora sequence* contains clusters of 3s duration each, that either contain one or two events (e.g. [1,2], [3], [2.25,0.75], …). Gaussian noise with a standard deviation of 0.04s (isochronous and rhythmic sequence) or 0.02s (stress and mora sequence) was added to the timing of the events.

2.3. Kolmogorov-Smirnov D: Does interval distribution differ from any hypothesized distribution?

The one-sample Kolmogorov-Smirnov test (K-S test, **Table 2**) can be used to evaluate how different an observed IOI distribution is from a particular, specified distribution, e.g. a normal distribution in the isochronous case, or a uniform distribution in the case of no actual structure (Lilliefors, 1967). The test uses the statistic D, measuring the maximum distance between two cumulative distribution functions. This measure can be used to compare e.g. different languages based on the normality of the intervals between syllable nuclei (Jadoul et al., 2016). When compared to a normal distribution, the isochronous pattern in **Figure 2.b** has a D of 0.085, about half of the 0.14 of the mora pattern (**Figure 2.e**). In order to statistically test an IOI distribution for normality, a modification of the K-S test by Lilliefors can be used instead (Lilliefors, 1967). Of the five example patterns in **Figure 2** only the random and rhythmic patterns are significantly non-normal according to the Lilliefors test (p=0.035 and p=0.024 respectively; n=24).

Table 2 – Analytical techniques to unveil temporal structure in one pattern, including their specific function.

Technique (References)	Function	Advantage	Disadvantage
Kolmogorov-Smirnov D (Jadoul et al., 2016)	One number describing the distribution of a sequence of IOI	Common and well-studied statistical tool	Not an absolute number, instead the relative distance from a specific hypothesized distribution
nPVI (Grabe & Low, 2002; Jadoul et al., 2016)	One number describing temporal variability of a sequence of IOI, taking only information about adjacent intervals into account	Summarizes temporal regularity with one number	Not very robust to different speakers and replications
Autocorrelation (Ravignani & Sonnweber, 2015; Hamilton, 1994)	Probes the existence of repeating subpatterns within a pattern	Few assumptions required	Provides only little information on the pattern (difficult to map autocorrelations to necessary/sufficient conditions on pattern structure)
Fourier analysis (Saar & Mitra, 2008; Norton & Scharff, 2016)	Decomposes signal into sum of isochronous pulses	Common, fast method in signal analysis	May return several best fitting pulses

GAT Pulse matching (Norton & Scharff, 2016)	Finds a metronomic pulse best fitting the pattern	Aimed and proficient at finding the slowest pulse in the pattern	Computationally intensive; less used than Fourier analysis
Phase portrait (Rothenberg et al., 2014; Wagner, 2007; Ravignani, in press)	Enables temporal regularities to be visualized as geometric regularities	Easy to plot, straightforward interpretation	Not an analytical, rather a visualization technique
Recurrence plots (Thiel et al., 2004; Coco & Dale, 2014)	Visualizes temporal regularities among non-adjacent elements	Provides a quick glance at higher-order regularities	Sensitive to the initial parameters (threshold for considering two IOIs similar), which can produce false positives/negatives
ARMA (Hamilton, 1994; Jadoul et al., in review)	Tests whether each IOI can be expressed as a linear combination of previous IOI	Most common method to analyse time series; based on much theoretical work; many statistical packages available; using Akaike sets increases robustness	Only captures linear relations

2.4. The normalized pairwise variability index: How variable are adjacent intervals?

The normalized pairwise variability index (nPVI) is a measure originally developed to quantify temporal variability in speech rhythm (**Table 2**; Grabe & Low, 2002; Toussaint, 2013). It is computed as follows. For each *pair of adjacent* IOIs, one calculates their difference and divides it by their average. The nPVI equals the average of all these ratios, multiplied by 100, namely:

$$nPVI = \frac{100}{n-1}\left[\left|\frac{IOI_2 - IOI_1}{\frac{IOI_2 + IOI_1}{2}}\right| + \cdots + \left|\frac{IOI_n - IOI_{n-1}}{\frac{IOI_n + IOI_{n-1}}{2}}\right|\right] \qquad [1]$$

The nPVI equals zero for a perfectly isochronous sequence, and it increases with an increasing *alternation* in onset timing. Some have suggested nPVI captures cross-linguistic rhythmic classes, with syllable-timed languages exhibiting low nPVI, stress-timed languages having high nPVI, and mora-timed languages lying between the two extremes (Grabe & Low, 2002). The nPVI values of our computer-generated example patterns align with this classification: The isochronous pattern, representing syllable-timing, has a low nPVI of 3.3, the stress pattern a high value of 92.74 (close to the random pattern with 94.54), while the mora pattern lies in between with 52.84. Notice however that nPVI has little explanatory power beyond simple, zeroth-order distributional statistics, such as Kolmogorov-Smirnov D (Jadoul et al., 2016; Ravignani, in press).

2.5. Autocorrelation: How similar is a pattern to itself at different points in time?

Autocorrelation is a technique that can help reveal higher order structure within a pattern, specifically repeating temporal sub-patterns (Hamilton, 1994; Ravignani, & Sonnweber, 2015). Autocorrelation correlates a time series (**Figure 3.b**) with a copy of itself at different time lags (**Figure 3.c–f**, red dotted line). It is calculated by having one copy of the signal slide stepwise across the other; at each step the products of all points in the two signals are calculated and added together. The result of the process is a function of these sums at the different time lags between the two signals (**Figure 3.h**). The autocorrelation starts at a time lag of zero, where the two copies overlap perfectly (**Figure 3.c**). At zero lag the autocorrelation function is normalized to 1 (**Figure 3.h**). Since the signal is non-zero only at the time points of the events, the function will be zero at the lags at which none of the events overlap (**Figure 3.d&h**). When one or more events in the two signals overlap a peak appears in the function (**Figure 3.e&h**). The more events overlap at a certain lag, the higher the peak in the function (**Figure 3.f&h**), suggesting that a sub-pattern might repeat after a duration that equals this lag. Note that unless the IOIs are exactly equal in duration, the events in the raw time series will not overlap. Real world patterns, however, are rarely isochronous and any rhythmic structure usually contains some amount of noise or jitter. To deal with noise one can convolute the time series with a narrow normal probability density function prior to autocorrelation, effectively replacing every single time point event with narrow Gauss curves (**Figure 3.a&b**). Several nearby, though not simultaneously overlapping, events can thereby partially overlap over a range of lags and together contribute to a single peak in the autocorrelation function (e.g. **Figure 3.e**).

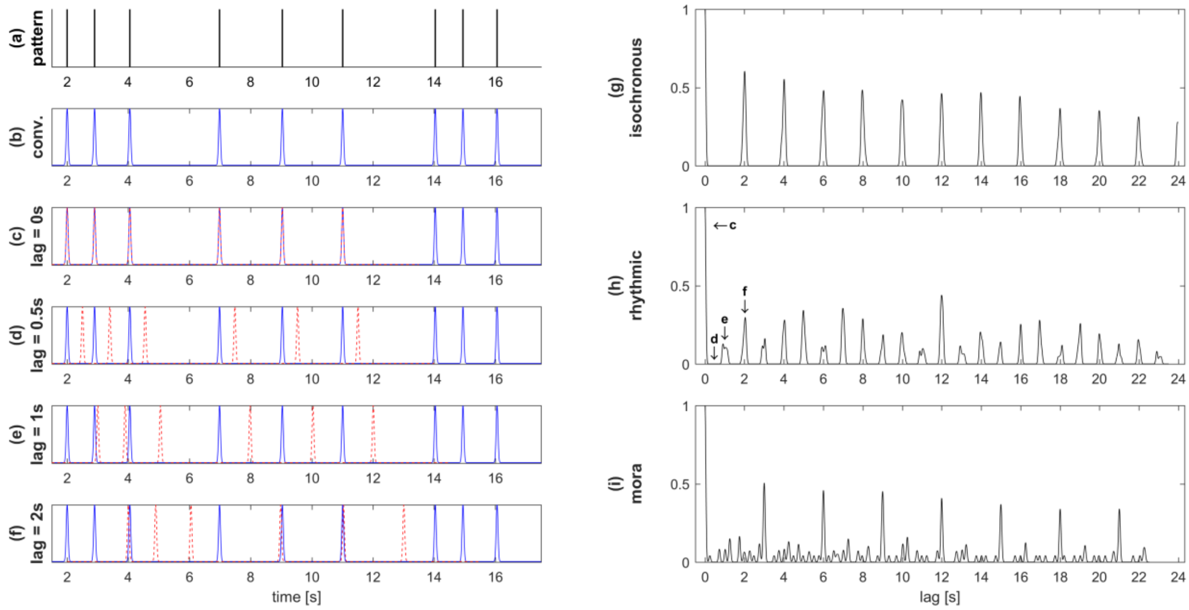


Figure 3 – Visualization of the autocorrelation process (a–f) and the resulting autocorrelation function for three of the example patterns (g–i).(a) The first nine events of the rhythmic sequence depicted in **Figure 2.c**. (b) The same events converted to a time series and convoluted with a narrow normal probability density function (npdf). (c) At zero lag the stationary copy of the time series (blue, continuous line) and the time-lagged copy (red, dotted line) overlap perfectly. For the sake of the example the time lagged copy is only shown for the first six events (i.e. the first repetition of the sub-pattern). (d) At a time lag of 0.5s none of the events overlap, i.e. across the whole time range at least one of the two signals is zero. (e) At 1s lag events 1 and 2 of the lagged copy partially overlap with events 2 and 3 of the stationary signal. (f) At a lag of 2s, three of the events overlap, resulting in a higher peak ("f" in h). The peak is also narrower than the one at 1s lag, indicating that the events overlap more closely. (g) Autocorrelation function of the isochronous sequence, depicted in **Figure 2.b**. (h) Autocorrelation function of the rhythmic sequence, seen in **Figure 2.c**. The letters c–f point to the lags that are depicted in (c–f). (i) Autocorrelation function of the mora sequence (see **Figure 2.e**).

Figure 2.b shows an isochronous pattern (IOI ≈ 2s) of 24 events with a small amount of Gaussian noise added. At a lag of 2s a peak close to one appears in the autocorrelation function, as events 2–24 of the signal overlap with events 1–23 of the lagged signal (**Figure 3.g**). The height of the following peaks – regularly occurring at n*2s – gradually decreases as the lagged signal moves beyond the original signal and fewer events overlap.

The “rhythmic sequence” contains 3 repetitions of a random sub-pattern consisting of 6 events with IOIs equaling 1, 2, or 3s (total duration = 12s; **Figure 2.c**). Looking at the autocorrelation plot of this pattern, the first peak appears at a lag of 1s, where some of the events start to overlap, and all following peaks are at lags 2,3,4,...s (**Figure 3.h**). This suggests 1s is the basic time unit of this pattern, i.e. all IOIs are multiples of 1s. The largest peak is located at 12s lag, because 12s is the total duration of the repeating sub-pattern. At this point all events of sub-patterns 2, 3 and 4 of the signal overlap with sub-patterns 1, 2 and 3 of the lagged signal. Note that a higher peak could potentially occur at a lower lag, if more than three quarters of events were to overlap. Relatively high peaks are likely to appear at harmonics (i.e. multiples) of the 12s lag, namely at 24 and 36s where the following repetitions of the sub-pattern overlap. In the “mora sequence” events occur every 3s as well as at varying time points in between (**Figure 2.e**). Accordingly, the autocorrelation function shows peaks at 3s, 6s, etc., as well as some noise stemming from less regularly occurring events (**Figure 3.i**).

2.6. Fourier transform: Decomposing the pattern into a sum of regular, clock-like oscillators

The Fourier transform is used to express any signal as a sum of sine waves of different frequencies and phases. In acoustics, one application is to decompose a sound wave into its constituent frequencies. When applied to the time series of a rhythmic pattern, the Fourier transform finds periodicities in the timing of the events. A particularly important periodicity is the pulse: the slowest isochronous sequence to which most events align. The Fourier transform can be visualized in a power spectrum, where the x-axis denotes the frequencies of the waves and the y-axis their magnitude, i.e. how much they contribute to the signal (**Figure 4**).

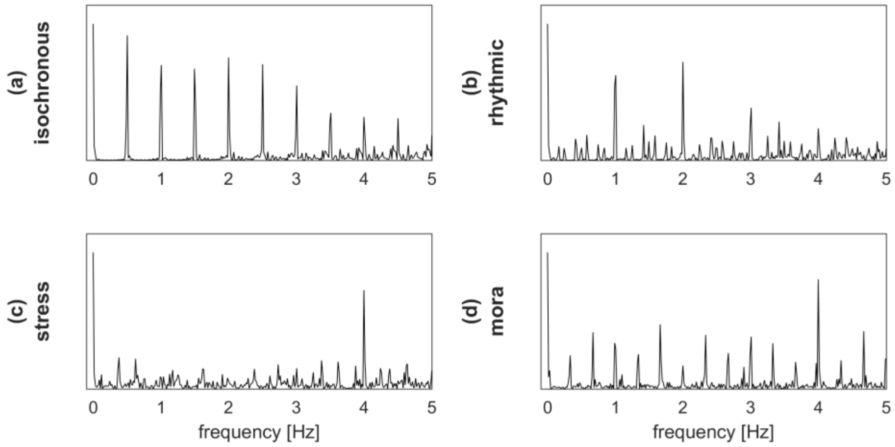


Figure 4 – Power spectra of four of the example patterns. The power spectrum is the result of a Fourier transform, which decomposes a signal into waves of different frequencies. It shows in what magnitude (y-axis) the different frequencies (x-axis) contribute to the signal.

In the “isochronous sequence” the first and largest peak in the power spectrum is located at 0.5Hz, corresponding to the pulse of the pattern, i.e. the frequency at which the events regularly occur, namely every 2 seconds (**Figure 4.a**). If a signal is a sine wave, the power spectrum will consist of only this peak. However, often the signal is a sequence of narrow spikes, hence additional waves are needed to reconstruct the signal. This explains the higher frequency waves in the power spectrum in **Figure 4.a**.

In the stress and mora patterns, the maximum peaks in the power spectra are located at 4Hz (**Figure 4.c&d**) because all IOIs are integer multiples of 0.25s, so a 4Hz pulse coincides with all events.

2.7. Pulse GAT: Which isochronous grid fits the pattern best?

A more direct, albeit computationally intensive, approach to finding the pulse of a sequence is “generate-and-test” (GAT). First, a low frequency, isochronous grid or pulse is created in the form of regularly spaced timestamps (**Figure 5.a**). The deviation of each event in the pattern to its nearest grid element is measured and the root-mean-squared deviation of the whole pattern, multiplied by the grid frequency (frmsd) is calculated. The grid is then shifted forward in time in small steps in order to find the best fit (i.e. lowest frmsd) for the grid of that particular frequency (**Figure 5.b**). Next, the grid frequency is increased in small steps and the minimal frmsd is again calculated for each step (**Figure 5.c&d**). If a rhythmic sequence has an underlying isochronous regularity, its frequency can be determined by finding the grid frequency with the overall lowest frmsd value. This GAT approach was recently used to find isochrony in zebra finch song (Norton & Scharff, 2016).

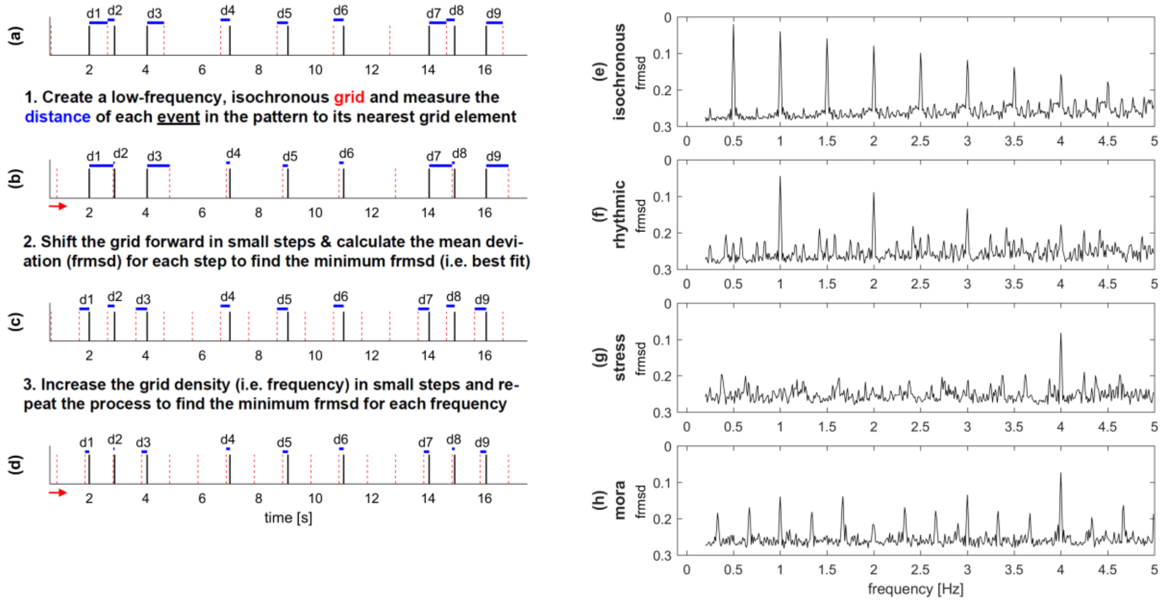


Figure 5 – Visualization of the pulse generate-and-test (GAT) method (a–d), and the result of its application to four of the example patterns (e–h). Note that the y-axis in (e–h) is inverted. (a–d) Depicted are the first nine events of the rhythmic sequence (black vertical lines) and the deviations (d1–d9, blue lines) to their nearest element of the isochronous grid (red dotted lines). Example grid frequencies are 0.5Hz (a&b) and 1Hz (c&d). (e) The isochronous sequence, where all events occur roughly 2s apart, best fits a pulse of 0.5Hz. (f) In the rhythmic sequence all IOIs are multiples of one (1, 2 and 3), so a pulse of 1Hz has the lowest frmsd (see also d). (g&h) In the mora and stress sequences all IOIs are multiples of 0.25s (0.25, 0.5, 0.75, …), so a pulse of 4Hz fits best to both.

In many cases, the GAT frequency with the lowest frmsd corresponds to the frequency of highest power in the Fourier transform. The GAT method, however, is specifically aimed at finding the slowest pulse that still fits all events in the sequence, also called the *tatum* (Bilmes, 1993). In our computer-generated “rhythmic sequence”, for example, the tatum has a frequency of 1Hz, because all IOIs (1s, 2s and 3s) are integer multiples of 1s (**Figure 5.f**). However, the frequency of highest power in the Fourier transform is located at twice the frequency, 2Hz (**Figure 4.b**). While a pulse of 2Hz covers all events as well, the slower pulse of 1Hz is enough to explain the underlying rhythmic structure. An advantage of the GAT method is that it provides the frmsd as a measure for goodness of fit. This can be used to statistically compare different patterns in terms of their pulse fidelity, albeit only to pulses of similar frequency, since the frmsd is frequency-dependent.

2.8. Phase portraits: How to visualize repeating rhythmic patters

Unlike IOI histograms, phase portraits visualize the durations of all IOIs of a pattern, while taking their first order sequential structure (i.e. adjacent IOIs) into account. Pairs of adjacent IOIs hereby serve as x- and y-coordinates respectively. A line connects coordinates of neighboring pairs. Together these lines form a continuous trajectory on a 2-dimensional plane. In the "rhythmic sequence" (**Figure 2.c**) for example, the first two IOIs are both 1s long, hence a dot is plotted at coordinates (1,1) (**Figure 6.a**). The next dot is plotted at coordinates (1,3), corresponding to the second and third IOI, which are 1s and 3s respectively. The trajectory connects these coordinates and moves from (1,1) to (1,3) to (3,2) and so on until it reaches (1,1) again at the end of the first sub-pattern (**Figure 6.b–d**). As the pattern repeats, the trajectory traces the same path three more times, albeit with slight deviations due to the noise in the timing of events (**Figure 6.e**).

Geometrical regularities in phase portraits correspond to rhythmic regularities in the acoustic patterns (Ravignani, in press; Ravignani, Delgado & Kirby, 2016). A repeated rhythmic pattern corresponds to similar superimposed polygons. A cyclical permutation of a rhythmic pattern (e.g. from 1,3,2,2 to 2,2,1,3) will produce the same polygon from a different starting point. Reversing the pattern (e.g. from 1,2,3 to 3,2,1) will result in the same, though rotated, polygon. Patterns that are palindromic on any phase (the "rhythmic sequence" for example has a palindromic cyclical permutation: 1,3,2,2,3,1) appear as polygons symmetrical with respect to the diagonal (**Figure 6.h**). An isochronous pattern or sub-pattern appears as a relatively dense cloud of nearby points (**Figure 6.g**). Some other structural regularities, like the fact that sets of four IOIs in the "stress pattern" add up to the same duration, might go undetected and resemble a "random pattern" (**Figure 6.f&i**).

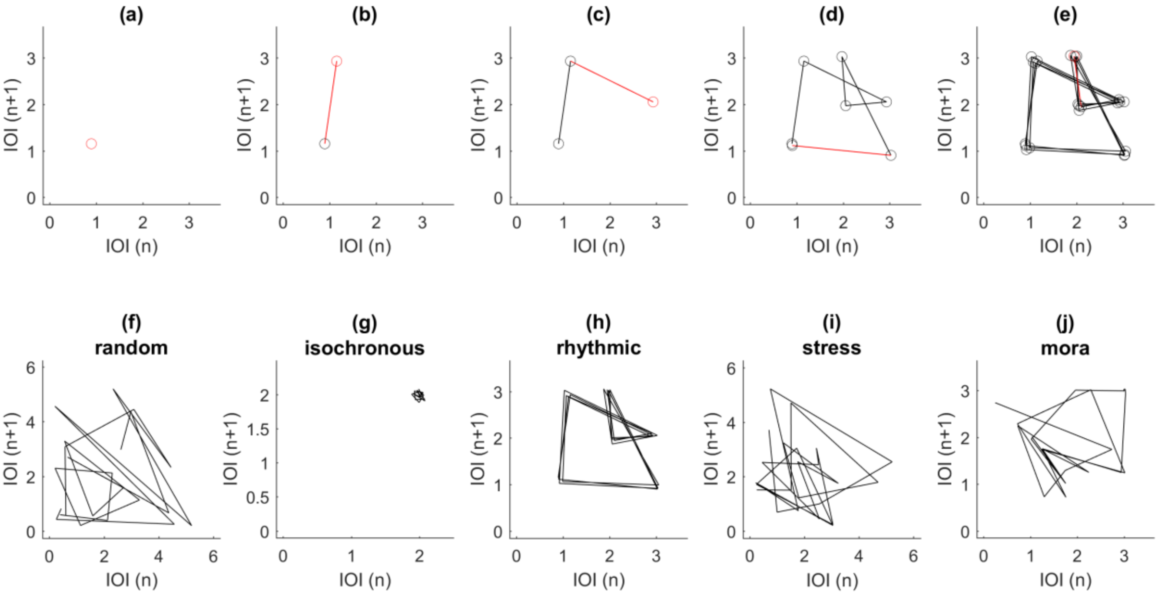


Figure 6 – Step-by-step visualization of the phase portrait based on the rhythmic sequence (a–e) and phase portraits of the five example patterns (f–j).Pairs of adjacent IOIs serve as x- and y-coordinates respectively and a line connects coordinates of neighboring pairs. (f) The random pattern shows no discernable geometrical structure. (g) The isochronous sequence appears as a dense cloud of nearby points. (h) Repetitions of a sub-pattern as in the rhythmic sequence appear as superimposed polygons, slightly shifted due to the jitter introduced by the Gaussian noise. (i&j) Higher order regularities like the constant cluster duration of the stress sequence do not appear in the phase portrait.

2.9. Recurrence plots: Visualizing pairwise similarities among all intervals in a pattern

Recurrence plots are another descriptive method for visualizing rhythmic structure, with the potential to reveal repeating sub-patterns at a glance. These plots show when a time series re-visits the same regions of phase space (Coco & Dale, 2014; Dale, Warlaumont & Richardson, 2011). Recurrence plots can be used to visualize any time series or – as shown in **Figure 7** – the sequence of IOIs. As such, the recurrence plot is a 2-dimensional matrix in which the similarity between any two IOIs is color-coded. The color code can be a gradient or, as shown here, monochromatic, where a black square represents similarity between two IOIs above a particular threshold. The IOI indices are noted on both axes of a recurrence plot in sequential order. Looking at the bottom row from left to right, the position of black squares indicates the sequential positions of IOIs that are similar to the first IOI, the second row to the second IOI, and so on.

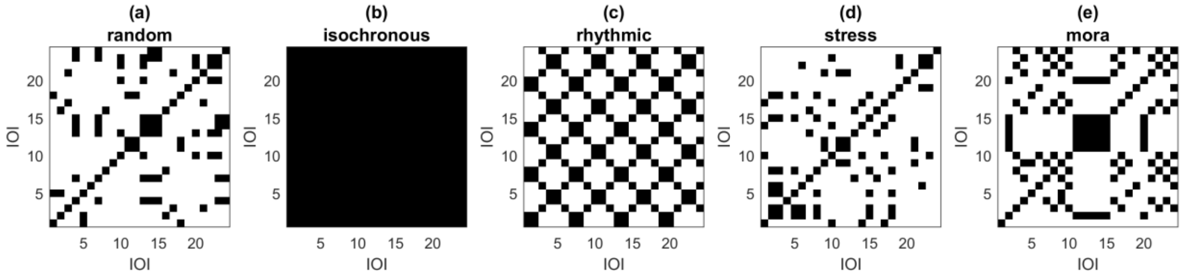


Figure 7 – Recurrence plots of the five example patterns (a–e). Each square reflects the similarity of two IOIs. Black squares represent pairs of IOIs whose difference is below a certain threshold (here: 0.3s). Recurrence plots are always symmetrical on the diagonal. (a) The random sequence shows no regular structure. (b) The plot for the isochronous sequence is all black, as all IOIs are similar to each other. (c) The repeating sup-patterns of the rhythmic sequence appear as repeating patterns in both dimensions. (d) Higher order regularities of the stress pattern do not appear in the recurrence plot. (e) Repeating IOIs of similar duration (3s, in the middle of the mora pattern) appear as larger black patches.

The plot for the "isochronous sequence" is all black, indicating that all IOIs are similar to each other (**Figure 7.b**). The repetition of the sub-pattern (6 IOIs) of the "rhythmic sequence" is plotted as repeating patterns in the horizontal and vertical directions. As the sub-pattern is palindromic, the plot is symmetrical with respect to both diagonals (**Figure 7.c**). Larger black patches, as seen in the "mora sequence" (**Figure 7.e**), indicate that a number of neighboring events share a similar IOI. Like phase portraits, recurrence plots visualize sequential structure, but fail to capture higher level structures, such as the equal cluster duration in the "stress pattern" (**Figure 7.d**).

2.10. Time series and regressions: Can the structure in a pattern be described by a linear equation?

Time series analysis denotes a broad range of statistical methods used to extract information from data points ordered in time. Time series analysis encompasses some of the techniques described above, namely autocorrelation, cross-correlation and Fourier analysis. In addition, autoregressive moving average models (ARMA) are promising time series techniques for speech rhythm (Jadoul et al., 2016), until now commonly employed in ecology, neuroscience, demography, and finance (Hamilton, 1994). While autocorrelation probes the existence of some linear relationship between intervals in a pattern, ARMA is used to test for, model, and quantify this linear relationship. In brief, an ARMA model is a parametric, linear description of a time series: One IOI is expressed as a linear combination (i.e. weighted sum) of previous IOI values, possibly previous values of another time series (e.g. intensity of previous units of speech/vocalization), and random noise. In principle, a number of metrical stress patterns in human speech could be captured by the equation:

$$IOI_t = a + b\, IOI_{t-1} + c\, IOI_{t-2} + e_t \qquad [2]$$

where $a, b,$ and c are the (empirically estimated) model parameters, and e_t is the random error. For instance, if $IOI_1 = 1$, $a = 4$, $b = -1$, and $c = 0$, the equation becomes

$$IOI_t = 4 - IOI_{t-1} \qquad [3]$$

and it describes both iambic and trochaic meters, namely an alternation of one short and one long interval, corresponding to a strong-weak, or weak-strong, alternation in stress. (The reader can try this herself by plugging $IOI_{t-1} = 1$ in Equation [3] and iterating, i.e. the resulting IOI_t becomes the next IOI_{t-1}.) Instead, if $c = -1$, then

$$IOI_t = 4 - IOI_{t-1} - IOI_{t-2} \qquad [4]$$

corresponding to the dactyl meter, that is one long interval followed by two short ones. ARMA models are extremely powerful but, apart from few exceptions (Gamba et al., 2016; Jadoul et al., 2016), still scarcely used in human phonology and animal communication.

3. Between-pattern analytical techniques: Comparing rhythms

3.1. Cross-correlation: Are patterns linearly related?

The process of cross-correlation is similar to the autocorrelation described above, except that it correlates two different signals instead of two copies of the same signal. This makes it a useful tool to find common rhythmic properties in multiple sequences.

The cross-correlation function of two isochronous sequences (**Table 3**), one with noise added (“noisy-isochronous”) and one without (“isochronous”), is similar to the autocorrelation function, except that the function is smaller than one at lag zero, as the events do not perfectly overlap (**Figure 8.b**, middle column). Cross-correlating two different “mora sequences” reveals a common rhythmic property: a subset of events occurs regularly every 3s in both sequences (**Figure 8.d**).

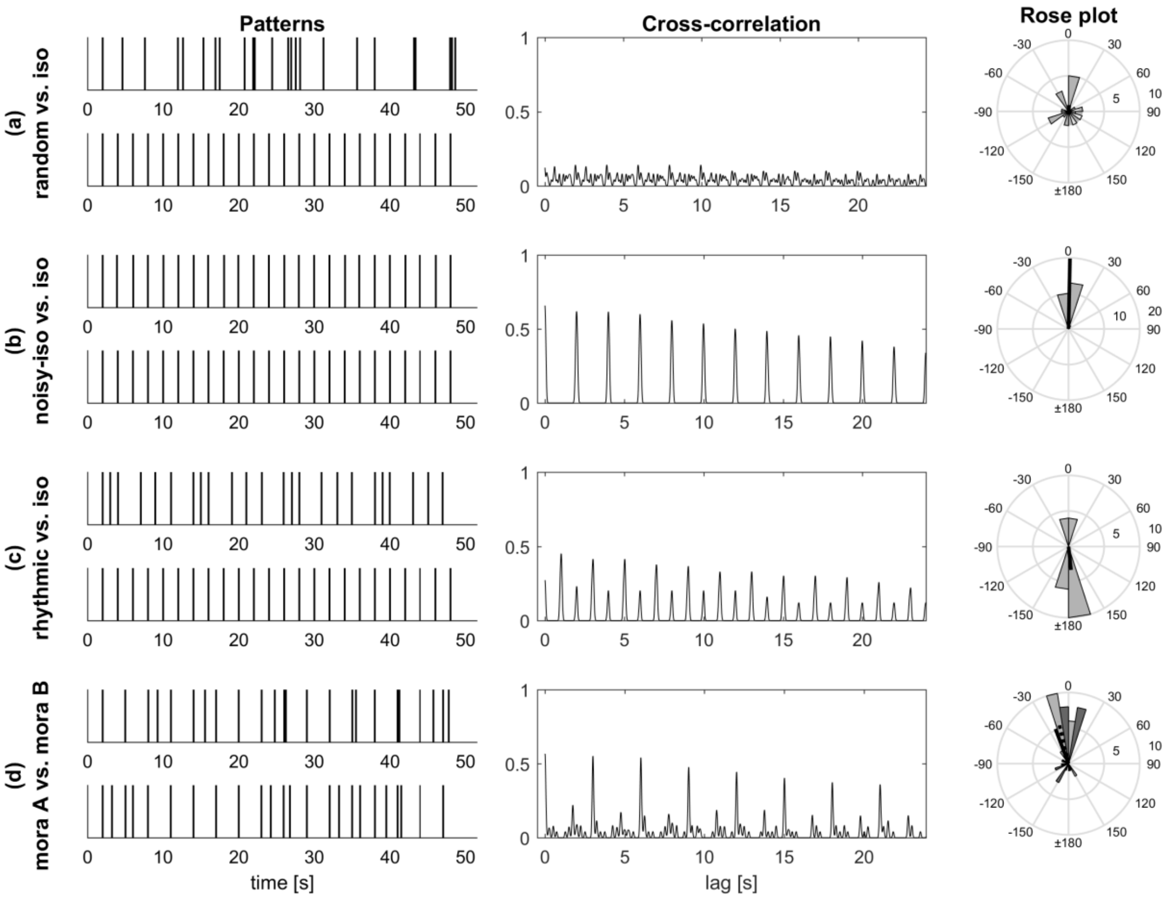


Figure 8 – Cross-correlation functions (middle) and rose plots (right) for four sets of two patterns each, with different rhythmic structure (a–d). Rose plots show the mean vector as a thick black line. The mean vector can have a length between 0 (uniform distribution) and the radius of the plot (all vectors have the same direction) (a) Comparison of a random pattern and an isochronous pattern (without added noise). (b) Comparison of two isochronous patterns, one with added noise (top, “noisy-iso”), the other without (bottom, “iso”). (c) Comparison of a ‘rhythmic’ pattern and an isochronous pattern (without added noise). (d) Comparison of two different mora patterns (mora A, top and mora B, bottom). In the rose plot for the mora patterns each of the 20 sections is split between values from mora A (light gray, dotted mean vector) and mora B (dark gray, solid mean vector). Both patterns are compared against an isochronous sequence with IOI=3s.

Table 3 – Analytical techniques to compare temporal structure between two or more patterns, including their specific function.

Technique (References)	**Function**	**Advantage**	**Disadvantage**
Cross correlation (Ravignani & Sonnweber, 2015)	Common subpatterns between two patterns	Few statistical assumptions required	Provides neither sufficient nor necessary conditions for two patterns to be structurally similar
ARMA (Hamilton, 1994)	IOIs in one pattern can be expressed as a linear combination of IOIs from another pattern	Common method for multidimensional time series	Only captures linear relations between patterns
Granger causality (Hamilton, 1994; Seth, 2010; Fuhrmann et al., 2014; Ravignani & Sonnweber, 2015)	Information from one pattern's IOIs helps better predict the other pattern	Useful test, powerful inference, easy to implement	Not a real test for logical causality, rather testing the added value of one time series in predicting another series.
Circular statistics (Fisher, 1995; Berens, 2009)	Testing hypotheses on periodically occurring events	Necessary for some data types; using classical statistics would be a mistake	A number of assumptions, concerning periodicity of events generating the data, are required

3.2. Multidimensional time series: How are patterns linearly related?

The ARMA models presented above for one pattern can be used to model the linear relationship between two patterns' IOIs. To test for temporal structure between two individuals vocalizing synchronously, performing turn-taking, etc., the ARMA equation for one individual can be enriched by past IOI terms of the other individual (e.g. IOI'_{t-1} and IOI'_{t-2}), leading to the equation:

$$IOI_t = a + b\ IOI_{t-1} + c\ IOI_{t-2} + b'\ IOI'_{t-1} + c'\ IOI'_{t-2} + e_t\ . \qquad [5]$$

Statistical estimation of parameters a, b, c, b', and c' allows in turn to draw inference about the relative contribution of the two individuals to each other's timing.

3.3. Granger causality: Can one pattern structure help predict the other?

Based on ARMA modelling, one can test whether one time series significantly affects the other, using a test for Granger causality (Hamilton, 1994). This technique tests whether future values of a target time series are better predicted by considering values of another time series, rather than using past values of the target series alone. In terms of Equation [5], Granger causality corresponds to testing whether $b'\ IOI'_{t-1} + c'\ IOI'_{t-2}$ contributes to the statistical prediction of IOI_t. Granger causality has been successfully applied in neuroscience and economics, and it is being increasingly employed to analyse production of rhythmic patterns in humans and other animals (Seth, 2010; Fuhrmann et al., 2014; Gamba et al., 2016).

3.4. Circular statistics: How does a pattern relate statistically to a fixed-pace clock?

Circular statistics (or directional statistics) offer a set of techniques that are useful when dealing with data from a circular distribution, like compass direction and time of day (Fisher, 1995). In such distributions both ends of the scale have the same value, i.e. they "wrap around" and can be mapped to a circle. On a linear scale, for example, the mean of 355° and 5° would be 180°, which makes little sense when calculating e.g. compass directions. The circular mean on the other hand is 360°, which is identical to 0°.

Circular statistics can be used to compare any sequence against a known isochronous pattern. In such cases the distance of all events in the sequence to their nearest event in the isochronous pattern can be expressed as a circular distribution. An event that exactly coincides with an isochronous event is defined to have an angular direction of 0°. Events that lag behind the nearest isochronous event have positive angles >0° (i.e. positive delay), and events that precede have negative angles (equivalently, positive angles of <360°).

The distribution of events can be visualized on a so-called rose plot, which is the circular equivalent of a linear histogram (**Figure 8**, right column). An arrow can be added to mark the circular mean vector, which indicates the mean direction and has a length between 0 (events are uniformly distributed) and 1 (events concentrate on a single direction).

Several significance tests exist to evaluate the circularity of a distribution (Zar, 2010). The Rayleigh z test can determine whether the sample data comes from a uniform distribution; For instance, the deviations of the random sequence from the isochronous pattern are uniformly distributed (**Figure 8.a**, p=82.7, z=216, n=24). The z test assumes that the data is unimodally distributed and sampled from a von Mises distribution, the circular equivalent of a normal distribution. Kuiper's test is the circular analogue of the Kolmogorov-Smirnov test and can be used to test the difference between two distributions, e.g. the sample data against a von Mises distribution. The bimodal distribution of the rhythmic pattern (**Figure 8.c**) for example differs significantly from a von Mises distribution (Kuiper's test, p<0.005, V=336, n=24). If an expected mean direction is known beforehand (e.g. 0° in synchronization experiments), the more precise V-test can be used to test against uniformity. If directions are not uniformly distributed and the mean direction does not significantly differ from 0°, the patterns are synchronous (V-test, p<0.001, V=23.8, n=24; **Figure 8.b**, isochronous pattern).

4. Where can these methods be fruitfully used?

Here we suggest some avenues of research where to apply the methods described above.

First, the field of birdsong bioacoustics should put more emphasis on temporal and rhythmic, not only spectral or positional, properties of the signal. Building on recent findings, songs from different birdsong species could be compared to temporal structures hypothesized a priori or across species (Benichov, Globerson & Tchernichovski, 2016; Norton & Scharff, 2016; Spierings & ten Cate, 2016, Janney et al. 2016), ultimately constructing 'rhythmic phylogenies'. Fourier analysis, GAT pulse matching and circular statistics might be particularly suitable to uncover isochronous structure in birdsong and human speech.

Second, the comparative study of vocal complexity has traditionally encompassed species capable of vocal production learning, putting relatively more emphasis on animals' abilities to modify spectral, rather than temporal, properties of the vocalizations (e.g. Kershenbaum et al., 2014). The methods presented here could be used to analyse rhythmic properties of acoustic behaviour in animals with limited vocal learning, which however have potential for temporal flexibility, for instance primates' chorusing (Fedurek et al., 2013; Gamba et al., 2016), sea lions barks (Schusterman, 1977, Ravignani et al., 2016), ape drumming (Dufour et al., 2015; Ravignani et al., 2013), etc. Phase portraits, autocorrelation, circular statistics, and recurrence plots could be used, among others, to uncover the possibly flexible temporal structures in these acoustic behaviours.

Third, conversational turn-taking across human languages and animal species is a topic gaining increasing scientific interest (Levinson, 2016; Vernes, 2016). This area is however missing an advanced and unified quantitative analytical framework. We suggest that the methods described here can be used for exactly such purpose, enabling comparisons of temporal structure in communicative interaction across human languages, modalities and species. Time series analysis, in particular ARMA models and Granger causality, might be particularly suitable to investigate turn-taking.

Fourth, the methods we present make almost no top-down assumptions about the structure present in the signal. Hence, they can be employed to investigate phonology, individual timing and rhythm in modalities and domains other than speech (e.g. Dale, Warlaumont & Richardson, 2011; Fan et al., 2016), most notably in sign languages (de Vos, Torreira & Levinson, 2015). Distributional indexes and all other structural measures in **Table 2** could be used to quantify temporal structure in a given modality or domain.

Fifth, and finally, the statistical methods presented can be used to test for cross-modal influences in multimodal communication (e.g. Tilsen, 2009). Temporal interdependencies between modalities could arise, for instance, in the co-evolution of human rhythmic vocalizations and coordinated movement (Laland, et al., 2016; Ravignani & Cook, 2016) or the simultaneous song and 'dance' display of some bird species (Dalziell et al., 2013; Ullrich, Norton & Scharff, 2016). Levenshtein distance (Post & Toussaint, 2011; Ravignani, Delgado & Kirby, 2016), and other dyadic techniques in **Table 3** could be used to relate temporal structure between modalities.

5. Conclusions

Quantitative cross-species comparison of rhythmic features can inform the evolution of human speech (Fitch, 2000; Yip, 2006; Petkov & Jarvis, 2012; Spierings & ten Cate, 2014, 2016; Lameira et al., 2015; Norton & Scharff, 2016; Ravignani et al., 2016), be it its origin, mechanisms or function.

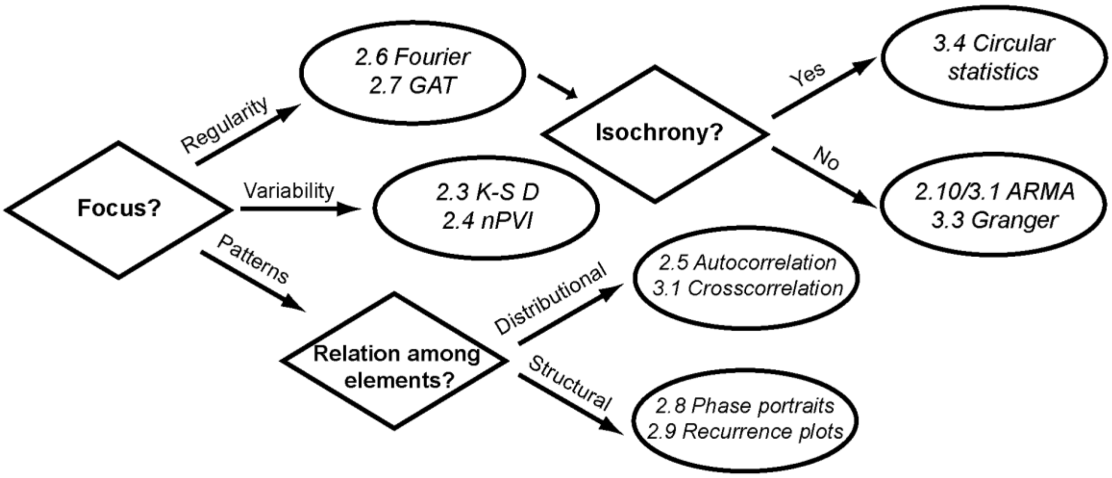


Figure 9 – Flux diagram to select the most appropriate technique for rhythm analysis. Answers (regular font) to specific questions (in bold) guide the reader towards the appropriate technique and corresponding subsection (in italics).

Here we present a number of techniques to analyse rhythmic temporal structure in animal vocalizations, including human speech. **Figure 9** provides an aid to select the appropriate technique in the behavioral analysis of timing. This figure guides researchers in their initial approach to rhythm analysis, and presents an overview of all techniques and methods described here.

Theoretically, comparative research should avoid treating rhythm as a monolithic entity (Fitch 2012; 2015). This theoretical advance can only be achieved via improvement of the methodological tools to discern and test for specific properties of rhythm. **Figure 9** and the techniques it refers to provide a roadmap to testing hypotheses on specific rhythmic characteristics in the data.

Finally, this paper focuses on temporal structure in rhythm. Timing is the main but not the only component underlying speech rhythm. Similarly, rhythm in animal phonology might derive from an interaction between temporal and spectral characteristics of the signal. New statistical techniques for comparative analyses should also be developed to incorporate spectral information like intonation, stress, etc. (Ramus et al., 1999).

6. Data accessibility

All sequences, figures and statistical results were generated in Mathworks Matlab R2012b. All code used in this paper can be freely downloaded from http://userpage.fu-berlin.de/phno/mrc/. It can be run in Matlab as well as the free and open-source alternative GNU Octave. The supplementary material accompanying this article contains a detailed explanation of the code and instructions on its usage. **Table 4** describes useful software (columns), packages (in square brackets), and functions, with their intended use.

Table 4 – Some available software to perform comparative speech/vocal production analyses of rhythmic features. Software (columns) aimed at specific measures (rows), specifying packages (in square brackets), and functions (in parentheses). All software discussed here is free except for Matlab (paid software, free packages). Note that the Octave implementation of the Kolmogorov-Smirnov test differs from that of Matlab and R (R core team, 2013).

Technique	**Matlab**	**Octave**	**R**
Histogram	histogram() (since version 2014b), hist()	hist()	hist()
Kolmogorov-Smirnov D	kstest()	kolmogorov_smirnov_test() [statistics]	ks.test()
Auto-/Cross-correlation	xcorr()	xcorr() [signal]	acf()
Fast Fourier transform	fft()	fft()	fft(); periodogram() [TSA]
Rose plots	polarhistogram() (since version 2016b), rose()	rose()	rose.diag() [ggplot2]
Circular Statistics	[CircStats] (Berens, 2009)	[CircStats]	[circular]
ARMA	arima() [Econometrics]; [GCCA toolbox] (Seth, 2010)	arma_rnd()	arima() [TSA]
Recurrence plots	plotRecurrence()	plotRecurrence()	[crqa] (Coco & Dale, 2014)

References

Bekius, A., Cope, T. E., and Grube M. (2016). The beat to read: A cross-lingual link between rhythmic regularity perception and reading skill. *Frontiers in Human Neuroscience*, 10:425.

Benichov, J. I., Globerson, E., and Tchernichovski, O. (2016). Finding the beat: From socially coordinated vocalizations in songbirds to rhythmic entrainment in humans. *Frontiers in Human Neuroscience*, 10:255.

Berens, P. (2009). CircStat: A Matlab toolbox for circular statistics. *Journal of Statistical Software*, 31:10, 1–21.

Bilmes, J. (1993). Techniques to foster drum machine expressivity. *International Computer Music Conference*, 93: 276–83

Boersma, P., and Weenink, D. (2013). *PRAAT: Doing phonetics by computer*, version 5.3.49. Amsterdam: Universiteit van Amsterdam. Retrieved from www.praat.org

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

Bowling, D. L., and Fitch, W. T. (2015). Do animal communication systems have phonemes? *Trends in Cognitive Sciences*, 19:10, 555–557.

Coco, M. I., and Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: An R package. *Frontiers in Quantitative Psychology and Measurement*, 5:510.

Collier, K., Bickel, B., van Schaik, C. P., Manser, M. B., and Townsend, S. W. (2014). Language evolution: Syntax before phonology? *Proceedings of the Royal Society B: Biological Sciences*, 281:20140263.

Dale, R., Warlaumont, A. S., and Richardson, D. C. (2011). Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *International Journal of Bifurcation and Chaos*, 21:04, 1153–1161.

Dalziell, A. H., Peters, R. A., Cockburn, A., Dorland, A. D., Maisey, A. C., and Magrath, R. D. (2013). Dance choreography is coordinated with song repertoire in a complex avian display. *Current Biology*, 23:12, 1132–1135.

de Boer, B. (2012). Air sacs and vocal fold vibration: Implications for evolution of speech. *Theoria et Historia Scientiarum*, 9, 13–28.

de Vos, C., Torreira, F., and Levinson, S. C. (2015). Turn-timing in signed conversations: Coordinating stroke-to-stroke turn boundaries. *Frontiers in Psychology*, 6:268.

Dufour, V., Poulin, N., Curé, C., and Sterck, E. H. (2015). Chimpanzee drumming: a spontaneous performance with characteristics of human musical drumming. *Scientific Reports*, 5:11320.

Fan, P. F., Ma, C. Y., Garber, P. A., Zhang, W., Fei, H. L., and Xiao, W. (2016). Rhythmic displays of female gibbons offer insight into the origin of dance. *Scientific Reports*, 6:34606.

Fedurek, P., and Slocombe, K. E. (2011). Primate vocal communication: A useful tool for understanding human speech and language evolution? *Human Biology*, 83:2, 153–173.

Fedurek, P., Schel, A. M., and Slocombe, K. E. (2013). The acoustic structure of chimpanzee pant-hooting facilitates chorusing. *Behavioral Ecology and Sociobiology*, 67:11, 1781–1789.

Fehér, O. (2016). Atypical birdsong and artificial languages provide insights into how communication systems are shaped by learning, use, and transmission. *Psychonomic Bulletin & Review*, 24, 97–105.

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

Fisher, N. I. (1995). *Statistical Analysis of Circular Data*. Cambridge, UK: Cambridge University Press.

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, 4:7, 258–267.

Fitch, W. T. (2011). The biology and evolution of rhythm: Unraveling a paradox. In P. Rebuschat, M. Rohrmeier, J. A. Hawkins, and I. Cross (eds), *Language and Music as Cognitive Systems*. Oxford, UK: Oxford University Press, 73–95.

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Review*, 11, 329–364.

Fitch, W. T. (2015). The biology and evolution of musical rhythm: An update. In *Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff*. Cambridge, MA: MIT Press, 293–324.

Fuhrmann, D., Ravignani, A., Marshall-Pescini, S., and Whiten, A. (2014). Synchrony and motor mimicking in chimpanzee observational learning. *Scientific Reports*, 4:5283.

Gamba, M., Torti, V., Estienne, V., Randrianarison, R. M., Valente, D., Rovara, P., et al. (2016). The indris have got rhythm! Timing and pitch variation of a primate song examined between sexes and age classes. *Frontiers in Neuroscience*. 10:249.

Ghazanfar, A. A. (2013). Multisensory vocal communication in primates and the evolution of rhythmic speech. *Behavioral Ecology and Sociobiology*, 67:9, 1441–1448.

Goswami, U., and Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives. *Laboratory Phonology*, 4:1, 67–92.

Grabe, E., and Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, 7, 515–546.

Gustison, M. L. and Bergman, T. (under review) The spectro-temporal properties of gelada calls resemble human speech. *Journal of Language Evolution*

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hannon, E. E., Lévêque, Y., Nave, K. M., and Trehub, S.E. (2016) Exaggeration of language-specific rhythms in English and French children's songs. *Frontiers in Psychology*, 7:939.

Jadoul, Y., Ravignani, A., Thompson, B., Filippi, P., and de Boer, B. (2016). Seeking temporal predictability in speech: Comparing statistical approaches on 18 world languages. *Frontiers in Human Neuroscience*, 10:586.

Janney, E., Taylor, H., Rothenberg, D., Scharff, C., Parra, L. C., and Tchernichovsky, O. (2016). Temporal regularity increases with repertoire complexity in the Australian pied butcherbird's song. *Royal Society Open Science*, 3: 160357.

Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., et al. (2014). Acoustic sequences in non-human animals: A tutorial review and prospectus. *Biological Reviews*, 91, 13–52.

Laland, K., Wilkins, C., and Clayton, N. (2016). The evolution of dance. *Current Biology*, 26:1, R5–R9.

Lameira, A. R., Hardus, M. E., Bartlett, A. M., Shumaker, R. W., Wich, S. A., and Menken, S. B. (2015). Speech-like rhythm in a voiced and voiceless orangutan call. PLoS ONE, 10:1, e116136.

Lameira, A. R., Maddieson, I., and Zuberbühler, K. (2014). Primate feedstock for the evolution of consonants. *Trends in Cognitive Sciences*, 18:2, 60–62.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5:3, 253–263.

Levinson, S. C. (2016). Turn-taking in human communication – origins and implications for language processing. *Trends in Cognitive Sciences*, 20:1, 6–14.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62:318, 399–402.

McAuley, J. D. (2010) Tempo and rhythm. In M. R. Jones, R. R. Fay, and A. N. Popper (eds), *Springer Handbook of Auditory Research*, Vol.36: *Music Perception*. New York, NY: Springer.

Norton, P., and Scharff, C. (2016). 'Bird Song Metronomics': Isochronous organization of zebra finch song rhythm. *Frontiers in Neuroscience*, 10:309.

Petkov, C. I., and Jarvis, E. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, 4:12.

Post, O., and Toussaint, G. (2011). The edit distance as a measure of perceived rhythmic similarity. *Empirical Musicology Review*, 6:3, 164–179.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73:3, 265–292.

Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Frontiers in Evolutionary Neuroscience*, 4:7.

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12:6, 718–724.

Ravignani, A., Olivera, V. M., Gingras, B., Hofer, R., Hernández, C. R., Sonnweber, R. S., and Fitch, W. (2013). Primate drum kit: A system for studying acoustic pattern production by non-human primates using acceleration and strain sensors. *Sensors*, 13:8, 9790–9820.

Ravignani, A., and Sonnweber, R. (2015). Measuring teaching through hormones and time series analysis: Towards a comparative framework. *Behavioral and Brain Sciences*, 38, 40–41.

Ravignani, A., and Cook, P. (2016). The evolutionary biology of dance without frills. *Current Biology*, 26:19, R878–R879.

Ravignani, A., Fitch, W., Hanke, F. D., Heinrich, T., Hurgitsch, B., Kotz, S. A., Scharff, C., Stoeger, A., and De Boer, B. (2016). What pinnipeds have to say about human speech, music, and the evolution of rhythm. *Frontiers in Neuroscience*, 10:274.

Ravignani, A., (in press). Visualizing and interpreting rhythmic patterns using phase space plots. *Music Perception.*

Ravignani, A., Delgado, T., and Kirby, S., (2016). Musical evolution in the lab exhibits rhythmic universals. *Nature Human Behaviour*, 1:0007.

Rothenberg, D., Roeske, T. C., Voss, H. U., Naguib, M., and Tchernichovski, O. (2014). Investigation of musicality in birdsong. *Hearing Research*, 308, 71–83.

Saar, S., and Mitra, P. P. (2008). A technique for characterizing the development of rhythms in bird song. *PLoS ONE*, 3:e1461.

Scharff, C., and Petri, J. (2011). Evo-devo, deep homology and FoxP2: Implications for the evolution of speech and language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366:1574, 2124–2140.

Schusterman, R. J. (1977). Temporal patterning in sea lion barking (Zalophus californianus). *Behavioral Biology*, 20:3, 404–408.

Seth, A. K. (2010). A MATLAB toolbox for Granger causal connectivity analysis. *Journal of Neuroscience Methods*, 186:2, 262–273.

Spierings, M., and ten Cate, C. (2014). Zebra finches are sensitive to prosodic features of human speech. *Proceedings of the Royal Society of London B: Biological Sciences*, 281:1787, 20140480.

Spierings, M., and ten Cate, C. (2016). Zebra finches as a model species to understand the roots of rhythm. *Frontiers in Neuroscience*, 10:345.

ten Cate, C. (2016). Assessing the uniqueness of language: Animal grammatical abilities take center stage. *Psychonomic Bulletin & Review*, 24:1, 91–96.

Thiel, M., Romano, M. C., and Kurths, J. (2004). How much information is contained in a recurrence plot? *Physics Letters* A, 330:5, 343–349.

Tilsen, S. (2009). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science*, 33:5, 839–79

Toussaint, G. T. (2013). *The Geometry of Musical Rhythm: What Makes a "good" Rhythm Good?* Boca Raton, FL: CRC Press.

Ullrich, R., Norton, P., and Scharff, C. (2016). Waltzing Taeniopygia: Integration of courtship song and dance in the domesticated Australian zebra finch. *Animal Behaviour*, 112, 285–300.

Vernes, S. C. (2016). What bats have to say about speech and language. *Psychonomic Bulletin & Review*, 24:1, 111–117.

Wagner, P. (2007). Visualizing levels of rhythmic organization. In J. Trouvain, and W. J. Barry (eds) *XVIth International Congress of the Phonetic Sciences*, Saarbrücken, 6–10 August 2007: pp. 1113–1117.

Yip, M. J. (2006). The search for phonology in other species. *Trends in Cognitive Sciences*, 10:10, 442–446.

Zar, J. H. (2010). *Biostatistical Analysis, 5th edn*. Upper Saddle River, NJ: Prentice Hall, Inc.

Supplementary Material

This document describes the code that was written for the article "Measuring rhythmic complexity: A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations." The code was used to carry out all analyses and produce most of the figures. It can be obtained from http://userpage.fu-berlin.de/phno/mrc/. With the help of this document the readers will be able to carry out the rhythmic analyses described in the article on their own.

The code was written in Mathworks Matlab. It was tested in Matlab versions 2012b, 2015b and 2016b, but should run in all modern Matlab versions. If you do not have access to Matlab, you can run the code in GNU Octave, a free and open source alternative that is mostly compatible with Matlab's syntax. Octave is available for Windows, OSX and Linux from this address: https://www.gnu.org/software/octave/download.html. The code was tested in Octave versions 3.8.1 and 4.0.3. For the full code to run in Octave the packages 'control', 'signal', 'statistics' and 'io' are needed. These can be downloaded from http://octave.sourceforge.net/packages.php. On Windows they should come pre-installed with Octave. To load all installed packages type the following in the command window after starting Octave (omitting the '>>'):

```
>> pkg load all
```

The code consists of several functions (**Table S1**), each contained in a separate text file with the extension '.m'. To access these functions, make sure that either the files are in your current folder (browse to the folder within Matlab/Octave in the left hand panel), or the folder containing the files is in your search path (type e.g. "addpath('C:/exampleFolder')" in the command window). You can call these functions from within the Matlab/Octave command window by typing the function name, followed by one or several arguments in parentheses, e.g.:

```
>> plotCorr(pattern, 0, 1000)
```

Multiple arguments are separated by commas. Most functions require as an argument either an array containing timepoints of events (e.g. 2,4,6,9,10, the "pattern"), or an array of inter-onset intervals, i.e. the time intervals between successive events (e.g.2,2,3,1, the "ioi"). Both these arrays can be created in one of three ways:

1. Typing in a pattern at the command line. To create a pattern from scratch, assign a comma-separated list of numbers in square brackets to a new array. The array name can be any string of letters (a-z), numbers (0-9) and underscores (_). They following example creates an array named "examplePattern" with a pattern consisting of events at 2, 4, 6 and 9 seconds:

```
>> examplePattern = [2, 4, 6, 9]
```

The corresponding array of IOIs would consist of the numbers 2 (difference between the second and the first event), 2 (difference between the third and the second event), and 3 (difference between the fourth and the third event):

```
>> exampleIOIs = [2, 2, 3];
```

Adding a semicolon (;) to the end of a line executes the command like normal, but suppresses the output.

2. Generating a pattern procedurally. The patterns that served as examples in the article were generated by the function generatePattern(). This function takes a number of arguments that affect the properties of the resulting pattern. It returns an array of events ("pattern") and an array of inter-onset intervals ("ioi"). To view a description of a function and its arguments and return values, the reader should type either "help" or "doc" followed by the function name in the command window, e.g.:

```
>> doc generatePattern
```

This prints the following text:

```
[pattern,iois] = generatePattern(n,nRep,type,stdev,seed)

generates random, isochronous, or one of three types of non-isochronous
rhythmic patterns.

n:      number of events (must be divisible by nRep wtithout remainder)
nRep:   number of repetitions of the rhythmic pattern
type:   either 'random', 'isochronous', 'rhythmic', 'stress' or 'mora'
stdev:  standard deviation for gaussian noise
seed:   seed for the random number generator

returns a 1-dimensional array of timestamps of events (pattern) and a
1-dimensional array of the corresponding inter-onset intervals (ioi).
```

The different pattern types (random, isochronous, rhythmic, stress and mora) follow certain rules as described in the main article. The stress pattern for example always produces clusters of four events that have a total duration of 8 seconds. The duration of the events within a cluster, however, is determined by a random number generator. The random number generator can be initialized into a certain state by the argument seed. This allows to reproduce a specific pattern. Calling the function with a seed value of 8 (see example below) will produce the same pattern each time the function is called, if all other arguments are the same. Calling the function with a different seed value will produce a different pattern. Setting the seed value to zero will result in a different pattern each time the function is called. The random number generator also affects the Gaussian noise added to the event timing. Isochronous patterns created with different seed values, for example, will therefore produce slightly different patterns, as the jitter of the events is slightly different. Readers are encouraged to experiment with different randomized patterns to see their effect on the different plots, for example how different permutations of a rhythmic pattern can lead to mirrored phase portraits.

The following command creates a 'rhythmic' pattern (and corresponding list of IOIs) similar to that used in the article, that has 24 events, 4 repetitions of a sub-pattern, an average inter-onset interval of 2s and Gaussian noise added with a standard deviation of 0.04:

```
>> [patternR, ioiR] = generatePattern(24, 4, 'rhythmic', 0.04, 8);
```

When a function has multiple return values, they are assigned to a comma-separated list of variables (in this case our two arrays) in square brackets. Again, the names of the variables can be any alphanumeric string.

3. Reading a pattern from an Excel table. The five example patterns from the article are included with the code as Excel table files (e.g. example_pattern_isochronous.xls). You can read these, or your own data, using the function readPatternFromXls(). Tables must contain the time values of several events either in a single row or a single column. The function takes the filename of the table as an argument and returns arrays for the pattern and the IOIs. If the table file is in your current folder or in a folder in the search path you can supply just its filename in single quotes (e.g. 'example_pattern_isochronous.xls'). Otherwise "filename" must contain the whole path to the file (e.g. 'C:/exampleFolder/example_pattern_isochronous.xls'). The following command reads the values of the mora sequence from the excel file:

```
>> [myPattern, myIOIs] = readPatternFromXls('example_pattern_mora.xls');
```

4. Plotting and analyzing patterns. The function plotAllSingle() calls several of the other functions and in doing so creates separate figures for all the single pattern analyses presented in the article. It takes as arguments the arrays "pattern" and "ioi" that you created through one of the three methods described above. The function also prints the Kolmogorov-Smirnov D (K-S D) and the normalized pairwise variability index (nPVI) to the command window. In addition, it returns a structure that contains the outputs of some of the functions called by plotAllSingle(), which can be used for further analyses. Type "doc plotAllSingle" or open the file "plotAllSingle.m" to view a list of all outputs contained in the results structure.

```
>> myResults = plotAllSingle(myPattern, myIOIs)
```

To access any of the outputs in the results structure, type the name of your results structure, followed by a period and the name of the output variable you are interested in, for example:

```
>> myResults.nPVI
```

All of the analysis functions can also be called separately. Each function takes either the pattern or the IOIs as an argument. The functions that transform the pattern to a time series – plotCorr() and plotFFT() – additionally take the temporal resolution of the time series as an argument. These functions transform the pattern into a time series by creating an array with a certain temporal resolution (e.g. one value per millisecond). This array has the value 1 at the time points of the events and the value 0 elsewhere. For the article a temporal resolution value of 1000 (time points per second) was used and the plotAllSingle() function uses 1000 as a default. Readers who wish to change this value should edit the appropriate line in the plotAllSingle.m file.

To only plot the autocorrelation function for the pattern "myPattern", type:

```
>> plotCorr(myPattern, 0, 1000);
```

This function takes either one or two patterns as arguments. If the user inputs a single pattern, this pattern will be autocorrelated. In this case set the second argument to zero. The correlation function can also be used to cross-correlate two different patterns. To do this, readers should supply a different pattern as the second argument, for instance:

```
>> plotCorr(patternOne, patternTwo, 1000);
```

The function plotRose() creates a circular histogram of deviations of a pattern from a previously defined strictly isochronous pattern. The following commands create a rhythmic pattern and an isochronous pattern without Gaussian noise added and creates a rose plot:

```
>> [patternR,ioiR] = generatePattern(24, 2, 4, 'rhythmic', 0.04, 0);
>> [patternI,ioiI] = generatePattern(24, 2, 0, 'isochronous', 0, 0);
>> plotRose(patternR, patternI);
```

To view any function, readers can open it in the Matlab or Octave editor or any other text editor. There, they will find a description of the function and its usage, as well as comments explaining each step (lines beginning with %). Users can change all functions according to their needs. Some internal parameters of the functions, for example, depend somewhat on the magnitude of your IOIs, like the standard deviation and width of the normal distribution in the plotCorr() function, or the threshold for the recurrence plots in plotRecurrence(). Other values that one might want to adjust include the width of the histogram bins in plotHistogram() and the freuquency limits in plotFFT() and plotGAT().

Supplementary Table 1 - Overview of all functions

Function	Arguments	Returns	Description
generatePattern	n, nRep, type, stdev, seed	pattern, ioi	Generates one of five different types of patterns (random, isochronous, rhythmic, stress, mora) and inter-onset intervals (IOI).
getKolmogorovSmirnovD	ioi	d	Calculates the Kolmogorov-Smirnov D (K-S D) from a list of IOIs.
getNPVI	ioi	nPVI	Calculates the normalized pairwise variability index (nPVI).
getPhaseFromPatterns	pattern, patternIso	phaseRad	Calculates the phase (in radians) of each event in a pattern, relative to a second, isochronous pattern. This function gets internally called by plotRose().
getTimeSeriesFromPattern	pattern, resolution	timeSeries	Constructs a time series from a pattern. This function gets internally called by plotCorr() and plotFFT().
plotAllSingle	pattern, ioi	results	Calls all plot functions consecutively for a pattern and prints the K-S D and nPVI.
plotCorr	patternOne, patternTwo, resolution	correlation, sigConvOne, sigConvTwo	Performs and plots either autocorrelation (if patternTwo = 0) or crosscorrelation. The pattern is first convoluted with a Gauss curve.
plotFFT	pattern, resolution	freq, power, timeSeries	Calculates the fast Fourier transform (FFT) of a pattern and plots the power spectrum.
plotGAT	pattern	freq, frmsd	Performs the pulse generate-and-test (GAT) method and plots the frmsd for the tested frequencies (default: 0.2-5Hz; set within the function).
plotHistogram	ioi	-	Plots a histogram of IOIs. (default bin width: 0.075)
plotPattern	pattern	-	Plots the events of a pattern as vertical lines on a timeline.
plotPhasePortrait	ioi	-	Creates a phase portrait of IOIs.
plotRecurrence	ioi	-	Creates a recurrence plot of IOIs. (default threshold: 0.3)
plotRose	pattern, patternIso	phaseRad	Creates a rose plot (circular histogram) of phases of a pattern relative to a pre-determined strictly isochronous pattern.
readPatternFromXls	filename	pattern, ioi	Reads a pattern from an Excel table file (*.xls/*.xlsx).

Publication B: ‘Bird Song Metronomics’

Norton, P., and Scharff, C. (2016). ‘Bird Song Metronomics’: Isochronous organization of zebra finch song rhythm. *Frontiers in Neuroscience* 10:309. doi:10.3389/fnins.2016.00309.

‘Bird Song Metronomics’: Isochronous Organization of Zebra Finch Song Rhythm

Philipp Norton & Constance Scharff

Department of Animal Behaviour, Freie Universität Berlin

Abstract

The human capacity for speech and vocal music depends on vocal imitation. Songbirds, in contrast to non-human primates, share this vocal production learning with humans. The process through which birds and humans learn many of their vocalizations as well as the underlying neural system exhibit a number of striking parallels and have been widely researched. In contrast, rhythm, a key feature of language and music, has received surprisingly little attention in songbirds. Investigating temporal periodicity in bird song has the potential to inform the relationship between neural mechanisms and behavioral output and can also provide insight into the biology and evolution of musicality. Here we present a method to analyze birdsong for an underlying rhythmic regularity. Using the intervals from one note onset to the next as input, we found for each bird an isochronous sequence of time stamps, a ‘signal-derived pulse’, or pulse^{S}, of which a subset aligned with all note onsets of the bird’s song. Fourier analysis corroborated these results. To determine whether this finding was just a byproduct of the duration of notes and intervals typical for zebra finches but not dependent on the individual duration of elements and the sequence in which they are sung, we compared natural songs to models of artificial songs. Note onsets of natural song deviated from the pulse^{S} significantly less than those of artificial songs with randomized note and gap durations. Thus, male zebra finch song has the regularity required for a listener to extract a perceived pulse (pulse^{P}), as yet untested. Strikingly, in our study, pulses^{S} that best fit note onsets often also coincided with the transitions between sub-note elements within complex notes, corresponding to neuromuscular gestures. Gesture durations often equaled one or more pulse^{S} periods. This suggests that gesture duration constitutes the basic element of the temporal hierarchy of zebra finch song rhythm, an interesting parallel to the hierarchically structured components of regular rhythms in human music.

Introduction

Rhythm is a key element in the structure of music and can be defined as the "systematic patterning of sound in terms of timing, accent and grouping" (Patel, 2008, p. 96). These patterns can be either periodic (i.e. regularly repeating) or aperiodic. A special case of a periodic pattern is an isochronous one, where the time intervals between successive events share the same duration. In many types of music across the world, including the Western European (Patel, 2008, pp. 97–99) and African (Arom, 1991, p. 211) traditions, the timing of sonic events, mostly note onsets, is structured by a perceptually isochronous pulse (Nettl, 2001). This pulse is a cognitive construct that is usually implicit rather than being materialized in the acoustic signal itself (Arom, 1991 p. 230; Fitch, 2013). For the purpose of this article we will call this the ‘perceived pulse’, or pulse^{P}. In all but the simplest of rhythms not all notes fall on the pulse and some pulses occur in the silence between notes. Therefore, the intervals between the notes in a piece are rarely isochronous, but many note onsets align to an isochronous pulse. In some musical styles, variations of tempo – and therefore pulse – are used for artistic effect (e.g. accelerando and ritardando in classical music), while in others the tempo remains constant throughout a piece or performance (e.g. Central African music; Arom, 1991, p. 20). Often the pulse is further organized by a metrical structure, the recurring hierarchical patterning of strongly and weakly accented events. In a waltz, for example, the pulse is perceptually divided into groups of three, of which the first one – the so-called downbeat – is perceived as more strongly accented than the following two (“*one*, two, three, *one*, two, three”). In this example, pulses on the lower level of the metrical hierarchy, i.e. every pulse, happen at three times the tempo of the higher level, consisting of only the strong pulses. The process of finding the pulse and frequently the subsequent attribution of meter allow us to infer the beat of a piece of music.

If you have ever danced or clapped your hands along to music, you have already encountered one function of a regular pulse: it facilitates the coordination of synchronized movements through a process called ‘beat perception and synchronization’. It also provides musicians with a common temporal reference that is necessary for coordinated ensemble performance (Arom, 1991, p. 179; Patel, 2008, pp. 99–100). Furthermore, expectations and the interplay of successful anticipations and surprises emerging from these expectations are thought to drive the “emotive power” of human music (Huron, 2006). Pulse and meter, as well as deviations thereof, can build anticipations in the time-domain that subsequently are either fulfilled or violated.

How did such an apparently universal aspect of human music evolve? Several authors have stressed the importance of a cross-species comparative approach to gain insights into the evolution of music (Carterette and Kendall, 1999; Fitch, 2006; Hauser and McDermott, 2003; Hulse and Page, 1988; Patel and Demorest, 2013). Crucial to this endeavor is the realization that the music faculty hinges on a variety of interacting perceptual, cognitive, emotional and motor mechanisms that may follow different evolutionary trajectories. It is therefore helpful to break down the music faculty into these different components and investigate which of them are present, either by homology or analogy, in non-human animals (Fitch, 2006, 2015; Honing et al., 2015; Ravignani et al., 2014).

One critical component is our capacity for vocal learning. It allowed us to develop speech as well as song, which is assumed to be universal to human music (Brown and Jordania, 2011; Nettl, 2001; Trehub, 2001). Of the many species that produce vocalizations or other acoustic signals of varying complexity, only a few are well known to rely on developmental learning to acquire some of their adult vocalizations, e.g. songbirds, hummingbirds and parrots as well as several species of bats, some marine mammals and elephants (rewieved by Petkov and Jarvis, 2012).

Birdsong in particular has caught the interest of researchers for its putative musical features (Baptista and Keister, 2005; Dobson and Lemon, 1977; Kneutgen, 1969; Marler, 2001; Rothenberg et al., 2014; Taylor, 2013). It has frequently inspired human music and prompted composers to incorporate it into their compositions (Baptista and Keister, 2005; Taylor, 2014). Birdsong and music might also share similar mechanisms and functions. For instance, the same regions of the mesolimbic reward pathway that respond to music in humans are active in female white-throated sparrows listening to conspecific song (Earp and Maney, 2012). Many bird species also coordinate their vocalizations by simultaneous or alternating chorusing (reviewed by Hall, 2009) or have been shown to temporally coordinate bodily movements in a dance-like manner with song during courtship (e.g. Dalziell et al., 2013; DuVal, 2007; Ota et al., 2015; Patricelli et al., 2002; Prum, 1990; Scholes, 2008; Soma and Garamszegi, 2015). Whether zebra finches (*Taeniopygia guttata*) coordinate singing among individuals has not been studied, but they do integrate song and dance during courtship in a non-random choreography (Ullrich et al., 2016; Williams, 2001). As in human ensemble music and dance, an isochronous pulse might serve as a temporal reference for duetting and dancing birds, facilitating the temporal coordination of vocalizations and movements. A recent study by Benichov et al. (2016) showed that zebra finches are also adept at coordinating the timing of unlearned calls in antiphonal interactions with a robot producing isochronously spaced calls.

When the robot produced some additional calls, timed to coincide with the birds’ response, both males and females quickly adjusted their calls to avoid jamming, successfully predicting the regular call pattern of the robot. The forebrain motor pathway that drives learned song production in male zebra finches seems to play a major role in this precise and flexible temporal coordination, not only in males but also in females that do not sing and have a much more rudimentary song system (Benichov et al., 2016). The capacity for ‘beat perception and synchronization’ that enables humans to extract the pulse from a complex auditory signal and move to it has so far been found only in several species of parrots (Hasegawa et al., 2011; Patel et al., 2009; Schachner, 2010) and a California sea lion (Cook et al., 2013). Since human music was used as a stimulus in these studies it is not clear how these findings relate to the animals’ own vocalizations: is there regularity in any learned natural vocalization signal that permits extraction of a regular pulse?

Song production in zebra finches has been successfully used as a model for studying vocal learning and production for several decades, motivated by its parallels to speech acquisition at behavioral, neural and genetic levels (reviewed by Berwick et al., 2012; Bolhuis et al., 2010; Doupe and Kuhl, 1999). Therefore a large body of knowledge exists about zebra finch song structure and development as well as their neurobiological basis. Zebra finch song learning and production is controlled by a neural network of specialized song nuclei (Bolhuis et al., 2010; Nottebohm et al., 1976). The nucleus HVC, cortical in nature, significantly contributes to the coding of song. Different ensembles of neurons fire short, sparsely occurring bursts of action potentials which, through a series of downstream nuclei, translate into a motor code controlling particular ensembles of muscles of the vocal organ (Fee et al., 2004; Hahnloser et al., 2002; Okubo et al., 2015). The level of resolution of our knowledge about how behavior is neurally coded is much finer grained in songbirds than in humans. So, while the present study in songbirds is guided by what we know about rhythm from human music it has the potential to shape our inquiry into the neural basis of human rhythm production and perception. The highly stereotypic structure of zebra finch song and the fact that it remains largely unchanged in the adult bird contributes to making it a good target for first investigations of periodicity, compared to more complex singers. We therefore analyzed zebra finch song rhythm, asking whether an isochronous pulse can be derived from the timing of its notes (signal-derived pulse; pulse^{S}).

Materials and Methods

Birds

This study used 15 adult male zebra finches, aged between 384 and 1732 days at the time of song recording. They were bred and raised at the Freie Universität Berlin breeding facility. Before entering this study, they were housed together with conspecific males, either in a large aviary or in a cage sized 90x35x45cm. In both cases they had acoustic and visual contact to female zebra finches held in other cages or aviaries in the same room. The rooms were kept under an artificial 12h/12h light/dark cycle at 25 ± 3°C. The birds had access to food, water, grit and cuttlebone ad libitum at all times. Birds in this study were solely used for song recording, a procedure for which the local authorities overseeing animal experimentation do not require a permit because it does not cause pain or discomfort. Information on the degree of relationship between the test subjects was only available for some of the birds. Of those, none were siblings, or had been raised by the same parents (3534, 4295, 4306, 4523 and g13r8). We cannot exclude dependencies in song structure arising from the possibility that pairs of birds were influenced by the same tutors.

Recording

For song recording, each male was transferred into a separate cage (40x30x40cm) inside a sound attenuation box (60x60x80cm), kept under a 12h/12h light/dark cycle. Audio was recorded through cardioid microphones, mounted at about 2cm distance from the center of the cage’s front wall in each box. These were connected to a single PC through an external audio interface. Audacity 2.0.3 was used to record a single-channel audio track (WAVE file, 44.1kHz, 16-bit) for each bird. Recording took place over a period of 3 years: 2013 (10 birds), 2014 (4) and 2015 (1) at varying times between 8 AM and 6 PM. In addition to song recorded in isolation (‘undirected song’) we also solicited so called ‘directed’ song from 9 of the 15 males by exposing them to the sound and sight of a female finch in a transparent plastic box placed in front of the recording cage. Directed and undirected songs of the same bird were recorded within one to three days.

Labeling

Recordings were segmented into smaller files of up to 10,000s (2h 46min) length and for each bird the segment containing most song was used for the analyses. Then, an IIR Chebyshev high-pass filter with a 1kHz cutoff was applied to remove low-frequency noise, using Avisoft Bioacoustics SASLab Pro 5.2.07 (henceforth SASLab). Note on- and offsets were determined by automatic amplitude threshold comparison in SASLab and saved as timestamps. All measurements obtained through this procedure were reviewed by visual examination of the song spectrogram and corrected by hand where necessary. Timestamps of falsely identified elements (i.e. above-threshold noise) were removed. In the rare cases in which notes could not be reliably measured by hand (due to overlapping noise or recording artifacts), all timestamps from the entire song were discarded. Introductory notes as well as calls preceding or following song were measured but not included in the subsequent analysis. All remaining timestamps were exported to MathWorks MATLAB R2012b 8.0.0.783 (henceforth Matlab), which was used for the rest of the analysis.

Song of zebra finches is composed of different notes, separated by silent intervals resulting from inhalation gaps. Notes consist of one or more sub-note elements, corresponding to neuromuscular gestures (hence called 'gestures'; Amador et al., 2013). For the analysis, we labeled notes with alphabetical letters. A string of recurrent note sequences is called a motif. Slight changes of note order can result in motif variants. For each bird, notes with the same bioacoustic features within a motif were labeled with the same alphabetical letter (for examples see **Figure 1**). The number of different notes sung by each individual ranged from four to seven, labeled *a* through *g*. The most commonly sung motif received the note labels in alphabetical order. The interval between notes (hence called 'gap') following note *a* was labeled *a*', following note *b b*' etc. Gaps were associated with the preceding syllable, as note duration correlates more strongly with the duration of the subsequent than the preceding gap (Glaze and Troyer, 2006). Introductory notes and calls were assigned different letters, and the corresponding timestamps were subsequently filtered out. When the first note of a motif was similar or identical to the introductory notes, it was considered the first note of the motif if it was present in each motif repetition and the gap between this note and the next was in the range typical of gaps within the motif. We used this criterion to distinguish between introductory notes and motif notes, because the former are separated by gaps of variable duration and the latter are not.

Rhythm analyses were performed on 'chunks', e.g. songs containing 1 to 10 continuously sung motifs. A new chunk started when a pause between two motifs lasted 300ms or more. Chunks containing fewer than four notes (e.g. *abc*) or fewer than three bioacoustically distinct notes (e.g. *ababab*) were discarded in order to avoid 'false positives', e.g. finding a regular pulse^{S} as a mathematical consequence of few notes or low complexity. For each bird we analyzed between 12 and 68 undirected song chunks, consisting of 4–34 notes each (9.1 ± 4.5; mean ±

std). Recordings of directed song contained 15–107 chunks, consisting of 4–42 notes (8.6 ± 5.9; mean ± std).

Pulse matching

We used a generate-and-test (GAT) approach to find the pulse^S (signal-derived pulse) that best fitted the note onsets. Essentially, isochronous pulses, i.e. strings of timestamps of equal intervals, were created for a range of different frequencies. To assess the goodness of fit of each of those pulses to a particular recorded song, the root-mean-square deviation (RMSD) of all notes in the song chunk from their nearest single pulse (i.e. timestamp) was calculated. Specifically, we aimed to determine the slowest regular pulse that could coincide with all note onsets of the particular song under investigation.

To numerically determine the lower range of pulse intervals we therefore used the shortest measured inter-onset interval (IOI) for each tested song chunk and added 10% to account for variability. Lower frequency limits calculated this way ranged from 5.5Hz (bird 4042) to 14.9Hz (bird 4669). Starting there, the pulse frequency was incremented in 0.01Hz steps up to 100Hz. Preliminary investigation revealed that the best fitting pulses very rarely had frequencies above 100Hz.

For each chunk, the pulse that fitted note onsets best was determined in the following way. Each of the pulses of incrementing frequency (by 0.01Hz steps) was displaced from the beginning of the recording by offsets ranging from zero to one period in 1ms steps. For each offset of each pulse the RMSD was calculated. The offset at which the RMSD was minimal was regarded as the “optimal offset”. The result of this process was a list of pulses of different frequencies (e.g. 5.50Hz, 5.51Hz, ..., 100Hz) for each chunk and their respective minimal RMSD.

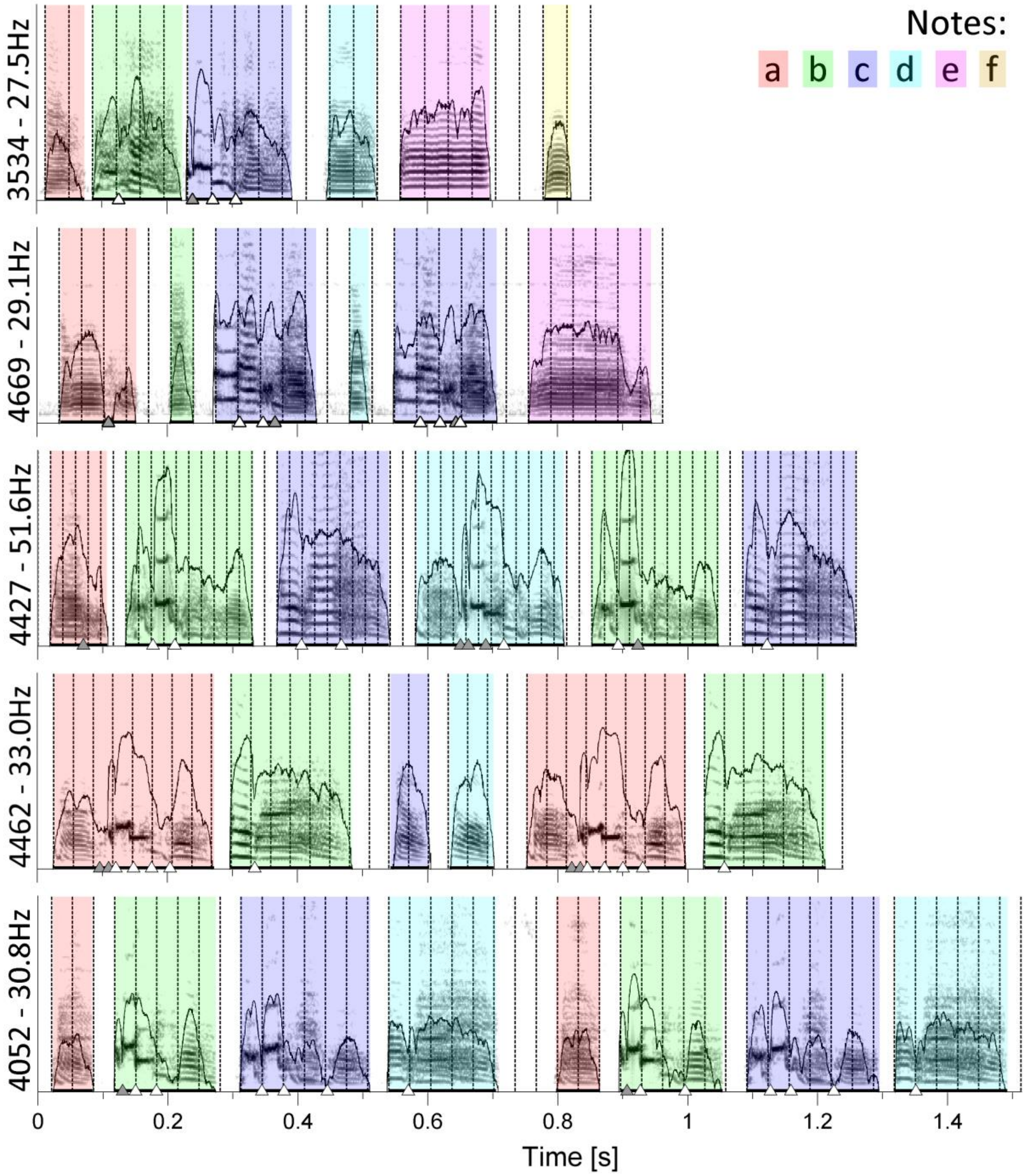


Figure 1 – Sonograms of song chunks from five different birds: 3534, 4669, 4427, 4462 and 4052 (top to bottom). For each song, note identity is indicated by color. Amplitude envelopes of the notes are outlined overlying the sonograms. Thicker black bars underneath the notes indicate note duration as determined by SASLab software. Isochronous pulses^{S} fitted to note onsets are marked as vertical dotted lines. Triangles indicate gesture transitions that either coincide with the pulse (white) or do not (grey, see Section Materials and Methods for details). Bird ID numbers and pulse frequencies are given to the left of each sonogram.

Because pulse frequency is mathematically related to RMSD, e.g. faster pulses are associated with lower RMSDs, we normalized the RMSD by multiplication with the pulse frequency, resulting in the ‘frequency-normalized RMSD’ (FRMSD). The FRMSD, unlike the RMSD, does not exhibit this long-term frequency-dependent decrease (**Supplementary Figure 1**). The RMSD on its own is an absolute measure of deviation. In contrast, the FRMSD was used in this study, measuring the deviation relative to pulse frequency. Essentially it indicates how well the pulse fits, taken into account its tempo. We selected the pulse with the lowest FRMSD as the best fitting pulse for each chunk.

Fourier analysis

A Fourier analysis was performed to confirm the results of the GAT pulse matching method (Saar and Mitra, 2008). To this end the note onset timestamps of each song were used to generate a point process, i.e. a number string with a 1ms time resolution, which was 1 at note onsets and 0 elsewhere. After performing a fast Fourier transform (FFT) on this string, we took the frequency of maximum power for each chunk (within the same Hz limits as above) and compared it to the frequency given by the GAT method.

Gesture transitions

Examination of the sonograms showed that not only note onsets, to which the pulses were fitted, but also onsets of distinct bioacoustic features within notes, corresponding to neuromuscular gestures, coincided with the pulse remarkably often. We identified possible time points of these gesture transitions quantitatively through a previously published algorithm that determines significant local minima in the amplitude envelope (Boari et al., 2015). Amplitude minima occur not only on gesture transitions, but also within gestures and notes of quasi-constant frequency (e.g. note *e* of 3534, **Figure 1**). Thus, we selected from the time points produced by the algorithm only those as gesture transitions that corresponded to clear discontinuities in the frequency trace, identified by visual examination of the sonograms. The percentage of gesture transitions that fell within certain ranges around the pulse, namely one tenth, one sixth and a quarter of the pulse period, were calculated. In **Figures 1 & 3** gestures with a distance of less than one sixth of the pulse period to the nearest pulse are highlighted.

Clustering

Visual examination revealed that the frequencies of the best fitting pulses of all song chunks from each bird tended to form clusters with individual values scattered between clusters (**Supplementary Figure 2**). To quantify this impression we used agglomerative hierarchical clustering, taking the group average of frequency distances as a dissimilarity measure. The dissimilarity threshold was set at 0.025 for all datasets. There was a significant positive correlation between cluster frequency mean and standard deviation (Linear regression; R^2 = 0.21; $p < 0.001$; n = 78), i.e. pulse frequency clusters were more tightly packed, the lower their frequency and vice versa. In order to obtain comparable clusters, different frequency transformations (square root, $\log_e$ and $\log_{10}$) were applied pre-clustering and their effect on this correlation was tested. Clustering in this study was done on the basis of $\log_{10}$-transformed frequency data because $\log_{10}$ transformation led to clusters with the least frequency-dependent standard deviation (Linear regression; R^2 = 0.0007; p = 0.824; n = 77).

Modeling

To address whether the pulse frequencies found through the GAT method could also be detected with similar goodness of fit in any arbitrary sequence of notes, we developed two sparse models of song with varying degrees of randomization. These models produce sequences of timestamps comparable to the ones obtained from the song recordings and consist of on- and offsets of virtual “notes”. The pulse deviation of the recorded bird songs was then compared to that of these artificial songs. We used the results to test the hypothesis that note onsets in zebra finch song align to an isochronous pulse more closely than expected by chance.

The first model, called “random sequence” model (Model R), creates virtual notes and gaps of random duration, albeit within a certain range. It ignores the note sequence of the original song, instead picking a new duration for each individual note. Therefore, the note sequence is not consistent across motif repetitions (e.g. natural song abcd abcd abcd compared to artifical song a’c’b’d’ g’i’h’k’ m’l’n’o’). Model R creates a pseudorandom value for each individual note in the analyzed song chunk and uses that as the duration of the corresponding modeled note. These pseudorandom values are drawn from a Pearson distribution using Matlab’s pearsrnd() function. The distribution’s mean, standard deviation, skewness and kurtosis are equal to the distribution of all observed note durations from either undirected or directed song, depending on which is to be modeled. The same is done for each gap, only this time the distribution is modeled on that of the observed gap durations. To be more conservative and avoid introducing high variability into the gaps of the model songs, outlier values and durations of gaps with unusually high mean and variability (gray points in **Figure 6**) were excluded in the creation of the pseudorandom number distribution.

Like model R, the second so called “consistent sequence” model (Model C) creates notes and gaps of random duration within the range of actually observed durations. Unlike model R though, model C takes the note sequence of the original song into account, keeping the duration of individual notes and their associated gaps in their sequence consistent across motif repetitions (e.g. natural song abcd abcd abcd compared to artificial song a’b’c’d’ a’b’c’d’ a’b’c’d’). In the first step of creating a virtual “song”, the different note types in the analyzed song chunk were determined (e.g. *a*, *b*, *c*, *d*). Then a set of 100 pseudorandom numbers were created for each note type of a bird, drawn from a standard normal distribution using Matlab’s randn() function. These sets were then transformed to have their respective means equal a random value (drawn from a uniform distribution) between the minimum and maximum of the means of all durations of each observed note type. The standard deviation of all sets equals the mean of the standard deviations of the durations of each note type in the database. The same was done for each gap, only this time using the standard deviations and range of means of the gap durations as the basis for the set transformation. Model C draws a random element from the appropriate set for each individual note in the analyzed song chunk and its associated gap, and uses that value as the duration of the corresponding modeled note/gap. The note/gap type durations in this model were kept consistent not only across motif repetitions within a chunk, but also across all analyzed chunks of a bird. This was achieved through seeding Matlab’s random number generator (RNG) before the creation of the duration sets during the modeling of each song chunk. The same seed value was used for all chunks of a single bird and different seed values were used for different birds. The RNG was seeded again before drawing the individual note/gap durations from the sets. Here, each chunk from a bird was assigned a different seed value. As a result, each modeled chunk used the same set of 100 durations for each note/gap type, but different values from that set were selected each time.

The deviations of two songs from their best fitting pulses cannot be compared if those pulses strongly differ in frequency. Just as the RMSD depends on the pulse frequency (described above), so does the FRMSD, as it measures deviation relative to pulse frequency. We therefore repeated the pulse matching process for both the recorded songs and the artificial songs, this time restricting the matched pulses to a certain frequency range that was different for each bird and identical for all recorded and artificial songs of one bird. Since we wanted to test whether we can find equally well fitting pulses for the artificial songs as we did for the recorded songs, we chose the mean of the largest frequency cluster of each bird as the center of the range. Furthermore, the upper bound of the range was twice the frequency of the lower bound. This assured that for any frequency outside of this range, either one integer multiple or one integer fraction of that fell within the range. We then compared the FRMSD values of all recorded songs and their best fitting pulse in their frequency range to those of the artificial songs. To exclude the possibility of the models producing particularly periodic or aperiodic songs by chance, the artificial song creation and subsequent FRMSD comparison were repeated 50 times for each song.

Statistics

To test the differences in pulse deviation between bird song and model song or between song contexts (directed and undirected song), a linear mixed effects analysis was performed (linear mixed model, LMM) using the statistical programming language R 3.0.2 (R Core Team, 2013) with the package lme4 (Bates et al., 2014). FRMSD was entered into the model as fixed effect. As FRMSD increases with the number of notes in a chunk (**Supplementary Figure 3**), the latter was used as a random intercept. P-values were obtained by likelihood ratio tests of the full model versus a reduced model without the fixed effect (FRMSD). One sample t-tests were used to test whether the percentage of gesture transitions occurred in certain ranges around the pulses significantly more often than expected by chance.

Results

For each of the 15 analyzed adult male zebra finches we found an isochronous pulse^S (signal-derived pulse) that coincided with all note onsets, using two independent analysis methods. For both, we used a continuous undirected song sample from each bird. The analyses were performed on segments, called 'chunks' that contained notes not separated by more than 300ms. Each analyzed chunk consisted of 1 to 10 motifs, composed of repeated unique notes, varying between 4 and 7 depending on the bird. Using a generate-and-test approach (GAT; see Section Materials and Methods) we identified for each chunk of a bird's recording a pulse^S that fitted best to the note onsets, i.e. had the lowest frequency-normalized root-mean-square deviation (FRMSD; **Figure 1**).

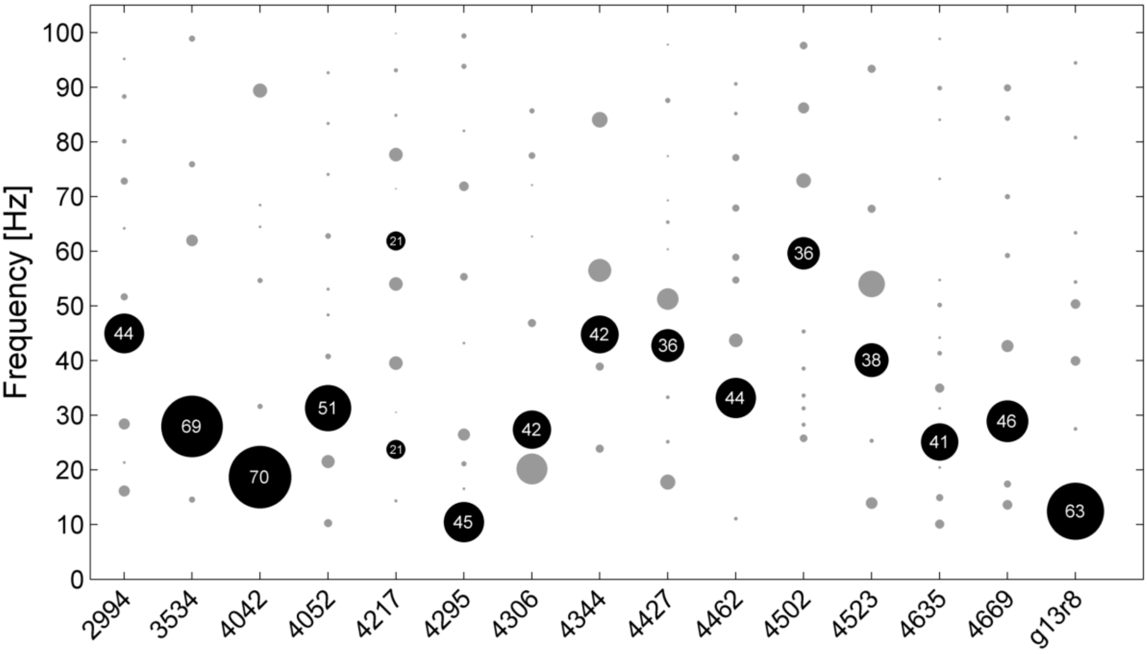


Figure 2 – Frequencies of the best fitting pulses^S for all analyzed chunks of undirected song for all 15 birds (bird ID numbers depicted on x-axis). Circles indicate frequency clusters as determined by hierarchical clustering analysis. Circle size corresponds to the percentage of chunks in the cluster relative to all chunks from the respective bird. The cluster containing the most chunks for each bird is black, the number inside the circle indicates the percentage of chunks within that cluster.

Pulse frequencies

For all birds except one, a particular best fitting pulse dominated, e.g. between 36% and 70% of analyzed chunks from each bird clustered around a particular frequency (black circles in **Figure 2**). For 11 of 15 birds, best fitting pulse frequencies lay between 25 and 45Hz. As a second analysis method to determine the best fitting pulses we applied fast Fourier transformations. We found that 91% of all chunks differed by less than 0.25Hz from the pulse frequencies identified by our GAT method.

In all birds a portion of songs were best fitted with pulses of different frequencies than those in the largest frequency cluster. Slight measurement inaccuracies may have led to different pulses having a lower deviation than the putative ‘real’ pulse in some songs. Song amplitude throughout the recordings varied slightly depending on the birds’ position in the cage and the orientation of their heads during singing. This is likely to have introduced some variability in the note onset measurement by amplitude threshold detection. The use of a dynamic time-warping algorithm for onset detection should provide more accurate measurements (e.g. Glaze and Troyer, 2006). Another factor that might tie into the variability in pulse deviation is the fact that zebra finches gradually slow down by a small degree during bouts of continuous song (Glaze and Troyer, 2006).

Gesture transitions

Often the best fitting pulse coincided not only with note onsets, but also with onsets of particular bioacoustic features within notes, corresponding to neuromuscular gestures. This was unexpected because the pulse was determined based on note onset times and not based on gesture transitions. To quantify this observation, we identified possible time points of gesture transitions through an algorithm that determines significant local minima in the amplitude envelope (Boari et al., 2015). Out of these time points we selected those that coincided with clear discontinuities in the frequency domain of the song spectrograms as gesture transitions. We did this for one song chunk from each of the 15 birds and found that overall 50.8% of gesture transitions fell within one sixth of the pulse period around single pulses (white triangles in **Figure 1**). If the gesture transitions were randomly distributed, 33.3% would be expected to fall in this range, as the range within one sixth of the period to either side of each pulse adds up to a third of total song duration. The percentage of gesture transitions that were within this range was significantly higher than the percentage expected by chance (one sample t-test, $t_{14} = 2.894$, $p = 0.0118$).

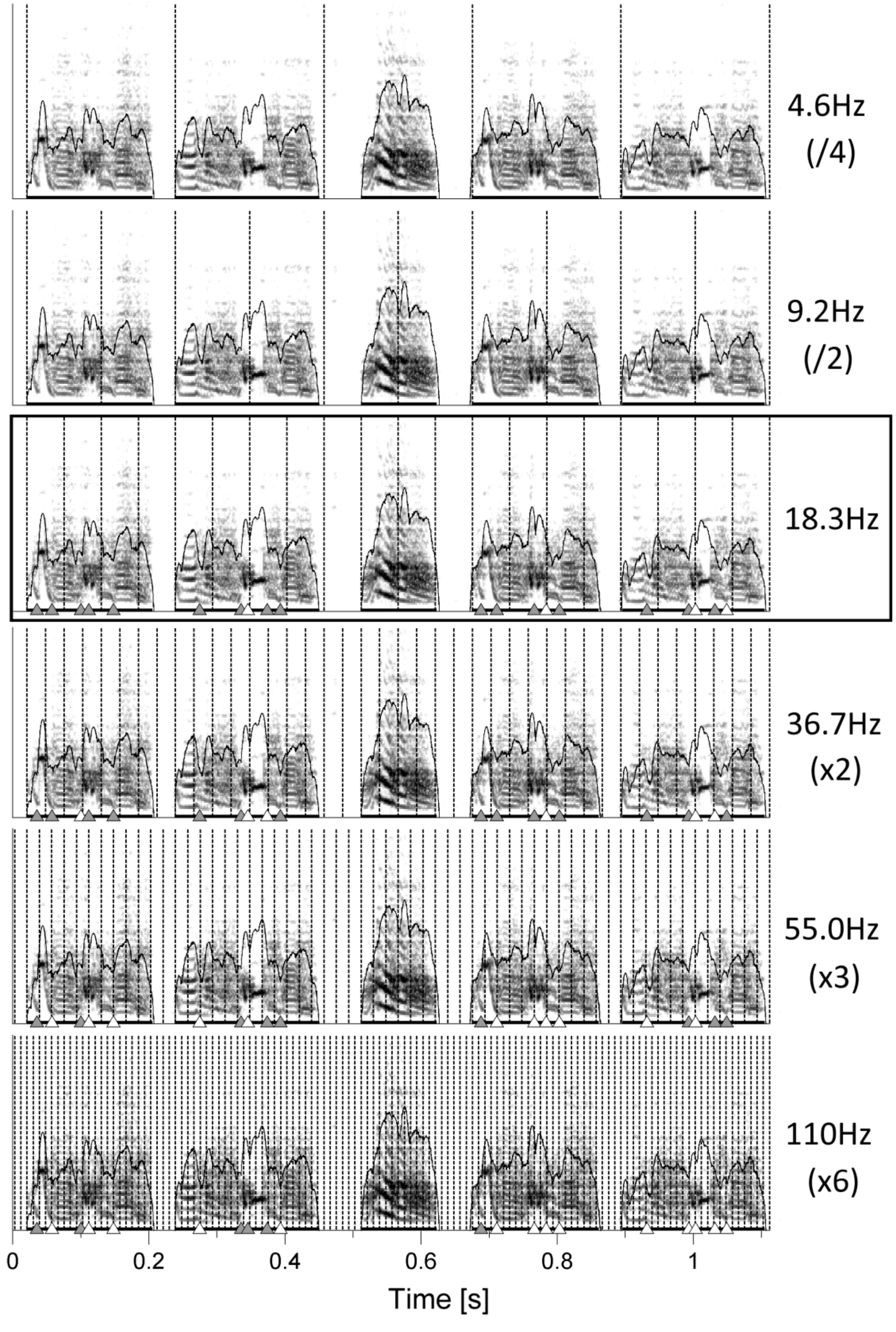


Figure 3 – Sonograms of one song chunk from bird 4042. Pulses^{S} are the one best fitting on note onsets (18.3Hz, black box) as well as integer fractions and multiples of that frequency. Triangles indicate gesture transitions that either coincide with the respective pulse (white) or do not (grey, see Section Materials and Methods for details).

We found that the pulses also had a significantly higher coincidence with the gestures than expected by chance when we chose other ranges. Within one tenth of the period around pulses lay 34.3% of the transitions, significantly more than the 20% expected by chance (t_{14} = 2.315, p = 0.0363). Within a quarter lay 65.9%, while 50% were expected if gesture transitions were randomly distributed (t_{14} = 2.639, p = 0.0195). Inspection of the spectrograms revealed many cases in which gesture duration equaled one or multiple pulse periods (for one pulse period see e.g. note *c* of bird 3534; *c* of 4669; *d* of 4427; *a* & *b* of 4462; *b* & *d* of 4052; for multiple pulse periods see *c* of 4427; **Figure 1**). In other cases multiple successive gestures added up to one pulse period (*c* of 3534; *c* of 4669; *b* of 4052). Note offsets did not systematically fall on the pulse, but in some cases notes consisting of a single gesture spanned one or more pulse periods (*d* of 3534; *b* & *e* of 4669; *c* of 4462; *a* of 4052). These observations imply a strong relationship between gesture durations and inter-onset intervals (IOI).

Motivated by the unexpected finding that the pulses fitted not only note onsets but also many of the gestures, we wondered whether even shorter gestures would coincide with faster pulses, corresponding to integer multiples of the slowest fitting one. Interestingly, inspection of one bird under five additional pulse frequencies revealed increasingly higher coincidence of pulses with all observable gesture transitions (**Figure 3**).

Directed song

Song directed by zebra finch males at females during courtship is less variable in various ways than when males sing so called ‘undirected’ song (Sossinka and Böhner, 1980). During courtship, zebra finches deliver their song slightly faster than during undirected singing (Cooper and Goller, 2006; Sossinka and Böhner, 1980). In addition, notes and the sequence in which they are sung are produced in a more stereotyped manner from rendition to rendition during directed singing. Whether the duration of notes is also less variable in the directed than the undirected singing context is not known (Glaze and Troyer, 2006). To find out whether directed song had a faster pulse or whether the pulse fitted better due to lower variability (i.e. lower FRMSD) we recorded 9 of the previously analyzed birds also in a directed song context.

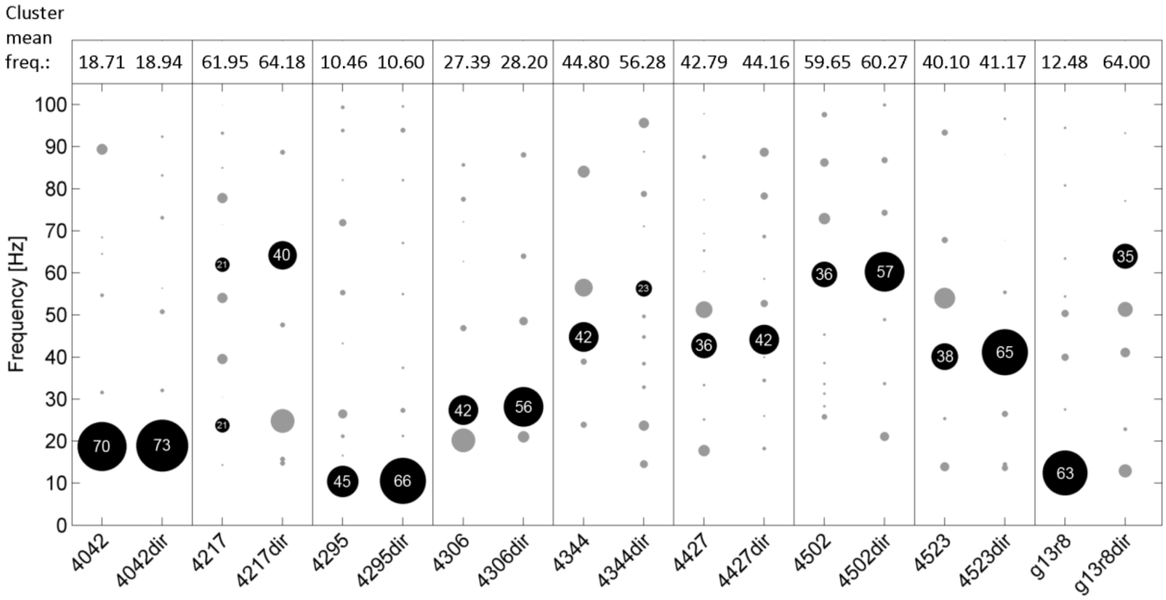


Figure 4 – Frequencies of the best fitting pulses^{S} for all analyzed chunks of undirected (left in each box) and directed song (right, e.g. “4042dir”) for 9 birds. Circles indicate frequency clusters as determined by hierarchical clustering analysis. Circle size corresponds to the percentage of chunks in the cluster relative to all chunks from the respective bird and condition. The cluster containing the most chunks for each bird and condition is black and the percentage of chunks in that cluster is given. Mean frequency of those clusters is shown on top of the figure.

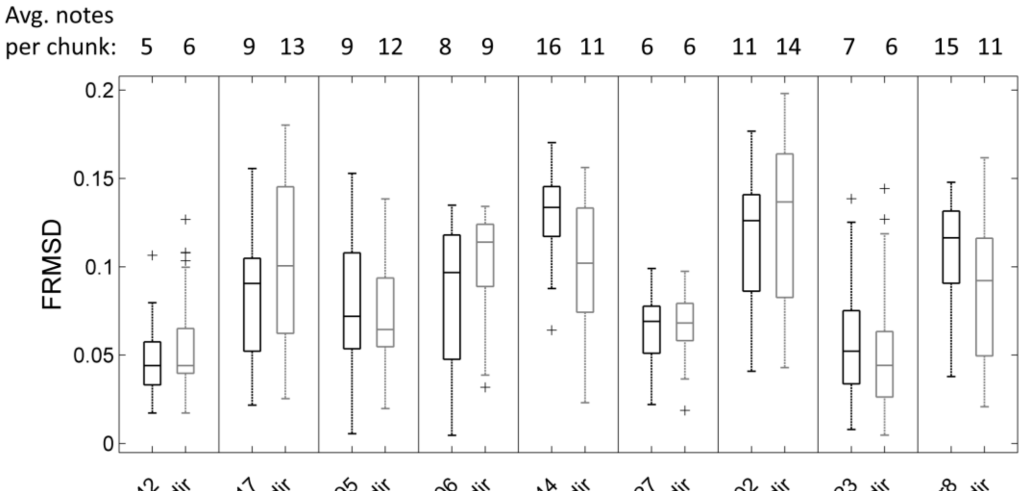


Figure 5 – Boxplot of frequency-normalized root-mean-square deviation (FRMSD) of note onsets to pulse^{S} for all undirected (left in each box) and directed song chunks (right in each box, e.g. “4042dir”) of 9 birds. Outliers are marked by crosses. There was no significant difference in FRMSD between undirected and directed song (LMM; $p>0.05$ for all 9 birds). Note that FRMSD increases with the number of notes in the chunk (**Supplementary Figure 3**), which was accounted for in the linear mixed model.

Mean pulse frequency of the largest cluster of undirected song was slightly lower than the nearest cluster in directed song in all birds (**Figure 4**). This is consistent with the fact that directed song is performed faster than undirected song (Kao and Brainard, 2006; Sossinka and Böhner, 1980; Woolley and Doupe, 2008), linked to a higher level of motivation during directed singing (Cooper and Goller, 2006). In 7 of 9 birds the pulse frequency best fitting most chunks was in the same range for undirected and directed songs. Interestingly, there was no significant difference in FRMSD between directed and undirected song (LMM; p>0.05 for all 9 birds; **Figure 5**), indicating that note onsets in directed song do not appear to have a stronger or weaker periodicity than those of undirected song.

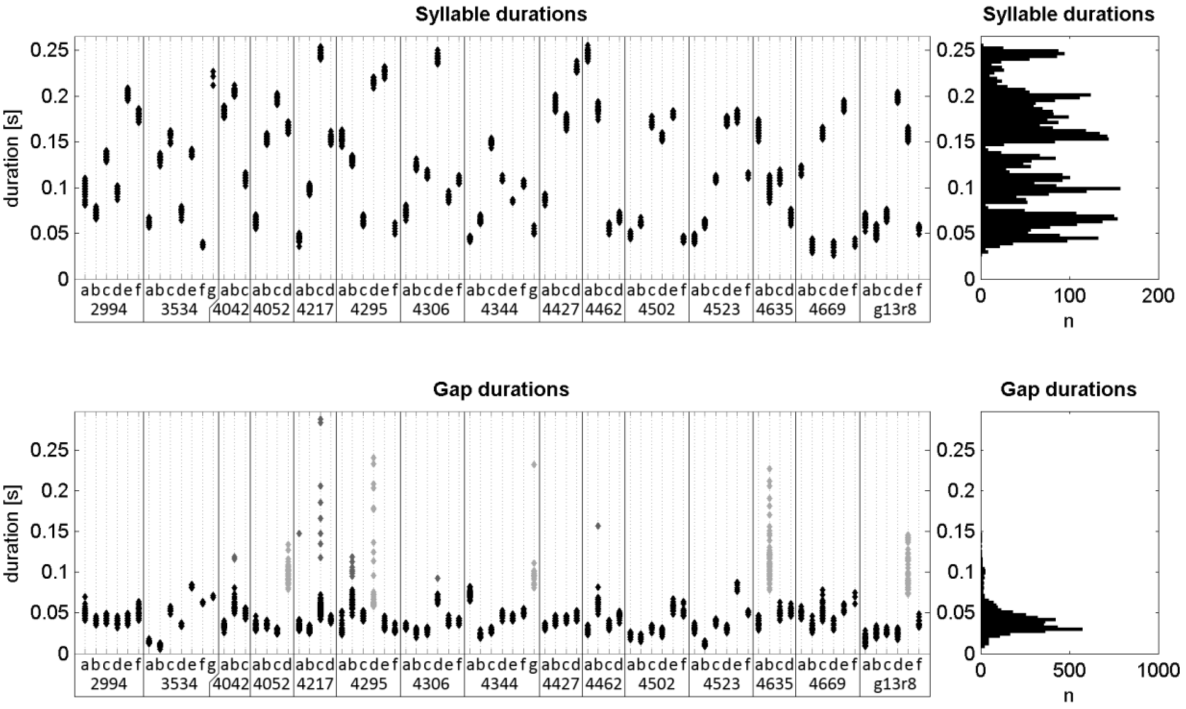


Figure 6 – Durations of all labeled notes and silent gaps of undirected song. Left: Scatterplots of durations separated by note type and bird ID (x-axis). Gaps are categorized by the preceding note. Right: Histograms of all note and gap durations. The durations of all 5006 notes ranged from 26–256ms (134 ± 60ms; mean ± std). Of all gaps, 4456 were within song chunks and had durations between 6 and 288ms (43 ± 23ms; mean ± std). Of those, 305 were either considered outliers or were part of gap types that had a high mean duration and large variance (dark gray and light gray points, respectively). Excluding those, the remaining gap durations ranged from 6–88ms (38 ± 12ms; mean ± std).

Comparison to randomized model “song”

To evaluate the fit of note onsets to the pulses, we created artificial “songs” consisting of randomized note and gap durations and compared the deviations of their note onsets from an isochronous pulse to those of the recorded birds.

The songs of the first model (“random sequence”, model R) do not replicate the note sequence of the recorded song. Instead a new pseudorandom duration is picked for each individual note and gap from a distribution modeled on that of the recorded notes and gaps. Through this comparison we could answer the question of whether a similar periodicity could be found in any arbitrary sequence of an equal number of (finch-like) song elements. We modeled the durations on the population of measured values of all birds in this study (**Figure 6**)
For each chunk we created 50 artificial songs with different randomized duration values each time and compared those to the recorded song chunks (see Section Materials and Methods for details). In overall 99% of comparisons bird songs had a lower pulse deviation (FRMSD) than the artificial songs created by model R (**Table 1**). In 88% of cases deviations were significantly lower compared to model songs, while the opposite never occurred (LMM; $p<0.05$). The analyzed natural songs therefore match a regular pulse significantly better than expected by chance. In other words, all inter-onset intervals (IOI) of one bird are proportional to each other (i.e. integer multiples of the pulse period), unlike an arbitrary sequence of (finch-like) durations.

In most cases IOIs within one chunk are not completely independent of each other, as notes or whole motifs are repeated, and repetitions of notes and associated gaps are mostly very similar in duration. Thus, we compared the recorded songs to a second model (“consistent sequence”, model C), that preserves the sequence of the recorded song. In all artificial songs produced by model C for one bird, for example, the notes based on note *a* have a similar duration. In 81% of comparisons, FRMSD was lower in the natural song than in the model C songs (**Table 1**). It was significantly lower in bird songs in 55% and significantly lower in model songs in 8% of comparisons (LMM; $p<0.05$). Model C songs performed better than model R songs in terms of pulse deviation, but still worse than the natural songs in the majority of cases. This leads us to conclude that the pulse is a result of the durations of the song elements as well as their sequence.

Table 1 – Results of the comparison between recorded undirected songs and model songs for all 15 birds. Artificial song creation was repeated 50 times for each song chunk with different pseudorandom values for each repetition. Values are the percent of repetitions in which the pulse^{S} deviation (FRMSD) was lower in natural songs compared to model songs and vice versa (first and third column in each block) and the percentage in which this difference was statistically significant (LMM; p < 0.05; second and fourth column of each block). Column mean is given at the bottom.

bird		**Model R (random sequence)**					**Model C (consistent sequence)**			
		deviation bird < model	**p < 0.05**	**deviation model < bird**	**p < 0.05**		**deviation bird < model**	**p < 0.05**	**deviation model < bird**	**p < 0.05**
2994		100	100	0	0		95	73	5	0
3534		98	70	2	0		98	50	2	0
4042		100	100	0	0		93	82	7	0
4052		100	100	0	0		84	66	16	0
4217		100	100	0	0		100	98	0	0
4295		100	100	0	0		77	57	23	11
4306		100	100	0	0		98	86	2	2
4344		100	60	0	0		95	50	5	0
4427		100	100	0	0		82	50	18	2
4462		100	100	0	0		77	43	23	7
4502		84	0	16	0		59	5	41	0
4523		100	98	0	0		98	77	2	0
4635		100	100	0	0		66	39	34	14
4669		100	94	0	0		70	39	30	9
g13r8		100	100	0	0		18	9	82	70
mean:		99	88	1	0		81	55	19	8

Discussion

We showed here for the first time that the song of a passerine songbird, the zebra finch, can be fitted to an isochronous pulse^{S} (signal-derived pulse). Note onsets coincided with pulses of frequencies between 10 and 60Hz (25–45Hz for most birds) and at different frequencies for each individual. In female-directed song this periodicity was not significantly different from undirected song. In addition to note onsets, many of the transitions between gestures within complex notes coincided with the same pulse as well, more so than expected by chance. Finding a pulse in zebra finch song raises questions about the underlying neural mechanism and its behavioral function. We cannot offer definite answers but some suggestions:

Song is coded in HVC neurons projecting to nucleus RA (HVC$_{RA}$) of the motor pathway. Different ensembles of those neurons fire at particular positions of each rendition of a song motif in a single, roughly 10 ms long, burst of action potentials (Hahnloser et al., 2002). Finding no connection between temporal firing of these neurons and note on- and offsets led to a working hypothesis, according to which HVC$_{RA}$ neurons act together like a clock, producing a continuous string of ticks (‘synfire chain’) throughout song on a 5–10ms timescale (Fee et al., 2004). Additional evidence for a clock-like signal in HVC controlling song production comes from experiments in which HVC was locally cooled (Long and Fee, 2008). This caused song to slow down up to 45% across all timescales, including gaps, while only slightly altering the acoustic structure. Since neural activity in RA gives rise to the motor code for song production (Mooney, 2009), one could expect to see the periodicity of the synfire chain reflected in the temporal structure of song. The frequency of this periodic activity would be in the range of 100–200Hz. The best fitting pulses found in this study, however, are between three and ten times slower. This suggests that the timing of song notes is organized on a slower timescale, occurring only at every nth clock tick, with n depending on the individual. Since we found these slower pulses in the songs of all birds and the songs were made up of several different notes, we propose that additional mechanisms must operate to orchestrate the timing signals of the internal clock into higher hierarchical levels giving rise to the slower pulse.

One such mechanism was proposed by Trevisan et al. (2006) to explain the diverse temporal patterns in the songs of canaries (*Serinus canaria*). They constructed a simple nonlinear model of respiratory control that could reproduce the air sac pressure patterns recorded during singing. This model, in which respiratory gestures emerge as different subharmonics of a periodic forcing signal, could predict the effects of local cooling of canary HVC on song notes (Goldin et al., 2013). As in zebra finches, canary song begins to slow linearly with falling temperature. At a certain point, however, notes begin to break into shorter elements, as forcing and respiration lock into a different integer ratio (e.g. from 2:1 to 1:1). Such a model might explain how a minimal time scale – e.g. in the form of an HVC synfire chain – could drive the timing of zebra finch notes on a subharmonic frequency. Zebra finch songs include more complex notes, in which several gestures of different duration are strung together in a single expiratory pulse. Our observation that gesture transitions preferentially coincided with the pulse on the note level, suggests that a similar mechanism might be responsible for periodic activation of the syringeal membrane.

Another study that recorded from $\mathrm{HVC}_{\mathrm{RA}}$ in zebra finches found that they fired preferentially at so called ‘gesture trajectory extrema’. These comprise gesture on- and offsets as well as extrema in physiological parameters of vocal motor control within gestures, namely air sac pressure and membrane tension of the syrinx (Amador et al., 2013). This suggests that gestures might be the basic units of song production and that their timing is coded early in the song-motor pathway. It cannot be ruled out in this scenario that a number of neurons continue to fire throughout the song, sustaining a clock-like functionality (Troyer, 2013). Our results imply that gestures transitions, like note onsets, contribute to song regularity. On average around half of the gesture transitions coincided with the pulse fitted to note onsets, significantly more than expected if they were randomly distributed. Those that did not, often occurred at the boundaries of gestures shorter than the pulse period, and successive short gestures often added up to one or multiple periods. These observations imply a strong relationship between gesture duration and inter-onset intervals (IOI), where gestures constitute the lowest level of the temporal hierarchy. Notes are on a higher level of this hierarchy, combining one or more gestures and the intervening inhalation gaps. In this sense the rhythmic structure in zebra finch song is reminiscent of the relationship between notes and phrases in metrical rhythms of human music.

What might be the behavioral function of the periodic organization of song? Temporal regularity in an auditory signal can facilitate the anticipation of events. In the wild, zebra finches live in large colonies that provide a very noisy environment. Females have to attend to the song of a single male against a backdrop of conspecific vocalizations as well as other sources of noise. Temporal predictability of an auditory signal has been shown to enhance auditory detection in humans (Lawrance et al., 2014), a phenomenon from which zebra finches could benefit as well. Humans are also thought to possess a form of periodic attention. When asked to judge the pitch difference of the last of an isochronous sequence of 10 tones of different pitches to the first, they were more successful when the last tone was on the beat than when it came slightly early or late (Jones et al., 2002). This supports the idea that accurate expectation (i.e. when a stimulus might occur) has a facilitating effect on attention, improving the ability to assess what the characteristics of the stimulus are (Huron, 2006; Seashore, 1938). The benefit of successful anticipation of events is that it allows the optimization of arousal levels and therefore the minimization of energy expenditure (Huron, 2006). When female zebra finches were given the choice between undirected and directed song from the same individual, they preferred to listen to the latter (Woolley and Doupe, 2008). In this study the strength of this preference was negatively correlated with the variability in fundamental frequency of multiple renditions of harmonic stacks (parts of notes with clear harmonic structure and little frequency-modulation, e.g. the first two gestures of note c in 4669’s song; **Figure 1**). This suggests that females attend to the pitch at specific times in a male’s song and show a preference for males that are able to consistently “hit the right note”. If that is the case, it would be advantageous for them to be able to anticipate the timing of these structures. Since these gestures seem to be periodically timed, females could benefit from a form of periodic attention. Instead of maintaining a constant high level of attention throughout the song or establishing a new set of expectations for each individual male, they could then simply adjust the “tempo” of their periodic attention to fit that of the singer. Females possess most of the nuclei of the song system, including HVC and RA, albeit much smaller. Until recently, the function of these nuclei was largely unknown, although in canaries HVC is implicated in song recognition and discrimination (Halle et al., 2002; Lynch et al., 2013). Benichov et al. (2016) showed that following disruption of the song system, the ability for precise, predictive timing of call coordination is greatly reduced in both males and females. It is therefore probable that females use some of the same structures that enable males to produce song with high temporal regularity, to either assess the quality of this regularity, or to use it for the anticipation of other song features.

Whether zebra finches perceive the apparent periodicity in song and if so, on what timescale, is still an open question that is crucial for our understanding of their function. A recent study showed that ZENK expression was found to be elevated in several auditory nuclei after exposure to arrhythmic song, where inter-note gaps were lengthened or shortened, compared to natural song (Lampen et al., 2014). The observed differences in neural response suggest that rhythm plays a role in auditory discrimination of songs. In another study, zebra finches learned to discriminate an isochronous from an irregular auditory stimulus (van der Aa et al., 2015). The birds did not generalize this discrimination well across tempo changes, suggesting that they discriminated based on differences in absolute time intervals rather than relative differences (i.e. equal intervals in the isochronous versus variable intervals in the irregular stimulus). Subsequently, zebra finches were asked to discriminate regular from irregular beat patterns, consisting of strongly accented tones with either a regular or a varying number of interspersed weakly accented tones. Here, some of the individuals were sensitive to the global pattern of regularity, but in general seemed to be biased towards attending to local features (ten Cate et al., 2016). The stimuli used in these studies, a series of metronome-like tones, lack features present in natural song – like timbre, pitch and amplitude modulation – which might be necessary for regularity detection, or for the birds to perceive it as a relevant signal. Further studies are needed to uncover whether zebra finches perceive a regular pulse^{P} in song.

The pulses^{S} fitted to song notes in the present study were faster by multiple factors than those humans preferentially perceive in musical rhythm. The latter are in a tempo range of around 500–700ms, which translates to a pulse^{P} frequency of 1.5–2Hz (Parncutt, 1994; van Noorden and Moelants, 1999). Zebra finches do possess a higher auditory temporal resolution than humans (Dooling et al., 2002). It is important to note, however, that pulses^{S} in the current study were fitted to all note onsets and represent the lowest level pulse in terms of note timing. In human music the perceived pulse^{P} is usually slower than this low level pulse with notes occurring between successive beats. If birds perceive a pulse^{P} in song, one could expect it to be on a longer timescale as well, e.g. integer multiples of the pulse^{S} period, where some but not all notes coincide with the pulse^{S} (see the top sonogram in **Figure 3** for an example).

Looking into the development of song regularity during song learning, especially in isolated juveniles, might provide further insights into whether periodicity is a result of song culture or whether it is neurally ‘hard-wired’.

Author Contributions

PN recorded songs and analyzed data; CS and PN designed study, prepared figures, interpreted results, drafted, and revised manuscript.

Funding

This research was funded by the BMBF project “Variable Töne” (FKZ: 01GQ0961) and the Excellence Cluster “Languages of Emotion” project “Do birds Tango?” (201).

Conflict of Interest Statement

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.

Acknowledgments

We thank Prof. Winfried Menninghaus and Julian Klein for fruitful discussions during the inception of this project.

References

Amador, A., Perl, Y. S., Mindlin, G. B., and Margoliash, D. (2013). Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature* 495, 59–64. doi:10.1038/nature11967.

Arom, S. (1991). *African Polyphony and Polyrhythm*. Cambridge, UK: Cambridge University Press doi:10.1017/CBO9780511518317.

Baptista, L. F., and Keister, R. A. (2005). Why birdsong is sometimes like music. *Perspect. Biol. Med.* 48, 426–443. doi:10.1353/Pbm.2005.0066.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., et al. (2014). Linear mixed-effects models using Eigen and S4.

Benichov, J. I., Benezra, S. E., Vallentin, D., Globerson, E., Long, M. A., and Tchernichovski, O. (2016). The forebrain song system mediates predictive call timing in female and male zebra finches. *Curr. Biol.* 26, 309–318. doi:10.1016/j.cub.2015.12.037.

Berwick, R. C., Beckers, G. J. L., Okanoya, K., and Bolhuis, J. J. (2012). A bird’s eye view of human language evolution. *Front. Evol. Neurosci.* 4, 5. doi:10.3389/fnevo.2012.00005.

Boari, S., Sanz Perl, Y., Amador, A., Margoliash, D., and Mindlin, G. B. (2015). Automatic reconstruction of physiological gestures used in a model of birdsong production. *J. Neurophysiol.* 5, 2912–2922. doi:10.1152/jn.00385.2015.

Bolhuis, J. J., Okanoya, K., and Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* 11, 747–759. doi:10.1038/nrn2931.

Brown, S., and Jordania, J. (2011). Universals in the world’s musics. *Psychol. Music* 41, 229–248. doi:10.1177/0305735611425896.

Carterette, E. C., and Kendall, R. A. (1999). “Comparative music perception and cognition,” in *The Psychology of Music*, ed. D. Deutsch (San Diego, CA: Academic Press), 725–791.

Cook, P., Rouse, A., Wilson, M., and Reichmuth, C. (2013). A California sea lion (Zalophus californianus) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *J. Comp. Psychol.* 127, 412–427. doi:10.1037/a0032345.

Cooper, B. G., and Goller, F. (2006). Physiological insights into the social-context-dependent changes in the rhythm of the song motor program. *J. Neurophysiol.* 95, 3798–3809. doi:10.1152/jn.01123.2005.

Dalziell, A. H., Peters, R. A., Cockburn, A., Dorland, A. D., Maisey, A. C., and Magrath, R. D. (2013). Dance choreography is coordinated with song repertoire in a complex avian display. *Curr. Biol.* 23, 1132–1135. doi:10.1016/j.cub.2013.05.018.

Dobson, C. W., and Lemon, R. E. (1977). Bird song as music. *J. Acoust. Soc. Am.* 61, 888–890.

doi:10.1121/1.381345.

Dooling, R. J., Leek, M. R., Gleich, O., and Dent, M. L. (2002). Auditory temporal resolution in birds: Discrimination of harmonic complexes. *J. Acoust. Soc. Am.* 112, 748–759. doi:10.1121/1.1494447.

Doupe, A. J., and Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annu. Rev. Neurosci.* 22, 567–631. doi:10.1146/annurev.neuro.22.1.567.

DuVal, E. H. (2007). Cooperative display and lekking behavior of the lance-tailed manakin (Chiroxiphia lanceolata). *Auk* 124, 1168–1185. doi:10.1642/0004-8038(2007)124[1168:CDALBO]2.0.CO;2.

Earp, S. E., and Maney, D. L. (2012). Birdsong: Is it music to their ears? *Front. Evol. Neurosci.* 4, 14. doi:10.3389/fnevo.2012.00014.

Fee, M. S., Kozhevnikov, A. a, and Hahnloser, R. H. R. (2004). Neural mechanisms of vocal sequence generation in the songbird. *Ann. N. Y. Acad. Sci.* 1016, 153–170. doi:10.1196/annals.1298.022.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition* 100, 173–215. doi:10.1016/j.cognition.2005.11.009.

Fitch, W. T. (2013). Rhythmic cognition in humans and animals: Distinguishing meter and pulse perception. *Front. Syst. Neurosci.* 7, 68. doi:10.3389/fnsys.2013.00068.

Fitch, W. T. (2015). Four principles of bio-musicology. *Philos. Trans. R. Soc. B Biol. Sci.* 370, 20140091. doi:10.1098/rstb.2014.0091.

Glaze, C. M., and Troyer, T. W. (2006). Temporal structure in zebra finch song: Implications for motor coding. *J. Neurosci.* 26, 991–1005. doi:10.1523/JNEUROSCI.3387-05.2006.

Goldin, M. a, Alonso, L. M., Alliende, J. a, Goller, F., and Mindlin, G. B. (2013). Temperature induced syllable breaking unveils nonlinearly interacting timescales in birdsong motor pathway. *PLoS One* 8, e67814. doi:10.1371/journal.pone.0067814.

Hahnloser, R. H. R., Kozhevnikov, A. A., and Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419, 65–70. doi:10.1038/nature00974.

Hall, M. L. (2009). "A review of vocal duetting in birds," in *Advances in the Study of Behavior - Vocal Communication in Birds and Mammals*, eds. M. Naguib, V. M. Janik, K. Zuberbühler, and N. S. Clayton (Oxford, UK: Academic Press), 67–122.

Halle, F., Gahr, M., Pieneman, A. W., and Kreutzer, M. (2002). Recovery of song preferences after excitotoxic HVC lesion in female canaries. *J. Neurobiol.* 52, 1–13. doi:10.1002/neu.10058.

Hasegawa, A., Okanoya, K., Hasegawa, T., and Seki, Y. (2011). Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Sci. Rep.* 1, 120. doi:10.1038/srep00120.

Hauser, M. D., and McDermott, J. (2003). The evolution of the music faculty: A comparative perspective. *Nat. Neurosci.* 6, 663–668. doi:10.1038/nn1080.

Honing, H., ten Cate, C., Peretz, I., and Trehub, S. E. (2015). Without it no music: Cognition, biology and evolution of musicality. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370, 20140088. doi:10.1098/rstb.2014.0088.

Hulse, S. H., and Page, S. C. (1988). Toward a comparative psychology of music perception. *Music Percept. An Interdiscip. J.* 5, 427–452. doi:10.2307/40285409.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: The MIT Press.

Jones, M. R., Moynihan, H., MacKenzie, N., and Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychol. Sci.* 13, 313–319. doi:10.1111/1467-9280.00458.

Kao, M. H., and Brainard, M. S. (2006). Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *J. Neurophysiol.* 96, 1441–1455. doi:10.1152/jn.01138.2005.

Kneutgen, J. (1969). “Musikalische” Formen im Gesang der Schamadrossel (Kittacincla macroura Gm.) und ihre Funktionen. *J. für Ornithol.* 110, 245–285. doi:10.1007/BF01671063.

Lampen, J., Jones, K., McAuley, J. D., Chang, S.-E., and Wade, J. (2014). Arrhythmic song exposure increases ZENK expression in auditory cortical areas and nucleus taeniae of the adult zebra Finch. *PLoS One* 9, e108841. doi:10.1371/journal.pone.0108841.

Lawrance, E. L. A., Harper, N. S., Cooke, J. E., and Schnupp, J. W. H. (2014). Temporal predictability enhances auditory detection. *J. Acoust. Soc. Am.* 135, EL357-EL363. doi:10.1121/1.4879667.

Long, M. A., and Fee, M. S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* 456, 189–194. doi:10.1038/nature07448.

Lynch, K. S., Kleitz-Nelson, H. K., and Ball, G. F. (2013). HVC lesions modify immediate early gene expression in auditory forebrain regions of female songbirds. *Dev. Neurobiol.* 73, 315–323. doi:10.1002/dneu.22062.

Marler, P. (2001). “Origins of Music and Speech: Insights from Animals,” in *The Origins of Music*, eds. N. K. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 31–48.

Mooney, R. (2009). Neural mechanisms for learned birdsong. *Learn. Mem.* 16, 655–669. doi:10.1101/lm.1065209.

Nettl, B. (2001). “An ethnomusicologist contemplates universals in musical sound and musical culture,” in *The Origins of Music*, eds. N. K. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 463–472.

Nottebohm, F., Stokes, T. M., and Leonard, C. M. (1976). Central control of song in the canary, Serinus canarius. *J. Comp. Neurol.* 165, 457–486. doi:10.1002/cne.901650405.

Okubo, T. S., Mackevicius, E. L., Payne, H. L., Lynch, G. F., and Fee, M. S. (2015). Growth and splitting of neural sequences in songbird vocal development. *Nature* 528, 352–357. doi:10.1038/nature15741.

Ota, N., Gahr, M., and Soma, M. (2015). Tap dancing birds: The multimodal mutual courtship display of males and females in a socially monogamous songbird. *Sci. Rep.* 5, 16614. doi:10.1038/srep16614.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Percept.* 11, 409–464. doi:10.2307/40285633.

Patel, A. D. (2008). *Music, Language, and the Brain*. New York, NY: Oxford University Press.

Patel, A. D., and Demorest, S. M. (2013). “Comparative music cognition: Cross-species and cross-cultural studies,” in *The Psychology of Music*, ed. D. Deutsch (Waltham, MA: Academic Press), 647–681.

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. *Curr. Biol.* 19, 827–830. doi:10.1016/j.cub.2009.03.038.

Patricelli, G. L., Uy, J. A. C., Walsh, G., and Borgia, G. (2002). Male displays adjusted to female’s response. *Nature* 415, 279–80. doi:10.1038/415279a.

Petkov, C. I., and Jarvis, E. D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* 4, 12. doi:10.3389/fnevo.2012.00012.

Prum, R. O. (1990). Phylogenetic analysis of the evolution of display behavior in the neotropical manakins (Aves: Pipridae). *Ethology* 84, 202–231. doi:10.1111/j.1439-0310.1990.tb00798.x.

R Core Team, R. (2013). R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria.*

Ravignani, A., Bowling, D., and Fitch, W. T. (2014). Chorusing, synchrony and the evolutionary functions of rhythm. *Front. Psychol.* 5, 1–15. doi:10.3389/fpsyg.2014.01118.

Rothenberg, D., Roeske, T. C., Voss, H. U., Naguib, M., and Tchernichovski, O. (2014). Investigation of musicality in birdsong. *Hear. Res.* 308, 71–83. doi:10.1016/j.heares.2013.08.016.

Saar, S., and Mitra, P. P. (2008). A technique for characterizing the development of rhythms in bird song. *PLoS One* 3, e1461. doi:10.1371/journal.pone.0001461.

Schachner, A. (2010). Auditory-motor entrainment in vocal mimicking species: Additional ontogenetic and phylogenetic factors. *Commun. Integr. Biol.* 3, 290–293. doi:10.4161/cib.3.3.11708.

Scholes, E. (2008). Evolution of the courtship phenotype in the bird of paradise genus Parotia (Aves: Paradisaeidae): Homology, phylogeny, and modularity. *Biol. J. Linn. Soc.* 94, 491–504. doi:10.1111/j.1095-8312.2008.01012.x.

Seashore, C. E. (1938). *The Psychology of Music*. New York, NY: McGraw-Hill.

Soma, M., and Garamszegi, L. Z. (2015). Evolution of courtship display in Estrildid finches: Dance in relation to female song and plumage ornamentation. *Front. Ecol. Evol.* 3, 1–11. doi:10.3389/fevo.2015.00004.

Sossinka, R., and Böhner, J. (1980). Song types in the zebra finch Poephila guttata castanotis. *Z. Tierpsychol.* 53, 123–132. doi:10.1111/j.1439-0310.1980.tb01044.x.

Taylor, H. (2013). Connecting interdisciplinary dots: Songbirds, “white rats” and human exceptionalism. *Soc. Sci. Inf.* 52, 287–306. doi:10.1177/0539018413477520.

Taylor, H. (2014). Whose bird is it? Messiaen’s transcriptions of Australian songbirds. *Twentieth-Century Music* 11, 63–100. doi:10.1017/S1478572213000194.

ten Cate, C., Spierings, M., Hubert, J., and Honing, H. (2016). Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Front. Psychol.* 7, 1–14. doi:10.3389/fpsyg.2016.00730.

Trehub, S. (2001). “Human processing predispositions and musical universals,” in *The Origins of Music*, eds. N. K. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 427–448.

Trevisan, M. A., Mindlin, G. B., and Goller, F. (2006). Nonlinear model predicts diverse respiratory patterns of birdsong. *Phys. Rev. Lett.* 96, 1–4. doi:10.1103/PhysRevLett.96.058103.

Troyer, T. W. (2013). The units of a song. *Nature* 495, 56–57. doi:10.1038/nature11957.

Ullrich, R., Norton, P., and Scharff, C. (2016). Waltzing Taeniopygia: Integration of courtship song and dance in the domesticated Australian zebra finch. *Anim. Behav.* 112, 285–300. doi:10.1016/j.anbehav.2015.11.012.

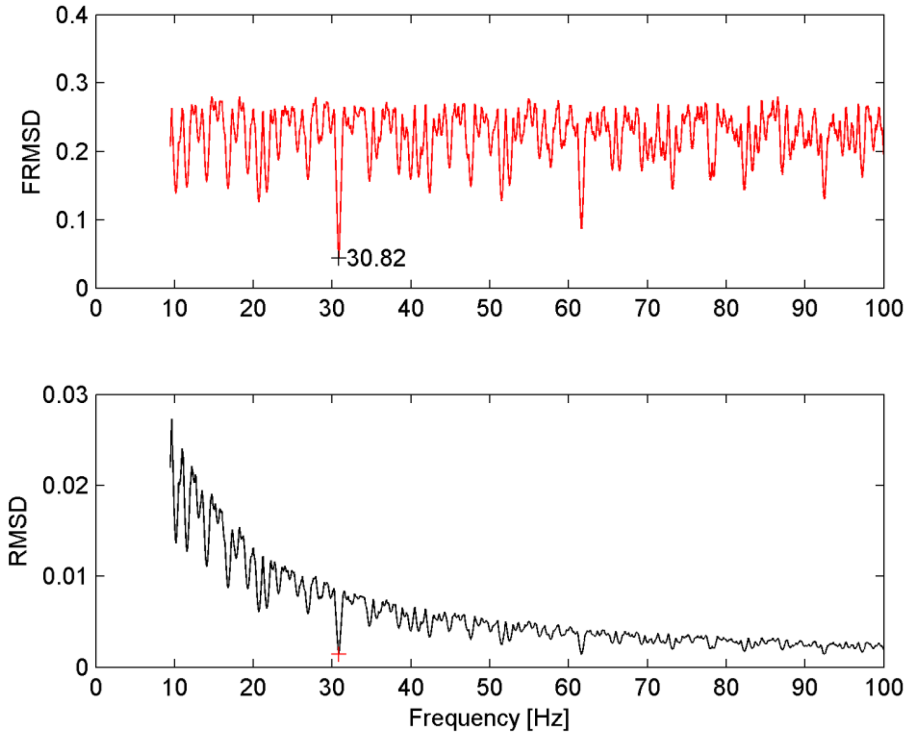
van der Aa, J., Honing, H., and ten Cate, C. (2015). The perception of regularity in an isochronous stimulus in zebra finches (Taeniopygia guttata) and humans. *Behav. Processes* 115, 37–45. doi:10.1016/j.beproc.2015.02.018.

van Noorden, L., and Moelants, D. (1999). Resonance in the perception of musical pulse. *J. New Music Res.* 28, 43–66. doi:10.1076/jnmr.28.1.43.3122.

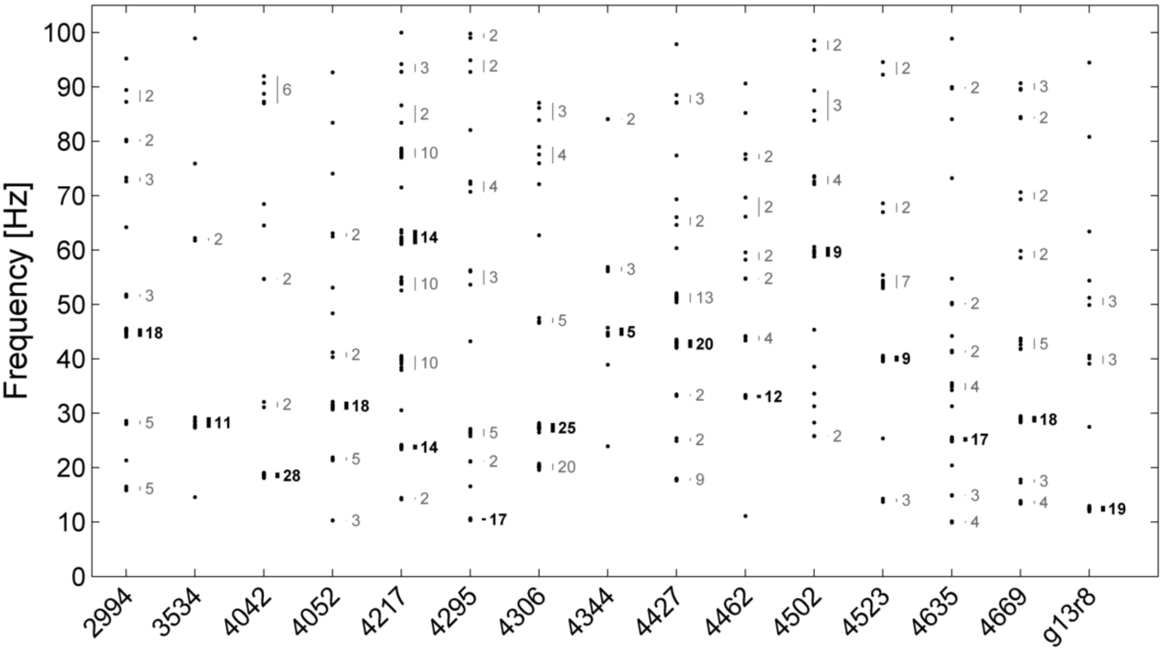
Williams, H. (2001). Choreography of song, dance and beak movements in the zebra finch (Taeniopygia guttata). *J. Exp. Biol.* 204, 3497–3506.

Woolley, S. C., and Doupe, A. J. (2008). Social context-induced song variation affects female behavior and gene expression. *PLoS Biol.* 6, e62. doi:10.1371/journal.pbio.0060062.

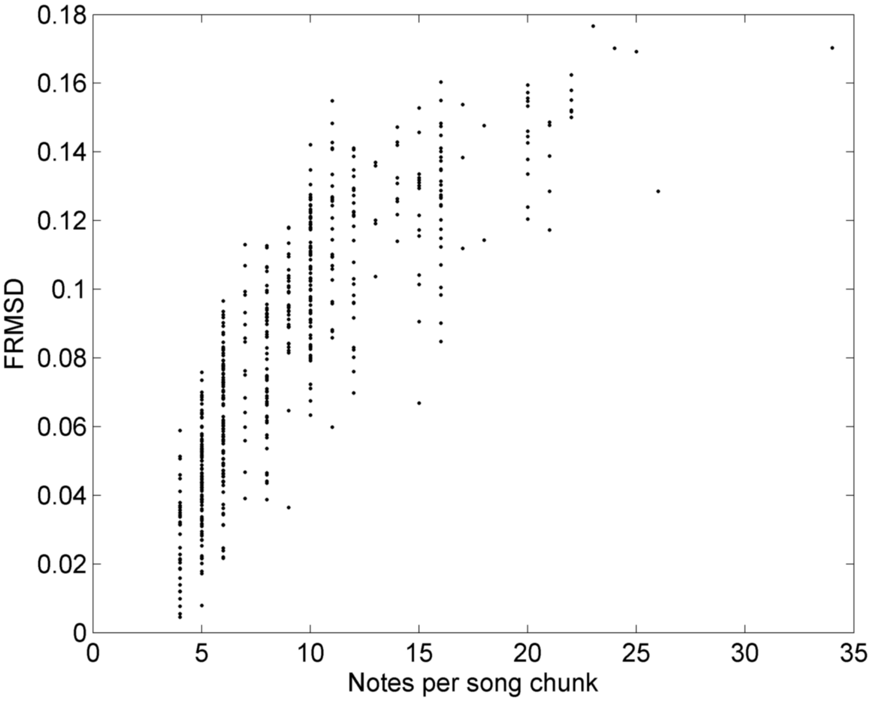
Supplementary Material



Supplementary Figure 1 – Frequency-normalized root-mean-square deviation (FRMSD, top) and root-mean-square deviation (RMSD, bottom) for one song chunk from bird 4052 (shown in the bottom sonogram in **Figure 1**). Isochronous pulses[S] for this chunk were created between 9.47 and 100Hz in 0.01Hz steps. RMSD of note onsets to nearest single pulse were calculated for each pulse frequency and multiplication with that frequency yielded FRMSD. The RMSD, unlike the FRMSD, decreases non-monotonically with increasing frequency. The pulse with the lowest FRMSD was selected as the best fitting pulse (30.82Hz in this case). Note that the pulse of double frequency – at 61.64Hz – has an RMSD equal to the 30.82Hz pulse but a higher FRMSD.



Supplementary Figure 2 - Frequencies of the best fitting pulses^S for all analyzed chunks of undirected song for all 15 birds (bird ID numbers depicted on x-axis). Lines next to points indicate which points belong to one cluster and numbers indicate the number of chunks in that cluster. This figure is identical to **Figure 2** in the main article, except that it shows individual points.



Supplementary Figure 3 – FRMSD versus number of notes for all analyzed song chunks.

Publication C: General isochronous rhythm in a neotropical bat

Burchardt, L., **Norton, P.**, Behr, O., Scharff, C., and Knörnschild, M. (accepted for publication). General isochronous rhythm in echolocation calls and social vocalizations of the bat Saccopteryx bilineata. *Royal Society Open Science*.

General isochronous rhythm in echolocation calls and social vocalizations of the bat *Saccopteryx bilineata*

Lara S. Burchardt[1], Philipp Norton[1], Oliver Behr[2],
Constance Scharff[1]*, Mirjam Knörnschild [1,3,4]*

[1]Department of Animal Behaviour, Freie Universität Berlin
[2]Chair of Sensor Technology, University of Erlangen-Nuremberg
[3]Smithsonian Tropical Research Institute, Barro Colorado Island, Panamá.
[4]Museum für Naturkunde Berlin - Leibniz Institute
for Evolution and Biodiversity Science
* Joint senior authors.

Abstract

Rhythm is an essential component of human speech and music but very little is known about its evolutionary origin and its distribution in animal vocalizations. We found a regular rhythm in three multisyllabic vocalization types (echolocation call sequences, male territorial songs, and pup isolation calls) of the neotropical bat *Saccopteryx bilineata*. The intervals between element onsets were used to fit the rhythm for each individual. For echolocation call sequences, we expected rhythm frequencies around 6-24 Hz, corresponding to the wingbeat in *S. bilineata* which is strongly coupled to echolocation calls during flight. Surprisingly, we found rhythm frequencies between 6 Hz and 24 Hz not only for echolocation sequences but also for social vocalizations, e.g. male territorial songs and pup isolation calls, which were emitted while bats were stationary. Fourier analysis of element onsets confirmed an isochronous rhythm across individuals and vocalization types. We speculate that attentional tuning to the rhythms of echolocation calls on the receivers' side might make the production of equally steady rhythmic social vocalizations beneficial.

Introduction

Music is widespread in all human cultures but its evolutionary origin is poorly understood (1). The field of biomusicology attempts to answer questions on the origin and purpose of music by focusing on the physiological, psychological, behavioral and evolutionary aspects of music in a comparative approach. That approach includes not only human music but musicality as a term for different traits that occur spontaneously and are based on and constrained by biology and cognition in animal vocalizations (2, 3). Music contains several key components – that can be separately investigated as musicality traits – such as pitch (governing melody and harmony), rhythm (defining temporal structure), and sonic qualities named timbre (1). Our study focuses on rhythm as a musicality trait.

Rhythm can be defined as a "systematic patterning of sound in terms of timing, accent, and grouping" (4). Overall, our intuitive understanding of rhythm concerns periodicity, which is the expectation of a recurrent event. One special kind of periodic rhythm is an isochronous beat, as produced e.g. by a metronome. In an isochronous beat, all beats have the same length and all beat-to- beat intervals have the same length (4). When it comes to analyzing animal vocalizations for rhythmicity, two questions need to be answered. (a) How well can an animal produce a certain rhythm and (b) are rhythmic patterns similar or different between vocalization types and between individuals? Another interesting comparison not regarded in this project would be between species. Furthermore, the relevance and biological constraints shaping an existing rhythm need to be discussed.

In a recent study on rhythm in song of zebra finches (*Taeniopygia guttata*) both questions were answered. Individual males had a distinct isochronous rhythm which fitted syllable onsets better than expected by chance. Distinct rhythms between individual males ranged from 10 to 60 Hz (5). Other examples of animals producing rhythmic signals include the Palm Cockatoo (*Probosciger aterrimus*) which uses tools to 'drum' a quasi-isochronous beat on branches in a consistent context (the rhythm frequencies were not analyzed in detail) (6) or chimpanzees cracking baobab fruits in a fashion probably eligible to generate individual signatures, which might help to recognize unseen companions (7). A subsequent question would be whether animals can distinguish between rhythms, isochronous or otherwise. Rats for example are able to discriminate between different isochronous rhythms in a habituation-dishabituation experiment (8) while European Starlings are able to discriminate between rhythmic and arrhythmic patterns (9). Moreover, the first instance for a biologically relevant rhythm in non-human mammalian vocalizations was found in the Northern Elephant Seal, where males can discriminate between familiar and unfamiliar male opponents using the temporal structure of vocalizations. Rhythms apparently differ between individuals in a way that facilitates discrimination of individuals (10). Nevertheless, compared to other aspects of vocal communication, studies on rhythmicality in animals are still sparse.

Our study aims to broaden the knowledge of rhythm in animal vocalizations by investigating whether isochronous rhythms can be found in different vocalization types of bats. Specifically, we investigated how well different vocalizations of bats fit an isochronous beat and whether the patterns are similar between individuals or vocalization types.

We studied the Neotropical greater sac-winged bat *Saccopteryx bilineata* which has a rich vocal repertoire (11) and is capable of vocal production learning (12). The species' vocal repertoire consists of distinct vocalization types that are uttered in different behavioral contexts. In this study, we focused on echolocation call sequences, isolation calls, and territorial songs, all of which are multisyllabic vocalizations with clear syllable onsets. Isolation calls are produced by pups to solicit maternal care and by adult males to appease dominant conspecifics (13-15). With a length of up to 2 seconds and a multisyllabic structure, isolation calls of S. *bilineata* are amongst the most acoustically complex bat isolation calls described (13, 14). Territorial songs are produced by adult males to repel rivals and attract mating partners (11, 16). They are acquired by imitating conspecifics' song during ontogeny (12, 16, 17). Echolocation calls are produced by male and female S. *bilineata* for orientation, navigation, and insect prey capture (18); in addition to their primary function, echolocation calls facilitate social communication among group members (19). We chose those three vocalization types to get insight into rhythmicity in both innate vocalizations (isolation calls, echolocation call sequences) and learned ones (territorial songs) as well as to investigate potential age differences in rhythmicity (in pup isolation calls).

The individual rhythms found in zebra finch song were discussed to be potentially advantageous for anticipating events, i.e. song syllables. Tuning attention to rhythmic production could reduce 'attentional energy' (sensu: (20)) and increase signal perception (5).Correspondingly, rhythmicity in bat vocalizations might be adaptive for saving metabolic energy since flight is energetically costly. In many bat species, echolocation calls are coupled to wingbeat and respiratory cycle (e.g. (21-24)), which is thought to be energy efficient. Moreover, not only behavioral correlates can be found but neuronal correlates: wingbeat and echolocation calls in *Roussettus aegyptiacus* are tightly coupled around theta frequencies (5 –

12 Hz, (25)), brain wave frequencies which are known to play a role in active movement and stimulus intake (26). Consistently, preliminary data on S. *bilineata* suggests a wingbeat of around 6- 12 Hz (pers. communication H.-U. Schnitzler). During search flight one or two echolocation calls might be uttered per wingbeat, which corresponds to echolocation call intervals of 6 to 24 Hz (wingbeat frequencies of around 6-12 Hz) found in other studies on S. *bilineata* (18, 27, 28).

Because of the coupling of wingbeat and echolocation pulses, we predicted isochronous pulses with frequencies between approximately 6 to 24 Hz in echolocation call sequences of S. *bilineata*. We assumed that echolocation call sequences would fit a specific isochronous pulse significantly better than random vocal sequences would. Moreover, we expected this rhythm to be similar between individuals due to common physiological constraints. Since social vocalizations (pup isolation calls and male territorial songs) are uttered by perched bats in the day roost, not coupled to wingbeat, we predicted to find individually different rhythms that might support vocal discrimination of different individuals, as previous research suggests.

Methods

Labeling of vocalization types

We analyzed three different vocalization types of S. *bilineata*, namely isolation calls, territorial songs, and echolocation call sequences (**Figure 1**). Isolation calls and territorial songs are multi- component vocalization types containing four different element types each, while echolocation call sequences are series of one element type with alternating frequencies.

For each vocalization, the on- and offset of its elements and the duration of the silent gaps between elements was determined for subsequent analyses. For isolation calls and territorial songs, element on- and offsets were determined manually based on oscillograms (see (13) and (29) for details). For echolocation call sequences, we used an automatized procedure in Avisoft SASLab Pro (based on amplitude detection threshold; - 20 dB relative to the call's peak frequency) to determine element on- and offsets.

We analyzed isolation calls from 25 pups (10 males, 13 females, 2 not sexed) belonging to a population of wild S. *bilineata* in Costa Rica (see (13) for details on study site and sound recordings). Each isolation call contained 5 – 26 elements (14 ± 3.5, mean ± SD) and was composed of 2 – 4 different element types (mean: 3 element types). We followed the nomenclature introduced in an earlier study (14) and labeled the element types (a-d) as introductory elements (a), simple variable elements (b), composite elements (c), and simple stereotyped elements (d). Data for each pup consisted of 20 isolation calls, recorded at two ontogenetic stages (non-volant and volant; 10 isolation calls each). Only one call per pup and day was selected to minimize temporal dependence among vocalizations.

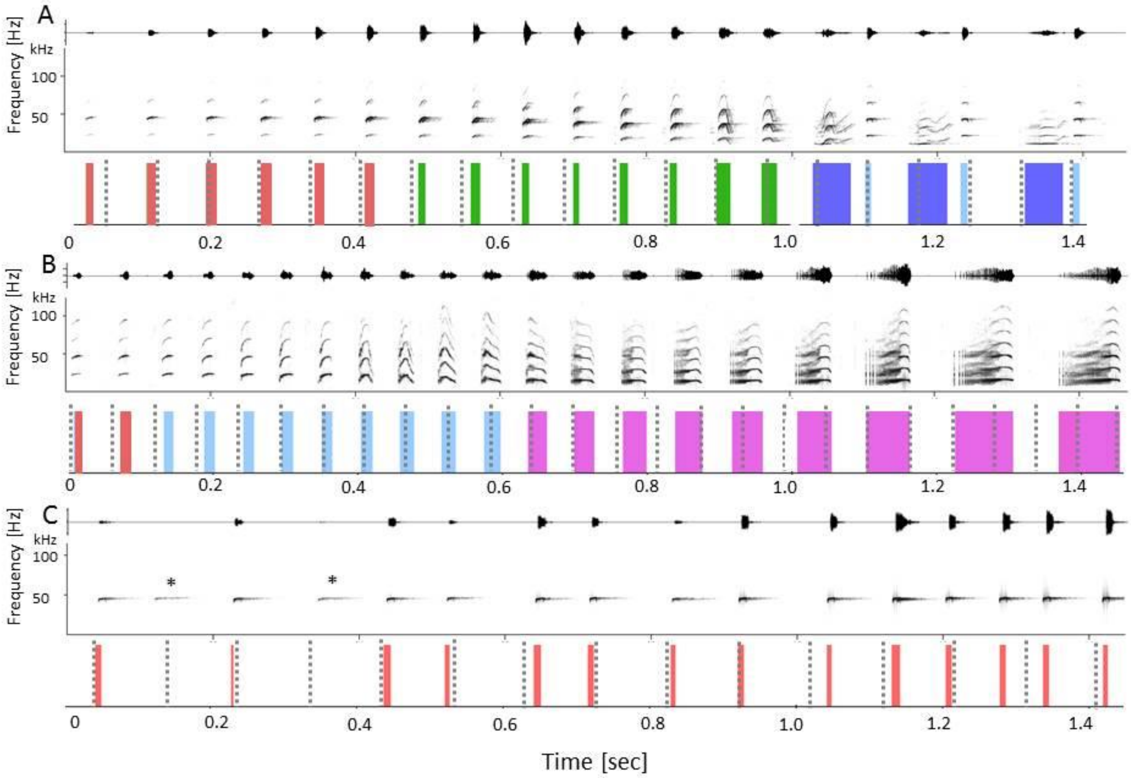


Figure 1 – Rhythm^S fits well on three vocalization types: Oscillograms (top rows in A-C) and spectrograms (middle rows) of vocalizations (A: isolation call, B: territorial song, C: echolocation call sequence) with fitted rhythm^S as dotted lines in the bottom row. Element durations are indicated by coloured bars, measured from the oscillograms. Note that echoes visible in the spectrograms may make the elements appear longer than they are in the oscillograms. Different colours indicate different element types (described in earlier studies (14, 29)). (A) Introductory elements, simple variable elements followed by composite elements and simple stereotyped elements in an alternating order. (B) Echolocation-like calls (comparable to the introductory elements in A), short tonal elements and buzz elements. (C) Echolocation calls. (*) indicates two elements not being labelled due to a low amplitude.

We analyzed territorial songs of 14 adult males belonging to a population of wild S. *bilineata* in Costa Rica (see (29) for details on study site and sound recordings). Data for each male consisted of 10 – 11 songs, which were recorded on different days. Each song contained 6 – 46 elements (20 ± 8.0 mean ±SD) and was composed of 1 – 5 five different element types (mean: 3

different element types). We followed the nomenclature introduced in an earlier study (29) and labeled the element types (a-e) as short tonal elements (a), buzz elements (b), trills (c), noise bursts (d) and echolocation-like calls (e) (**Figure 1**).

Sequences of echolocation calls were recorded from 33 wild Costa Rican S. *bilineata* (15 males, 18 females) when they were released after capture, i.e. in a non-foraging context. Calls of known individuals were recorded in standardized release situations in relatively open space (e.g. at a forest clearing). Recorded calls resembled normal search calls (see (19) for details on study sites and sound recordings). Echolocation call sequences consisted of 11 – 38 elements (21± 6.95, mean ± SD) with no further differentiation into different elements types. One echolocation call sequence per bat was used for further analysis.

Assessment of best-fitting rhythms

Simply analyzing inter-onset intervals of social vocalizations, as is often done for echolocation call sequences (e.g. (18, 27, 28)), is problematic since this would oversimplify the temporal structure of multisyllabic social vocalizations with strongly varying syllable durations. Other approaches to analyze temporal structure of animal vocalizations include generate-and-test approaches or Fourier Analysis (30). We chose a generate-and-test approach (GAT approach) originally developed for rhythm analysis in zebra finch song (5). The GAT approach allowed us to find an isochronous rhythm (i.e. a pattern with equal time intervals) that best fitted the onsets of elements in a given sequence. We named this best fitting rhythm 'signal-derived rhythm' or rhythm^S (same as pulse^S in (5)). The GAT approach was performed by a custom MATLAB program (see (5) for more details). It creates isochronous pulse trains in 0.01 Hz increments in a predefined frequency window of 5-100 Hz (i.e. 5-100 pulses per second). The lower range of rhythm frequencies was determined by expected values (18, 27, 28) the upper range experimentally by testing different ranges. 100 Hz was deemed appropriate because, when testing for up to 200 Hz only very few values for best fitting rhythms lay above 100 Hz. Restricting the frequency window was a question of minimizing computing time. For each rhythm, temporal deviations of each element to the nearest pulse gave an overall root-mean-square deviation (RMSD). Pulses were offset (+ one phase in 1 ms steps, see (5)) to minimize the RMSD. Since RMSD is negatively correlated with rhythm frequency (i.e. faster rhythms generally result in lower RMSD values; see **Figure 2**, bottom), we normalized the RMSD by multiplying it by the respective rhythm frequency, yielding a measure for deviation relative to the rhythm period. The resulting frequency-normalized RMSD (FRMSD) was used to assess the goodness-of-fit for each rhythm: the lowest FRMSD indicated the best-fitting rhythm frequency. This way the slowest isochronous rhythm, coinciding best with element onsets, was found (**Figure 2**).

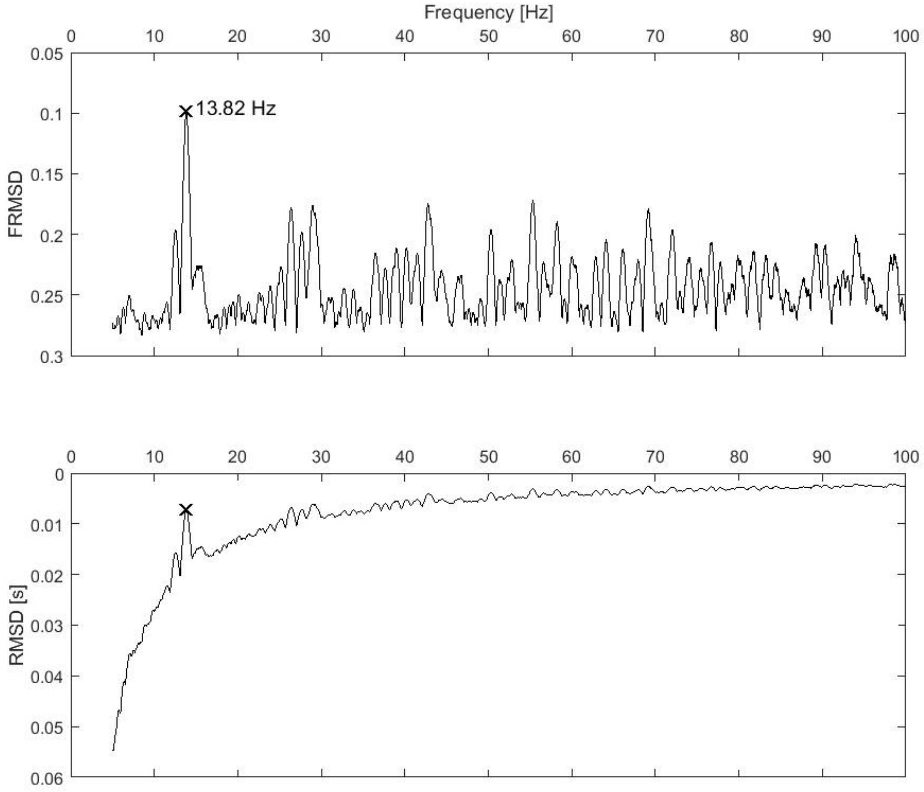


Figure 2 – Optimization Process: Best fitting rhythms were found by selecting the rhythm with the lowest corresponding FRMSD (black cross with corresponding rhythm^S), the frequency normalized root-mean-square deviation (upper panel). The lower panel shows the corresponding RMSD values.

Clustering

A visual examination of the resulting best-fitting rhythm^S indicated an accumulation of certain frequency values for each individual and vocalization type. Rhythm frequencies showed a strong right skewness, which is why common measures such as mean, or median would have been inaccurate. Therefore, we performed a cluster analysis to assess whether specific rhythm frequency clusters existed. We applied an agglomerative, hierarchical clustering algorithm which used the group average of frequency distances as a dissimilarity measure (dissimilarity threshold was set to 0.05 for all data sets). The frequency data were log10-transformed before clustering because an earlier study (5) showed that log10-transformation resulted in comparable clusters for different frequencies since these clusters had the least frequency- dependent standard deviation.

Modeling

To confirm that the rhythm frequencies obtained by the GAT approach are an inherent property of the respective vocalization type and cannot be found in arbitrary element sequences, we created artificial temporal vocalization patterns based on the previously measured element and gap durations, assessed their FRMSD values and compared them to the FRMSD values of the original vocalization types. We created two different types of artificial vocalization patterns that were used in different models: In Model 1 we used artificial vocalization patterns with randomized element and gap duration but intact sequence information (i.e. the correct order of consecutive elements). In Model 2, we used artificial vocalization patterns where each element and gap were replaced with a random duration, irrespective of element type and sequence. Model 2 did not apply to echolocation call sequences because they consisted of only one element type repeated in series, thus making the dismissal of sequence information pointless. Element and gap durations for both models were drawn randomly out of the pool of original recorded durations of the same type from all individuals (elements a–e and gaps following elements a–e respectively). The respective pool from which durations were drawn contained only element and gap durations of the vocalization type (isolation calls, territorial songs, or echolocation call sequences) to be modeled.

For each vocalization, we ran both Model 1 and 2 (not for echolocation call sequences, see above) 50 times. For every iteration, a new FRMSD value was obtained. We calculated the means of all model FRMSD values per individual and compared them to the means of all original FRMSD values per individual.

Fourier Analysis

Results of the GAT approach were compared to FFT analyses of all sequences (following (5, 31)). Timestamps of element onsets were used to form a binary point process. We created strings with a time resolution of 5 ms in which only events (i.e. element onsets) were represented by '1', everything else in the string was represented by '0'. The higher the temporal resolution of the input data, the lower the frequency resolution of the FFT output will be. With the sequence lengths available to us, a time resolution of 5 ms proved to be the best compromise between the two constraints. After calculating a fast Fourier analysis, frequencies of maximum power were selected as 'best fitting rhythms and the pattern compared to GAT-results. A customized Matlab script was used for the analysis.

Statistics

Data distribution was assessed using a Shapiro-Wilks test for all datasets. Artificial data from both randomizations (1 and 2) were compared to original data with repeated measures ANOVA (Tukey's post hoc comparison) for isolation calls and territorial songs. Echolocation call sequences were tested against randomization 1 via a Welch-corrected t-test because variances differed significantly. A paired t-test was used to compare the results of different ontogeny stages in isolation calls. Statistical differences were considered significant for $P<0.05$ ($^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$). When random numbers were needed those were generated using the R-function 'runif'.

Software

For analyses and preparing figures, we used Matlab (Version 2016b & 2015b), R (Version 3.5.1), GraphPad Version 5 and AvisoftSASLab Pro Version 5.2.10. Customized Matlab programs written by Philipp Norton (PN) and Lara Sophie Burchardt (LSB) were adjusted and used for the rhythm optimization (PN), model calculations (PN & LSB), FFT analysis (LSB) and cluster visualization (PN).

Results

Isochronous rhythm

For each vocalization, we found an isochronous rhythm (rhythm^S) that coincided best with the onsets of elements (supplementary audio files A1-3). A rhythm^S between 6 – 20 Hz dominated across individuals as well as across vocalization types: 49.4% of isolation calls (247 out of 500 calls), 41% of territorial songs (59 out of 143 songs) and 57% of echolocation call sequences (19 out of 33 sequences) had a best fitting rhythm of 6-20 Hz (**Figure 3**). Corresponding results were obtained when focusing on individuals instead of vocalization types. 20 out of 25 pups produced isolation calls which clustered predominantly in the frequency range of 10-20 Hz; the largest clusters contained 25-70% of calls per pup (**Figure 4A**). 9 out of 14 males produced territorial songs which clustered predominantly in the frequency range of 10-20 Hz; the largest clusters contained 30-60% of songs (**Figure 4B**). We considered clusters with their mean falling into the range between 10-20 Hz and the cluster comprising at least 25% of data (clusters are marked in red in **Figure 4A-C**). Echolocation call sequences clustered predominantly in the frequency range of 6-20 Hz, 39% of sequences making up the strongest cluster (between 6 and 10 Hz), adding up to 57% between 6 and 20 Hz. Note that echolocation call sequences were pooled over all individuals (**Figure 4C**, see Methods).

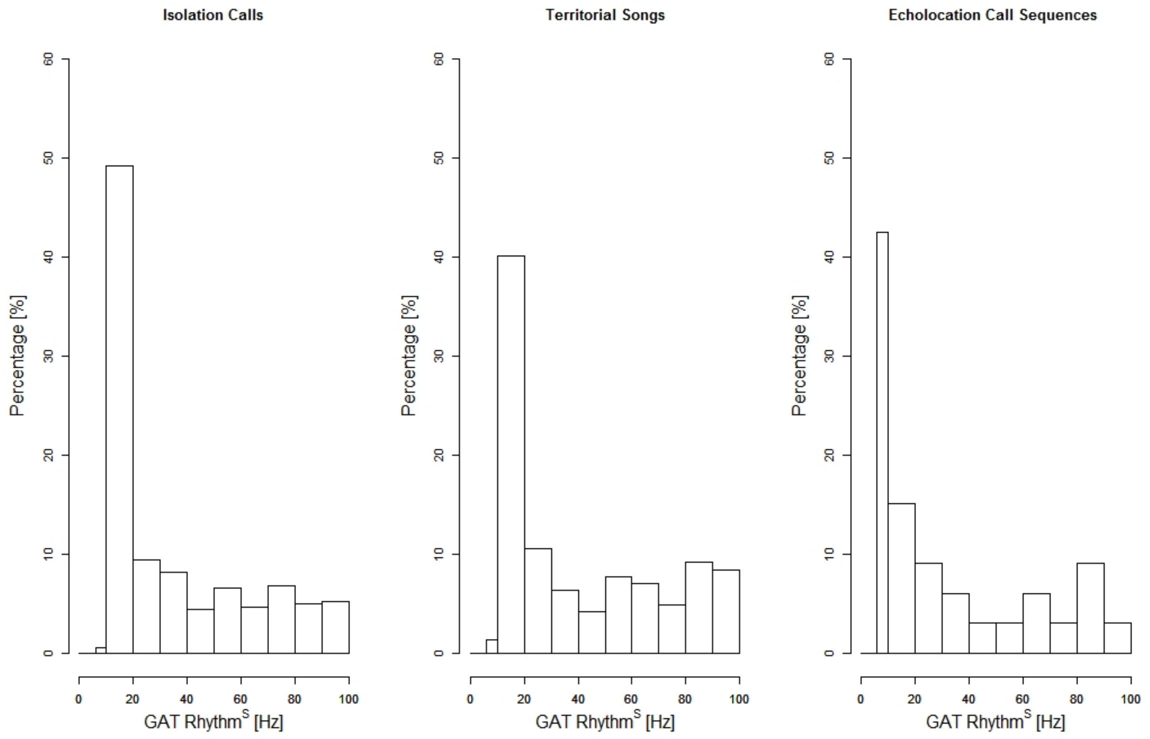


Figure 3 – GAT Analysis: Regular rhythms in S. *bilineata* vocalizations. The relative majority of calls/songs occurred in rhythm frequencies below 20 Hz for all vocalization types.

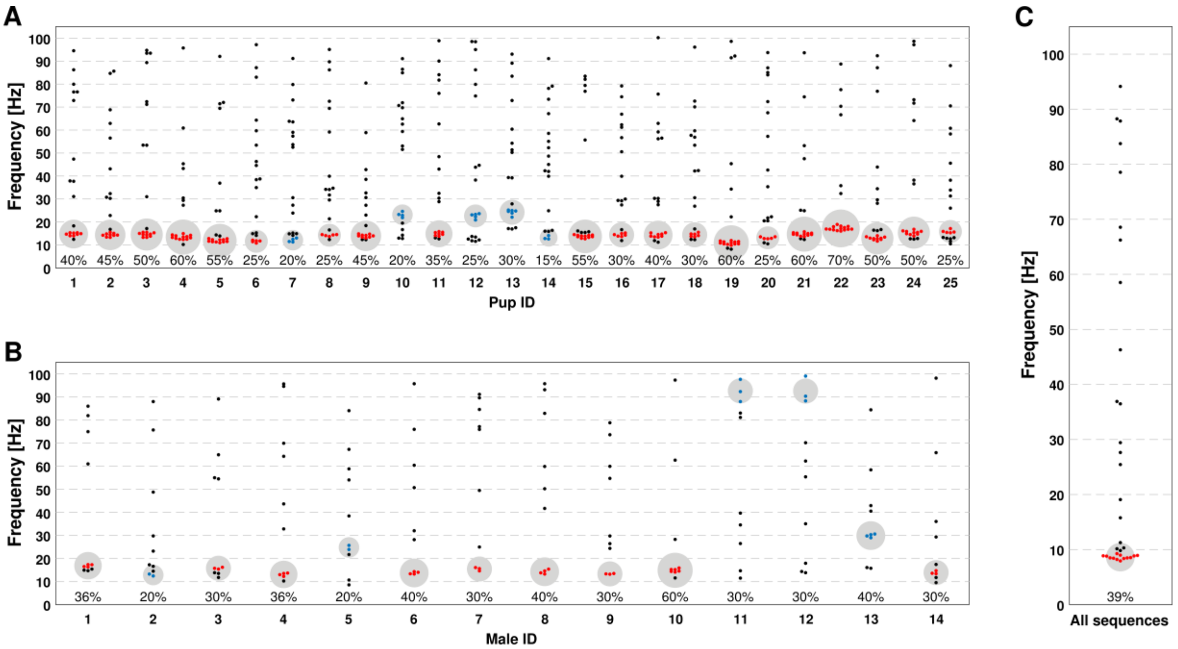


Figure 4 – Isochronous beat in bat vocalization: (A) Rhythm clusters in isolation calls of S. *bilineata* pups. (B) Rhythm clusters in territorial songs of S. *bilineata* males. (C) Rhythm clusters in echolocation call sequences of S. *bilineata* adults. Marked in red are the data belonging to the largest cluster containing at least 25% of songs/calls, within the range of 6 to 20 Hz. Marked in blue are the data belonging to the largest cluster that were not considered. The percentage of data in the largest cluster is shown in the bottom of each column. The area of circles is scaled to the percentage of calls/songs in the respective clusters.

Comparison to artificially randomized vocalizations

To confirm that the observed element onsets in S. *bilineata* vocalizations aligned to an isochronous rhythm well and more closely than expected by chance, we compared the FRMSD values of artificial vocalization types to the FRMSD values of the original vocalization types. All artificial vocalization types had randomized element and gap durations; sequence information, i.e. the consecutive order of elements was either preserved (Model 1) or ignored (Model 2).

As expected, original vocalizations had significantly lower FRMSD values than artificial model 1 or model 2 vocalizations (Repeated Measures ANOVA: isolation calls: F=71.17, df= 74, p<0.0001; territorial songs: F=30.38, df= 41, p<0.0001; unpaired t-test (with Welch correction): echolocation call sequences: t=2.35,df=33, p= 0.0023), indicating that the element onsets of original vocalizations matched an isochronous rhythm more closely than expected by chance (**Figure 5**).

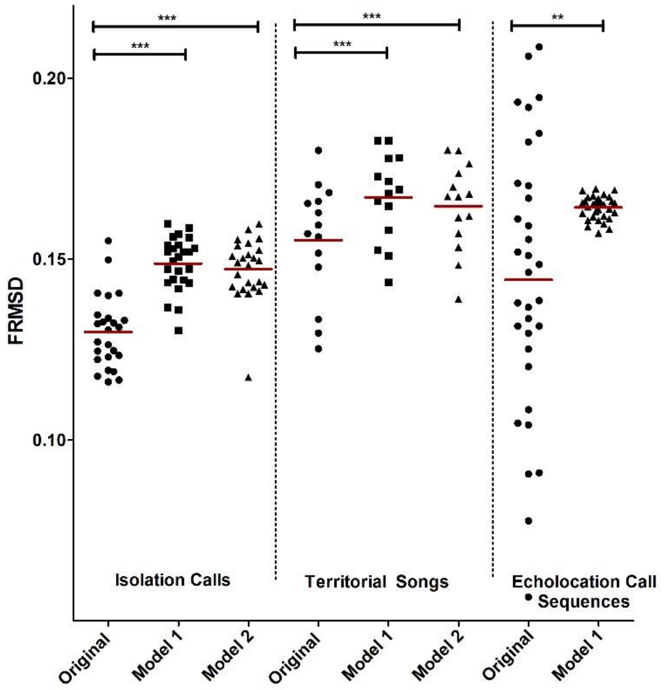


Figure 5 – Model validation: Mean values for FRMSD, comparing original data to 'bat-like' artificial data (Model 1: intact sequence information, Model 2: random sequences). Original data showed significantly lower deviations (*P<0.05; **P<0.01; ***P<0.001). Depicted are means per individual for isolation calls and territorial song and best fitting rhythms of single sequences for echolocation call sequences, explaining the higher spread. Red lines indicate the respective mean of a dataset.

Fourier analysis

Results of the fast Fourier analysis of a binary point process string where element onsets were represented by '1' resulted in the same if not stronger picture at the level of vocalization types. 55.4% of isolation calls, 47.8% of territorial songs and 66% of echolocation call sequences showed a dominant rhythm between 6 and 20 Hz (54% between 6-10Hz) (**Supplementary Figure 1**).

Ontogeny effect

Furthermore, we ran statistical analyses to investigate the effect of ontogeny on rhythmicity for pup isolation calls. For each pup, we compared the frequencies of isochronous rhythms of the first and last two isolation calls recorded during ontogeny (non-volant phase and volant phase). Rhythm^{S} frequencies in isolation calls did not change significantly during the pups' ontogeny (paired t-test: t=1.31, df= 49 p=0.20, **Supplementary Figure 2**).

Discussion

The novel aspect presented in this study is the documentation of isochronous rhythm patterns in different vocalization types of the bat S. *bilineata*. With a generate-and-test approach (GAT) as well as an FFT analysis, vocalizations were analyzed to find a best fitting rhythm over a wide frequency range of 5-100 Hz (i.e. pulses per second). Even though the three analyzed vocalization types (pup isolation calls, male territorial songs, and echolocation call sequences) differed in their acoustic structure and the behavioral situation they were produced in, their best-fitting rhythms fell in a quite narrow frequency window. Element onsets coincided best with rhythm frequencies between 6-20 Hz, independent of vocalization type and vocalizing individual. Analyses showed that rhythm frequencies were most abundant between 6-10 Hz for echolocation call sequences and between 10-20 Hz for territorial songs and isolation calls. The same picture was found with an FFT analysis at the level of vocalization types.

Therefore, the best fitting rhythms were comparatively similar across vocalization types and vocalizing individuals in S. *bilineata*, with social communication signals showing rhythms twofold of echolocation call sequences. Other studies on rhythmicality in animal vocalizations so far did show patterns that differed between individual animals (5), and temporal structure, namely the rhythm, may be used by conspecifics for individual discrimination (10). A biological constraint shaping rhythms to be more alike between individuals is not apparent. Since there are not many comparable studies yet our results might prove to be the rule rather than an exception.

Nevertheless, the pattern of rhythm^{S} in the analyzed vocalizations, could be caused by physiological constraints and/or mechanisms to save energy. The production of echolocation calls when a bat is searching for prey items but has not detected anything yet is correlated with respiration which, in turn, is tightly coupled to wing beat. For many bat species, a 1:1 relation has been found (e.g. (22)). The soprano pipistrelle (*Pipistrellus pygmaeus*), for example, produces one or two echolocation calls per wingbeat and respiratory cycle (24). In other pipistrelle bats (*P. pipistrellus*, *P. kuhlii*, *P. nathusii* (23)), greater horseshoe bats (*Rhinolophus ferrumequinum*), little brown bats (*Myotis lucifugus*), Parnell's mustached bats (*Pteronotus parnellii rubiginosus*) and Seba's short-tailed bats (*Carollia perspicillata* (21)) wingbeat and echolocation calls are also coupled. Coupling was also found in the tongue-clicking Pteropodid bat *Rousettus aegyptiacus*, indicating that a strong coupling of wing beat, respiration and sonar emission is widespread in bats regardless of sound production mechanism.

In S. *bilineata*, respiratory cycle and wing beat are between 6 and 12 Hz during search flight (pers. communication H.-U. Schnitzler). Our results suggest that in the release situation the echolocation call sequences were recorded, bats mainly uttered one call per wingbeat, which fits the low sensory needs in the relatively open space in which releases took place. In a situation with higher sensory needs, expected rhythm frequencies should be doubled, i.e. lie between 12 and 24 Hz, most of which overlaps strongly with the rhythm frequencies found in the social vocalizations. Therefore, we argue that the rhythm frequencies most abundant in social vocalizations (10-20 Hz) and in echolocation call sequences during search flight (6-10 Hz and, to a lesser degree, 10-20 Hz) can be regarded as comparatively similar.

During prey capture, however, echolocation call sequences contain not only search flight calls but also approach flight calls (when prey has been detected and is approached) and a so-called final buzz (immediately before prey capture, very short and broadband echolocation calls with extremely short IOIs are produced), which enhances the sensory information available for the foraging bat. Even though wing beat, respiratory cycle and sonar emission are tightly coupled during search flights (likely to increase energy efficiency), this might not provide sufficient sensory information during prey capture, i.e. in a situation where high temporal resolution is needed (per wing beat and respiratory cycle up to 10-15 pulses can be emitted (23). A larger ratio between wing beat, respiratory frequency, and emitted echolocation calls could result in a weaker rhythmic pattern in our analyses. In the approach phase, the number of echolocation calls per wing beat can vary widely, depending on the current sensory needs of a foraging bat. Therefore, it seems reasonable to assume that echolocation call sequences during prey capture do not follow any clear rhythm but strongly depend on the bats' current sensory needs. This could easily be tested on echolocation call sequences recorded in foraging situations. Correspondingly, a previous study on the big brown bat *Eptesicus fuscus* showed that the strict 1:1 synchronization of wing beat, respiration, and call emission was not found during complex navigation tasks, where freely behaving individuals had to search for prey (tethered mealworms, suspended at about 1.5 m height) in a flight room, equipped with various obstacles, such as artificial houseplants (32). During search flights, however, metabolic needs, e.g. being energy efficient, may play a more important role (33). To investigate the task/situation dependence of the coupling of wing beat, respiration, and call emission it would be worthwhile to analyze rhythm^{S} of echolocation call sequences produced in a feeding context in bat species in which a strict 1:1 coupling has been found during search flight.

The determination of rhythm^S (method developed by (5)) could be a valuable addition to currently used methods since it is not dependent on a laboratory setting. Knowledge of wing beat, and/or respiratory rates could be combined with analyses of rhythm^S of echolocation call sequences and social vocalizations recorded from freely behaving, wild bats to gain insights on coupling relations in natural situations. Especially for more complex vocalization types with variable element durations, the GAT approach and FFT analysis provide an advantage over simply assessing IOIs. The latter method ignores the sequential structure of vocalizations and their variable element durations, potentially concealing higher order regularity.

To assess the goodness-of-fit for our analyses of rhythm^S, we compared deviations from rhythm^S of original and artificially created vocalizations that were randomly drawn from a pool of typical element and gap durations. Original vocalizations deviated significantly less from rhythm^S than did artificial vocalizations (i.e. element onsets of original vocalizations coincided significantly better with an isochronous rhythm than artificial vocalizations), indicating that the rhythm^S found in S. *bilineata* vocalizations was not an artifact of the typical duration and sequence of this species' vocalizations.

One aspect worthy of discussion is the relation between rhythm frequencies of echolocation call sequences produced by S. *bilineata* during search flight (which were coupled to wing beat frequencies) and social vocalizations produced by individuals hanging in their day-roost (pup isolation calls and male territorial songs). We doubt that rhythm frequencies of isolation calls and territorial songs are caused by a coupling of sound emission to respiration since echolocation calls produced by roosting bats can occur at any point in the respiratory cycle (22). Taken this into account, it seems reasonable to assume that social calls can be emitted at any point in the respiratory cycle as well. Nevertheless, as stated before, we argue there is a relation between the dominant frequencies of the three vocalization types, and we regard them as being comparatively similar. The similarity of rhythm frequencies could suggest a common evolutionary background, which might be the coupling between respiration, wingbeat and echolocation call emission. However, increasing evidence suggests that flight preceded echolocation (34, 35), which would indicate that vocal communication preceded echolocation as well (assuming that bats' predecessors communicated with social calls, as many small mammals do). It is therefore possible that social calls, despite being probably phylogenetically older than echolocation, adopted the rhythm frequencies of echolocation calls at some point.

It is interesting to compare the strength of rhythms between isolation calls and territorial songs since isolation calls are produced within minutes after birth (14) while territorial songs are learned during ontogeny (12). Generally, a higher variability in rhythm^{S} may be expected when comparing learned vocalizations to innate vocalizations. In our study, rhythm frequencies predominantly clustered between 6-20 Hz, but cluster strength of individuals was on average lower in territorial song than in isolation calls (37% in territorial song compared to 44.7% in isolation calls; GAT approach). This difference in individual cluster strength resembled the overall difference between both vocalization types, since only 41.6% of all territorial songs had rhythm frequencies between 6-20 Hz, while 49.8% of isolation calls did.

Rhythmic properties of echolocation could represent the same neuronal correlates underlying production of social vocalizations. In the Egyptian fruit bat (*R. aegyptiacus*) wingbeat and tongue clicks are tightly coupled around 10 Hz (25), as we found for *S. bilineata*. These rhythm frequencies show a resemblance to the frequency of theta brain waves. Thought to be important for movements, spatial memory and active stimulus intake (26) amongst others, theta waves might be a promising neural correlate explaining the production of the detected rhythms.

It might be advantageous to produce rhythmic vocalizations because 'rhythmic attention' (*sensu* (36)) helps receivers to decode rhythmic signals easier and faster (37). The attention of receivers cycles in an oscillatory way when a rhythm exists (e.g. (38, 39)). Since rhythmic signals are predictable, 'rhythmic attention' enables receivers to provide most 'attentional energy' at a time point where the next stimulus is to be expected. This is advantageous because cognitive capacities are limited (40) and an optimization of attention timing is helpful to not miss relevant stimuli. For example, when humans were asked to assess the difference in pitch of two focal tones separated by regularly timed tones, the assessment of pitch difference was better when the second focal tone followed the regular timing of the separating tones than when was slightly displaced from the regular timing (41). Another example from macaques shows that neuronal oscillations in the primary visual cortex entrain to a stimuli stream (visual stimuli) when the stream is rhythmic, a mechanism resulting in decreased reaction time and an increase in the response gain for events that are task relevant (42). Bats' attention as well as the auditory system collectively could be tuned to echolocation rhythms, because bats are exposed to those rhythms for large parts of their lives (43). Therefore, it might be advantageous to produce vocalizations in the same frequency window to increase detection by receivers. At the moment, we do not know whether rhythmic attention plays a role in S. *bilineata*. Playback experiments violating expected rhythmic patterns in social vocalizations would be a valuable avenue for future research. A switch from a rhythm determined by physiological constraints to a rhythm decoupled from its original production constraints but still with an adaptive function (e.g. rhythmic attention) might have been one step during evolution that paved the way to develop music as we know it.

In summary, this study demonstrates an isochronous rhythm in three bat vocalization types in which metabolic constraints leading to rhythmic patterns are more (echolocation calls) or less (isolation calls, territorial songs) likely. The two methods used in this study (GAT and FFT) enable the analysis of best fitting rhythms in a corresponding way. Future studies should profit by complementary use of both methods. To further study the coupling or decoupling of wing beat, respiration and sound emission in animals as well as its biological relevance, it would be highly beneficial to compare different species of bats and birds which sing in flight as well as other echolocating mammals. Such a comparative approach could provide valuable insights into the origin and relevance of rhythmicality in animals.

Ethics

All experiments and protocols for capturing and handling bats comply with the current laws of Costa Rica. Permit numbers are given in the original publications from which the data were drawn (13, 19, 29).

Data accessibility

The dataset supporting this article has been uploaded as part of the electronic supplementary material. The code (GAT approach) with detailed explanations was already published in a previous publication (30).

Authors’ contributions

OB and MK collected the data. PN and CS developed the computational framework (GAT approach). LSB analyzed the data and wrote the manuscript. CS and MK supervised the project. All authors discussed the results and contributed to the final manuscript. All authors gave final approval for publication.

Competing interests

We have no competing interests to declare.

Funding

This work was supported by a stipend from the Elsa-Neumann Foundation to LSB and a Heisenberg Fellowship (DFG KN935 3-1) from the German Research Foundation to MK.

Acknowledgements

We would like to thank Ahana Aurora Fernandez and Martina Nagy, as well as Hans-Ulrich Schnitzler for helpful discussions.

References

1. Honing H, ten Cate C, Peretz I, Trehub SE. Without it no music: cognition, biology and evolution of musicality. Philosophical Transactions of the Royal Society B: Biological Sciences. 2015;370(1664):20140088. http://dx.doi.org/10.1098/rstb.2014.0088

2. Wallin NL. Biomusicology - Neurophysiological, Neuropsychological, and Evolutionary Perspectives on the Origins and Purpose of Music. Stuyvesant, NY: Pendragon Press; 1991.

3. Ravignani A, Thompson B, Filippi P. The evolution of musicality: what can be learned from language evolution research? Frontiers in Neuroscience. 2018;12(20). http://dx.doi.org/10.3389/fnins.2018.00020

4. Patel AD. Music, Language and the Brain. New York, NY: Oxford University Press; 2008.

5. Norton P, Scharff C. "Bird song metronomics": isochronous organization of zebra finch song rhythm. Frontiers in Neuroscience. 2016;10(309). http://dx.doi.org/10.3389/fnins.2016.00309

6. Heinsohn R, Zdenek CN, Cunningham RB, Endler JA, Langmore NE. Tool-assisted rhythmic drumming in palm cockatoos shares key elements of human instrumental music. Science Advances. 2017;3(6):e1602399. http://dx.doi.org/10.1126/sciadv.1602399

7. Merguerditchian A, Vuillemin A, Pruetz JD. Identifying the ape beat in the wild: rhythmic individual signatures from sounds of manual fruit cracking in Fongoli Chimpanzees. In: Proceedings of the 12th International Conference on the Evolution of Language (Evolang12). Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika; 2018. http://dx.doi.org/10.12775/3991-1.072

8. Celma-Miralles A, Toro JM. Beat perception in a non-vocal learner: rats can identify isochronous beats. In: Proceedings of the 12th International Conference on the Evolution of Language (Evolang12). Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika; 2018. http://dx.doi.org/10.12775/3991-1.015

9. Hulse SH, Humpal J, Cynx J. Discrimination and generalization of rhythmic and arrhythmic sound patterns by European starlings (Sturnus vulgaris). Music Perception: An Interdisciplinary Journal. 1984;1(4):442-464. http://dx.doi.org/10.2307/40285272

10. Mathevon N, Casey C, Reichmuth C, Charrier I. Northern elephant seals memorize the rhythm and timbre of their rivals voices. Current Biology. 2017;27(15):2352-2356.e2. http://dx.doi.org/10.1016/j.cub.2017.06.035

11. Behr O, Helversen O. Bat serenades - complex courtship songs of the sac-winged bat (Saccopteryx bilineata). Behavioural Ecology and Sociobiology. 2004;56(2):106-115. http://dx.doi.org/10.1007/s00265-004-0768-7

12. Knörnschild M, Nagy M, Metz M, Mayer F, von Helversen O. Complex vocal imitation during ontogeny in a bat. Biology Letters. 2009;6(2):156-159. http://dx.doi.org/10.1098/rsbl.2009.0685

13. Knörnschild M, Nagy M, Metz M, Mayer F, von Helversen O. Learned vocal group signatures in the polygynous bat Saccopteryx bilineata. Animal Behaviour. 2012;84(4):761-769. http://dx.doi.org/10.1016/j.anbehav.2012.06.029

14. Knörnschild M, von Helversen O. Non-mutual vocal mother-pup recognition in the sac-winged bat, Saccopteryx bilineata. Animal Behaviour. 2008;76(3):1001-1009. http://dx.doi.org/10.1016/j.anbehav.2008.05.018

15. Fernandez AA, Knörnschild M. Isolation calls of the bat Saccopteryx bilineata encode multiple messages. Animal Behavior and Cognition. 2017;4(2):169-186. http://dx.doi.org/10.12966/abc.04.05.2017

16. Knörnschild M, Blüml S, Steidl P, Eckenweber M, Nagy M. Bat songs as acoustic beacons - male territorial songs attract dispersing females. Scientific Reports. 2017;7(1):13918. http://dx.doi.org/10.1038/s41598-017-14434-5

17. Eckenweber M, Knörnschild M. Social influences on territorial signaling in male greater sac-winged bats. Behavioural Ecology and Sociobiology. 2013;67(4):639-648. http://dx.doi.org/10.1007/s00265-013-1483-z

18. Jung K, Kalko EKV, von Helversen O. Echolocation calls in Central American emballonurid bats: signal design and call frequency alternation. Journal of Zoology. 2007;272(2):125-137. http://dx.doi.org/10.1111/J.1469-7998.2006.00250.X

19. Knörnschild M, Jung K, Nagy M, Metz M, Kalko EKV. Bat echolocation calls facilitate social communication. Proceedings of the Royal Society B: Biological Sciences. 2012;279(1748):4827-4835. http://dx.doi.org/10.1098/rspb.2012.1995

20. Bermeitinger C, Frings C. Rhythm and attention: does the beat position of a visual or auditory regular pulse modulate T2 detection in the attentional blink? Frontiers in Psychology. 2015;6(1847). http://dx.doi.org/10.3389/fpsyg.2015.01847

21. Schnitzler H-U. Fledermäuse im Windkanal. Zeitschrift für vergleichende Physiologie. 1971;73(2):209-221. http://dx.doi.org/10.1007/BF00304133

22. Suthers RA, Thomas SP, Suthers BJ. Respiration, wing-beat and ultrasonic pulse emission in an echo-locating bat. The Journal of Experimental Biology. 1972;56:37-48.

23. Kalko EKV. Coupling of sound emission and wingbeat in naturally foraging european pipistrelle bats (Microchiroptera: Vespertilionidae). Folia Zoologica. 1994;43(4):363-376.

24. Wong JG, Waters DA. The synchronisation of signal emission with wingbeat during the approach phase in Soprano Pipistrelles (Pipistrellus Pygmaeus). The Journal of Experimental Biology. 2001;204:575-583.

25. Yartsev MM, Ulanovsky N. Representation of three-dimensional space in the hippocampus of flying bats. Science. 2013;340(6130):367-372. http://dx.doi.org/10.1126/science.1235338

26. Colgin LL. Mechanisms and functions of theta rhythms. Annual Review of Neuroscience. 2013;36:295-312. https://doi.org/10.1146/annurev-neuro-062012-170330

27. Bayefsky-Anand S. Echolocation calls of the greater sac-winged Bat (Saccopteryx bilineata) in different amounts of clutter. Bat Research News. 2006;47(1):7-10.

28. Ratcliffe JM, Jakobsen L, Kalko EKV, Surlykke A. Frequency alternation and an offbeat rhythm indicate foraging behavior in the echolocating bat, Saccopteryx bilineata. Journal of Comparative Physiology A. 2011;197(5):413-423. http://dx.doi.org/10.1007/s00359-011-0630-0

29. Behr O, Helversen O, Heckel G, Nagy M, Voigt CC, Mayer F. Territorial songs indicate male quality in the sac-winged bat Saccopteryx bilineata (Chiroptera, Emballonuridae). Behavioural Ecology. 2006;17(5):810-817. https://doi.org/10.1093/beheco/arl013

30. Ravignani A, Norton P. Measuring rhyhtmic complexity: a primer to quantify and compare temporal structure in speech, movement, and animal vocalizations. Journal of Language Evolution. 2017;2(1):4-19. http://dx.doi.org/10.1093/jole/lzx002

31. Saar S, Mitra PP. A technique for characterizing the development of rhythms in bird song. PLoS ONE. 2008;3(1):e1461. http://dx.doi.org/10.1371/journal.pone.0001461

32. Moss CF, Bohn K, Gilkenson H, Surlykke A. Active listening for spatial orientation in a complex auditory scene. PLoS Biology. 2006;4(4):e79). http://dx.doi.org/10.1371/journal.pbio.0040079

33. Suthers RA, Fattu JM. Mechanisms of sound production by echolocating bats. American Zoologist. 1973;13(4):1215-1226.

34. Simmons N, Geisler JH. Phylogenetic relationships of Icaronycteris, Archaeonycteris, Hassianycteris, and Palaeochiropteryx to extant bat lineages, with comments on the evolution of echolocation and foraging strategies in Microchiroptera. Bulletin of the American Museum of Natural History. 1998;235:4-169.

35. Speakman JR. The evolution of flight and echolocation in bats: another leap in the dark. Mammal Review. 2001; 31(2):111-30. https://doi.org/10.1046/j.1365-2907.2001.00082.x

36. Jones MR, Kidd G, Wetzel R. Evidence for rhythmic attention. Journal of Experimental Psychology: Human Perception and Performance. 1981;7(5):1059-1073. http://dx.doi.org/10.1037/0096-1523.7.5.1059

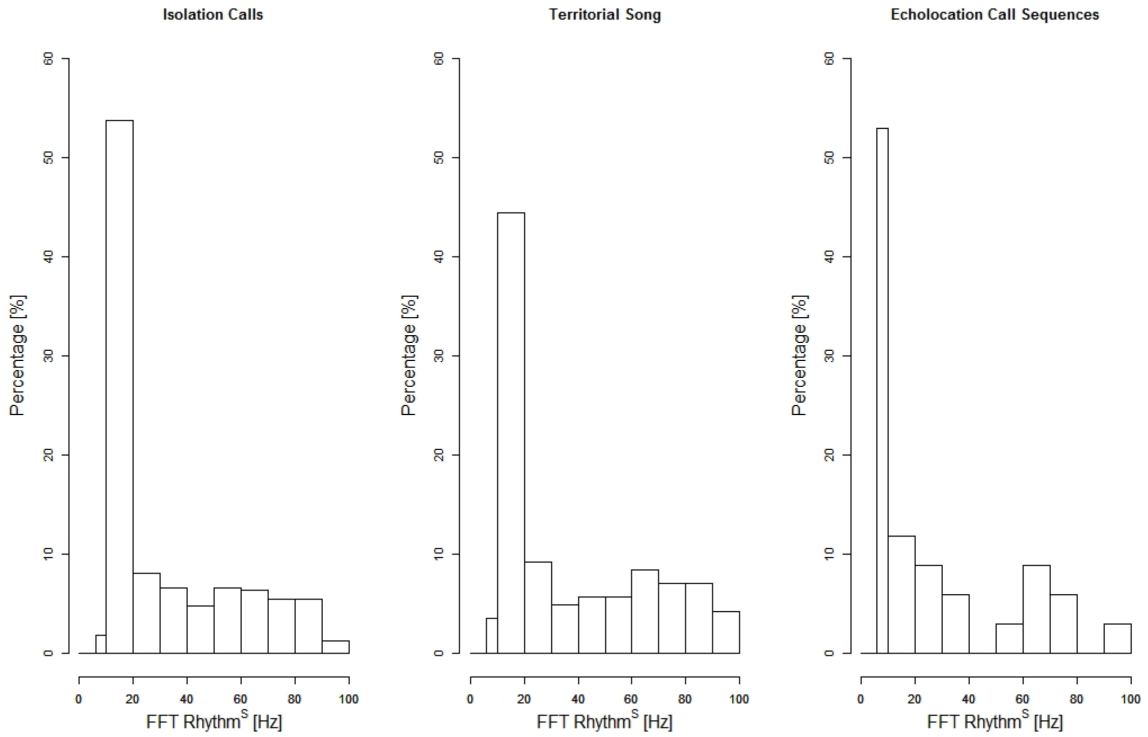
Supplementary Material

Validation analysis - Tempo changes

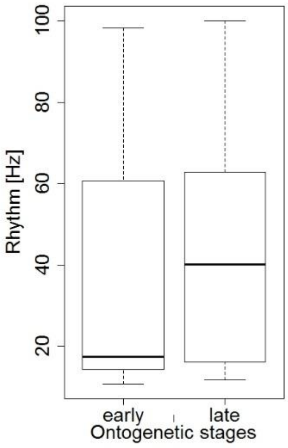
To assess tempo changes within vocalizations, we calculated linear regressions for the Inter-Onset-Intervals (IOI) sequences to test whether these were significantly different from zero (using an F-test), which would indicate a significant change in tempo. We conducted this analysis for a random subset of each vocalization type (two sequences per individual for isolation calls and territorial songs, all data for echolocation call sequences). To corroborate results from the tempo analysis, individual syllable deviations of the first, middle and last syllable were compared per vocalization type by means of a Friedman test; this was done to test whether deviations changed throughout a syllable sequence. This analysis was conducted on a subset of the data, chosen in the same way as for the tempo analysis.

The majority of isolation calls (74%) had a stable tempo, 22% of calls showed a decrease in tempo and 4% of calls an increase. On the contrary, the majority of territorial songs (79%) decreased in tempo, especially in the last fifth of songs (**Supplementary Figure 3**). However, inter-onset intervals did not increase continuously but rather abruptly, often doubling and quadrupling. These multiples of inter-onset intervals make it unlikely that the observed change in tempo had a negative effect on rhythm^{S} in our study. Furthermore results were confirmed by the FFT analysis, which is stable against tempo changes.

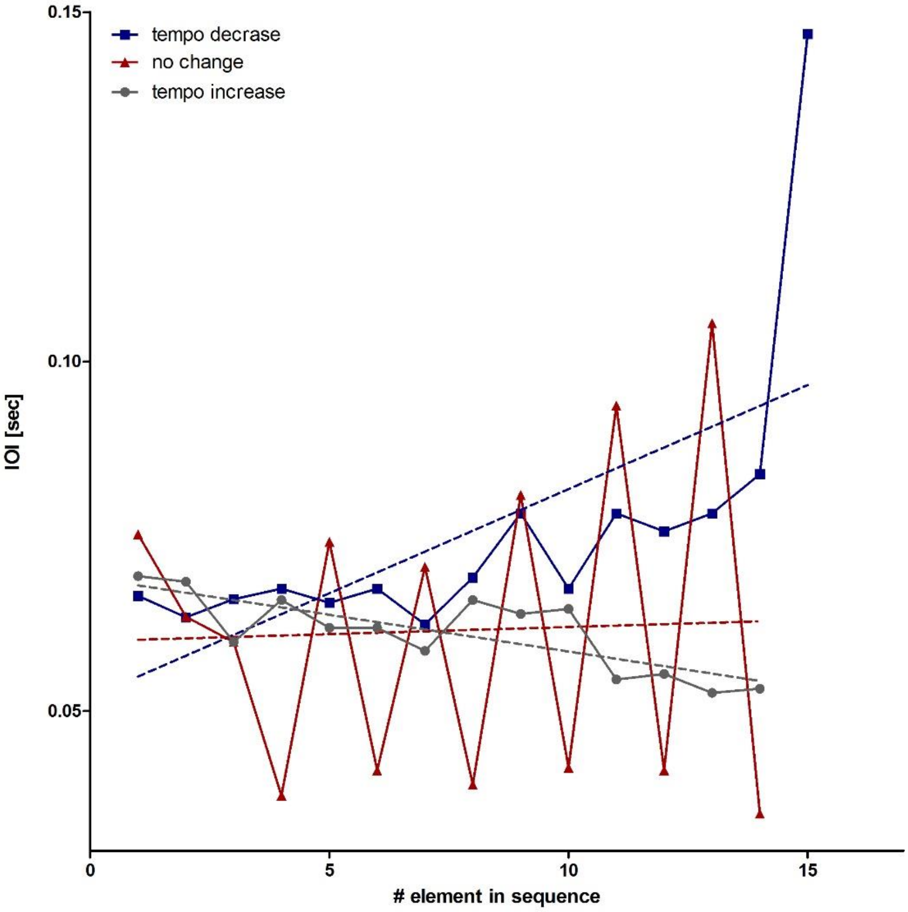
To corroborate that changes in tempo did not affect rhythm^{S}, we calculated individual element deviations to the nearest single pulse. Element deviations did not change throughout vocalizations, since a best fitting rhythm was found by an optimization task regarding all elements of a sequence. Nevertheless, individual element deviations of vocalizations with tempo changes (territorial songs) did not differ from vocalizations without tempo changes (isolation calls) (Kruskal-Wallis, p=0.78, F=2.47, df=6, **Supplementary Figure 4**), suggesting that changes in tempo played a negligible role in our study. Another argument for this interpretation is the results from FFT analysis. Since the same pattern was found with a method in which tempo changes cannot affect the outcome, it is reasonable to say that tempo changes did not influence the results from GAT analysis in a crucial way.



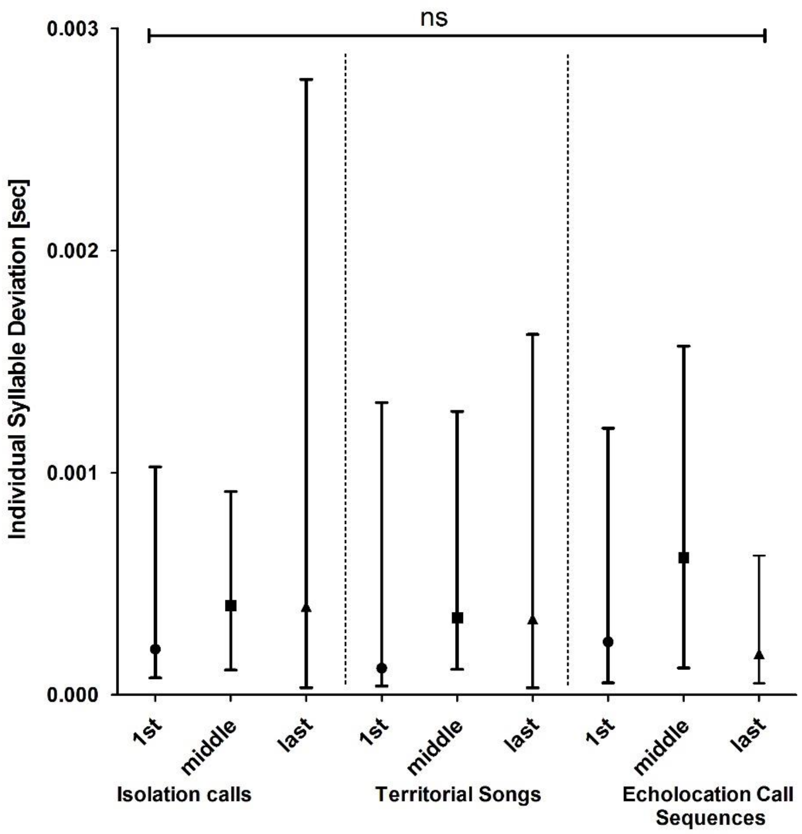
Supplementary Figure 1 – FFT analysis: Regular rhythm^S in bat vocalizations. The relative majority of calls/songs occur in rhythm^S frequencies below 20 Hz for all vocalization types.



Supplementary Figure 2 – Effect of ontogenetic stage on rhythm^S in pup isolation calls. Early ontogeny did not differ from late ontogeny. Medians, interquartile range (25-75%) and whiskers (0-100%) are shown.



Supplementary Figure 3 – Tempo changes in sequences: Three IOI sequences are shown as solid lines; dashed lines show corresponding linear regressions. Slopes of regression lines were tested against zero. Significant difference from zero was interpreted as tempo change. In red (triangle) an isolation call with no tempo change is shown, in grey (circle) an isolation call increasing in tempo and in blue (square) a territorial song decreasing in tempo rather abruptly are shown.



Supplementary Figure 4 – Syllable deviations of individual syllables. Individual deviations from rhythm^{S} of first, middle and last syllable of calls/songs were compared. Median and interquartile range are shown. No significant differences were found.

Publication D: Differential song deficits after knockdown of FoxP1/2/4

Mendoza, E., **Norton, P.**, Barschke, P., and Scharff, C. (in preparation). Effects on song learning differ after lentivirally mediated knockdown of FoxP1, FoxP2 or FoxP4 in Area X of zebra finches. *Journal of Neuroscience*.

Effects on song learning differ after lentivirally mediated knockdown of FoxP1, FoxP2 or FoxP4 in Area X of zebra finches

Ezequiel Mendoza, Philipp Norton,
Peggy Barschke, Constance Scharff

Department of Animal Behaviour, Freie Universität Berlin

Abstract

Mutations in the transcription factors FOXP1[1] and FOXP2 are associated with speech impairments. FOXP1 is additionally linked to cognitive deficits, as is FOXP4. These FoxP proteins are highly conserved in vertebrates and expressed in comparable brain regions including the striatum. In male zebra finches, experimental manipulation of FoxP2 in Area X, a striatal song nucleus essential for vocal production learning, affects song development, adult song production, spine density and dopamine-regulated synaptic transmission of striatal neurons. We previously showed that in the majority of Area X neurons FoxP1, FoxP2, and FoxP4 are co-expressed, can di- and multimerize with each other and differentially regulate the expression of target genes. These findings raise the possibility that FoxP1, FoxP2, and FoxP4 (FoxP1/2/4) affect neural function differently and in turn vocal learning. To address this directly, we downregulated FoxP1 or FoxP4 in Area X of juvenile zebra finches and compared the resulting song phenotypes to the previously described inaccurate and incomplete song learning after FoxP2 knockdown. We found that experimental downregulation of FoxP1 and FoxP4 led to impaired song learning with partly similar features as those reported for FoxP2 knockdowns. However, there were also specific differences between the groups leading us to suggest that specific features of the song are impacted differentially by developmental manipulations of FoxP1/2/4 expression in Area X.

[1] FOXP denotes human protein, Foxp rodent and FoxP all other species (Kaestner et al., 2000)

Significance Statement

We compared the effects of reduced amounts of the transcription factors FoxP1, FoxP2 and FoxP4 in a striatal song nucleus, Area X, on vocal production learning in juvenile male zebra finches. We show for the first time that these temporally and spatially precise manipulations of the three FoxPs affect spectral and temporal song features differentially. This is important because it raises the possibility that the different FoxPs control different aspects of vocal learning through combinatorial gene expression or by acting in different microcircuits within Area X. These results are consistent with the deleterious effects of human FOXP1 and FOXP2 mutations on speech and language and add FOXP4 as a possible candidate gene for vocal disorders.

Introduction

Heterozygous mutations of the FOXP2 transcription factor are associated with a speech deficit called developmental verbal dyspraxia (DVD) (Lai et al., 2001) or Childhood Apraxia of Speech (Morgan and Webster, 2018). FOXP1 mutations cause a wider spectrum of impairments including speech problems (Fisher and Scharff, 2009; Bacon and Rappold, 2012; Siper et al., 2017; Sollis et al., 2017). A FOXP4 mutation is associated with delayed development, laryngeal hypoplasia and feeding problems (Charng et al., 2016). FOXP1/2/4 are expressed in diverse brain regions, including the striatum (Bowers and Konopka, 2012). The striatum in patients carrying FOXP2 mutations differs structurally and functionally from that of their unaffected siblings (Watkins et al., 2002; Liegeois et al., 2003). FoxP1/2/4 are expressed in the striatum in mice and other vertebrates (Shu et al., 2001; Ferland et al., 2003; Takahashi et al., 2003; Haesler et al., 2004; Teramitsu et al., 2004; Bonkowsky and Chien, 2005; Takahashi et al., 2008a; Takahashi et al., 2008b; Takahashi et al., 2009; Mashiko et al., 2012; Mendoza et al., 2015; Spaeth et al., 2015). In striatal neurons of mice carrying a mutant allele of Foxp2, similar to one reported in patients, synaptic plasticity is impaired and ultrasonic vocal communication is altered (Groszer et al., 2008; Castellucci et al., 2016; Chabout et al., 2016).. While the latter may also be due to the crucial functions of Foxp2 in the development of craniofacial cartilage (Xu et al., 2018), striatal-specific deletion of Foxp2 (French et al., 2018) causes mice to execute rapid motor sequences more variably, emphasizing the importance of the striatum for fine control of motor behaviors. Together these findings implicate the striatum as an important site of integrated FoxP1/2/4 neural function.

We study FoxP function in songbirds because birdsong and speech share many features (Doupe and Kuhl, 1999). Both are learned during critical developmental periods through auditory-guided vocal imitation. Speech learning in people and song learning in birds are constrained by innate predispositions and are also strongly affected by social factors. Birdsong and speech depend on analogous neural pathways that are functionally lateralized (Petkov and Jarvis, 2012; Pfenning et al., 2014). Thus songbirds provide a genuine model for behavioral, neural and molecular analyses of genes relevant for vocal communication (Bolhuis et al., 2010).

In zebra finches, FoxP2 expression levels in Area X, the striatal song nucleus required for learning, discrimination and maintenance of song (Sohrabji et al., 1990; Scharff and Nottebohm, 1991; Scharff et al., 1998; Aronov et al., 2008) vary with age and singing activity (Haesler et al., 2004; Miller et al., 2008; Teramitsu et al., 2010; Thompson et al., 2013; Adam et al., 2016). Experiments disrupting the dynamic regulation of FoxP2 impair song learning, social modulation of song variability and dopamine-sensitive signal transmission through the cortical-basal ganglia-thalamic forebrain song circuit (Haesler et al., 2007; Murugan et al., 2013; Heston and White, 2015). In many Area X medium spiny neurons (MSNs) FoxP2 can dimerize and oligomerize with FoxP1 and FoxP4 (Mendoza et al., 2015; Mendoza and Scharff, 2017). In cell culture, FoxP proteins of mice (Li et al., 2004a) and humans (Estruch et al., 2018) also dimerize. Dimerization is prerequisite for the transcriptional function of FoxP proteins (Li et al., 2004a; Chae et al., 2006; Li et al., 2007; Song et al., 2012). Dimerization may also be important for the phenotype of FOXP human mutations (Mizutani et al., 2007; Sollis et al., 2015; Sollis et al., 2017).

In summary, FoxP2 in humans and songbirds are clearly relevant for vocal communication. Given the recent implications of FoxP1 and FoxP4 in related phenotypes and the co-expression of all three FoxPs and their molecular interaction we hypothesized that FoxP1 and FoxP4 in Area X are also relevant for song behavior. To address this, we experimentally down-regulated either FoxP1 or FoxP4 in zebra finch Area X and compared the resulting song phenotypes to the previously described inaccurate and incomplete song learning after the FoxP2 knockdown.

Materials & Methods

Subjects

All experiments were performed in accordance with the guidelines of the governmental law (TierSchG). 60 male zebra finches (*Taeniopygia guttata*) were used in this study under the project approved by the Landesamt für Gesundheit und Soziales (LaGeSo) G0117/12. Animals were housed under a 12h:12h light:dark cycle with food and water provided *ad libitum*. Birds were non-invasively sexed aged between 7-14 post-hatch days (PHD) (Adam et al., 2014).

Generation of lentivirus against zebra finch FoxP1 and FoxP4.

Short hairpins against FoxP1 and FoxP4 were generated as described for FoxP2 (Haesler et al., 2007). The structure of the linear DNA encoding shRNA hairpins was sense-loop-antisense. The sequence of the loop was GTGAAGCCACAGATG. We tested the sequence specificity of 12 short hairpins against FoxP1 and 11 short hairpins against FoxP4. To do so we over-expressed in HeLa cells each one either with FoxP1 or with FoxP2 or with FoxP4. All FoxP over-expression constructs were cloned from adult zebra finch brain cDNA and tagged with the Flag epitope (Mendoza et al., 2015). Subsequent Western blot analysis using a Flag antibody (Flag-M2 Sigma-Aldrich Cat# F3165, RRID: AB_259529, previously Stratagene) revealed three hairpins that strongly reduced FoxP1 expression levels but not the expression of FoxP2 or FoxP4 (FoxP1-sh1, target sequence AACAGTATACCTCTATAC, FoxP1-sh2, target sequence TGCATGTCAAAGAAGAAC, and FoxP1-sh3, target sequence CCATTAGACCCAGATGAAA). Using the same approach for FoxP4, we identified two hairpins that strongly reduced FoxP4 expression but not the expression of FoxP1 or FoxP2 (FoxP4-sh1, target sequence CCAGAATGTGACGATCCCC, FoxP4-sh2, target sequence CCCGTGCACGTGAAGGAGGAG). We used beta-actin as loading controls for all western blots (detected with antibody Sigma-Aldrich Cat# A5441, RRID: AB_476744). The DNA fragments encoding the hairpins FoxP1-sh1, FoxP1-sh2, FoxP1-sh3 and FoxP4-sh1, FoxP4-sh2 were subcloned into a modified version of the lentiviral expression vector pFUGW containing the U6 promoter to drive their expression. As a control, we used the previously described non-targeting hairpin (Control-sh, sequence AATTCTCCGAACGTGTCACGT) cloned into the modified pFUGW (Haesler et al., 2007). All viral constructs expressed GFP under the control of the human ubiquitin C promoter. Recombinant lentivirus was generated as described previously (Haesler et al., 2007). Titers of virus solution were usually in the range of 1-3x10^6 IU/µl.

Stereotaxic neurosurgery

Birds used subsequently for song analysis were injected with one of the different lentiviral vectors, e.g. one of the three FoxP1 knockdown (kd) constructs or one of the two FoxP4 kd constructs, or the control constructs (Haesler et al., 2007). Injections were performed as described (Haesler et al., 2007; Adam et al., 2016). Briefly, at PHD 23 birds were injected bilaterally with approximately 200nL each into 8 sites per Area X. Injection side, order and the type of construct, were randomized. To determine kd efficiency via qRT-PCR (see below) we injected additional birds into Area X in one hemisphere with the vector carrying one of the different kd constructs, and Area X of the other hemisphere with a non-silencing Control-sh construct (**Figure 1a-c**)(Haesler et al., 2007).

Quantification of FoxP1 or FoxP4 mRNA knockdown efficiency

To test whether FoxP1 or FoxP4 contribute to song learning in zebra finches the levels of both genes were reduced separately in Area X *in vivo*, using lentivirus-mediated RNA interference (RNAi; FoxP1-sh2/3 or FoxP4-sh7/19). The rationale and overall procedure followed previously published protocols (Haesler et al., 2007; Adam et al., 2016). Briefly, 6 birds for each FoxP for follow-up by qRT-PCR were transferred to their home cages after surgery and grew up in the presence of their biological parents and siblings. All birds were sacrificed at 50±2 PHD and did not sing for two hours prior to it (for a timeline of experiment see **Figure 1a**). Each hemisphere was embedded in Tissue-Tek O.C.T. compound in a mold and immediately shock-frozen in liquid nitrogen or dry ice and stored at -80°C. Brains were cut by cryostat as described (Olias et al., 2014; Adam et al., 2016). Microbiopsies (0,5-1,5 mm diameter) of Area X from both hemispheres were excised and stored individually at -80 °C (**Figure 1c-d**). Remaining sections were stored in 4 % (w/v) paraformaldehyde/PBS solution (PAF) and used to verify successful targeting and to assess the location of GFP signal in the surroundings of the punched out Area X (**Figure 1d**). For the RNA extraction from these small amounts of tissue (approximating 1 mm^3 per hemisphere), we used 200µl of TRIZOL for each punch. To digest remaining DNA we used Turbo DNAse from AMBION following the manufactures instructions. cDNA synthesis was carried out using random hexamer primers and 100ng total RNA of the combined microbiopsies of each bird. Reverse-transcriptase free reactions were included to control for genomic DNA contamination. All cDNAs were diluted with nuclease free water (5-fold for individual microbiopsies).

For the quantification of FoxP1 and FoxP4 mRNA expression levels in Area X of kd animals, we used the real-time PCR system Mx3005P and the MxPRO QPCR program (Stratagene; Agilent Technologies, U.S.A.). qRT-PCR reactions were run in triplicates in a total reaction volume of 25 µL as described (Olias et al., 2014; Adam et al., 2016). The efficiency of all

primer pairs ranged from 2±10%. We used the following primer pairs:
FoxP1 (5´ CGTTAAAGGGGCAGTATGGA 3´ / 5´ GCCATTGAAGCCTGTAAAGC 3´),
FoxP4 (5´ TGACAGGGAGTCCCACCTTA 3´ / 5´ AGCTGGTGTTGATCATGGTG 3´),
HMBS (5’ GCAGCATGTTGGCATCACAG 3’ / 5’ TGCTTTGCTCCCTTGCTCAG 3’)
(Haesler et al., 2007),
GFP (5’ AGAACGGCATCAAGGTGAAC 3’ / 5’ TGCTCAGGTAGTGGTTGTCG 3’)
(Adam et al., 2016, 2017). Reactions were run with the following times and temperatures: 10‘ at 95 °C followed by 40 cycles of 30” at 95 °C, 30” at 65 °C, 30´´ at 72°C (60 °C for HMBS and FoxP1); and a melting curve to check for amplification specificity. The mean Ct for each sample was derived from the run data and used to calculate relative gene expression for the gene of interest (GOI) (FoxP1 or FoxP4). We used HMBS as a reference gene, as it is the most stable of all tested potential reference genes for our experiments (Haesler et al., 2007; Adam et al., 2016, 2017). Relative expression values were averaged per animal and hemisphere. Only cDNA from GFP-positive biopsies in both hemispheres were used to measure the expression of FoxP1 or FoxP4 and HMBS. Data were normalized to the Control-sh hemisphere and set to 100%.

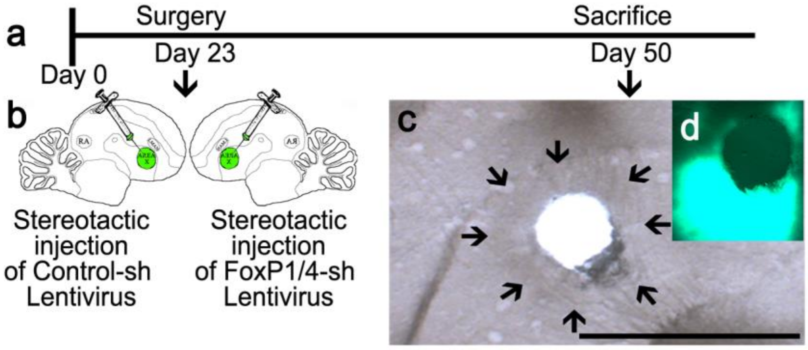

a
Surgery
Day 23
Sacrifice
Day 50
Day 0
b
RA
Area X
Stereotactic injection of Control-sh Lentivirus
Stereotactic injection of FoxP1/4-sh Lentivirus
c
d

Figure 1 – Timeline of *FoxP1* and *FoxP4* qPCR quantification using lentiviral-mediated RNAi *in vivo*. (a) 23 day old birds were injected bilaterally into Area X. One hemisphere received a Control-sh virus, the other hemisphere a sh-knockdown virus against FoxP1 (FoxP1-sh2 or FoxP1-sh3) or against FoxP4 (FoxP4-sh7 or FoxP4-sh19). (b) After surgery, birds were kept with their parents until PHD 50. Brains were extracted, frozen and stored at -80°C. 200 µm slices were cut by cryostat, Area X microbiopsies were punched (c) and stored at -80°C for subsequent mRNA extraction. Correct targeting was assessed by PAF-fixing the slices from which punches were taken and assessing GFP expression in the surrounding tissue (d) and determining the location of Area X by phase contrast (arrows in c). Scale bar 2mm.

Quantification of the percentage of targeted neurons

Because it is not possible to verify the efficiency of knockdown via QPCR from microbiopsies of Area X and simultaneously to determine the percentage of infected neurons histologically in the same animals, we checked the percentage of neurons infected in 3 additional animals. To do so, we quantified the number of medium spiny neurons in Area X that were infected by the Control-sh virus (GFP). We assessed the number of MSN by FoxP1 immuno-reactivity (Abcam, Foxp1 mouse monoclonal, ab32010; RRID: AB_1141518) because we determined previously that FoxP1 mostly co-localizes with FoxP4 Area X neurons (Mendoza et al., 2015), and because the FoxP4 antibody used in this article did not work in perfused brains. Sections were analyzed with a 40x oil objective on a Zeiss Axiovert 200M Digital Research Microscopy System. The Slidebook Digital Microscopy software package (Intelligent Imaging Innovations) was used for fluorescence image acquisitions. Per Area X in each hemisphere we acquired 4 images at 40x magnification using the AxioVision 4.6 program and manually counted all neurons in which GFP and FoxP1 were co-localized.

Quantification of the volume of Area X infected in birds whose song was analyzed

Birds were overdosed with Isoflurane (Forane-ABBVIE (B5068)) and subsequently perfused with 4%PFA/PBS. Brains were dissected and post-fixed overnight in 4% PFA/PBS paraformaldehyde. Brains were sectioned sagittally at 40µm thickness with a vibratome (Leica, Wetzlar, Germany) and sections stored in PBS at 4°C in the dark. Every fourth slice was stained with Acetylcholinesterase (AChE) (Karnovsky and Roots, 1964) to visualize and measure the size of Area X. Sections were mounted on Chromalum (Chromium(III) potassium sulfate)/gelatin coated slides and embedded with Mowiol (6 g Glycerin, Merck, Darmstadt, Germany; 1.04092.1000; 2.4 g Mowiol 488; Calbiochem, La Jolla, CA; 475904; and 12 ml 0.2 M Tris-HCl, pH 8.5). The remaining sections were stored in cryoprotectant and stored at -20°C. To calculate the targeted area we quantified Area X as well as the GFP targeted area using ImageJ following the procedure of (Tramontin et al., 1998).

Song tutoring, recording, and analysis

Tutoring – Juveniles were raised in their respective family cohorts until PHD20. Between PHD20 and PHD30 the adult male was removed to prevent song exposure before tutoring (Roper and Zann, 2006). After surgery at PHD23 birds were returned to their home cages with their mother and sibling females and remained there until PHD30. Subsequently each experimental juvenile was tutored by one adult male in a sound-isolated recording box, because under these conditions the pupil learns to produce a song that most resembles the song of his tutor (Tchernichovski and Nottebohm, 1998). Song was recorded continuously throughout this period using Sound Analysis Pro [SAP (Tchernichovski et al., 2000)]. A day before sacrifice (at PHD95 or later) a minimum of 50 motifs (see next paragraph for definition) of undirected singing from the experimental bird was recorded in the absence of the tutor for up to 5 days for subsequent bioacoustic analysis (**Figure 2**). To be able to directly compare the effects of experimental reduction of FoxP1 or FoxP4 in Area X on song development to those of FoxP2 we analyzed the recordings obtained in this study (FoxP1, FoxP4) and re-analyzed the recordings from Haesler et al., 2007 using the same bioacoustic parameters for all groups. This modus operandi served to minimize experimenter-induced variability and also to assess replicability of the present data and those of the two previous reports on developmental song deficits as a consequence of FoxP2 kd in Area X (Haesler et al., 2007; Murugan et al., 2013).

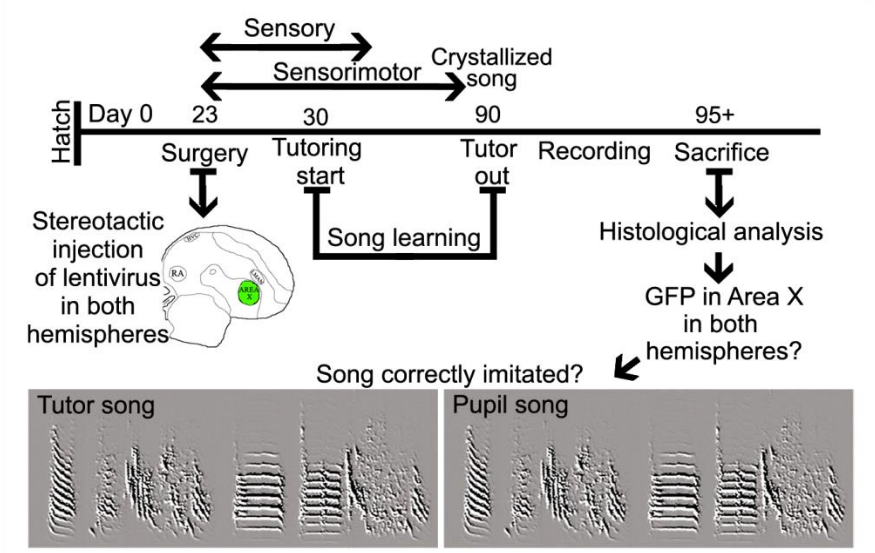


Figure 2 – Timeline of FoxP1 or FoxP4 knockdown in Area X and vocal learning success. In the first two weeks after hatching the birds were sexed. On day 23, at the beginning of the sensory learning period, either Control-sh, or FoxP1-sh2/3, or FoxP4-sh7/19 virus was bilaterally injected into Area X of male zebra finches. From day 30 on, injected birds were housed in sound-recording chambers together with an adult male zebra finch as a tutor. After reaching 90 days of age, the tutor was removed and adult song was recorded. Before song analysis, we verified correct targeting by analysis of GFP expression in in Area X of both hemispheres.

Song terminology – Zebra finch song is individual-specific and consists of a series of acoustically distinct elements (3-9 in this study) separated by silent gaps. The song elements are arranged in a repeated order, called 'motif'. The order of song elements can vary slightly, resulting in slightly different motifs. The most frequently sung motif is the 'typical' motif.

Analysis of motif imitation – We quantified how well pupils copied the motif of their tutor using a similarity score and an accuracy score obtained in SAP from ten asymmetric pairwise comparisons of the pupil's typical motif with the tutor motif, similar as described previously (Haesler et al., 2007). We report the similarity and accuracy scores of an M x N batch similarity analysis (which compares 10 motifs of the tutor to the 10 motifs of the pupil, resulting in 100 independent comparisons). Similarity values reflect how much of the tutor's song material was detected in the pupil's motif. Accuracy is a measure of how well the sound elements were copied by the pupil.

Song analysis – We investigated different aspects of pupils' song learning success and song performance. (i) How many song elements of the tutor did the pupil imitate. (ii) How many elements of a pupil's song were not recognized in the tutor's song. Pupils' song can contain elements that are sufficiently different from the tutor's song to not be recognized as an imitated element by SAP. (iii) How accurate was the imitation of pupils' song elements. (iv) How variable was the performance of individual song elements of pupils compared to variability of tutors. (v) How stereotyped was the sequential delivery of multiple song motifs of pupils' songs compared to the stereotypy of tutors. Did pupils repeat elements ('stutter')? (vi) How were the durations of song elements and the inter-element intervals ('gaps') distributed in the tutors' and pupils' songs. (vii) Did the delivery of multiple song motifs of pupils differ in rhythmic isochronicity from that of fathers.

To address (i-iii) we compared each song element of the tutor to all song elements of the pupil with a symmetric batch MxN analysis in SAP. The element of a pupil with the highest similarity and accuracy score (in SAP) to an element of the tutor was considered imitated and thus 'shared' by tutor and pupil. When two pupil elements had similar scores to an element of the tutor, we took the order within the motif also into consideration. The scores of shared elements between tutor and pupils ranged between 70 and 100 in similarity or accuracy comparisons. To quantify (i) whether FoxP-sh birds imitated fewer elements of their tutors than did control-sh birds we quantified the number of elements shared by tutor and pupil. We then counted how many elements the pupil shared with the tutor and expressed this as the fraction of all elements specific to the tutor. A value of 1 indicates that all tutor elements were found in the pupil's song. As the value approaches zero, increasingly fewer elements of the tutor are represented in the pupils' songs. To quantify (ii) the fraction of elements the pupils sang that were not found in the tutor's song was as the number of elements unique to the pupil divided by the total number of elements of the tutor. A value of zero reflects that there are no different or additional elements in the pupil's song.

(iv) Element delivery – To assess the rendition to rendition variability in element performance we chose 32 motifs randomly, took 10 of each of the elements of the typical motif from tutors and pupils and measured the similarity and accuracy in a symmetric M x N batch analysis. We thus compared how similar to itself an element was in each rendition of a song of a pupil to the self-similarity of an element in the tutor song. Results of these comparisons between elements are expressed in a single measure which is the product of

similarity and accuracy to obtain the element identity score (as reported by (Haesler et al., 2007).

(v) Stereotypy of song performance and stuttering – Stereotypy is a measure that addresses whether the bird sings the same order of elements each time. We quantified stereotypy as described previously (Scharff and Nottebohm, 1991) from the same 32 randomly chosen motifs used to quantify element performance (see above) of each bird, Stereotypy scores range between 0 and 1, with 1 reflecting that the birds sang the same sequence of elements in the same order in all 32 motifs. Lower scores indicate more sequence variability in a motif from rendition to rendition. We also quantified the propensity of birds to repeat song elements by calculating the percentage of all elements sung by each bird that was preceded by an element of the same type (e.g. AA).

(vi) Duration of song elements and silent gaps – We measured the overall distribution of all durations of song element and inter-element intervals ('gaps') from 48 ± 24 (mean ± SD) song motifs per bird. We then compared the distributions of element durations between pupils and tutors, using the Jensen-Shannon distance (the square root of the Jensen-Shannon divergence) as a dissimilarity metric between the two probability distributions (Lin, 1991; Endres and Schindelin, 2003; Sasahara et al., 2015). To estimate how similar/dissimilar song element and gap durations between two groups of untreated adult zebra finches are, we also compared the distributions of element and gap durations of a cohort of 15 adult males that were analyzed previously by Norton & Scharff (2016) to the tutors of the current study. In order to quantify the similarity in shape of the distributions of the gap duration independently of their position on the x-axis (i.e. their absolute duration), we shifted the tutor distribution in 0.002s steps and calculated the Jensen-Shannon distance (JSD) between the lagged tutor distribution and the stationary pupil distribution for each step. We report the JSD and the lag at which the JSD was minimal.

(vii) Rhythm analysis – We determined the isochronous pulse that best fit to the song element onsets of each motif in the 42 motifs used for duration analysis (above) using the method described in (Norton and Scharff, 2016). The frequencies of the best fitting pulses for all songs of each bird were clustered (see (Norton and Scharff, 2016) for details) and the percentage of songs in the largest cluster of each bird was determined. The higher this percentage is, the more songs have the same pulse. The best fitting pulses of tutor songs clustered in a range from 20 to 60Hz. To compare the rhythmicity of the pupil songs to that of the tutors, we restricted the analysis of the pulses to this frequency range. Pulse fit was quantified as the root-mean-square of the deviations of each song element onset to its nearest pulse, multiplied by the pulse frequency (Frequency-normalized Root-Mean-Square Deviation or FRMSD). To assess whether the rhythmic regularity (i.e. pulse fit) could just be a

by-product of zebra finch specific song element and gap durations independent of the birds' individual song elements and their order, we compared each pupil's song rhythm to the rhythm of 50 model songs with an identical number of song elements and identical sequence, but different randomized element and gap durations ("Model C" in Norton & Scharff, 2016). The FRMSD (pulse fit) of each of the 50 sets of model songs were compared to the bird songs in a separate linear model. Of the comparisons that detected a significant difference in FRMSD ($p<0.05$), the percentage of these comparisons in which the bird songs had a lower FRMSD (i.e. a better pulse fit and therefore a higher degree of isochronous organization of their song rhythm) is reported here (**Figure 13d**).

Linear discriminant analysis (*LDA*) – To test whether the 4 groups (FoxP1-sh, FoxP2-sh, FoxP4-sh, and Control-sh) could be discriminated by differences in their song phenotype alone we selected 5 features of song structure from different domains: One spectral feature (the amount of frequency modulation of song elements), one measure of song learning success (number of copied song elements from tutor), two measures of temporal variability (CV of duration of the most variable inter-onset-interval and average CV of gap durations) and one rhythmic parameter (average FRMSD). Discrimination success was evaluated by prediction of the treatment group of each bird through leave-one-out cross-validation. To do so one individual after another is removed from the set, the discriminant functions are calculated each time and used to classify the missing individual.

Statistics – All statistical tests were performed using the data analysis software R (R Development Core Team, 2013) and/or GraphPad Prism 4.0. All graphs were prepared with GraphPad Prism 4.0 (GraphPad Software, San Diego, CA) or R.

Results

Selection of specific short hairpins to downregulate zebra finch FoxP1 or FoxP4

To determine efficacy and specificity of different short hairpins against FoxP1 and FoxP4 we overexpressed FoxP1 or FoxP4 in HeLa cells. Three (FoxP1-sh1, -sh2, -sh3) out of the twelve FoxP1 short hairpins tested strongly reduced FoxP1 protein levels (**Figure 3**) but did not affect the expression of FoxP2 (**Figure 3b**) or FoxP4 (**Figure 3c**). This is interesting because the FoxP1-sh2 differed only at 2 nucleotides from the FoxP2 gene and at 5 nucleotides from the FoxP4 gene, whereas FoxP1-sh1 and FoxP1-sh3 ranged from 57% to 63% in sequence similarities to the other FoxP members. The FoxP1-sh1 affected the expression of the protein least and was therefore not further used in this study.

Two (FoxP4-sh7 and -sh19) of the eleven short hairpins tested reduced FoxP4 protein levels (**Figure 3f**). FoxP4-sh7 and FoxP4-sh19 were 23-71% similar when compared to the other FoxP subfamily members and did not alter the expression of FoxP1 (**Figure 3d**) or FoxP2 (**Figure 3e**). We used both for further studies. In a previous study (Haesler et al., 2007) a nontargeting short hairpin control (Control-sh) was shown not to affect FoxP2 expression. We used the same Control-sh in this study and showed that it did not alter the expression of either FoxP1 (**Figure 3g**) or FoxP4 (**Figure 3h**).

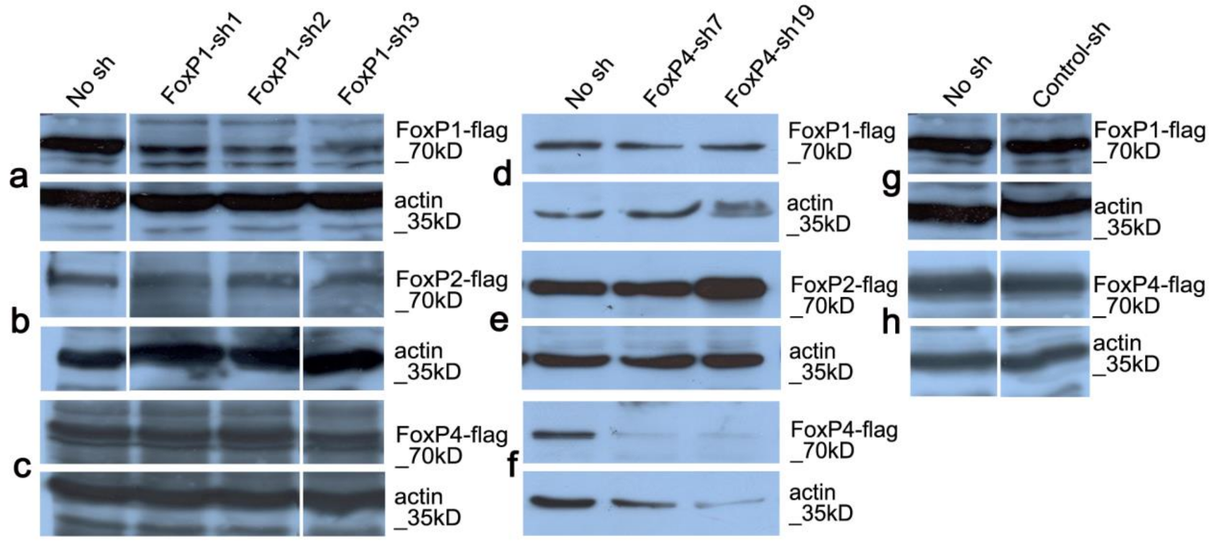


Figure 3 – Western blots showing specific downregulation of FoxP1 or FoxP4 using short hairpins (sh). Overexpression of zebra finch FoxP1 (a, d and g), or FoxP2 (b and e), or FoxP4 (c, f and h), each tagged with a Flag-epitope, and one of different hairpin constructs against FoxP1 (FoxP1-sh1, FoxP1-sh2, or FoxP1-sh3; a-c), or FoxP4 (FoxP4-sh7 and FoxP4-sh19; d-f), or control short hairpin (g-h) in HeLa cells. Western blot analysis using the Flag antibody (top panels in a-h) revealed that all short hairpins against FoxP1 (a-c) efficiently reduced FoxP1 levels (a, upper panel), but did not downregulate FoxP2 (b) or FoxP4 (c); all short hairpins against FoxP4 efficiently reduced FoxP4 levels (f, upper panel), but did not downregulate FoxP1 (d) or FoxP2 (e); the control short hairpin did not downregulate FoxP1 (g) or FoxP4 (h). Immunostaining with actin antibody shows comparable loading of protein samples in all cases (a-h; bottom panels). Westerns shown in panels (a, b, c, g and h) were run in the same membrane but due to different loading order some were cut to arrange them in the same order for all panels.

Efficacy of cellular infection by lentivirus in Area X

To assess how many MSN in Area X can on average be infected we injected GFP-expressing control virus stereotaxically into Area X of three 23-day old birds and at PHD 50 quantified the number of cells in which the GFP signal was co-localized with FoxP1 immunoreactivity (**Figure 4a-d**). We chose FoxP1 because most MSN in Area X express FoxP1, either in combination with FoxP2 and/or FoxP4 or alone (Mendoza et al., 2014). 89% of GFP positive cells were also immunoreactive against FoxP1 (**Figure 4a-e**), consistent with previous studies

(Wada et al., 2006; Haesler et al., 2007). Off the total FoxP1-expressing neuron population in Area X on average 16% of the cells also expressed GFP, indicating virus infection (**Figure 4a-d, f**).

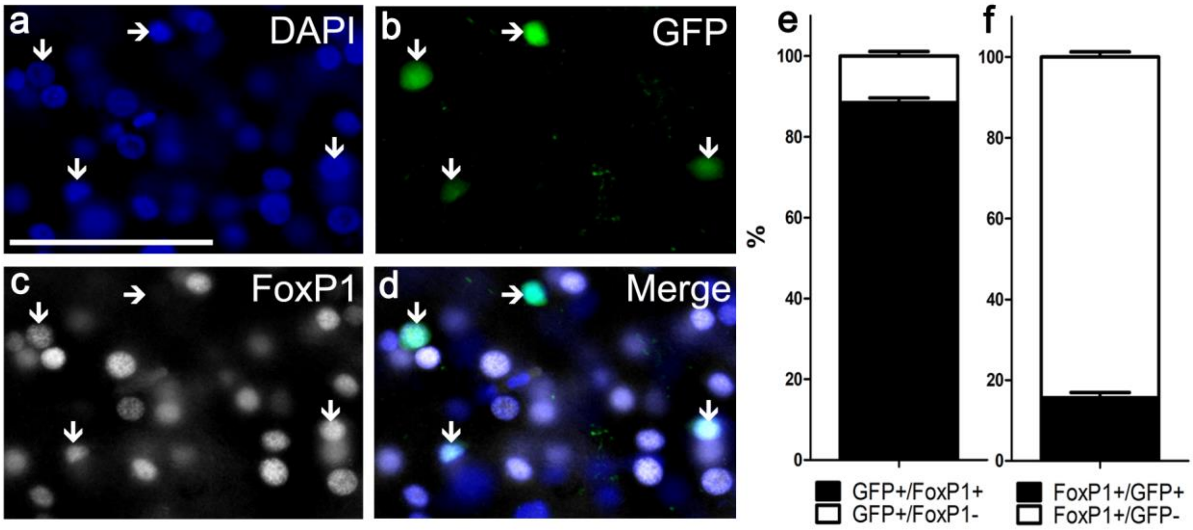


Figure 4 – Efficacy of infection. Quantification of Area X MSN neurons at PHD50 expressing GFP as a result of virus injection (control short hairpin) at PHD23. Photomicrograph at 40x magnification shown in a Z stack projected photo (a) blue fluorescence of DAPI stained cell nuclei, (b) green fluorescent GFP expression indicating virus infected cells (c), FoxP1 immunoreactivity revealed by a secondary Alexa 568 antibody (red) false-color-coded in white and (d) overlay with vertical arrows pointing to neurons co-expressing FoxP1 and GFP. One GFP positive cell that does not express FoxP1 is indicated by a horizontal arrow. (e) Infected neurons co-expressing GFP and FoxP1 expressed as a percentage of the total number of GFP expressing neurons. (f) Virus-infected GFP-expressing and FoxP1 immunoreactive neurons expressed as percentage of the total number of FoxP1 expressing neurons. In (e) and (f) bars refer to mean of means + standard error of the mean [SEM]. Error bar in (a) applies to panels (a-d) 50 µm.

Efficacy of FoxP1 or FoxP4 mRNA downregulation in Area X

We evaluated the reduction of FoxP1 or *FoxP4* mRNA expression at PHD50 by QPCR after injections of the respective knockdown viruses in Area X of PHD23 males (**Figure 1a-d**). The amount of knockdown was quantified by comparing FoxP expression in the knocked down hemisphere with the control injected one, as described (Haesler et al., 2007; Olias et al., 2014; Adam et al., 2016, 2017).

FoxP1 mRNA levels in Area X were on average 20% lower in the hemispheres injected with the knockdown FoxP1-sh2 or FoxP1-sh3 viruses than in the control injected hemispheres (**Figure 5a**). Comparable results were obtained for FoxP4-sh7 or sh19 (**Figure 5c**) and controls (**Figure 5b**). In contrast, *GFP* mRNA levels did not differ statistically between control and knockdown-injected hemispheres (data not shown), as reported previously reports (Haesler et al., 2007).

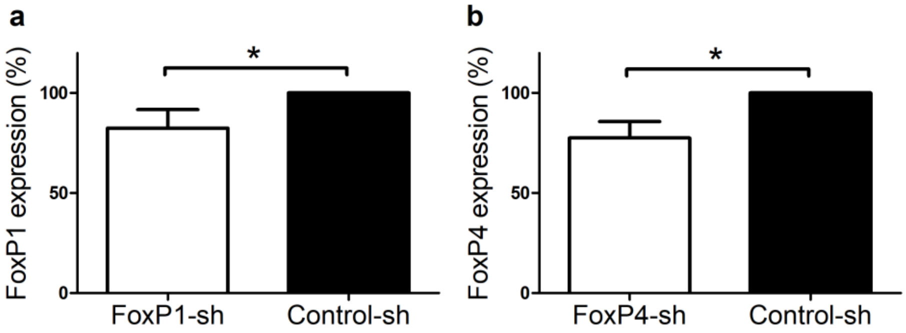


Figure 5 – *In vivo* downregulation of FoxP1 or FoxP4 in Area X. mRNA levels of FoxP1 (a) or FoxP4 (b) assessed by qRT-PCR in Area X tissue was significantly lower in the FoxP1-sh or FoxP4-sh injected hemisphere than in the Control-sh injected hemisphere of the same animal (Wilcoxon signed rank test, W=-21, p=0.03, n=6).

Quantification of virus-infected Area X volume

Before analyzing the adult songs of birds that were injected as juveniles with knockdown viruses in Area X bilaterally, or with corresponding controls, we assessed the percentage of Area X tissue that was infected, as judged by GFP fluorescence in tissue sections, and compared this to the previously published results on FoxP2 (Haesler et al., 2007) (**Figure 6a-c**). The volume of the infected area was similar across hemispheres in FoxP1, FoxP4 and control birds (One way ANOVA; p>0.05; F=2.71; DF=2; **Figure 6c**). On average, the GFP fluorescence in both hemispheres covered 34.8% of Area X for FoxP1 (SEM 17.16), 28,6% for FOXP4 (SEM 16.27), and 19.6% for the controls (SEM 8.92), i.e. were in the same range as the 20.4% reported for FoxP2 knockdown birds (Haesler et al., 2007).

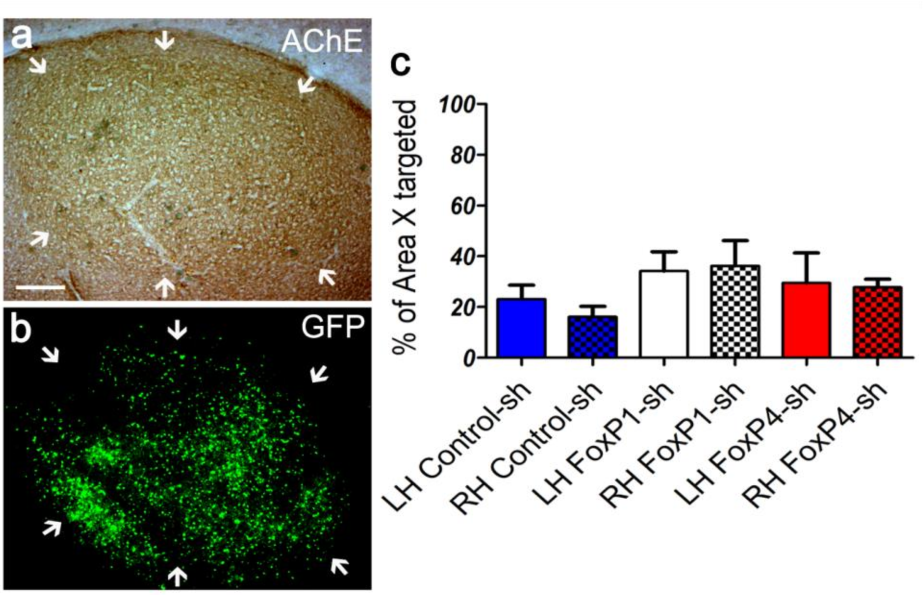


Figure 6 – Quantification of Area X volume targeted by the viral infection in birds whose song learning was assessed. Representative photomicrographs of Area X (a,b). (a) bright-field photo of a sagittal section stained for AChE delineating Area X (white arrows) scale bar 200μm; (b) same section under fluorescence illumination showing GFP signal. (c) volume of the virus induced GFP-expressing tissue within Area X, expressed as percentage of total Area X volume in left and right brain hemispheres. Both hemispheres were infected to similar degrees in all groups s(average for each hemisphere ±SEM).

Knockdown of FoxP1/2/4 in juveniles affects adult song in multiple ways.

Comparing sonograms from tutors and pupils in the different treatment groups we noticed striking deficits in the adult songs of pupils that had received FoxP1 or FoxP4 knockdown injections as juveniles (**Figure 7b,d**) in contrast to control injected birds (**Figure 7a**). The song deficits of birds with FoxP1 and FoxP4 knockdowns were partly similar to the ones reported for FoxP2 knockdowns (Haesler et al., 2007) but there were also differences between the FoxP1/2/4 knockdown animals. To exemplify the type of deficits observed, **Figure 7** provides two song motifs each of tutor-pupil pairs per treatment group.

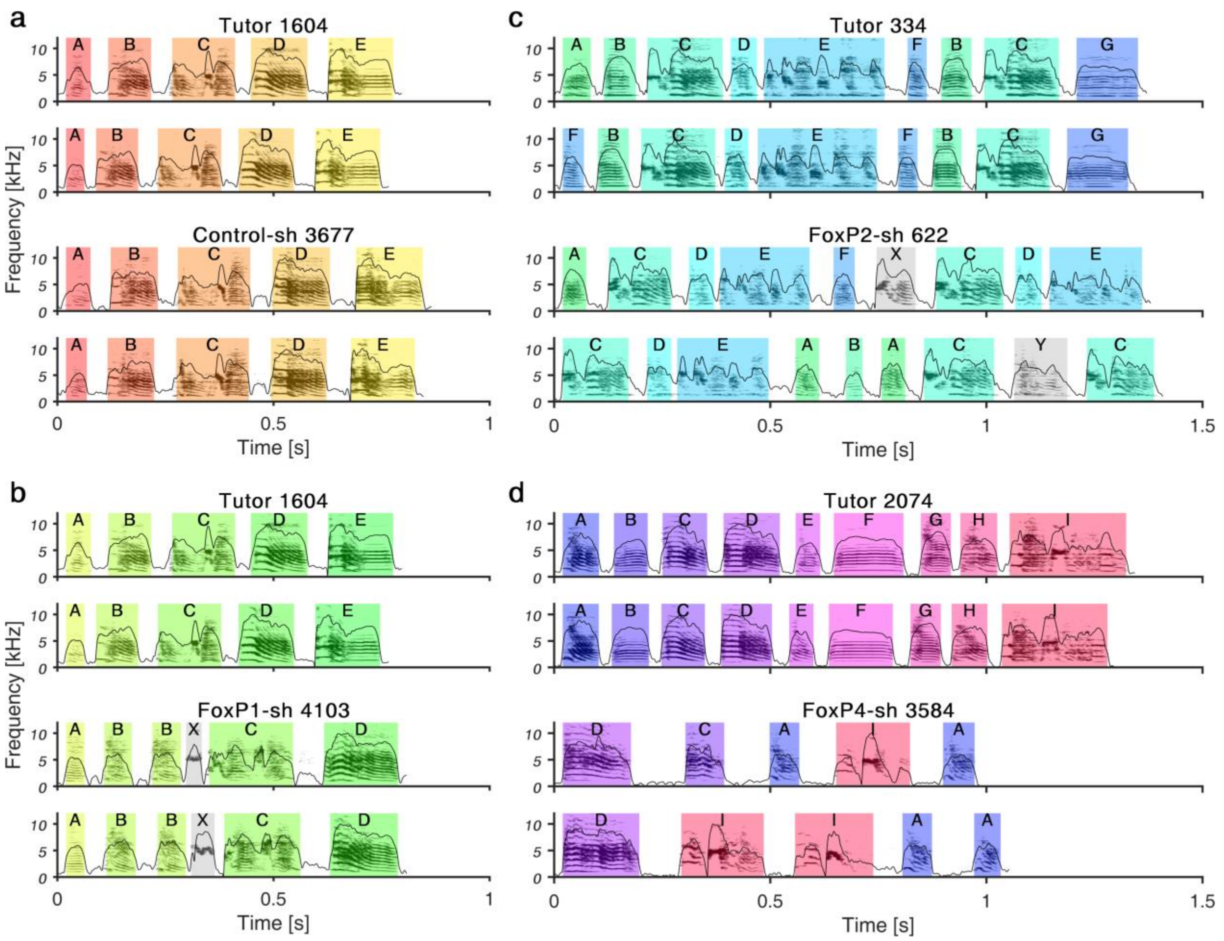


Figure 7 – Representative sonograms with amplitude envelops overlaid illustrating different deficits in the experimental groups (bottom two rows in a-d) and their respective tutors (top two rows in a-d). Song elements of the same type in each tutor-pupil pair are indicated by the same color and identified by the same letter. The identity of song elements was determined by systematic similarity comparison between pupil and tutor elements using Sound Analysis Pro software (Tchernichovski et al., 2000). Song elements are separated by silent inhalation gaps. (a) Control-sh injected pupil 3677 imitated all elements from his tutor 1604 and delivered them in the same order. (b) FoxP1-sh injected pupil 4103 had the same tutor as the control injected juvenile in (a). In contrast to the control-sh, the FoxP1-sh pupil did not copy element E, added an element that was not recognized by SAP as matching any tutor element between B and C (highlighted in gray, X) and copied element C less accurately. (c) FoxP2-sh injected pupil 622 only copied elements C, E and G from the tutor 334, included an element not recognized in the tutor song and the sequence of elements varied from rendition to rendition. (d) FoxP4-sh injected pupil 3584 only copied elements A, C, D and I from tutor 2047, the sequence as well as durations of song elements and gaps were altered. Delivery from rendition to rendition was not stereotyped and elements were repeated often (I I A A in the second example)

The tutor birds (**Figure 7a-d** top two panels) produced the stereotyped song that is characteristic for zebra finches, singing their song elements mostly in the same order in every motif rendition. The Control-sh injected bird (**Figure 7a** bottom two panels) copied all elements, kept them in the same sequence as the tutor and sang them consistently from rendition to rendition. This high copy fidelity is typical when one pupil grows up in the presence of one tutor (Tchernichovski and Nottebohm, 1998; Tchernichovski et al., 1999). In contrast, none of the FoxP1/2/4 knockdown birds copied the songs of their tutors as faithfully. While there were differences in degree and kind between the treatment groups, some song deficits were observed in all knockdown conditions. For instance, pupil songs were only partly composed of song elements that were recognizable as tutor imitations, whereas other pupil elements could not be matched to the tutor (**Figure 7b, c**). Even when elements were clearly imitations of the tutor's elements, the copy fidelity was often lower in knockdown pupils than in controls (**Figure 7d** element C and I). There was also a higher incidence of pupils not singing the copied song elements in the same order as the tutor (e.g. **Figure 7c, d**). Moreover, knockdown pupils had a higher tendency to repeat the same song element multiple times, resulting in a stutter (e.g. **Figure 7b, d**) and to change the order in which song elements were delivered from rendition to rendition (e.g. **Figure 7b-d**). The latter was particularly evident in FoxP4 knockdown pupils, in which we also noted a tendency for atypical timing of song.

Taken together, visual inspection of sonograms indicated that reduced levels of FoxP1 or FoxP4 in Area X during the song learning phase impaired song along multiple dimensions, mirroring some of the previously described song deficits resulting from FoxP2 knockdown in Area X (Haesler et al., 2007). Because other features were not seen before and seemed to segregate with the particular treatment group, we analyzed the songs of all FoxP-sh birds and their tutors in more detail.

Similarity of motifs is affected in all FoxP-sh groups, accuracy only in the FoxP2-sh group

First, we compared all pupils' songs to the songs of their tutor's to quantify overall song learning success. We analyzed undirected song of birds after they had reached 90 days when song is well learned and does not change much thereafter (Williams, 2004). Song learning success was quantified using Sound Analysis Pro software (SAP) (Tchernichovski et al., 2000). SAP analyzes the acoustic features of song along multiple dimensions and provides 'similarity' values, a measure for the amount of song material copied by the pupil and 'accuracy' values that indicate how well the copied song material is imitated. To get a comprehensive view of how well pupil and tutor motifs matched acoustically, we compared 10 motifs each of Control-sh with FoxP1-sh, FoxP2-sh, and FoxP4-sh birds to their tutors using SAP MxN batch processing and asymmetric (used for songs of different birds) comparison, resulting in 100 independent comparisons (FoxP2-sh song data from Haesler et al., 2007).

Confirming our impression from the visual analysis of sonograms, the SAP similarity scores were significantly lower in all FoxP-sh birds compared to Control-sh animals (**Figure 8a**), reflecting the fact that knockdown birds copied the tutor material incompletely

(**Figure 7**). Examining the copied portions of the song revealed that lower accuracy of imitation was found more often in birds of the knockdown groups than in control birds, but this was statistically significant only in the FoxP2-sh group (**Figure 8b**).

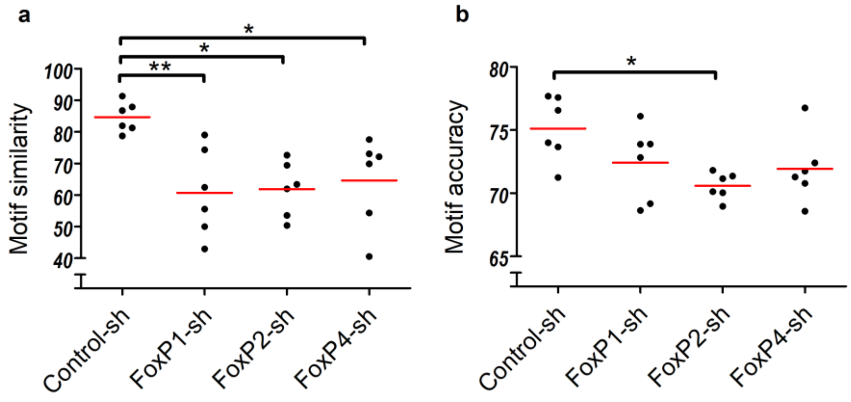


Figure 8 – Pupils in all three knockdown groups imitate tutor song incompletely ((a) Kruskal-Wallis test, p=0.0053, Kruskal-Wallis statistic = 12.73; Dunn`s Multiple Comparison Test *p<0.05;**p<0.005), but only FoxP2-sh pupils are significantly more inaccurate in the imitation fidelity of the copied song material ((b) Kruskal-Wallis test, n.s. p=0.054, Kruskal-Wallis statistic = 7.640; Dunn's Multiple Comparison Test *p<0.05). Scatter dot plots, each dot represents the mean similarity or accuracy score for each animal, the red line indicates the mean of means.

Frequency modulation (FM) is altered only in FoxP4-sh birds

More detailed analysis of spectro-temporal features of song at the motif level revealed no significant differences for pitch, goodness of pitch, amplitude modulation and entropy (data not shown), but frequency modulation was significantly different in the group of FoxP4-sh birds (**Figure 9**).

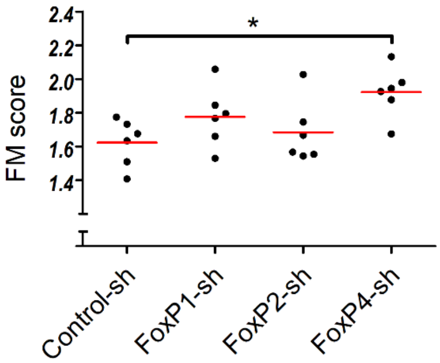


Figure 9 – Knockdown of FoxP4 in Area X affects frequency modulation (FM). Scatter dot plot, each dot represents the mean scores of an asymmetric M x N batch comparison between tutor and their respective pupil using SAP, the red line is the mean of means (Kruskal-Wallis test, p=0.0467, Kruskal-Wallis statistic = 7.967; Dunn's Multiple Comparison Test *p<0.05)

Knockdown FoxP1/2/4 copy fewer song elements from their tutors than control birds

To gain further insight into the exact nature of the lower motif imitation success and the reduced accuracy of copying in the different knockdown groups, we quantified whether pupils a) copied all tutor elements or improvised/invented some; b) copied tutor elements accurately; c) copied the sequential order of tutor elements; d) copied the duration of tutor elements, and e) copied the duration of the silent gaps between elements.

Knockdown animals copied fewer song elements from their tutors than did control animals (**Figure 10a**). The majority of Control-sh birds copied all elements of the tutor (4 of 6 birds), whereas none of the FoxP down-regulation birds copied all elements from their tutors. FoxP2 knockdown birds copied significantly fewer song elements from their tutors than did Control-sh birds (**Figure 10a**).

In all experimental and control groups some song elements could not be matched to any elements present in the tutors' song (**Figure 10b**). This was most prominent in FoxP1 knockdown birds (4 of 6 birds), but also occurred to different degrees in the other groups.

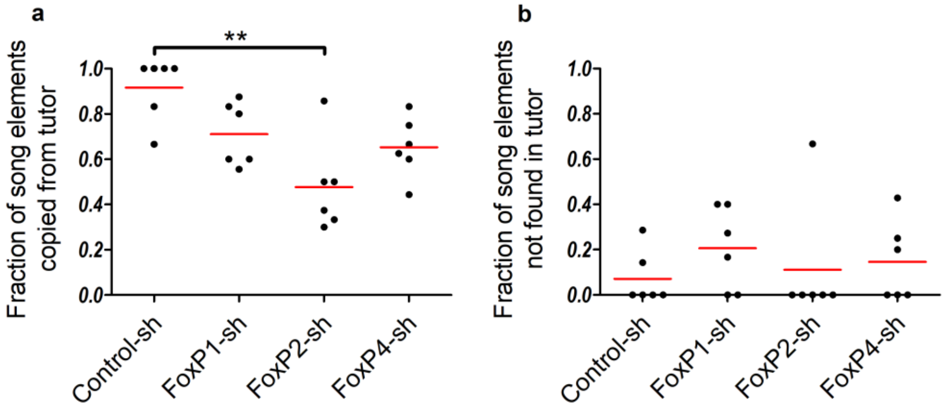


Figure 10 – Fraction of pupil elements copied from the tutor(a) and fraction of pupil elements not present in the tutor (b). FoxP knockdown birds copied fewer elements than control birds. (a) All FoxP knock-down birds copied fewer elements from their tutors than control birds, Foxp2-sh birds significantly so. Values 0 to 1 calculated as the number of copied elements in the pupil divided by the number of elements present in the tutors' song (Kruskal-Wallis test p=0.0075, Kruskal-Wallis statistic = 11.98; Dunn's Multiple Comparison Test **p < 0.005). (b) The song of some pupils in all groups contained elements not matched to any element in the tutors' songs. Values 0 to 1 calculated as the number of elements in the pupil that are not found in the tutor divided by the number of total elements present in the tutor (Kruskal-Wallis test, n.s. p=0.4494, Kruskal-Wallis statistic = 2.640; Dunn's Multiple Comparison Test n.s.).

FoxP2/4-sh birds' elements are less self-similar

To see how consistently elements were reproduced from rendition to rendition, we compared the similarity and accuracy of copied elements in ten renditions of the same element. We multiplied the resulting similarity and accuracy scores and called that product the identity score (Haesler et al., 2007). We found that the identity score between the Control-sh and FoxP1-sh birds did not significantly differ from their tutors (**Figure 11a-b**). In contrast, the FoxP2-sh and FoxP4-sh birds had a significantly lower identity score of their elements than their tutors (**Figure 11c-d**).

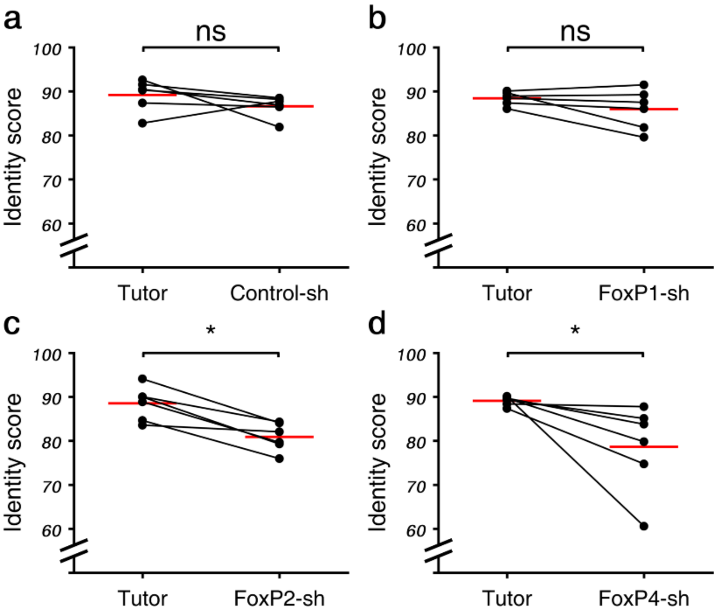


Figure 11 – Consistent reproduction of copied song elements is impaired in FoxP2-sh and FoxP4-sh birds. Scatter dot plot, each dot represents the mean identity score ((similarity*accuracy)/100) for each animal of a symmetric batch MxN analysis of ten renditions of each element in SAP, the red line is the mean of means (Wilcoxon matched pairs signed rank test, FoxP2-sh and FoxP4-sh p=0.0313, n=6, W=21).

Sequence stereotypy and Stuttering in FoxP-sh birds

To follow up our initial impression that some FoxP-sh birds varied the sequence of elements in subsequent motifs more than is typical for zebra finches (**Figure 7b-d**), we chose 32 random motifs of each bird and calculated a stereotypy score as described previously (Scharff and Nottebohm, 1991). Here a value of 1 means that birds sang the same element sequence in all 32 motifs without any variations, and with increasing sequence variability the stereotypy score approaches zero (**Figure 12a-d**). We found the FoxP2-sh and FoxP4-sh pupils to be significantly more variable than their tutors (**Figure 12c-f**). In addition, some birds in each of the knockdown groups repeated song elements, which was not the case in the tutor or control groups (**Figure 12g-h**). This stuttering-like behavior, measured as the percentage of elements that are preceded by an element of the same type (e.g. AA), was most pronounced in FoxP2-sh birds with 4 birds having an element repetition rate of at least 4% (2 each in FoxP1-sh and FoxP4-sh, none in tutors and Control-sh (**Figure 12g**)).

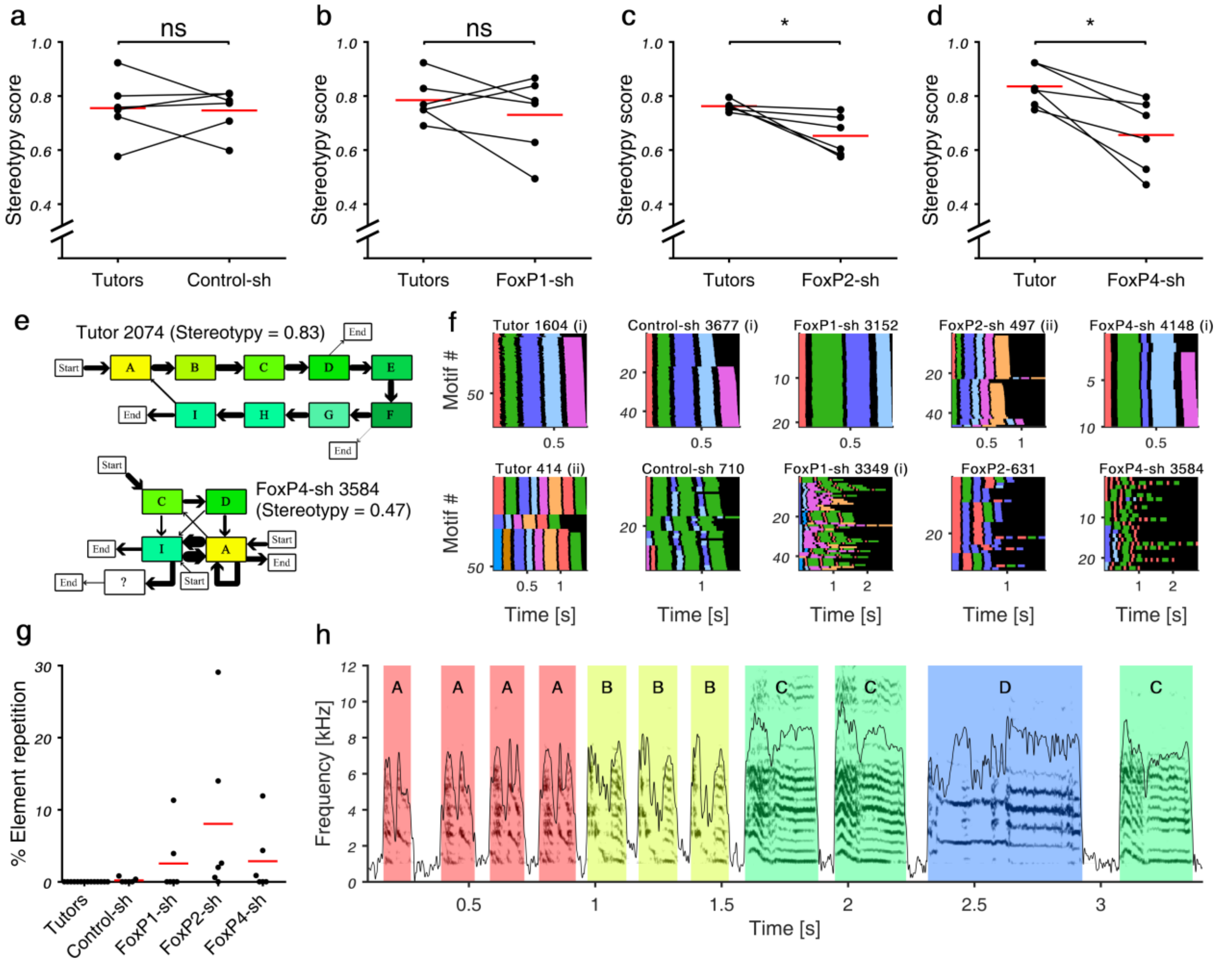


Figure 12 – After FoxP2 and FoxP4 knockdown the sequential delivery of song elements was more variable in pupils than in their tutors. In addition, some FoxP1-sh, FoxP2-sh, and FoxP4-sh have a high rate of song element repetitions (stuttering) not found in tutors and control animals. (a-d) Paired scatter dot plot. Each dot represents the stereotypy score for one animal, the red line is the mean. Tutor-pupil pairs are connected by black lines (Wilcoxon matched pairs signed rank test, FoxP2-sh and FoxP4-sh exact rank p=0.0313, n=6, W=21). (e) Sequence diagrams of the songs of FoxP4 knockdown bird 3584 (bottom) and its tutor 2074 (top). Boxes with letters A to I represent song elements, a song element not found in the tutor song is marked by a question mark. The arrows represent transitions between subsequent song elements. The size of an arrow is proportional to the relative frequency of occurrence. Arrows that do not originate at a song element mark the start of a song and arrows that do not point to an element mark the end of a song (e.g. all songs of 2074 start with A and end mostly with I, rarely with D or F). (f) Representative examples of sequence variability in 20-50 sequentially sung motifs (y-axis), indicated as thin color coded lines, sorted and stacked. The duration of each song elements is indicated by one color, song elements of the same type have the same color, silent gaps are shown in black (x-axis). Motifs are sorted alphabetically by element sequence and within identical sequences by motif duration. For each experimental group, we show one bird with high (top row) and one with low sequence stereotypy (bottom row) from each group (left-to-right: Tutors, Control-sh, FoxP1-sh, FoxP2-sh, FoxP4-sh). Pupils that were tutored by one of the tutors shown in the first column are indicated by (i) and (ii). (g) Quantitative representation of stuttering. Each dot represents the percentage of all song elements of one animal that are preceded by an element of the same type. (h) Qualitative representation of stuttering. Sonogram of an example song of FoxP2 knockdown bird 628, showing an element repetition rate of 55 percent (6 out of 11 elements are preceded by an element of the same type).

Isochronus pulse in FoxP-sh birds

We also evaluated the isochronous organization of song in all four groups and compared it to that of the tutors. We determined the isochronous pulse that best fit the song element onsets for each song (**Figure 13a, b**). As observed previously (Norton and Scharff, 2016), frequencies of the best fitting pulses formed well-defined clusters. The largest frequency cluster of each tutor bird contained on average 56% of songs in contrast to pupils (34%) (**Figure 13c**). All but 3 of the pupils had a smaller percentage of their songs in their largest cluster than their tutor (exceptions were one bird each of Control-sh, FoxP2-sh and FoxP4-sh, **Figure 13c**). The same pulse was, therefore, less consistently detected in pupil songs than in tutor songs. This suggests a looser isochronous organization in the pupil songs. A direct comparison of pulse deviation between the songs of different birds (unlike a comparison of pulse frequencies) is problematic, as deviation depends on a number of factors that differ between individuals, such as pulse frequency and the number of song elements. We therefore created simple model songs based on each of the analyzed bird songs and compared pulse deviation between bird and model songs. The latter featured the same number of elements in the same sequence as the birdsong they were modeled on but element and gap durations were randomized (see Methods). For each song of one bird a model song was created, the best fitting pulse for that song determined and the pulse deviation (FRMSD) between all songs of one bird and their respective model songs tested for a significant difference in a linear model. This process was repeated 50 times with different randomized element and gap durations in the model songs. In an average of 77% of the comparisons of tutor versus a model song that reported a significant difference in FRMSD, tutors had a lower FRMSD, i.e. a better pulse fit than the pupils, including the control pupils (**Figure 13d**). The poor rhythm of control birds might be due to their age, a point we will take up in the discussion.

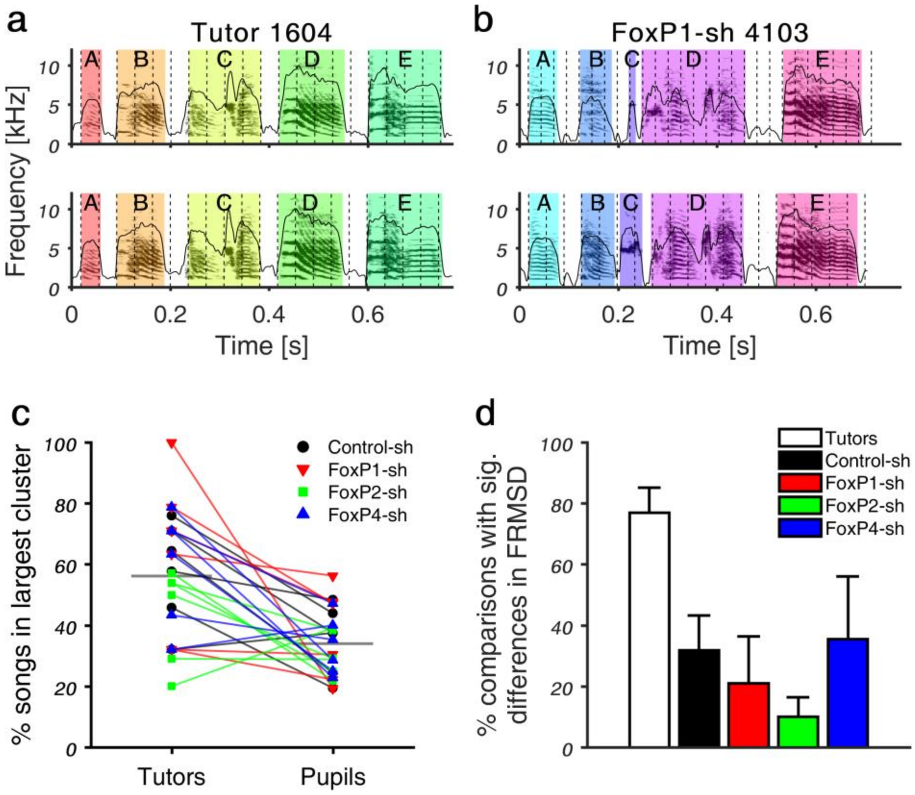


Figure 13 – All pupil groups have lower song isochronicity than tutors. (a-b) Two example motifs each of FoxP1-sh bird 4103 (b) and its Tutor bird 1604 (a) with the isochronous pulse best fitting to song element onsets overlaid as vertical dashed lines. (a) Tutor bird 1604 had almost the same pulse frequency for both renditions (top: 27.49Hz; bottom: 27.64Hz). Pulse fit as measured by frequency-normalized root-mean-square deviation of the pulse from element onsets (FRMSD, see methods) was relatively high (top: FRMSD=0.019; bottom: FRMSD=0.024). (b) FoxP1-sh bird 4103 had pulses of different frequencies best fitting the two motifs (top: 39.91Hz; bottom: 27.92Hz) and a relatively low pulse fit (top: FRMSD=0.086; bottom: FRMSD=0.086). (c) Paired plot of the percentage of all songs that were in the largest pulse frequency cluster for each bird. Lines connect each tutor (left) with his pupil (right). Gray horizontal lines show the mean. Except for 3 birds (one Control-sh, one FoxP2-sh, one FoxP4-sh), all pupils had a lower percentage compared to their tutor. (d) Bargraph of the percentage of bird-to-model comparisons with significant differences ($p<0.05$) in pulse deviation (FRMSD), in which the bird had a lower deviation (±SEM) than the model, e.g. a better rhythm.

Analysis of element and GAP durations in FoxP-sh birds

In search for possible explanations of the impaired song rhythm of Control-sh birds, we looked at the overall distribution of element and gap durations in the different treatment groups and their tutors (**Figure 14**) by quantifying the dissimilarity between the distributions. To do so, we calculated the Jenson-Shannon distance (JSD); the higher the JSD, the more dissimilar the two distributions are. Song element distributions were about equally dissimilar to the tutors in all treatments (Control-sh: JSD = 0.43; FoxP1-sh: 0.41; FoxP2-sh: 0.47; FoxP4-sh: 0.46; **Figure 14c**), as was the distribution of a cohort of 15 different previously analyzed adult males (JSD = 0.43; duration data from Norton & Scharff, 2016). As expexted the gap distribution of the cohort of adult males was very similar to that of the tutors (JSD = 0.25). Distributions of the gap durations of Control-sh, FoxP1-sh and FoxP2-sh had higher, but comparable dissimilarities (Control-sh: JSD = 0.38; FoxP1-sh: 0.41; FoxP2-sh: 0.44; **Figure 14d**). In contrast, FoxP4-sh had a considerably higher JSD (0.66), likely due to the increased overall durations of gaps. FoxP knockdown birds had an increased gap duration variability compared to control birds (Control-sh: Standard deviation = 0.017; FoxP1-sh: 0.021; FoxP2-sh: 0.025; FoxP4-sh: 0.022). Among the pupil birds, the percentage of element repetition was positively correlated with the coefficient of variation of inter-onset-intervals (Pearson, R = 0.79, $p < 0.001$), indicating that birds that stutter also have problems with the accurate timing of song elements.

Pupil birds were 96 ± 6 days of age at the time of recording (mean ± std). While song learning is largely completed by ~90 days, some song changes occur beyond that age. Among those is a gradual shortening of the gaps, while the song element duration remains unchanged on average (Glaze and Troyer, 2013). In order to quantify the similarity of the shape of the gap duration distributions independently of their position on the x-axis (i.e. leaving aside the overall higher duration of pupil gaps), we therefore shifted the tutor distribution in 2ms steps and calculated the Jensen-Shannon distance (JSD) between the lagged tutor distribution and the stationary pupil distribution for each step. Even after shifting the tutor distributions towards the pupil distributions to the point of smallest dissimilarity, knockdown birds still showed a relatively high JSD (FoxP1-sh: minimal JSD = 0.41 at lag 2ms; FoxP2-sh: 0.38 at 8ms; FoxP4-sh: 0.44 at 18ms). Control bird gap distribution, on the other hand, was as similar to tutors as the adult cohort after shifting (Control-sh: JSD = 0.22 at lag 8ms; adult cohort: 0.22 at lag 2ms). This result suggests that the songs of Control-sh birds – like those of normal untreated birds – would have acquired the level of isochronous rhythmic organization found in the tutor birds with age.

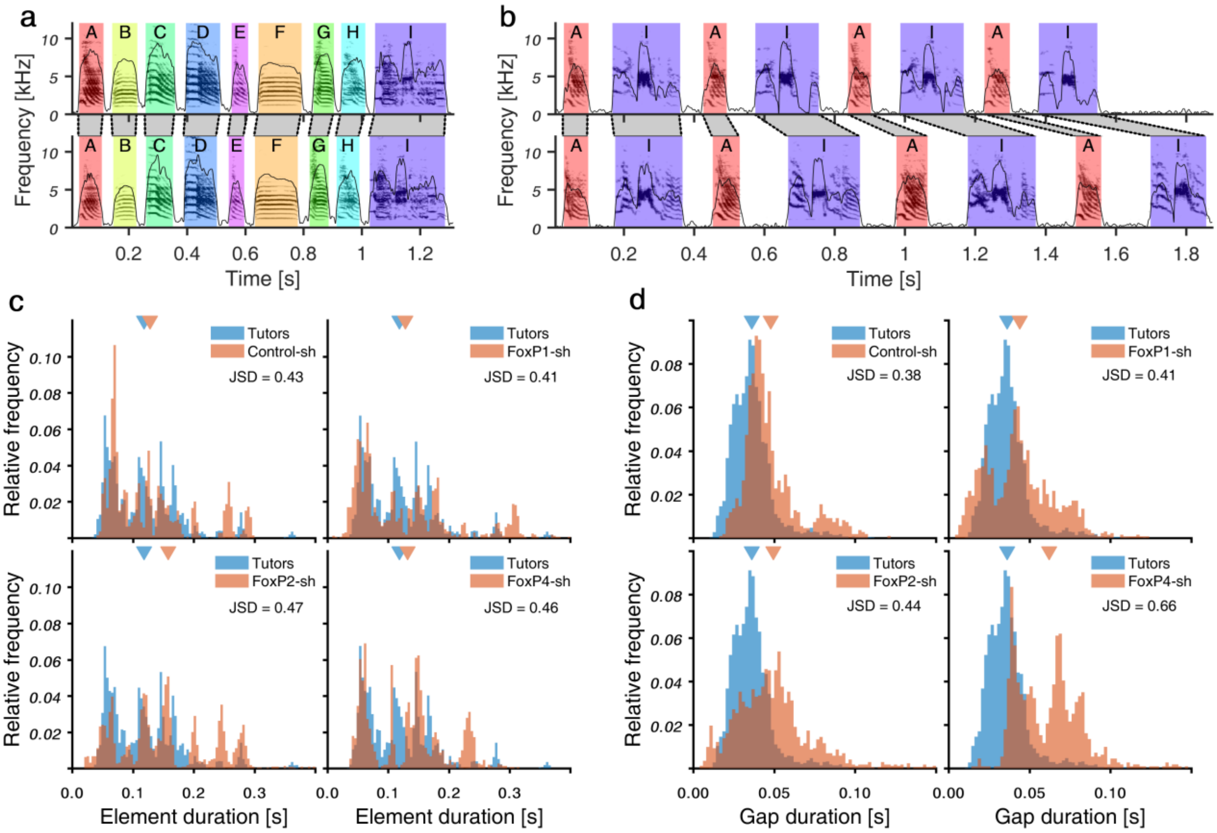


Figure 14 – The duration of song gaps is abnormally variable in FoxP knockdown birds. (a-b) Spectrograms of two example songs each for FoxP4-sh bird 3584 (b) and its tutor 2074 (a). Black dotted lines connect the song element on- and offsets of the two songs. The duration of gaps in the songs of the FoxP4 knockdown bird is abnormally variable. (c-d) Histograms of the durations of all song elements (c) and song gaps (d) of tutors (blue) as well as Control-sh, FoxP1-sh, FoxP2-sh and FoxP4-sh (red, left to right, top to bottom). Triangles on the top show the means. JSD = Jensen-Shannon distance between tutor and pupil distribution.

Segregation of the phenotypes of FoxP-sh birds

As we found the treatment groups differentially affected in various aspects of song learning, we wanted to find out if the four groups could be discriminated by their song phenotype alone. To that end we performed a linear discriminant analysis (LDA) using 5 features of song structure from different domains: Two spectral measures (the amount of frequency modulation of song elements and the tutor-to-pupil identity score difference), one measure of song learning success (number of copied song elements from tutor) and two temporal measures (CV of duration of the most variable inter-onset-interval and pulse deviation vs. model, see **Figure 13d**). Birds of the same group cluster together in the signal space, with very little overlap (**Figure 15**). Control-sh birds, FoxP2-sh and FoxP4-sh are well separated. FoxP1-sh is closest in space to the control birds, consistent with song deficits occurring in the fewest number of measures (e.g. not significantly affected in identity and stereotypy scores, copied notes, and frequency modulation). To test discrimination by these features, we applied leave-one-out cross-validation. Following this procedure, 54% of the birds were correctly classified as belonging to their respective treatment group. Classification rate as expected by chance was 25%, as there are four possible classes.

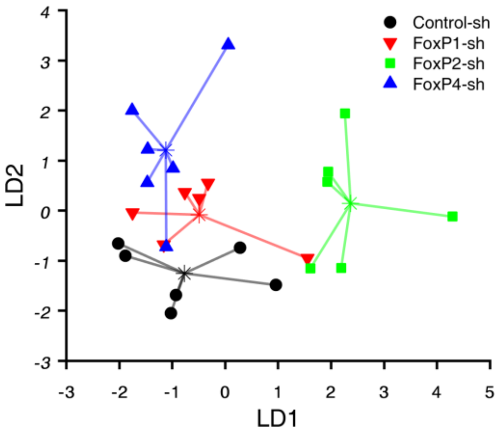


Figure 15 – The different treatment groups cluster together in the signal space of a linear discriminant function analysis, indicating that they can be discriminated by their song phenotype above chance. Each dot represents one bird in the signal space of the first two linear discriminant functions (LD1 & LD2, arbitrary units). Asterisks mark group centroids and lines connect each animal to the centroid of its group.

To summarize, FoxP1/2/4 knockdown in Area X affects song learning. Reduced motif similarity (**Figure 8a**), scrambled order of song elements (**Figure 7**) and a smaller fraction of elements copied (**Figure 10**) is a common phenotype of all FoxP knockdown pupils. FoxP1 knockdown seems to result in the mildest impairment of all FoxPs. Although FoxP1 knockdown birds do not copy all the elements of their tutors, the material that they do copy has high spectro-tempral fidelity. FoxP2 and FoxP4 knockdown resulted in a more severe phenotype than FoxP1, affecting most of the features studied. FoxP2-sh was more severely affected in motif accuracy (**Figure 8b**), fraction of copied notes (**Figure 10**) and temporal regularity (**Figure 13d**), while FoxP4 was most affected in frequency modulation (**Figure 9**). Although there is some overlap between the knockdown groups, taken together each group has a specific combination of impairments that makes most members of the group more similar to each other than to the other groups (**Figure 15**).

Discussion

We previously showed that many medium spiny neurons of zebra finches co-express FoxP1, FoxP2, and FoxP4 in Area X, a region important for vocal learning (Mendoza et al., 2015). Furthermore, FoxP1/2/4 can dimerize and oligomerize with each other in those neurons and can share the same binding sites and target genes (Mendoza and Scharff, 2017). In the present study, we addressed whether and how the knockdown of FoxP1 or of FoxP4 in Area X of juvenile male zebra finches affects song development. To do so, we used the lentivirally mediated siRNA methodology previously employed to precisely reduce FoxP2 spatially and temporally, which resulted in incomplete and inaccurate song development (Haesler et al., 2007; Murugan et al., 2013). Of note, homozygous deletions of FoxP1 (Li et al., 2004b) and FoxP4 (Wang et al., 2004) in mice are embryonic lethal, so that postnatal manipulations are essential to test for post-developmental effects.

A limitation of the local knockdown technology is the variability in the targeted area reached and the efficacy of knockdown. The maximum volume reached was 80% of Area X but on average, the knockdown of FoxP1 was 36 % and FoxP4 of 29% was sufficient to cause learning deficits. Importantly, Area X expands considerably in both size and cell number between the injection at day 23 and analysis at day 90 or later, so that the fraction of Area X infected during the song learning period was likely larger than that measured at 90 days (Nordeen and Nordeen, 1988; Haesler et al., 2007). These results are in line with a previous study on virally injected rats, in which blocking neural plasticity in 10-20% of lateral amygdala neurons was sufficient to impair memory formation (Rumpel et al., 2005). In zebra finches the same approach for manipulating FoxP2 expression resulted in song learning deficits (Haesler et al., 2007), reduction of spine density (Schulz et al., 2010), and affected the speed of signal propagation through the cortico-striatal pathway (Murugan et al., 2013) and song performance (Heston and White, 2015).

The protein reduction of FoxP1 and FoxP4 in cell culture was evident (**Figure 3a,f**), but the down-regulation in the brain (**Figure 5a,c**) was not as strong as the one reported previously for FoxP2 with the same virus (Haesler et al., 2007). A reason for that could be that many more neurons express FoxP1 or FoxP4 than FoxP2 (Mendoza et al., 2015), so that the fraction of neurons with reduced FoxP1 or FoxP4 protein levels compared to the entire population of neurons that express FoxP1 of FoxP2 is lower (**Figure 4**). In addition, the majority of FoxP2 neurons express low levels of FoxP2 (Thompson et al., 2013), whereas FoxP1 and FoxP4 neurons express these proteins at higher levels (Mendoza et al., 2015). Therefore experimental reduction of FoxP2 is more readily achieved than reduction of FoxP1 or FoxP4.

Regardless of the slightly less efficient reduction FoxP1 or FoxP4 expression compared to FoxP2, experimental lowering of all three neutrally expressed FoxP proteins in Area X affects song development significantly so, but with interesting differences and also differently than lesioning Area X (Scharff and Nottebohm, 1991). In contrast to the present results, adult male song after juvenile Area X lesions contains unusual long elements and reduced stereotypy, ruling out that our approach is damaging Area X neurons in a non-specific way.

The cortico-basal ganglia circuits promotes learning of action sequences through trial-and-error learning and that basal ganglia drive the variability necessary for this reinforcement-based learning. This learning could be driven by the reward-related dopamine signaling that projects to the basal ganglia from the VTA and SNpc (Graybiel, 2005). In the striatum of the zebra finch, there is co-expression of D1A, D1B and D2 receptors and FoxPs (Kubikova et al., 2009; Mendoza et al., 2015). A down-regulation of FoxP2 in Area X affected dopamine receptor and DARPP-32 expression (Murugan et al., 2013) that might affect the dopaminergic reinforcement signals in the medium spiny neurons. Thus, the regulation of the FoxP subfamily members during times of vocal plasticity could be functionally related to dopamine signaling. When FoxP2 was manipulated to resemble the human FoxP2 in a mouse, a decrease in dopamine levels was reported (Enard et al., 2009; Enard, 2011), further suggesting a link between FoxP2 and dopamine. That FoxP2 is involved in plasticity of neurons was shown in the zebra finch using the same FoxP2 short hairpins. In this work, it was reported that neurons expressing the FoxP2-sh virus had fewer spine density (Schulz et al., 2010). Furthermore, target genes regulated by FoxP2 affect neurite outgrowth, synaptic plasticity and axon guidance (Spiteri et al., 2007; Vernes et al., 2007; Vernes et al., 2011). In addition, results of mouse FoxP2 manipulations support this showing alterations in dendrite length and synaptic plasticity (Groszer et al., 2008; Enard et al., 2009; Reimers-Kipping et al., 2010; French et al., 2011). Foxp1 and Foxp2 manipulations in mouse also resulted in abnormal vocalizations (Shu et al., 2005; Fujita et al., 2008; Gaub et al., 2010; Fischer and Hammerschmidt, 2011; Gaub et al., 2015). Genetic manipulations in Area X were also reported for FoxP2 (down- (Haesler et al., 2007) and up-regulation (Heston and White, 2015)) and mir-9 (Shi et al., 2018). This microRNA down-regulates FoxP1 and FoxP2 in zebra finches.

All Manipulations of FoxP2 in songbirds lead to song impairments and low motif similarity. All of them report syllable omissions and adding elements not found in the tutor (improvisations in (Heston and White, 2015)). Syntax similarity, a measure similar to stereotypy, was reported to be normal when FoxP2 is overexpressed (Heston and White, 2015) and also affected by mir-9 downregulation (Shi et al., 2018) which coincides with our results of FoxP2 downregulation. The frequency modulation of the fundamental (FM) is abnormal after FoxP2 over-expression (Heston and White, 2015), but when we down-regulated FoxP2, or

FoxP1 or after mir-9 induced FoxP2 down-regulation (Shi et al., 2018), FM was not different from the tutor, whereas downregulation of FoxP4 affected FM.

Human phenotypes related to FOXP1 and FOXP2 mutations support a role for these transcription factors in vocal learning (Bacon and Rappold, 2012). FOXP4 mutations lead to developmental delay (Charng et al., 2016) but whether vocal learning was affected was not reported. FOXP4 is expressed more widely and homogeneously in the brain than FOXP1 and FOXP2 are, therefore a mutation might affect more brain functions than FOXP1 or FOXP2. In our experiments we down-regulated FoxPs in one area important for vocal learning, without the possible effects of the down-regulation in other brain regions that may contribute to a more severe phenotype and without developmental effects.

Possible explanations for the fact that all FoxP downregulations in Area X impacted song learning might be that heterodimers of the FoxP subfamily are important for regulating pathways important for vocal learning, so the absence or mutation of one affects the whole machinery affecting song learning. This would suggest that either heterodimers have a different binding site, which is not known, or bind to a specific co-factor that is needed for controlling target genes important for vocal learning, which is also not known. FoxP subfamily members regulate different target genes, that could all be needed for vocal learning, and the absence of any (or a set) of these targets could affect vocal learning in specific ways. FoxP members can bind to, and regulate the same genes, and therefore affect song learning. This is supported by the fact that all FoxP subfamily proteins regulate the SV40 and *VLDLR* promoter, but it was also shown that they do not always bind to the same site (Sin et al., 2015; Mendoza and Scharff, 2017). The Area X equilibrium is likely to be affected no matter which gene is down-regulated and thereby song learning is affected. On a cautionary note, we cannot rule out that the observed effects re due to the induction of the same off-target effect, since we used the same virus and short hairpins against the same conserved subfamily members However, this is not probable because: i) we used different short hairpins for each FoxP subfamily members; ii) we demonstrated that down-regulation a possible cross-reaction to the closest homologs, and short hairpins are specific even if compared to the same subfamily; iii) we used small short hairpins proofed not to be toxic or induce other side effects; iv) not all gene down-regulations in Area X lead to impaired song (unpublished data); v) the specific song impairments differ after downregulation of FoxP1 from those due to downregulation of FoxP2 and those due to downregulation of FoxP4.

Together, our data suggest that the neutrally expressed proteins of the FoxP subfamily, FoxP1/2/4, act in the basal ganglia in concert and regulate pathways important in song learning in the zebra finch. Thus, all three FoxPs are needed for the proper regulation of their target genes and in turn, behavior.

References

Adam I, Scharff C, Honarmand M (2014) Who is who? Non-invasive methods to individually sex and mark altricial chicks. J Vis Exp.

Adam I, Mendoza E, Kobalz U, Wohlgemuth S, Scharff C (2016) FoxP2 directly regulates the reelin receptor VLDLR developmentally and by singing. Mol Cell Neurosci 74:96-105.

Adam I, Mendoza E, Kobalz U, Wohlgemuth S, Scharff C (2017) CNTNAP2 is a direct FoxP2 target in vitro and in vivo in zebra finches: complex regulation by age and activity. Genes Brain Behav 16:635-642.

Aronov D, Andalman AS, Fee MS (2008) A specialized forebrain circuit for vocal babbling in the juvenile songbird. Science 320:630-634.

Bacon C, Rappold GA (2012) The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. Hum Genet.

Bolhuis JJ, Okanoya K, Scharff C (2010) Twitter evolution: converging mechanisms in birdsong and human speech. Nat Rev Neurosci 11:747-759.

Bonkowsky JL, Chien CB (2005) Molecular cloning and developmental expression of foxP2 in zebrafish. Dev Dyn 234:740-746.

Bowers JM, Konopka G (2012) The role of the FOXP family of transcription factors in ASD. Dis Markers 33:251-260.

Castellucci GA, McGinley MJ, McCormick DA (2016) Knockout of Foxp2 disrupts vocal development in mice. Sci Rep 6:23305.

Chabout J, Sarkar A, Patel SR, Radden T, Dunson DB, Fisher SE, Jarvis ED (2016) A Foxp2 Mutation Implicated in Human Speech Deficits Alters Sequencing of Ultrasonic Vocalizations in Adult Male Mice. Front Behav Neurosci 10:197.

Chae WJ, Henegariu O, Lee SK, Bothwell AL (2006) The mutant leucine-zipper domain impairs both dimerization and suppressive function of Foxp3 in T cells. Proc Natl Acad Sci U S A 103:9631-9636.

Charng WL et al. (2016) Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. BMC Med Genomics 9:42.

Chi Z, Margoliash D (2001) Temporal precision and temporal drift in brain and behavior of zebra finch song. Neuron 32:899-910.

Chokas AL, Trivedi CM, Lu MM, Tucker PW, Li S, Epstein JA, Morrisey EE (2010) Foxp1/2/4-NuRD interactions regulate gene expression and epithelial injury response in the lung via regulation of IL-6. J Biol Chem.

Delic S, Streif S, Deussing JM, Weber P, Ueffing M, Holter SM, Wurst W, Kuhn R (2008) Genetic mouse models for behavioral analysis through transgenic RNAi technology. Genes Brain Behav 7:821-830.

Di Benedetto B, Wefers B, Wurst W, Kuhn R (2009) Local knockdown of ERK2 in the adult mouse brain via adeno-associated virus-mediated RNA interference. Mol Biotechnol 41:263-269.

Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. Annu Rev Neurosci 22:567-631.

Enard W (2011) FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. Curr Opin Neurobiol 21:415-424.

Enard W et al. (2009) A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell 137:961-971.

Endres DM, Schindelin JE (2003) A new metric for probability distributions. Ieee T Inform Theory 49:1858-1860.

Estruch SB, Graham SA, Quevedo M, Vino A, Dekkers DHW, Deriziotis P, Sollis E, Demmers J, Poot RA, Fisher SE (2018) Proteomic analysis of FOXP proteins reveals interactions between cortical transcription factors associated with neurodevelopmental disorders. Hum Mol Genet.

Ferland RJ, Cherry TJ, Preware PO, Morrisey EE, Walsh CA (2003) Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. J Comp Neurol 460:266-279.

Fischer J, Hammerschmidt K (2011) Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: insights into the evolution of vocal communication. Genes Brain Behav 10:17-27.

Fisher SE, Scharff C (2009) FOXP2 as a molecular window into speech and language. Trends Genet 25:166-177.

French CA, Jin X, Campbell TG, Gerfen E, Groszer M, Fisher SE, Costa RM (2011) An aetiological Foxp2 mutation causes aberrant striatal activity and alters plasticity during skill learning. Mol Psychiatry.

French CA, Vinueza Veloz MF, Zhou K, Peter S, Fisher SE, Costa RM, De Zeeuw CI (2018) Differential effects of Foxp2 disruption in distinct motor circuits. Mol Psychiatry.

Fujita E, Tanabe Y, Shiota A, Ueda M, Suwa K, Momoi MY, Momoi T (2008) Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. Proc Natl Acad Sci U S A 105:3117-3122.

Garza JC, Kim CS, Liu J, Zhang W, Lu XY (2008) Adeno-associated virus-mediated knockdown of melanocortin-4 receptor in the paraventricular nucleus of the hypothalamus promotes high-fat diet-induced hyperphagia and obesity. J Endocrinol 197:471-482.

Gaub S, Fisher SE, Ehret G (2015) Ultrasonic vocalizations of adult male Foxp2-mutant mice: behavioral contexts of arousal and emotion. Genes Brain Behav.

Gaub S, Groszer M, Fisher SE, Ehret G (2010) The structure of innate vocalizations in Foxp2 deficient mouse pups. Genes Brain Behav.

Glaze CM, Troyer TW (2006) Temporal structure in zebra finch song: implications for motor coding. J Neurosci 26:991-1005.

Glaze CM, Troyer TW (2013) Development of temporal structure in zebra finch song. Journal of Neurophysiology 109:1025-1035.

Graybiel AM (2005) The basal ganglia: learning new tricks and loving it. Curr Opin Neurobiol 15:638-644.

Groszer M et al. (2008) Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. Curr Biol 18:354-362.

Haesler S, Rochefort C, Georgi B, Licznerski P, Osten P, Scharff C (2007) Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. PLoS Biol 5:e321.

Haesler S, Wada K, Nshdejan A, Morrisey EE, Lints T, Jarvis ED, Scharff C (2004) FoxP2 expression in avian vocal learners and non-learners. J Neurosci 24:3164-3175.

Heston JB, White SA (2015) Behavior-Linked FoxP2 Regulation Enables Zebra Finch Vocal Learning. J Neurosci 35:2885-2894.

Karnovsky MJ, Roots L (1964) A "Direct-Coloring" Thiocholine Method for Cholinesterases. J Histochem Cytochem 12:219-221.

Kubikova L, Wada K, Jarvis ED (2009) Dopamine receptors in a songbird brain. J Comp Neurol 518:741-769.

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519-523.

Li B, Samanta A, Song X, Iacono KT, Brennan P, Chatila TA, Roncador G, Banham AH, Riley JL, Wang Q, Shen Y, Saouaf SJ, Greene MI (2007) FOXP3 is a homo-oligomer and a component of a supramolecular regulatory complex disabled in the human XLAAD/IPEX autoimmune disease. Int Immunol 19:825-835.

Li S, Joel Weidenfeld, Morrisey EE (2004a) Transcriptional and DNA Binding Activity of the Foxp1/2/4 Family Is Modulated by Heterotypic and Homotypic Protein Interactions. MOLECULAR AND CELLULAR BIOLOGY, Vol. 24:809-822.

Li S, Zhou D, Lu MM, Morrisey EE (2004b) Advanced cardiac morphogenesis does not require heart tube fusion. Science 305:1619-1622.

Liegeois F, Baldeweg T, Connelly A, Gadian DG, Mishkin M, Vargha-Khadem F (2003) Language fMRI abnormalities associated with FOXP2 gene mutation. Nat Neurosci 6:1230-1237.

Lin JH (1991) Divergence Measures Based on the Shannon Entropy. Ieee T Inform Theory 37:145-151.

Mashiko H, Yoshida AC, Kikuchi SS, Niimi K, Takahashi E, Aruga J, Okano H, Shimogori T (2012) Comparative anatomy of marmoset and mouse cortex from genomic expression. J Neurosci 32:5039-5053.

Mendoza E, Scharff C (2017) Protein-Protein Interaction Among the FoxP Family Members and their Regulation of Two Target Genes, VLDLR and CNTNAP2 in the Zebra Finch Song System. Front Mol Neurosci 10:112.

Mendoza E, Tokarev K, During DN, Retamosa EC, Weiss M, Arpenik N, Scharff C (2015) Differential coexpression of FoxP1, FoxP2, and FoxP4 in the Zebra Finch (Taeniopygia guttata) song system. J Comp Neurol 523:1318-1340.

Miller JE, Spiteri E, Condro MC, Dosumu-Johnson RT, Geschwind DH, White SA (2008) Birdsong decreases protein levels of FoxP2, a molecule required for human speech. J Neurophysiol 100:2015-2025.

Mizutani A, Matsuzaki A, Momoi MY, Fujita E, Tanabe Y, Momoi T (2007) Intracellular distribution of a speech/language disorder associated FOXP2 mutant. Biochem Biophys Res Commun 353:869-874.

Morgan AT, Webster R (2018) Aetiology of childhood apraxia of speech: A clinical practice update for paediatricians. J Paediatr Child Health 54:1090-1095.

Murugan M, Harward S, Scharff C, Mooney R (2013) Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. Neuron 80:1464-1476.

Nordeen EJ, Nordeen KW (1988) Sex and regional differences in the incorporation of neurons born during song learning in zebra finches. J Neurosci 8:2869-2874.

Norton P, Scharff C (2016) "Bird Song Metronomics": Isochronous Organization of Zebra Finch Song Rhythm. Front Neurosci 10:309.

Olias P, Adam I, Meyer A, Scharff C, Gruber AD (2014) Reference Genes for Quantitative Gene Expression Studies in Multiple Avian Species. PLoS One 9:e99678.

Petkov CI, Jarvis ED (2012) Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. Front Evol Neurosci 4:12.

Pfenning AR et al. (2014) Convergent transcriptional specializations in the brains of humans and song-learning birds. Science 346:1256846.

Reimers-Kipping S, Hevers W, Paabo S, Enard W (2010) Humanized Foxp2 specifically affects cortico-basal ganglia circuits. Neuroscience 175:75-84.

Rumpel S, LeDoux J, Zador A, Malinow R (2005) Postsynaptic receptor trafficking underlying a form of associative learning. Science 308:83-88.

Salahpour A, Medvedev IO, Beaulieu JM, Gainetdinov RR, Caron MG (2007) Local knockdown of genes in the brain using small interfering RNA: a phenotypic comparison with knockout animals. Biol Psychiatry 61:65-69.

Sasahara K, Tchernichovski O, Takahasi M, Suzuki K, Okanoya K (2015) A rhythm landscape approach to the developmental dynamics of birdsong. Journal of the Royal Society Interface 12.

Scharff C, 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. J Neurosci 11:2896-2913.

Scharff C, Nottebohm F, Cynx J (1998) Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway. J Neurobiol 36:81-90.

Schulz SB, Haesler S, Scharff C, Rochefort C (2010) Knockdown of FoxP2 alters spine density in Area X of the zebra finch. Genes Brain Behav 9:732-740.

Shi Z, Piccus Z, Zhang X, Yang H, Jarrell H, Ding Y, Teng Z, Tchernichovski O, Li X (2018) miR-9 regulates basal ganglia-dependent developmental vocal learning and adult vocal performance in songbirds. Elife 7.

Shu W, Yang H, Zhang L, Lu MM, Morrisey EE (2001) Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J Biol Chem 276:27488-27497.

Shu W, Cho JY, Jiang Y, Zhang M, Weisz D, Elder GA, Schmeidler J, De Gasperi R, Sosa MA, Rabidou D, Santucci AC, Perl D, Morrisey E, Buxbaum JD (2005) Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. Proc Natl Acad Sci U S A 102:9643-9648.

Sin C, Li H, Crawford DA (2015) Transcriptional regulation by FOXP1, FOXP2, and FOXP4 dimerization. J Mol Neurosci 55:437-448.

Siper PM, De Rubeis S, Trelles MDP, Durkin A, Di Marino D, Muratet F, Frank Y, Lozano R, Eichler EE, Kelly M, Beighley J, Gerdts J, Wallace AS, Mefford HC, Bernier RA, Kolevzon A, Buxbaum JD (2017) Prospective investigation of FOXP1 syndrome. Mol Autism 8:57.

Sohrabji F, Nordeen EJ, Nordeen KW (1990) Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch. Behav Neural Biol 53:51-63.

Sollis E, Graham SA, Vino A, Froehlich H, Vreeburg M, Dimitropoulou D, Gilissen C, Pfundt R, Rappold GA, Brunner HG, Deriziotis P, Fisher SE (2015) Identification and functional characterization of de novo FOXP1 variants provides novel insights into the etiology of neurodevelopmental disorder. Hum Mol Genet.

Sollis E, Deriziotis P, Saitsu H, Miyake N, Matsumoto N, Hoffer MJV, Ruivenkamp CAL, Alders M, Okamoto N, Bijlsma EK, Plomp AS, Fisher SE (2017) Equivalent missense variant in the FOXP2 and FOXP1 transcription factors causes distinct neurodevelopmental disorders. Hum Mutat.

Song X, Li B, Xiao Y, Chen C, Wang Q, Liu Y, Berezov A, Xu C, Gao Y, Li Z, Wu SL, Cai Z, Zhang H, Karger BL, Hancock WW, Wells AD, Zhou Z, Greene MI (2012) Structural and biological features of FOXP3 dimerization relevant to regulatory T cell function. Cell Rep 1:665-675.

Spaeth JM, Hunter CS, Bonatakis L, Guo M, French CA, Slack I, Hara M, Fisher SE, Ferrer J, Morrisey EE, Stanger BZ, Stein R (2015) The FOXP1, FOXP2 and FOXP4 transcription factors are required for islet alpha cell proliferation and function in mice. Diabetologia 58:1836-1844.

Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, Ou J, Vernes SC, Fisher SE, Ren B, Geschwind DH (2007) Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. Am J Hum Genet 81:1144-1157.

Takahashi H, Takahashi K, Liu FC (2009) FOXP genes, neural development, speech and language disorders. Adv Exp Med Biol 665:117-129.

Takahashi K, Liu FC, Hirokawa K, Takahashi H (2003) Expression of Foxp2, a gene involved in speech and language, in the developing and adult striatum. J Neurosci Res 73:61-72.

Takahashi K, Liu FC, Hirokawa K, Takahashi H (2008a) Expression of Foxp4 in the developing and adult rat forebrain. J Neurosci Res 86:3106-3116.

Takahashi K, Liu FC, Oishi T, Mori T, Higo N, Hayashi M, Hirokawa K, Takahashi H (2008b) Expression of FOXP2 in the developing monkey forebrain: comparison with the expression of the genes FOXP1, PBX3, and MEIS2. J Comp Neurol 509:180-189.

Tchernichovski O, Nottebohm F (1998) Social inhibition of song imitation among sibling male zebra finches. Proc Natl Acad Sci U S A 95:8951-8956.

Tchernichovski O, Lints T, Mitra PP, Nottebohm F (1999) Vocal imitation in zebra finches is inversely related to model abundance. Proc Natl Acad Sci U S A 96:12901-12904.

Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP (2000) A procedure for an automated measurement of song similarity. Anim Behav 59:1167-1176.

Teramitsu I, Poopatanapong A, Torrisi S, White SA (2010) Striatal FoxP2 is actively regulated during songbird sensorimotor learning. PLoS One 5:e8548.

Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA (2004) Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. J Neurosci 24:3152-3163.

Thompson CK, Schwabe F, Schoof A, Mendoza E, Gampe J, Rochefort C, Scharff C (2013) Young and intense: FoxP2 immunoreactivity in Area X varies with age, song stereotypy, and singing in male zebra finches. Front Neural Circuits 7:24.

Tramontin AD, Smith GT, Breuner CW, Brenowitz EA (1998) Seasonal plasticity and sexual dimorphism in the avian song control system: stereological measurement of neuron density and number. J Comp Neurol 396:186-192.

Vernes SC, Spiteri E, Nicod J, Groszer M, Taylor JM, Davies KE, Geschwind DH, Fisher SE (2007) High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. Am J Hum Genet 81:1232-1250.

Vernes SC, Nicod J, Elahi FM, Coventry JA, Kenny N, Coupe AM, Bird LE, Davies KE, Fisher SE (2006) Functional genetic analysis of mutations implicated in a human speech and language disorder. Hum Mol Genet 15:3154-3167.

Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, Ho J, Mombereau C, Brewer A, Lowy E, Nicod J, Groszer M, Baban D, Sahgal N, Cazier JB, Ragoussis J, Davies KE, Geschwind DH, Fisher SE (2011) Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet 7:e1002145.

Wada K, Howard JT, McConnell P, Whitney O, Lints T, Rivas MV, Horita H, Patterson MA, White SA, Scharff C, Haesler S, Zhao S, Sakaguchi H, Hagiwara M, Shiraki T, Hirozane-Kishikawa T, Skene P, Hayashizaki Y, Carninci P, Jarvis ED (2006) A molecular neuroethological approach for identifying and characterizing a cascade of behaviorally regulated genes. Proc Natl Acad Sci U S A 103:15212-15217.

Wang B, Weidenfeld J, Lu MM, Maika S, Kuziel WA, Morrisey EE, Tucker PW (2004) Foxp1 regulates cardiac outflow tract, endocardial cushion morphogenesis and myocyte proliferation and maturation. Development 131:4477-4487.

Watkins KE, Vargha-Khadem F, Ashburner J, Passingham RE, Connelly A, Friston KJ, Frackowiak RS, Mishkin M, Gadian DG (2002) MRI analysis of an inherited speech and language disorder: structural brain abnormalities. Brain 125:465-478.

Williams H (2004) Birdsong and singing behavior. Ann N Y Acad Sci 1016:1-30.

Xu S et al. (2018) Foxp2 regulates anatomical features that may be relevant for vocal behaviors and bipedal locomotion. Proc Natl Acad Sci U S A 115:8799-8804.

General Discussion

This thesis presents the first evidence for an isochronous rhythmic structure in the learned vocalizations of two distantly related species: the zebra finch, *Taeniopygia guttata*, and the greater sac-winged bat, *Saccopteryx bilineata*. The differential influence of FoxP1, FoxP2 and FoxP4 during the zebra finch's learning phase on this structure and on the other temporal and spectral song features is reported. Additionally it provides a broad overview of analytical methods for quantifying rhythmic regularity and complexity in vocalizations, movements and other behaviors unfolding in time.

One of the many questions that was on my mind since the first discovery of the isochronous rhythmic structure in zebra finch song was: To what degree is this regularity 'hard-wired' (i.e. an emergent property of the way song is neurally coded), and to what degree is it more maturation dependent, shaped by an interplay of innate and acquired perceptional dispositions, potentially guided by feedback from other conspecifics? While I could not answer this question with my thesis, I will discuss our findings in context of the current state of knowledge about the neural and behavioral mechanisms of song and its development, and I will speculate about potential answers within this context.

One finding from the study reported in Publication D was particularly unexpected. Zebra finches in the control group produced songs with a markedly reduced isochronous rhythmic structure compared to their tutors. FoxP levels were not experimentally reduced in these birds. Other than that they received the same treatment as birds in the FoxP1/2/4 knockdown groups. There are several possible interpretations for this result. Arguably the most promising, as briefly discussed in the article, is that their song rhythm might not have been fully developed at the time of recording, between 90 and 100 PHD (age in post-hatch days). Their tutors, as well as all of the birds analyzed in Publication B were older when they were recorded, most having been well over a year in age. The shape of the overall durational distribution of gaps in the control birds was very similar to that of their tutors, but gaps were longer on average. The isochronous rhythmic structure of zebra finch song might therefore only emerge after the gradual reduction and increasing stereotypy of the between-note gaps that takes place in zebra finches after song crystallization (Glaze and Troyer, 2013).

Unlike gap durations, notes stabilize with crystallization in both absolute duration and variability (Glaze and Troyer, 2013). A similar development has been reported in Bengalese finches (*Lonchura striata domestica*; James and Sakata, 2014) and Java sparrows (*Lonchura oryzivora*; Ota and Soma, 2014), hinting at a more general pattern in Estrildid finches. The spectral structure of notes is actively maintained in adulthood through a process that requires auditory feedback (Leonardo and Konishi, 1999; Lombardino and Nottebohm, 2000; Nordeen and Nordeen, 1992; Vallentin et al., 2016) and likewise remains highly stereotyped (Tchernichovski et al., 2001). The fact that note structure, both spectral and temporal, consolidates with crystallization indicates that the rhythmic structure on the level of gestures is established for individual notes around the time of sexual maturity (~90 PHD).

At this time the neural code of song in HVC appears to consolidate as well. In adult male zebra finches, individual HVC neurons projecting to downstream motor areas fire short (roughly 10 ms long) bursts of action potentials at precise time points during different song renditions (Hahnloser et al., 2002). This temporal precision gradually emerges during song development, as an increasing fraction of neurons fire single bursts time-locked relative to the onset of a specific note (Okubo et al., 2015). A subset of these successively bursting HVC neurons has been argued to form chains that act as an internal clock that underlies song timing ('chain model'; Fee et al., 2004; Li and Greenside, 2006; Long et al., 2010; Troyer, 2013; c.f. Amador et al., 2013; Boari et al., 2015; see also Publication B Discussion). While this clock was originally hypothesized to ‘tick’ throughout the motif, several observations suggest that multiple discrete chains of neurons are responsible for the temporal structure of different notes (Danish et al., 2017). Durational variability of gaps is higher than that of notes in normal adult song (Glaze and Troyer, 2006). When locally manipulating the temperature of HVC, song is performed more slowly, the more HVC is cooled down. Interestingly, notes stretch uniformly on all timescales in this situation, whereas stretching of inspiratory gaps is less pronounced and not uniform (Andalman et al., 2011; Long and Fee, 2008). Inspiration during singing thus seems to be influenced not exclusively by top-down ‘cortical’ control of the song system (Schmidt and Goller, 2016). Instead, nonlinear dynamics of the brain stem respiratory control system could interact with the song control system and contribute to the respiratory pattern of song (Hamaguchi et al., 2016; Schmidt and Ashmore, 2008). Taken together, these findings support the view that the isochronous structure on the level of individual notes is indeed in a sense ‘hard-wired’ in the adult zebra finch.

Although Glaze and Troyer (2013) reported that gaps shortened on average, different gaps did not shorten to the same degree and some even increased in duration. The adjustment of the gap duration might act as ‘tuning screws’ for the birds and enable to fine-tune the temporal regularity on the motif-level. Modification of the intervals connecting the now ‘hard-coded’ song notes could give rise to the pulse^S that we detected in older birds. Further studies are needed to determine whether gap modification in adulthood actually leads to song rhythm converging on an isochronous rhythmic structure. To that end, song development could be tracked over the months after crystallization. Given what we know about the pulse^S in adult birds, testable predictions can be made regarding the hypothesized endpoint of this development already at the time of crystallization. I often observed that integer ratios of the pulse^S period equaled the durations of the shortest gestures, around 10ms in duration. For an example see the repeated gesture (middle of note 1 and 4) and the single gesture (middle of note 2 and 5) in the bottom row of **Figure 3** in Publication B. These pulses around or above 100Hz, the equivalent of the tatum on the level of gestures (i.e. the pulse fitting all gesture onsets), are in the range of the tempo of the hypothesized clock in the chain model. A single 'tick' of the clock could therefore correspond to the shortest possible gesture, while other gestures may last two, three, or more ticks. Given that this gesture-level pulse is present at crystallization, a finite number of possibilities for the tempo of the final note-level pulse can be predicted (2x, 3x, 4x, … the gesture-level pulse; 6x in the example in Publication B, **Figure 3**). Some of these should be more likely, as most adult birds tested so far had pulses^S in the 25–45Hz range. Further constraints may arise from limits in the flexibility of gap duration modification. For example, if a proposed pulse would necessitate a gap to double in speed, it might be an unlikely candidate due to physiological limits. The comparison of predicted and observed gap modifications can then provide evidence for or against the hypothesis of a maturation dependent process guiding the emergence of the pulse^S and of a hierarchical rhythmic structure through an interaction of different 'cortical' and 'subcortical' neural systems. Repeating this process under different experimental conditions (e.g. isolation from adult males during different phases pre- and post-crystallization) may furthermore allow to entangle innate from acquired rhythmic predispositions.

Another intriguing opportunity to decipher neural coding of song rhythm and its development may lie in a phenomenon that has received relatively little experimental attention since its discovery. While most bird vocalizations are exclusively produced during expriations, some zebra finches also produce tonal notes during inspiration (Goller and Daley, 2001). During these phonations, inspiratory pressure is increased two-fold compared to other minibreaths in the same song, and respiratory and syringeal motor patterns apparently differ. They seem to develop from broadband inspiratory sounds into high-frequency whistles in juveniles (Veit et al., 2011). Interestingly, these can consist of multiple gestures (Goller and Daley, 2001). Examining the temporal development of inspiratory notes and the underlying neural pattern has the potential to further disentangle the contributions of song system and respiratory system to the development of the temporal isochronous structure in zebra finch song. For example, a continuing development of the internal structure of multi-gesture inspiratory notes past crystallization would suggest a different control mechanism compared to expiratory notes. If this further gesture modification then results in an overall hierarchically structured song regularity comparable to songs without inspiratory notes, this implies a process of rhythm development in adult life that includes a form of phonetic plasticity. Tape-tutored juveniles have been shown to readily learn these types of notes also produce them during inspiration (Goller and Daley, 2001). This allows for the creation of artificial song learning targets containing inspiratory notes with various spectro-temporal parameters.

The fast respiratory patterns during zebra finch song can cause significant hyperventilation, leading to almost complete apnea in some individuals (Franz and Goller, 2003). The observation that gaps gradually become shorter and less variable might be an outcome of the birds' increasing proficiency in producing the demanding minibreaths with continuing practice. A study showed that food restrictions in early life significantly reduce the number of notes sung per second (note rate) in zebra finches (Zann and Cash, 2008). Note rate may thus be an honest signal for condition during mate choice, and males could further increase this rate through practice. Support for this idea comes from the observation that the proportion of sound versus silence within song plays a positive role in female song preferences (Holveck and Riebel, 2007; Leadbeater et al., 2005).

In this sense, the song would 'mature' well past the onset of reproductive activity. Considering song in the context of mate selection, what could explain female preferences for older males? Zebra finches in the wild face a high mortality rate (Zann, 1996, p. 142-145). Increased age can be an indicator for disease resistance and other traits that contribute to survivability. Studied immune traits in the zebra finch either increase with age or increase in the first years, before decreasing again (Noreen et al., 2011). Additionally, the likelihood of previous pairings increases with age, bringing with it experiences in parental care. Females with previous breeding experience are more successful at rearing chicks from subsequent pairings, even in a captive colony with minimal foraging demands (Baran and Adkins-Regan, 2014). Since both male and female zebra finches invest in parental care, the male's previous experience is likely to increase the reproductive success of a pair as well. For several other bird species age is in fact positively related to quality in terms of survival and reproduction (Martin, 1995). In addition to its function in mate choice, zebra finch song also plays a role in pair bonding. Zebra finches often form life-long pairs, and males frequently direct their song performance at their partner (e.g. Ikebuchi and Okanoya, 2006). The attractiveness of song should therefore also play a role in established pairs, and females might even influence further developments of song through feedback.

Lack of feedback might provide an alternative explanation for the reduced isochronous rhythmic structure in the songs of the control birds in Publication D. In order to limit their auditory exposure to a single model and to reduce confounding variables, the juveniles were housed with only one adult male conspecific, their song tutor. Zebra finch song plays an important role in mate choice and females show specific preferences for certain song features (reviewed by Riebel, 2009). It is conceivable that they use these preferences in assisting the song learning process of their offspring, thereby increasing its chances of mating and thereby their own reproductive success.

As discussed in Publication B, little is known about rhythmic perception in zebra finches. In the years since this publication no new reports on this specific question have come to my attention. The question of whether zebra finches perceive a pulse^{P} in their conspecific songs (as opposed to in an artificial stimulus; ten Cate et al., 2016) remains open. Further research in this area has the potential to uncover shared and diverging principles in the relation of pulse and melodic content in animal vocalizations and human music. In my analyses I came across several songs with a relatively low pulse^{S} fit, that had the following properties. They contained a multi-gesture note with an initial short high-frequency gesture followed by a gesture of much lower frequency (see last note in Publication D, **Figure 14a** for an example). In most of these cases a much better fitting pulse^{S} could be found if I used the first gesture transition (i.e. the onset of the second, low-frequency geture) as the event for this note instead of the note onset. Although anecdotal, I mention this observation for two reasons. First, as a reminder that any attempts at relating the pulse^{S} as measured in this thesis and the pulse^{P} must be based on the assumption that note onsets are of particular (but not necessarily exclusive) salience for rhythm perception of the studied animal. The second point is also related to the question of salience in pulse perception, as well as to another rhythmic phenomenon observed in music. Low-pitched sounds (e.g. bass instruments and low voices) often carry the musical beat and provide a stronger support for the synchronization of movement to music (such as in dance) than high-pitched sounds (Burger, 2013; Burger et al., 2018; Large, 2000; Lerdahl and Jackendoff, 1996). There are first implications of underlying auditory and cortical neurophysiological mechanisms (Hove et al., 2014; Lenc et al., 2018), providing further opportunities for species comparative research.

We chose the zebra finch as a starting point for our search for the equivalent of a musical beat in learned animal vocalization due to the species' short and highly stereotyped song and for the wealth of knowledge about its behavior, neurobiology and genetics that accrued in the past decades. I validated the generate-and-test method for detection of an isochronous rhythmic structure and reported the existence of such a structure in zebra finch song. This paves the way for future comparative research into the development and production of this form of rhythmic regularity, as well as the biological substrates that facilitate it. I propose the European starling (*Sturnus vulgaris*) as one of many promising songbird species for further studies. Starlings have demonstrated the ability to discriminate isochronous and hierarchical temporal patterns of metronome-like pulses from heterochronous patterns (Hulse et al., 1984). They also generalized this learned discrimination across tempo changes to a degree (Hulse et al., 1984). This tentatively suggests an intriguing cognitive ability to abstract the concept underlying metrical rhythm.

Similarly *Saccopteryx*, one of the few bat species for which vocal learning has been demonstrated conclusively so far, has only been a first target of examination of isochronous structure in bat vocalizations. Only a fraction of the many bat species have been investigated for the capacity of vocal learning, and there are several promising families for future studies in this regard (Knörnschild, 2014). **Publication C** reported a comparable rhythmic tempo of wingbeat-coupled echolocation calls and non-coupled vocalizations in the greater sac-winged bat. This finding provides compelling incentive for further studies on the influence of diverse physiological constraints on rhythmic predispositions in different bat species and perhaps echolocating marine mammals.

Taken together, this thesis underlines the importance of a broad cross-species comparative approach to uncover how the different traits of musicality came together to form music as we know it today.

References

Amador, A., Perl, Y. S., Mindlin, G. B., and Margoliash, D. (2013). Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature* 495, 59–64. doi:10.1038/nature11967.

Andalman, A. S., Foerster, J. N., and Fee, M. S. (2011). Control of vocal and respiratory patterns in birdsong: dissection of forebrain and brainstem mechanisms using temperature. *PLoS One* 6, e25461. doi:10.1371/journal.pone.0025461.

Baran, N. M., and Adkins-Regan, E. (2014). Breeding Experience, Alternative Reproductive Strategies and Reproductive Success in a Captive Colony of Zebra Finches (Taeniopygia guttata). *PLoS One* 9, e89808. doi:10.1371/journal.pone.0089808.

Boari, S., Sanz Perl, Y., Amador, A., Margoliash, D., and Mindlin, G. B. (2015). Automatic reconstruction of physiological gestures used in a model of birdsong production. *J. Neurophysiol.* 5, 2912–2922. doi:10.1152/jn.00385.2015.

Burger, B. (2013). Relationships between spectral flux , perceived rhythmic strength , and the propensity to move. *Proc. Sound Music Comput. Conf.*, 179–184. doi:10.1002/(ISSN)1099-0542.

Burger, B., London, J., Thompson, M. R., and Toiviainen, P. (2018). Synchronization to metrical levels in music depends on low-frequency spectral components and tempo. *Psychol. Res.* 82, 1195–1211. doi:10.1007/s00426-017-0894-2.

Danish, H. H., Aronov, D., and Fee, M. S. (2017). Rhythmic syllable-related activity in a songbird motor thalamic nucleus necessary for learned vocalizations. *PLoS One* 12, 1–28. doi:10.1371/journal.pone.0169568.

Fee, M. S., Kozhevnikov, A. a, and Hahnloser, R. H. R. (2004). Neural mechanisms of vocal sequence generation in the songbird. *Ann. N. Y. Acad. Sci.* 1016, 153–170. doi:10.1196/annals.1298.022.

Franz, M., and Goller, F. (2003). Respiratory patterns and oxygen consumption in singing zebra finches. *J. Exp. Biol.* 206, 967–978. doi:10.1242/jeb.00196.

Glaze, C. M., and Troyer, T. W. (2006). Temporal structure in zebra finch song: Implications for motor coding. *J. Neurosci.* 26, 991–1005. doi:10.1523/JNEUROSCI.3387-05.2006.

Glaze, C. M., and Troyer, T. W. (2013). Development of temporal structure in zebra finch song. *J. Neurophysiol.* 109, 1025–1035. doi:10.1152/jn.00578.2012.

Goller, F., and Daley, M. A. (2001). Novel motor gestures for phonation during inspiration enhance the acoustic complexity of birdsong. *Proc. R. Soc. B Biol. Sci.* 268, 2301–2305. doi:10.1098/rspb.2001.1805.

Hahnloser, R. H. R., Kozhevnikov, A. A., and Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419, 65–70. doi:10.1038/nature00974.

Hamaguchi, K., Tanaka, M., and Mooney, R. (2016). A Distributed Recurrent Network Contributes to Temporally Precise Vocalizations. *Neuron* 91, 680–693. doi:10.1016/j.neuron.2016.06.019.

Holveck, M.-J., and Riebel, K. (2007). Preferred songs predict preferred males: Consistency and repeatability of zebra finch females across three test contexts. *Anim. Behav.* 74, 297–309. doi:10.1016/j.anbehav.2006.08.016.

Hove, M. J., Marie, C., Bruce, I. C., and Trainor, L. J. (2014). Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proc. Natl. Acad. Sci.* 111, 10383–10388. doi:10.1073/pnas.1402039111.

Hulse, S. H., Humpal, J., and Cynx, J. (1984). Discrimination and generalization of rhythmic and arrhythmic sound patterns by european starlings (Sturnus vulgaris). *Music Percept. An Interdiscip. J.* 1, 442–464. doi:10.2307/40285272.

Ikebuchi, M., and Okanoya, K. (2006). Growth of pair bonding in Zebra Finches: physical and social factors. *Ornithol. Sci.* 5, 65–75. doi:10.2326/osj.5.65.

James, L. S., and Sakata, J. T. (2014). Vocal motor changes beyond the sensitive period for song plasticity. *J. Neurophysiol.* 112, 2040–2052. doi:10.1152/jn.00217.2014.

Knörnschild, M. (2014). Vocal production learning in bats. *Curr. Opin. Neurobiol.* 28, 80–85. doi:10.1016/j.conb.2014.06.014.

Large, E. W. (2000). On synchronizing movements to music. *Hum. Mov. Sci.* 19, 527–566. doi:10.1016/S0167-9457(00)00026-9.

Leadbeater, E., Goller, F., and Riebel, K. (2005). Unusual phonation, covarying song characteristics and song preferences in female zebra finches. *Anim. Behav.* 70, 909–919. doi:10.1016/j.anbehav.2005.02.007.

Lenc, T., Keller, P. E., Varlet, M., and Nozaradan, S. (2018). Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proc. Natl. Acad. Sci.* 115, 8221–8226. doi:10.1073/pnas.1801421115.

Leonardo, A., and Konishi, M. (1999). Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* 399, 466–70. doi:10.1038/20933.

Lerdahl, F., and Jackendoff, R. (1996). *A Generative Theory of Tonal Music*. doi:10.1525/mts.1985.7.1.02a00120.

Li, M., and Greenside, H. (2006). Stable propagation of a burst through a one-dimensional homogeneous excitatory chain model of songbird nucleus HVC. *Phys. Rev. E* 74, 011918.

doi:10.1103/PhysRevE.74.011918.

Lombardino, A. J., and Nottebohm, F. (2000). Age at deafening affects the stability of learned song in adult male zebra finches. *J. Neurosci.* 20, 5054–5064. doi:10.1523/JNEUROSCI.20-13-05054.2000.

Long, M. A., and Fee, M. S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* 456, 189–194. doi:10.1038/nature07448.

Long, M. a, Jin, D. Z., and Fee, M. S. (2010). Support for a synaptic chain model of neuronal sequence generation. *Nature* 468, 394–399. doi:10.1038/nature09514.

Martin, K. (1995). Patterns and mechanisms for age-dependent reproduction and survival in birds. *Am. Zool.* 35, 340–348. doi:10.1093/icb/35.4.340.

Nordeen, K. W., and Nordeen, E. J. (1992). Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behav. Neural Biol.* 57, 58–66. doi:10.1016/0163-1047(92)90757-U.

Noreen, E., Bourgeon, S., and Bech, C. (2011). Growing old with the immune system: a study of immunosenescence in the zebra finch (Taeniopygia guttata). *J. Comp. Physiol. B* 181, 649–656. doi:10.1007/s00360-011-0553-7.

Okubo, T. S., Mackevicius, E. L., Payne, H. L., Lynch, G. F., and Fee, M. S. (2015). Growth and splitting of neural sequences in songbird vocal development. *Nature* 528, 352–357. doi:10.1038/nature15741.

Ota, N., and Soma, M. (2014). Age-dependent song changes in a closed-ended vocal learner: Elevation of song performance after song crystallization. *J. Avian Biol.* 45, 566–573. doi:10.1111/jav.00383.

Riebel, K. (2009). Song and female mate choice in zebra finches: A review. *Adv. Study Behav.* 40, 197–238. doi:10.1016/S0065-3454(09)40006-8.

Schmidt, M. F., and Ashmore, R. C. (2008). "Integrating breathing and singing: Forebrain and brainstem mechanisms," in *Neuroscience of Birdsong*, eds. H. Zeigler and P. Marler (New York, NY: Cambridge University Press), 115–135.

Schmidt, M. F., and Goller, F. (2016). Breathtaking songs: Coordinating the neural circuits for breathing and singing. *Physiology* 31, 442–451. doi:10.1152/physiol.00004.2016.

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. doi:10.1126/science.1058522.

ten Cate, C., Spierings, M., Hubert, J., and Honing, H. (2016). Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Front. Psychol.* 7, 1–14. doi:10.3389/fpsyg.2016.00730.

Troyer, T. W. (2013). The units of a song. *Nature* 495, 56–57. doi:10.1038/nature11957.

Vallentin, D., Kosche, G., Lipkind, D., and Long, M. A. (2016). Inhibition protects acquired song segments during vocal learning in zebra finches. *Science* 351, 267–271. doi:10.1126/science.aad3023.

Veit, L., Aronov, D., and Fee, M. S. (2011). Learning to breathe and sing: development of respiratory-vocal coordination in young songbirds. *J. Neurophysiol.* 106, 1747–1765. doi:10.1152/jn.00247.2011.

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

Zann, R., and Cash, E. (2008). Developmental stress impairs song complexity but not learning accuracy in non-domesticated zebra finches (Taeniopygia guttata). *Behav. Ecol. Sociobiol.* 62, 391–400. doi:10.1007/s00265-007-0467-2.

Erklärung über die eigenständige Verfassung der vorgelegten Dissertation

Hiermit versichere ich, dass die vorgelegte Dissertation gemäß §7 Abs. 4 der Promotionsordnung des Fachbereichs Biologie, Chemie, Pharmazie der Freien Universität Berlin vom 25. April 2018 eine selbstständig verfasste Forschungsleistung darstellt und ich keine anderen als die angegebenen Hilfsmittel verwendet habe. Die Arbeit hat weder in dieser noch in ähnlicher Form einem anderen Promotionsausschuss vorgelegen.

Ort, Datum

Philipp Norton