The role of vocal consistency in bird communication, a case study in the blue tit (*Cyanistes Caeruleus*)

Javier Sierro Martínez

This dissertation is submitted for the degree of Doctor of Philosophy

July 2021

Lancaster Environment Centre

Declaration

I declare that the work presented in this thesis is my own work and that it has not been submitted elsewhere for the award of any other degree.

Table of Contents

Acknowledgements

Nearly four years ago, I decided to come to Lancaster without really knowing where I was coming to. I met my first supervisor, Ian, through skype and it soon became clear that it would be stimulating (and fun) to work under his supervision. Not later I met my second supervisor, Selvino, and since the first day he has been engaging and devoted me as much time as a main supervisor. Through talking and arguing they brought the best of me and provided the academic and personal support to get all the way here, in the end of my PhD. For this I thank them and hope we will still meet for more discussions about the life of birds.

But, before arriving in Lancaster, I had to go through many more life stages that wouldn’t have been possible without my family. My mother, as a history teacher and passionate humanist always encouraged my curiosity and my need to seek answers. My father, as a professor in Palaeontology, gave me the perspective to understand and integrate the tiny details into a larger picture, learn about the small to understand the big patterns of nature. I can’t forget my siblings that have always showed me their support, always confident that I could get to where I am now.

But one person that has been there in every situation was Irene, my partner. She has supported me throughout the entire process. As most PhD students and other academic nerds, I have been nearly obsessed with my work most often than not. She has been supportive, building a confortable and loving place to go even if we were many kilometres apart.

I can’t forget all the people that were with me during most of the time at Lancaster. With Margherita, I spent many hours of conversations about humans, birds, political systems, social constructs or the holistic view of human kind. And of course I will always remember Friday drinks and hikes with Esti, Rob, Tom, Aimee, Katharina and other people. From the office, I have an especial place in my mind for Irene, a nature, social and music enthusiast that is also a beautiful person, talking to her was always fulfilling and motivating.

Abstract

Birdsong is a multidimensional acoustic signal used for communication during intra- and intersexual interactions. The anatomical and behavioural adaptations related to sophisticated phonation skills indicate there are directional selection pressures shaping phonation mechanisms in songbirds. The level of neuro-motor challenge displayed while singing is normally referred to as vocal performance of song. Theoretical and empirical work has shown that vocal performance is related to individual quality and is meaningful during intra- and intersexual communication. Here, I use the blue tit (*Cyanistes caeruleus*) as a model species to investigate the ecological and communicative implications of vocal performance in birdsong with particular emphasis on vocal consistency. Vocal consistency, as the ability to reproduce the same song element with minimal variation, has been proposed as an honest signal of individual quality in songbirds. In line with this hypothesis, I found that higher vocal consistency in males was preferred by females and was associated with increased reproductive success. I also provide evidence to validate the Spectral Cross-Correlation method used to measure vocal consistency, by showing it is sensitive to acoustic variation in natural birdsong. Although vocal consistency may be meaningful by itself, I built a multi-species database to investigate possible interactions of vocal consistency with other aspects of song. The results suggest there is a general trade-off between vocal performance within song and song length in birds. Following this finding, I conducted two playback experiments that showed that 1) long songs elicited a stronger response from territorial males than short songs but 2) that long songs with reduced vocal consistency triggered a lower response from territorial males, compared with short songs of high vocal consistency. These playback experiments indicated there is communicative value in the performance trade-off between song length and vocal consistency. Finally, I provide evidence that female song is frequent in blue tits and contextual behaviour data suggest it plays a role in intrasexual competition, similar to male song. However, during dawn chorus, males but not females were observed singing. These contextual differences indicate that female song may not be subjected to strong selection pressures related to seeking copulations. I hypothesize that such functional differences may relate to sexual differences in song parameters as females sang with lower song diversity and lower vocal performance than male.

List of Figures

List of Tables

Chapter 1 – Introduction

Birdsong

Bird vocalisations are ubiquitous and can be heard in nearly all natural soundscapes. Functionally, birdsong refers to those vocal signals associated with reproductive activities such as finding a mate or securing breeding resources (Collins, 2004). From a developmental point of view, ontogeny of birdsong includes a learning phase, during which, young birds must hear, interact and practice their song to acquire the species-specific features (Catchpole and Slater, 2008). These characteristics of birdsong make it an ideal model system to study many aspects of evolutionary ecology, especially the impact of sexual selection on behaviour and vocal communication. They also allow for comparative studies on evolution of anatomy and physiology of vocal learning and sound production (Berwick and Chomsky, 2013; Marler and Slabbekoorn, 2004).

One of the most striking features of birdsong is the wide variety of sounds produced even within one individual, from single sound elements to whole syntactic structures (Catchpole and Slater, 2008). For this reason, song complexity was soon compared to the striking shapes and colours of male ornaments found in many species, suggesting it had a function in mate attraction (Darwin 1872) . In some species, larger repertoires of songs are indeed preferred by mates and correlate with higher reproductive success (i.e., Catchpole 1996) . In other species, such ecological correlations seem absent or song repertoires are much reduced, rendering mixed evidence in support of the hypothesis that song repertoires have evolved as a signal of fitness due to sexual selection (Byers and Kroodsma, 2009; Soma and Garamszegi, 2011). Regardless of song complexity, some acoustic parameters of song have also been suggested to play an important role during communication. For instance, in many taxa, the fundamental frequency of vocal signals is associated with body size since the size of the vocal tract is a physical limitation to the lowest frequencies that can be produced (Fletcher, 2004; Ryan and Brenowitz, 1985; Wallschläger, 1980) and body size can be a relevant characteristic when it comes to assess an opponent’s fighting ability or a partner’s reproductive capacity. But yet another striking aspect of birdsong, with important communicative value, lies in the mechanisms of sound production, the complex phonatory system and the sophisticated neuro-muscular motor skills necessary to produce functional song (Suthers, 2004).

Bird phonation

Despite an early interest in the mechanisms of sound production in birds (Miskimen, 1951; Nottebohm, 1971; Stein, 1968), it was not until the last decades of the 20th century that we gained important insights into the mechanisms of bird phonation (Suthers, 2004) although it is still limited to few species (Düring et al., 2013). The phonatory organ of songbirds is the syrinx, a sophisticated vocal apparatus with two independent sound sources (Düring et al., 2013; Nottebohm, 1972; Nowicki and Capranica, 1986). Laboratory experiments using muscle denervation techniques, together with the monitoring of electric pulses in syringeal muscles, revealed that each of the two sources contributes asymmetrically to sound production (Nowicki and Capranica, 1986; Ritchison, 1986). Furthermore, there seems to be motor lateralization as each side of the syrinx is specialized in producing sound in different ranges of the frequency spectrum (Allan and Suthers, 1994; Suthers et al., 1999). During singing, a complex coordination between respiratory and syringeal muscles allows birds to take short mini-breaths in between notes, even at a delivery rate of 30 notes per second (Hartley and Suthers, 1989; Suthers et al., 1999). In fact, several anatomical features in the syrinx have been associated with very fast muscle movements and high stabilization during phonation, minimizing alterations to the motor patterns due to body movements (Düring et al., 2013). Through a precise coordination and coupling of sound production in both sides of the syrinx, birds can produce two-voiced sounds (Nowicki and Capranica, 1986; Zollinger et al., 2008), quick frequency shifts (Christie et al., 2004), or even switch the sound production from one side to the other without interrupting the sound flow of a continuous frequency sweep (Ritchison, 1986).

Despite the detailed description of sophisticated sound production of songbirds, such descriptive studies did not lead immediately to theoretical or empirical research on the evolutionary implications of such anatomical and behavioural adaptations. While the singing behaviour of songbirds is highly variable, in terms of size of song repertoire or singing style, the presence of sophisticated phonation skills seems common in songbirds (Byers and Kroodsma, 2009; Price, 2013). Hence, a general theoretical framework was put forward, proposing that sexual selection on motor performance has led to anatomical and behavioural adaptations of phonation in songbirds (Botero and de Kort, 2013; Byers et al., 2010; Podos, 1997; Price, 2013).

Song performance

The study of the communicative value of motor performance in animal displays is recent, compared to the traditional focus on function of ornaments in courtship (i.e., Bostwick and Prum 2003; Fusani et al. 2007) . Very often, the complex motor patterns in animal displays are fast movements that, in real time, seem imperceptible to the human eye (Rome, 2006). Observational and descriptive studies of animal courtship displays gave way to quantitative studies that aimed to measure motor displays with precision using new technologies (e.g. high speed video recorders, Bostwick and Prum 2003). The findings showed that several aspects of motor performance were important during mate choice (Barske et al., 2011). Motor performance has been proposed as an honest signal of whole-organism condition, not necessarily because it is costly to the individual, but because executing challenging motor patterns provides information on the functioning of multiple, if not all, fundamental organ systems (Barske et al., 2011; Byers et al., 2010; Irschick et al., 2007). The execution of challenging motor displays involves the central nervous system, the muscle-skeletal system, the cardiovascular systems and can be metabolically demanding, sometimes reaching physiological limits. For instance, Barske et al. (2011) described some of fastest heartbeat rates recorded in birds at 1374 b.p.m during male courtship displays in golden-collared manakin (*Manakus vitellinus*). Two major aspects are distinguished in motor performance: vigour and skill. The term vigour refers to those acts that impose a strong metabolic demand on the individual, increasing energy consumption and overall metabolic rate (Byers et al., 2010). Skill, on the other hand, refers the "degree of precision in the activation and coordination of motor units" (Byers et al., 2010).

The fact that birdsong involves a complex motor performance was not obvious, perhaps because most movements take place inside the body, hidden from view (Rome, 2006). In fact, the fastest muscle movements found in vertebrates are associated with sound production (Bostwick and Prum, 2003; Elemans et al., 2004; Rome et al., 1996). Furthermore, singing does not seem energetically costly, compared to resting metabolic rates (Oberweger and Goller, 2001; Ward et al., 2003), if we exclude the metabolic demand to build and maintain brain song nuclei. Even if birdsong does not necessitate a major motor challenge in terms of energy expenditure, it has been postulated as a challenging motor display that requires learning and executing precise movements and coordination of phonatory muscles at high speeds (Allan and Suthers, 1994; Byers et al., 2010; Podos and Nowicki, 2004; Zollinger et al., 2008).

Podos (1996,1997) showed that birdsong could indeed be challenging to produce, suggesting a performance limit in the trade-off between the rate of note delivery and the bandwidth of trill notes (but see Prather et al. (2012) for a sensory constraint explaining the same phenomenon). During trilled songs, composed of several repetitions of the same sound in quick succession, birds need to readjust the vocal apparatus in between notes, from the syringeal position at the end of one note to the position needed to start the next note. Therefore, the time gap between notes, related with the trill rate, limits the extent of mechanical reconfiguration in the syrinx, related to frequency bandwidth of the trill. There are multiple combinations of trill rate and trill bandwidth that are challenging, and this is measured as the vocal deviation (a song produced near the theoretical physiological limit has ‘low vocal deviation’). Such a trade-off appears to be a general phenomenon in birdsong (Podos, 1997, 2001) and some studies showed vocal deviation to be meaningful during communication (Amy et al., 2015; Ballentine et al., 2004; de Kort et al., 2009a), although not all (Cramer, 2013b).

Several features of song performance have been shown to correlate with higher social status (Botero et al., 2009; Christie et al., 2004), increased reproductive output (Byers, 2007; Kempenaers et al., 1997; Poesel et al., 2001), longevity (de Kort et al., 2009b) and sexual attractiveness (Ballentine et al., 2004; Vehrencamp et al., 2013; Wasserman and Cigliano, 1991). Some of the most important song performance features are: 1) rapid changes in frequency - vocal deviation (Podos, 1997), frequency excursion index (Podos et al., 2016) or frequency jumps (Christie et al., 2004; Zollinger et al., 2008), 2) producing high power vocalisations - high amplitude songs (Brumm and Todt, 2004) or peak percentage performance (Forstmeier et al., 2002), and 3) the singing effort in terms of sound produced per time unit - sound density (Holveck et al., 2008; Leadbeater et al., 2005). One important feature of song performance is vocal consistency, defined as the ability to repeat the same song structure with precision and minimal variation, and has been suggested as an important indicator of whole-organism condition associated with quality of the neuro-motor systems (Botero and de Kort, 2013; Botero et al., 2009; Clark et al., 1987; Sakata and Vehrencamp, 2012).

Vocal consistency

Vocal consistency is the ability to produce highly stereotyped repetitions of the same song element (Sakata and Vehrencamp, 2012). Even though it can refer to various parameters of song (e.g. regularity of silent intervals between songs, Byers 2007), the term is mostly used to refer to the accurate control during phonation itself, i.e., the ability to produce the same sound with the same spectro-temporal structure, multiple times with minimal variation (Sakata and Vehrencamp, 2012). Singing with high vocal consistency would fall into the category of motor performance skill, as it requires high precision in the execution of neuro-motor patterns. However, it is not a signal of vigour as it does not imply a significant increase in metabolic rate. Several lines of evidence have demonstrated that vocal consistency plays an important role in many aspects of bird communication. During the learning phase, vocal consistency increases and reaches maximum values in crystalized song in swamp sparrows (*Melospiza georgiana*) (Clark et al., 1987). Several playback experiments have shown that birds are able to discriminate between high and low consistency songs during simulated territorial interactions in banded wrens, *Thryothorus pleurostictus*, (de Kort et al. 2009b), in blue tits (Porcedda unpublished) and in great tits, *Parus major* (Riviera-Gutierrez et al. 2011). Furthermore, vocal consistency has been shown to vary seasonally in banded wrens (Vehrencamp et al., 2013), associated with changes in size of brain song nuclei in song sparrows (*Melospiza melodia*) (Smith et al., 1997a), and has also been shown to vary with age in banded wrens (de Kort et al., 2009b; Vehrencamp et al., 2013) and great tits (Rivera-Gutierrez et al., 2012). Finally, Botero et al (2009) found a significant, positive, association between vocal consistency and reproductive success, which makes vocal consistency a signal of fitness in the tropical mockingbird (*Mimus gilvus*). Importantly, vocal consistency may also interact with other aspects of vocal performance, as Vehrencamp et al.

(2013) showed that singing close to a physiological limit (low vocal deviation) results in a decrease in vocal consistency.

Prior to measure vocal consistency, bird sounds must be classified within individual birds. Within the same sound type, it is assumed that birds executes the same motor pattern in order to produce the same spectro-temporal sound structure (Allan and Suthers, 1994; Suthers et al., 1996). This premise leads to the assumption that discrepancies in the acoustic structure between two renditions of the same sound are due to imprecisions in the execution of such motor pattern (Allan and Suthers, 1994; Suthers et al., 1996). Therefore, vocal (in)consistency can be assessed by estimating the variation of a single acoustic parameter across multiple renditions of the same sound type, for instance the variation in peak frequency (i.e., Byers, 2007). But bird sounds present complex spectro-temporal shapes with quick frequency and amplitude modulations in time, which are difficult to characterise using single, static variables. For this reason, other methods that consider the entire spectrographic shape are preferred (Cramer, 2013a). To measure vocal consistency, accounting for small discrepancies between renditions in the time-frequency-amplitude domains, researches often use the Spectrogram Cross-Correlation (SPCC) algorithm (Botero et al., 2009; Clark et al., 1987; Cramer, 2013a). Every single rendition is compared with all other renditions of the same sound type in pair-wise comparisons. The spectrograms are computed for each sound and the SPCC estimates the maximum correlation coefficient between paired spectrograms at various temporal alignments, rendering a value between 0 (entirely distinct) and 1 (identical).

Study species

The blue tit is a common model species in the study of evolutionary ecology (Mainwaring and Hartley, 2019) and was the species chosen to conduct the field studies presented in this thesis. The blue tit is a small passerine, weighing approximately 11 grams that is found in deciduous and mixed woodlands across Europe (Cramp and Perrins, 1993). Along its geographic range, it is a common, resident species, breeding and wintering in the same areas. Importantly, as in other temperate species, its social behaviour changes seasonally (Cramp and Perrins, 1993; Hinde, 1952). During autumn and winter, blue tits form flocks, sometimes with other tit species, and do not show territorial behaviour, but despite that, many individual birds can be found near their breeding territories (Cramp and Perrins, 1993; Hinde, 1952). Most pairs are formed in late winter before flocks break up and partners move in to settle in their breeding territories (Beck et al., 2020; Hinde, 1952). In Britain, egg laying starts in early-April and breeding pairs raise a single brood, as second clutches are rare (Cramp and Perrins, 1993).

Blue tits generally breed in monogamous pairs but are occasionally socially polygynous and frequently genetically polyandrous (Leech et al. 2001; Schlicht and Kempenaers 2021). In fact, most breeding pairs stay together as long as they both survive (Dhondt and Adriaensen, 1994). However, during the mating season, both male and female partners engage in extra-pair copulations (EPC) (Kempenaers et al., 1997). This leads to approximately half of all nests having at least one extra-pair offspring (Charmantier and Blondel, 2003; Kempenaers et al., 1992; Leech et al., 2001). Importantly, polygyny is also found in most populations as some males attract a secondary female once the primary female is in the egg laying state (Kempenaers, 1994, 1995).

Given that copulations are under female control (Kempenaers et al., 1995), polygyny and EPC rates cause large inter-individual variation in reproductive success among males. This particular source of selection due to female choice has been related to some of the sexual difference in plumage colouration (Andersson et al., 1998; Hunt et al., 1998; Hunt et al., 1999). However, while male blue tits gain benefits by attracting a secondary female, females show strong aggressive behaviour towards intruding females, presumably to avoid polygyny (Breiehagen and Slagsvold, 1988; Gorissen et al., 2002; Hinde, 1952; Kempenaers, 1994, 1995; Midamegbe et al., 2011; Slagsvold et al., 1999). Polygyny is usually detrimental for females since male parental effort is reduced at the secondary nest (Hartley unpublished data). Hence, female-female competition for mates to avoid polygyny (Slagsvold and Lifjeld, 1994) may create a particular source of selection that can generate or maintain secondary sexual signals in females (Midamegbe et al., 2011).

Aims and objectives

The general objective of this thesis was to understand the communicative value of vocal performance in birdsong, with emphasis on vocal consistency, using the blue tit as a model species. In Chapter 2, I test and validate the main method to measure vocal consistency in birdsong: the SPCC algorithm. I explore how the measure of vocal consistency may be affected by variation in acoustic structure of song and speculate on the impact of these parameters on the perception of vocal consistency by birds. In Chapter 3, I present observational data investigating the association between vocal consistency and reproductive success in male blue tits, as well as empirical results studying female preferences for male song in the wild. Moreover, in this chapter I explore the co-evolution of two important aspects of birdsong that are seemingly in conflict: repetition and diversity. Using a playback experiment, I investigate the communicative value and the possible selection pressures that select simultaneously for consistency (repetition) and song repertoires (diversity) in blue tits. In Chapter 4, I explore the trade-off between vocal performance and song length using a multi-species data set. Based on those observations I propose the stable performance hypothesis in which I predict that birds adjust song length to avoid a decline in vocal performance. To test the communicative value of such a trade-off, I conduct two playback experiments to examine the response of male blue tits during simulated agonistic interactions. Finally, in Chapter 5, I present a comprehensive study of previously overlooked female singing behaviour, including the context, the occurrence and the acoustic parameters of female song in relation to the song of their male partners. Based on these observations, I suggest that female song functions mainly as a territorial or mate-guarding signal during intrasexual competition. I then hypothesize whether functional differences may have led to differences in acoustic structure between male and female song, with emphasis on vocal performance. Finally, in Chapter 6, I put all these findings into perspective and provide an integrative view of the role of vocal performance in communication and the possible evolutionary forces that have shaped vocal performance in birdsong.

Chapter 2 - Sound properties affect measurement of vocal consistency in birdsong: validation of spectrogram cross-correlation (SPCC)

Abstract

Vocal consistency, defined as the similarity between subsequent repetitions of a song element, has been proposed to reflect whole-organism performance and thus be indicative of individual quality in birds. It is measured as the acoustic similarity between subsequent renditions of the same sound, normally using a Spectrogram Cross-Correlation (SPCC) algorithm. Here, I test whether the SPCC is sensitive to the acoustic discrepancies found within multiple renditions of the same sound of a bird's song. I created a set of 31,000 synthetic sounds, including 1,000 reference sounds and 30,000 inexact variants with known differences in frequency, bandwidth or duration based on within-song variation of blue tits. I then assessed the SPCC scores for acoustic similarity in relation to the known differences between reference-variant pairs. I found that SPCC is sensitive to acoustic discrepancies within the natural range of vocal consistency found in blue tit song. Importantly, I also found that the sensitivity of SPCC is significantly affected by the bandwidth of the sound, an important acoustic parameter that can differ greatly between singing styles; the same difference between two sounds means lower SPCC score for lowband than for broadband sounds. I propose that the bias in measuring vocal consistency, due to frequency bandwidth, mirrors a perceptual bias in sound discrimination in humans and perhaps also in birds. Finally, I validated my findings using natural song recordings of 99 male blue tits (*Cyanistes caeruleus*) confirming that frequency bandwidth of song has an impact on measuring vocal consistency in natural song of birds.

Introduction

Birdsong is a challenging display for an individual's neuro-motor system (Suthers, 2004) and is likely to relay important information about the signaller's condition or other qualities (Botero and de Kort, 2013; Byers et al., 2010; Sakata and Vehrencamp, 2012). One important aspect of song performance is vocal consistency, the ability to repeat precisely a note, syllable, phrase or whole song (Botero and de Kort, 2013; Sakata and Vehrencamp, 2012). The diversity of ecological correlates associated with vocal consistency highlights the importance of this song trait in bird communication (Ballentine et al., 2004; Botero et al., 2009; Cramer, 2013a; de Kort et al., 2009b; Smith et al., 1997a; Vehrencamp et al., 2013). Although some authors have used the term vocal consistency to measure different specific aspects of song (e.g. consistent inter-song intervals), it is most often used to measure the precision in producing the same phonetic structure many times without variation. Hence, to measure vocal consistency within individuals it is required to compute pair-wise comparisons between renditions of the same type of sound and measure the acoustic similarity between them.

Producing subsequent renditions of the same acoustic pattern is thought to be the result of a bird executing the same motor pattern (Allan and Suthers, 1994; Suthers et al., 1996). This premise leads to the assumption that discrepancies in the acoustic structure among renditions are due to imprecisions in the execution of such motor patterns (Allan and Suthers, 1994; Suthers et al., 1996). By measuring the acoustic similarity between two renditions of the same sound, we assess the precision in which the same motor pattern has been executed (Cardoso, 2017). Several studies have shown that this is a signal of quality associated with increased fitness, theoretically related to the neuro-motor skills of the individual (Botero and de Kort, 2013; Sakata and Vehrencamp, 2012).

Therefore, in the study of vocal consistency, the sounds being compared are pre-classified beforehand. This is a crucial difference with other areas of bioacoustics research, such as song learning and species recognition, that use acoustic similarity scores for sound classification (i.e., Kogan and Margoliash 1998) . In that case, the sounds to be contrasted may belong to different individuals, different species and often may be recorded under different recording conditions. To achieve a robust classification of sounds that are only partially similar it is often necessary to omit the fine details of sound structure (Lachlan et al., 2010). On the contrary, measuring vocal consistency is focused on measuring those minute differences between sounds that have been classified *a priori* as "the same type". Hence, the methods used to measure acoustic similarity in different areas of research are not necessarily be the same.

Since birds are highly sensitive to minute differences in the acoustic structure of two sounds (Fishbein et al., 2019; Lawson et al., 2018; Margoliash, 1983; Theunissen and Doupe, 1998), the method to measure vocal consistency also must be sensitive to those acoustic differences. A commonly used method is the Spectrogram Cross-Correlation (SPCC) algorithm that measures the acoustic similarity between two sounds represented by two spectrograms. A spectrogram is essentially a double matrix with frequency in the Y-axis, time in the X-axis and the sound amplitude of each time-frequency bin. Two spectrogram matrices can be overlaid to estimate a correlation coefficient, as a measure of similarity between the two sounds, but there are many options for how these two spectrograms are overlaid, a common problem when comparing two time series. In the SPCC, this problem is solved by the second step in the method, the cross-correlation algorithm. The cross-correlation algorithm computes multiple correlations between both spectrograms at different temporal alignments, selecting the maximum correlation of all correlations computed. This algorithm is essentially an optimizer (in the temporal dimension) that provides the maximum similarity score possible between two sounds, rendering an acoustic similarity score from 0 (highly dissimilar) to 1 (identical).

Although the SPCC is considered a suitable method to measure vocal consistency (Cramer, 2013a), it is not clear how sensitive the method is to acoustic discrepancies found in individual birds. There are also reservations to whether it provides an objective, universal index of vocal consistency regardless of the singing style (Cardoso, 2017). This a common problem in the study of vocal performance since different singing styles might impose different physiological challenges and therefore the assessment of vocal performance is difficult to generalize (Cardoso, 2017). The bounded, standardized and unit-less nature of the SPCC similarity score has been an argument for the universality of the index, but it is still possible that the temporal or spectral properties of the sounds influence the SPCC response to acoustic discrepancies.

Here, I investigate the response of SPCC to acoustic discrepancies in a controlled set of synthetic sounds that can be defined and manipulated. These synthetic sounds emulate whistle-like vocalizations of songbirds, common in many species, when upper harmonics are filtered in the vocal tract (Nowicki, 1987; Nowicki et al., 1989). I used this set of synthetic sounds to test 1) if the SPCC method is sensitive to acoustic discrepancies within the range of natural variation found in birdsong, 2) whether the SPCC response is influenced by the spectral or temporal properties of sound and 3) if the SPCC response is linear along the natural range of vocal consistency found in birdsong. Because the cross-correlation algorithm acts as an optimizer in the temporal dimension, I predict that the SPCC sensitivity to temporal discrepancies will be lower than to spectral discrepancies. Finally, I test these findings and predictions derived from the analysis of synthetic sounds with a data base of natural song recordings from blue tits.

Methods

Study design and sound synthesis

I built a database of synthetic sounds to investigate the response of the SPCC score to acoustic discrepancies in frequency, bandwidth and duration. To that end I created a set of reference sounds that were periodic sounds with linear or sinusoidal frequency modulation and no harmonics, simulating whistle-like, blue tit song (mean fundamental frequency of 5.1 kHz, bandwidth 0-1.7 kHz and duration of 15-120 ms, Table 2.1). For each reference sound I synthesized 30 inexact copies in three treatment groups, hereafter frequency, bandwidth and duration treatments (Figure 2.1). For the frequency treatment, I created inexact variants that had the same spectrographic shape as the reference sound but were only higher or lower in mean fundamental frequency, ranging in ± 217 Hz around the reference mean fundamental frequency (Figure 2.1A). For the bandwidth treatment, I created inexact copies that were different in the bandwidth by expanding or shrinking the reference sound in the frequency spectrum while keeping the duration and fundamental frequency unchanged, ranging ± 415 Hz in bandwidth around the reference bandwidth (Figure 2.1B). Finally, in the duration treatment, I stretched or shrank the reference sound in the temporal dimension to create an inexact variant that differed only in duration, but with the same bandwidth and fundamental frequency, ranging ± 0.29% around the reference duration (Figure

2.1C). The range of differences between reference sounds and inexact variants was based on the variation found in blue tit song (see below). The full synthesis process as well as the following acoustic analyses were conducted in R software (Ligges, 2013; R Development Core Team, 2016; Sueur et al., 2006).

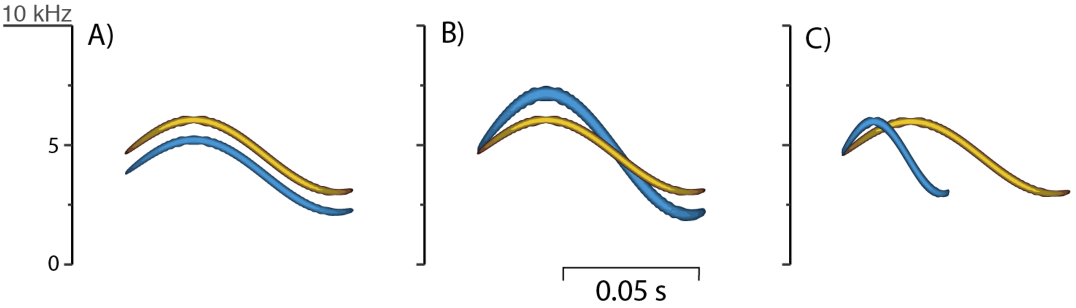


Figure 2.1 – Spectrograms of a synthetic sounds built as a reference (gold) and inexact variants (blue). The reference sound in gold could have frequency modulation, as shown here, or not. The blue sound in Figure 2.1A is an inexact variant created in the frequency treatment, only different in the height along the frequency axis. The blue sound in Figure 2.1B is a synthetic variant of the bandwidth treatment, with the same duration and mean fundamental frequency but different bandwidth. In Figure 2.1C, the blue sound is the variant of the duration treatment, with the same bandwidth and mean fundamental frequency.

Natural variation in birdsong

To create the synthetic sounds that simulated real birdsong, I used blue tit song as a model. Songs were obtained from a large data base of recordings that were collected daily from January to May, 2018 to 2020, in a blue tit, nest-box breeding population at Lancaster University campus, UK (Mainwaring and Hartley, 2009), using a Marantz PMD661 recorder (48kHz sampling rate and 24-bit depth) and a Sennheiser ME67 microphone. Blue tit song is normally composed of 2 or 3 introductory notes followed by a trill (Figure 2.2, (Bijnens and Dhondt, 1984; Cramp and Perrins, 1993; Latimer, 1977). A trill was defined as a group of notes where the same sound type is repeated several times in quick succession. A note was defined as a continuous trace in the spectrogram. Blue tits are discontinuous singers, they repeat the same song type structure many times with silent pauses in between and this is known as a song type. Normally, blue tits repeat the same song type between 10 to more than a 100 times before they switch to a different song type (Poesel and Kempenaers, 2000). Individual male blue tits have relatively small repertoires, ranging from 3-8 different song types and, within a population, most individuals share the same song types (Bijnens and

Dhondt, 1984; Doutrelant et al., 2000a). I measured the duration of each note manually, measured the bandwidth as the distance between the maximum and the minimum frequency that were measured in the power-spectrum as the highest and lowest frequency above a -20 dB amplitude threshold (window size: 1024 samples; see Podos, 1997). I also tracked the fundamental frequency of each note (window size: 512 samples; amplitude threshold; 15%) and estimated the mean, fundamental frequency of each note. Within each trill, I calculated the difference of each note to the mean duration, mean bandwidth and mean fundamental frequency of all notes in the trill to measure the within-trill variation in these three parameters.

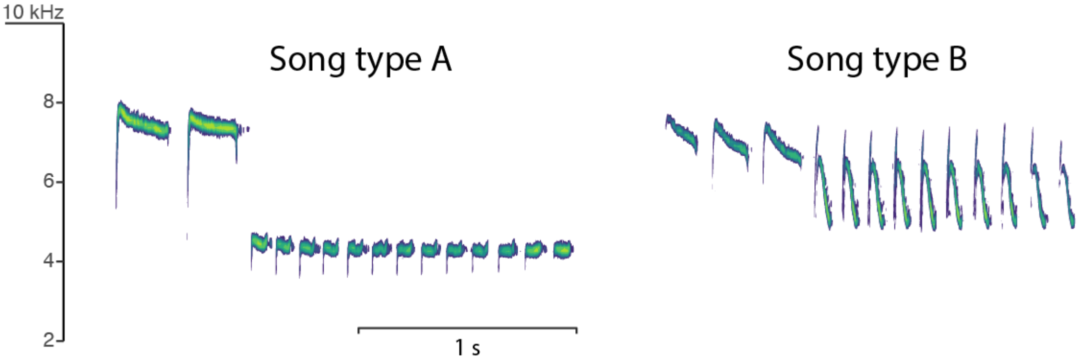


Figure 2.2 – Spectrograms of two blue tit songs from the same individual. These two song types exemplify the extremes of trill bandwidth found across the population with narrowband (song type A) and broadband trills (song type B).

Estimating a similarity score: Spectrogram cross-correlation algorithm

I measured the acoustic similarity within the synthetic sound set between each variant-reference pair using the SPCC algorithm, with a temporal resolution of 1 ms and a frequency resolution of 93.75 Hz (Clark et al., 1987; Cortopassi and Bradbury, 2000). The same SPCC procedure was used to measure vocal consistency in all blue tit songs analysed, as the acoustic similarity of every pair-wise note comparison within trill. The output of the SPCC is a curve with all the correlation coefficients at each time offset (Figure 2.3). The maximum correlation coefficient is taken as the acoustic similarity score derived from the SPCC (Figure 2.3).

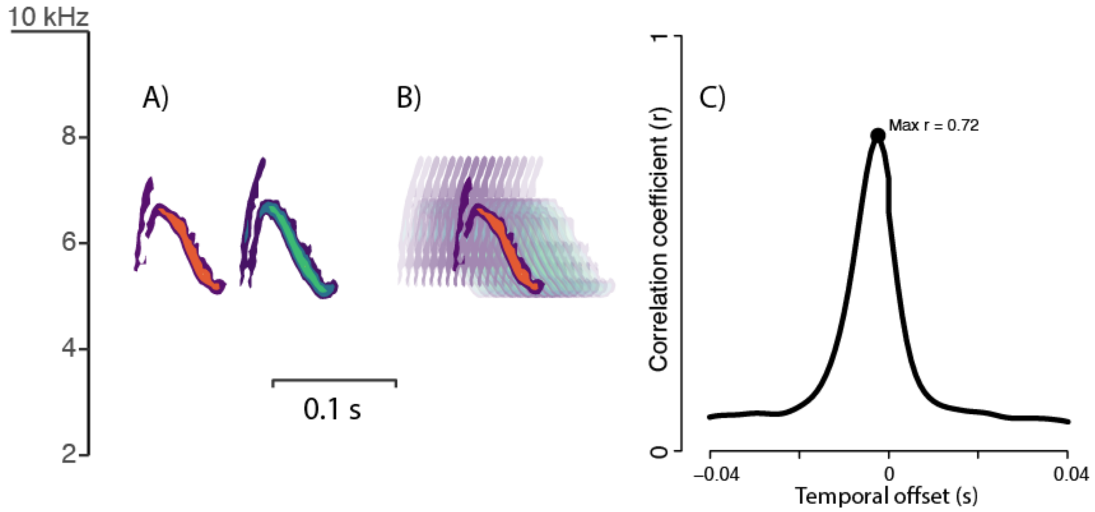


Figure 2.3 – Example of SPCC algorithm used to compare two notes of the trill from song type B shown in Figure 2.2. The two notes to be compared are shown in Figure 2.3A, Figure 2.3B is a visual explanation of the multiple alignments of both notes during the SPCC algorithm and Figure 2.3C shows the correlation coefficients calculated at each of these alignments. The maximum correlation is taken as the SPCC score.

Statistical modelling of SPCC response

All measures are presented as mean ± one standard deviation (SD), unless otherwise indicated. Statistical analysis was carried out in R software (R Development Core Team, 2016). I used Linear Mixed Effects Models (LMMs) with the SPCC score as the response variable as a function of the difference between variant-reference sound. In this way I could test whether the SPCC score was correlated with the known differences between sounds, namely, the sensitivity of SPCC algorithm to detect acoustic discrepancies. Moreover, I included the absolute bandwidth and absolute duration of the reference sound as explanatory variables in the model to investigate a possible effect of song characteristics in the sensitivity of SPCC to detect acoustic differences. I fitted a separate statistical model for each treatment (bandwidth, duration and frequency) and each model was fitted with a sample of 10,000 pair-wise comparisons between reference-variant sounds, since each of the 1,000 reference sounds had 10 inexact variants for each treatment. Hence, to avoid pseudo-replication, the identity of the reference sound was included as a random effect in the model. I considered a variable to have a significant impact on the SPCC score if the 95% Confidence Interval (CI) did not overlap with zero. The range of SPCC scores is naturaly restricted to values between 0 and 1, indicating that a linear response of the SPCC could not be suitable. Hence, I also fitted a linear model with an arcsine and a logit transformation of the response variable, as these functions represent probability distributions that are bounded between 0 and 1. I used the goodness of fit (R^2) to select the best model, as well as the model diagnostics to check for normality and homoscedasticity of the residuals. All numerical variables were scaled and centred so model estimates are standardized (Gelman, 2008) allowing us to compare the SPCC response in the three treatments.

Testing the results in blue tits

Finally, I tested whether the conclusions derived from the analysis of synthetic sounds were reflected in a real case using the blue tit song data. To this end, I chose two blue tit song types that were at the extremes of trill bandwidth range across the population. Song type A presents notes of very narrow frequency modulation, hereafter narrowband trill, while song type B presents trills with notes of broadband frequency modulation, hereafter broadband trill (Figure 2.2). I tested whether the vocal consistency measured in narrowband notes was similar to that measured in broadband notes. Moreover, I then categorized all pairwise comparisons as similar in frequency, if the difference in their mean fundamental frequency was less 25 Hz, or different in frequency, in the difference was greater. This arbitrary threshold was chosen because, in real birdsong, very few notes are identical in frequency. Hence, I could see the impact of bandwidth in measuring vocal consistency when two notes are different, at least, in frequency. I used a Mann-Whitney U test to compare the SPCC scores of broad and narrowband trills in the blue tit with the same and with different mean fundamental frequency.

Results

SPCC response in synthetic sounds

The complete synthetic data set consisted of 31,000 sounds, including 1,000 reference and 30,000 variant sounds. Based on the goodness of fit of the three models, the SPCC response was not linear within the range of natural variation of acoustic discrepancies in fundamental frequency, bandwidth and duration found in blue tit song (Figure 2.4 and 2.5 and Table 2.1). Instead, I found that the relation between the SPCC score and the acoustic differences between reference-variant sound best fitted an arcsine curve (Figure 2.5), as suggested by the goodness of fit and the pattern of distribution of the residual *versus* the fitted values (Figure 2.4).

The SPCC method was sensitive to acoustic discrepancies between reference-variant pairs, as the SPCC score showed a significant negative correlation with the acoustic discrepancies in frequency, bandwidth and duration (Figure 2.5, Table 2.2). The SPCC method was most sensitive to spectral differences (frequency and bandwidth), but less sensitive to temporal differences (duration), (Figure 2.5, Table 2.2). This means that, for a 1% difference in frequency and the same 1% difference in duration between two sounds, the SPCC provides a lower score for the difference in frequency. I also found that the SPCC score was influenced by the bandwidth of the sounds being compared in the frequency and the duration treatment (Figure 2.5, Table 2.2). For example, for a 1% difference in frequency, the SPCC score was higher if the reference sound had a broad frequency bandwidth than if it was narrowband (Figure 2.5). This is reflected in the steeper down slope in the SPCC response for narrowband notes than for broadband notes, in the frequency treatment (shown by blue gradient in Figure 2.5, Table 2.2). In the duration treatment, this effect was opposite as the same difference in duration rendered a higher SPCC score in narrowband notes (Figure 2.5B). In general, shorter sounds rendered higher SPCC scores but this effect was very small (Table 2.2).

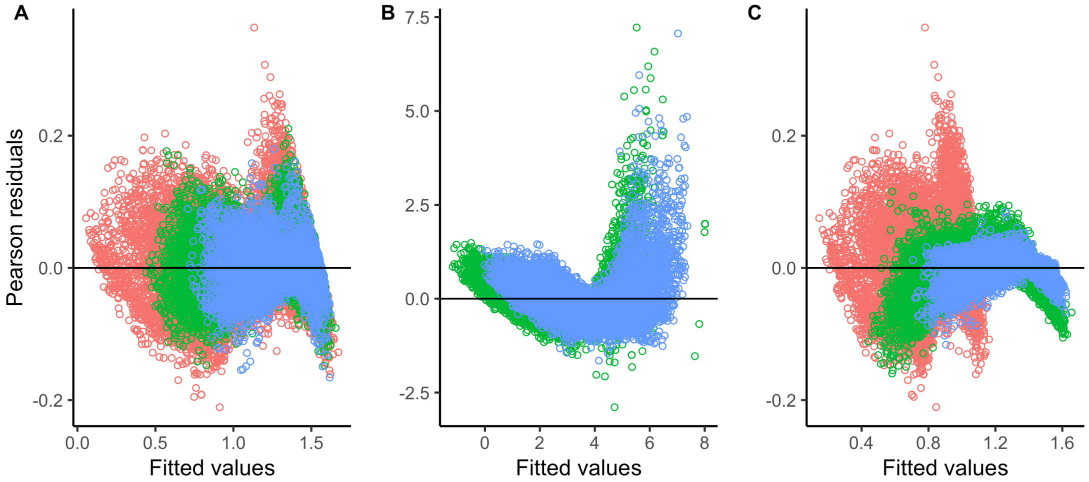


Figure 2.4 – Fitted-vs-residual plots of the models with an arcsine transformation (2.4A), with a logit transformation (2.4B) and with no transformation (2.4C). The frequency treatment is shown in red, the bandwidth treatment in green and the duration treatment in blue. Residuals in 2.4A, after applying the arcsine transformation, show the highest homoscedasticity and the best fit.

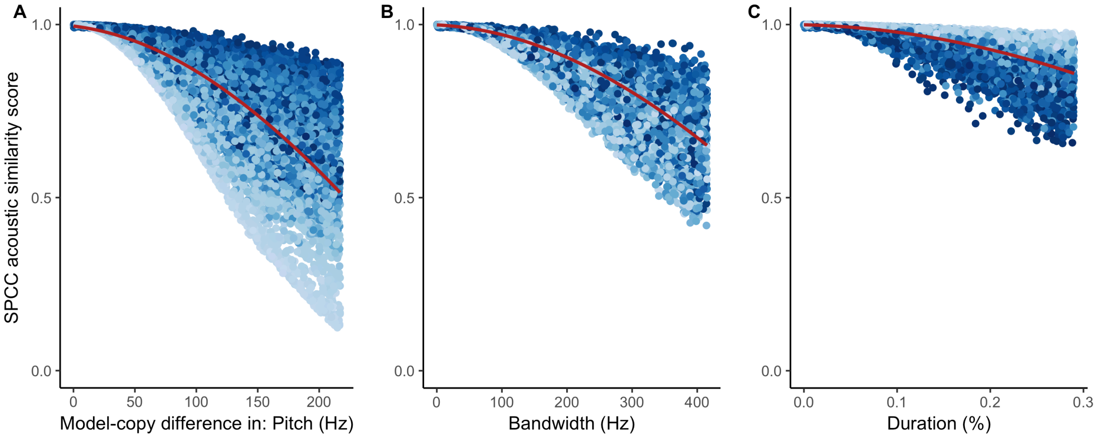


Figure 2.5 – Response of the SPCC score to acoustic discrepancies in mean fundamental frequency (Figure 2.5A), bandwidth (Figure 2.5B) and duration (Figure 2.5D). The colour gradient shows the bandwidth of the reference sound from 0 kHz, i.e., a pure tone, (light blue) to 1.7 kHz (dark blue). SPCC algorithm is most sensitive to frequency discrepancies in sounds, as shown by the steeper down slope in Figure 2.5A regarding the frequency treatment. The duration treatment in Figure 2.5C shows the shallowest slope, indicating that SPCC is least sensitive to temporal discrepancies. Figure 2.5A shows the impact of bandwidth in the SPCC response when dealing with frequency discrepancies. Here, SPCC score of narrowband notes (light blue) decrease in a steeper fashion than broadband sounds (dark blue). This effect is opposite in the case of SPCC response to duration discrepancies, where narrowband sounds (light blue) have a very shallow slope compared to broadband sounds (dark blue).

Table 2.1 – Goodness of fit shown by the marginal R^2 for all three models fitted in each treatment group. In all cases, the best fitting model was the arcsine model

Treatment	**Arcsine model R^2**	**Logit model R^2**	**Linear model R^2**
Frequency	0.86	0.78	0.82
Bandwidth	0.91	0.81	0.84
Duration	0.90	0.80	0.82

Table 2.2 – Output of the model investigating the SPCC response to acoustic differences in frequency, duration and bandwidth. For each fixed effect I present the model estimate, the lower and higher CI of the estimate, the T statistic and the conditional R squared. The estimate of the parameter of reference-variant difference indicates the slope in the correlation between the SPCC score and the programmed difference between synthetic sounds, i.e., the sensitivity of the SPCC to detect acoustic differences in each treatment. The slope is significantly lower than zero, indicating that the acoustic similarity form SPCC is sensitive to acoustic variation in the three treatments. The SPCC is most sensitive to spectral differences (frequency and bandwidth) than in temporal differences (duration). The bandwidth of the sounds being compared has a significant impact on the SPCC score, especially in the frequency and duration treatment but with opposite effects. The duration of the sound shows a significant impact on the SPCC score as shorter sounds tend to have higher SPCC values, but the effect size is relatively small.

Treatment	**Parameters**	**Estimate**	**T value**	**CI 5%**	**CI 95%**	**R^2**
Frequency	Intercept	1.009	0.993	1.025	125.039	0.954
	Model-variant difference	-0.268	-0.271	-0.264	-163.158	
	Bandwidth	0.108	0.093	0.124	13.425	
	Duration	-0.042	-0.058	-0.027	-5.252	
	Model-variant difference : Bandwidth	0.049	0.046	0.053	29.74	
	Model-variant difference : Duration	-0.016	-0.019	-0.013	-9.798	
Duration	Intercept	1.285	1.279	1.291	411.734	0.965
	Model-variant difference	-0.142	-0.144	-0.141	-184.833	
	Bandwidth	-0.065	-0.071	-0.059	-20.94	
	Duration	-0.039	-0.045	-0.032	-12.35	
	Model-variant difference : Bandwidth	-0.03	-0.032	-0.029	-39.592	
	Model-variant difference : Duration	-0.022	-0.024	-0.021	-28.767	
Bandwidth	Intercept	1.13	1.121	1.14	233.152	0.975
	Model-variant difference	-0.236	-0.238	-0.234	-249.223	
	Bandwidth	0.038	0.029	0.048	7.845	
	Duration	-0.038	-0.048	-0.029	-7.925	
	Model-variant difference : Bandwidth	0.021	0.019	0.023	22.342	
	Model-variant difference : Duration	-0.022	-0.024	-0.02	-22.934	

SPCC and bandwidth in blue tit song

The full database of blue tit song included 42,431 notes of 7,595 trills in 99 individual male blue tits (77 ± 56 trills per individual). In this set, I measured the variation in mean fundamental frequency, bandwidth and duration of each note within the trill with respect to the mean per trill, for all the 7,595 trills. I chose the 95% quantile in variation within trill as the maximum acoustic differences introduced between synthetic reference sounds and their variants. From the entire blue tit data set, I selected the song type A (narrowband) and song type B (broadband) of the population, that included 8,462 notes from 1,091 trills in 52 individuals (21 ± 14 trills per individual). As predicted by my analysis of synthetic sounds, I found that SPCC scores were significantly higher in broadband trills than in narrowband trills (Broadband: 0.76 ± 0.15, Narrowband: 0.85 ± 0.07 SPCC score, W = 1056, P < 0.001, 5% CI = -0.083, 95% CI = -0.025, Figure 2.6 and 2.7). Such a difference was not significant if they were similar in frequency (Broadband: 0.85 ± 0.09, Narrowband: 0.87 ± 0.06 SPCC score, W = 345, P = 0.017, 5% CI = -0.057, 95% =0.010, Figure 2.6 and 2.7) but SPCC scores were ten times lower in narrowband trills if the sounds were different in frequency (Broadband: 0.71 ± 0.16, Narrowband: 0.84 ± 0.07 SPCC score, W = 174, P < 0.001, 5% CI = -0.138, 95% CI = -0.0504; Figure 2.6 and 2.7). The distribution of SPCC scores in both the synthetic sounds and the real birdsong was largely skewed to the upper limit of the distribution, near 1 (Figure 2.5).

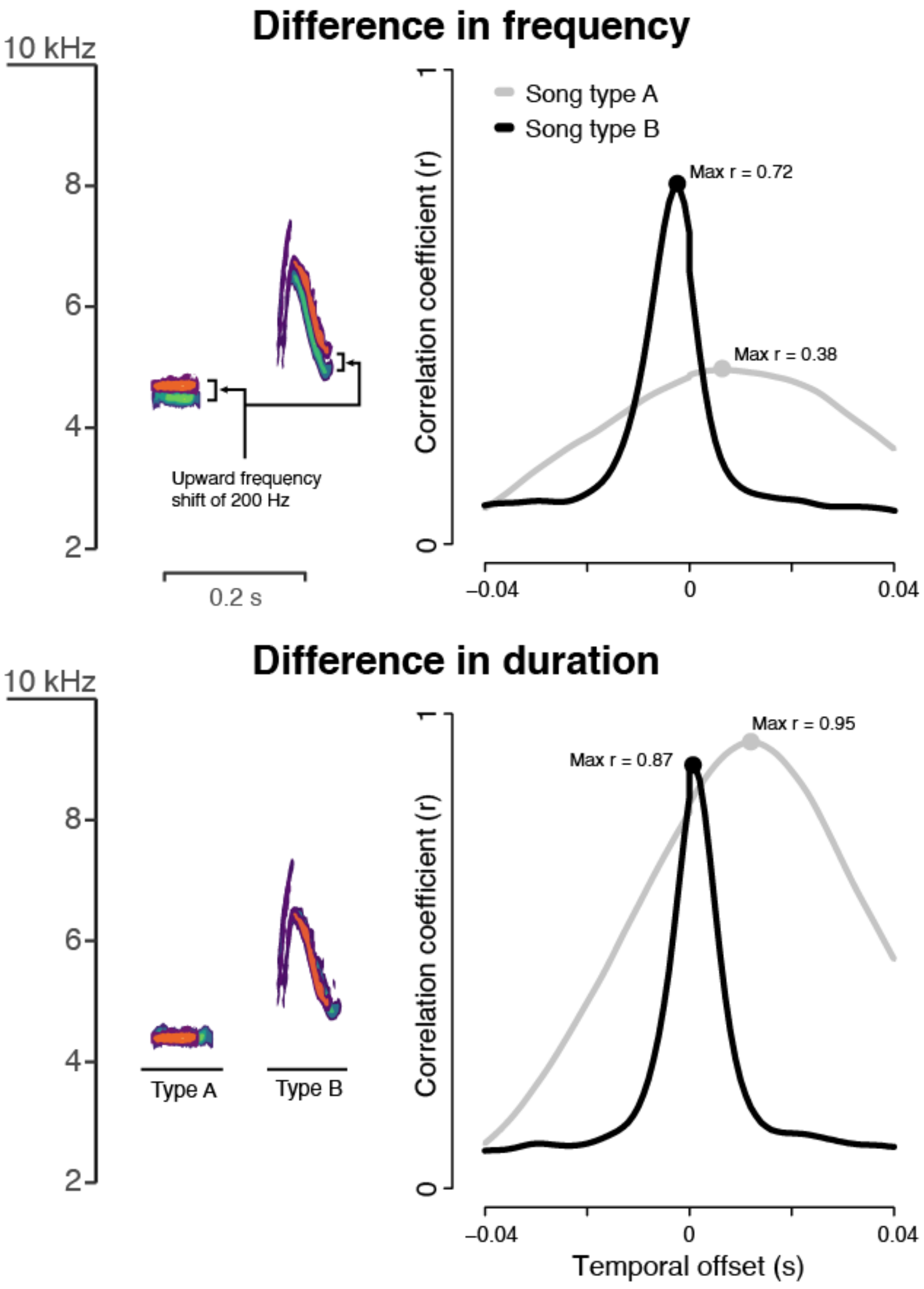
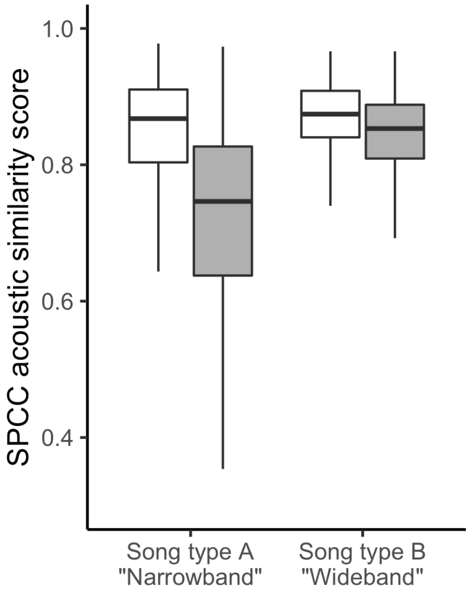

Difference in frequency
10 kHz
8
6
4
2
0.2 s
Upward frequency shift of 200 Hz
Correlation coefficient (r)
1
0
Song type A
Song type B
Max r = 0.72
Max r = 0.38
–0.04
0
0.04
Difference in duration
10 kHz
8
6
4
2
Type A
Type B
Correlation coefficient (r)
1
0
Max r = 0.95
Max r = 0.87
–0.04
0
0.04
Temporal offset (s)

Figure 2.6 - SPCC curves derived from comparing two notes of both song type A and B. Blue-green spectrograms show two note types arbitrarily used as a reference in an example. Overlaid there are two other renditions of each note type using red colours, varying in frequency and duration. The upper left plot depicts the two note types (green) and variants (red) that differed mainly in frequency. On the right, the respective SPCC curves derived from comparing both pairs of notes. For the same difference in frequency, the SPCC score is much lower in narrowband notes (song type A), compared to broadband notes (song type B). The lower left plot depicts a difference in duration, but not in frequency, in both pairs of notes, with the respective SPCC curves on the right. For the same difference in duration, narrowband notes (song type A) render a slightly higher SPCC score than broadband notes (song type B).



2.7 – Impact of bandwidth in the SPCC score as seen in natural song of blue tits. In general, the comparison of broadband notes using SPCC renders higher similarity scores. But this difference is only marginal and not significant when both notes have similar frequency (white boxes). On the contrary, when two notes being compared have different frequency, the SPCC scores are drastically lower in narrowband trills, compared to broadband trills.

Discussion

My results support the use of Spectrogram Cross-Correlation (SPCC) to measure vocal consistency in birds, as the acoustic similarity score derived from SPCC correlated significantly with the known acoustic discrepancies between synthetic sounds. The sensitivity of SPCC to the acoustic differences was not linear, i.e., the change in SPCC was not equal across the range of vocal consistency found in birdsong. This means that, a change from 1 to 2 % in frequency difference between two sounds led to a smaller decrease in SPCC than a change from 20 to 21% of difference in frequency between sounds. This may reflect a biologically meaningful pattern since the increase in the challenge of motor performance is likely higher when approaching physiological limitations (Cardoso, 2017). For instance, increasing vocal consistency from 0.95 to 0.96, given that 1 is the maximum, may imply a greater performance challenge than the same increase from 0.75 to 0.76. I also found that, in the case of spectral discrepancies (frequency and bandwidth), the sensitivity of SPCC decreased as the note bandwidth increased. These findings are also reflected in the analysis of blue tit song, as SPCC scores in narrowband trills were significantly lower than in broadband trills.

SPCC method was sensitive to small discrepancies in acoustic structure, within the range of natural variation found in vocal consistency of blue tit song. Given that birds are sensitive to small discrepancies in acoustic structure (Fishbein et al., 2019), this finding validates the use of SPCC to measure vocal consistency. The temporal optimization process, associated with the cross-correlation algorithm within the SPCC, is likely responsible of the reduced sensitivity to differences in the temporal dimension. However, using an algorithm to find the correct alignment is crucial in measuring acoustic similarity because this task is highly subjective. There are other alternatives to solve the alignment problem (i.e., Dynamic Time Warping; DTW) but, unless the optimization process acts in the three dimensions (temporal, spectral and amplitude dimensions), this step will always cause differences in the sensitivity. The cross-correlation along the X-axis further implies that the frequency bandwidth (Y-axis) influences the similarity scores, given the same difference in frequency or duration. Low bandwidth sounds, i.e., a pure tone, are characterized by one single frequency in the Y-axis, represented by a straight, horizontal line in the spectrogram. Hence, two pure tones of different frequency are essentially two parallel lines and thus they would never ‘meet’ at any time offset in the cross-correlation process, rendering a low similarity score. On the other hand, broadband tones of different frequency can be partly matched during SPCC if the difference in frequency is smaller than the bandwidth. This exemplifies the impact of bandwidth on the SPCC response, indicating that the same difference in frequency is not reflected with a similar decrease in SPCC if measured in two pairs of sounds with different bandwidth. Furthermore, if two pure tones of the same frequency have different duration, they are essentially two overlapping lines, leading to a high SPCC score. In the case of frequency modulating sounds with the same difference in duration, the overlapping is reduced at any time offset, rendering a much lower SPCC score. Again, the bandwidth of the sounds influences the response of SPCC to differences in duration.

At first, the impact of note bandwidth on the sensitivity of SPCC may seem inadequate, implying that measurements of vocal consistency in songs with different bandwidth may not be comparable. But the bandwidth bias in SPCC sensitivity may not be a defect of the method as we do not know how vocal consistency (i.e., acoustic discrepancy) is assessed and perceived by birds. In humans, the threshold of frequency discrimination increases significantly with increasing frequency modulation (increasing bandwidth) (Dooley and Moore, 1988). Similarly, when two pure tones of different frequency are presented in sequence, the threshold of frequency discrimination is lower than when those two tones are presented by modulating the first frequency into the second frequency (Fastl, 1978). These studies show that assessing acoustic discrepancies is more difficult when the sounds to be compared have broadband frequency modulations, but it is easier when comparing narrowband sounds. Hence, the impact of bandwidth in the SPCC score seems to mirror the perception of consistency in humans (De Cheveigne, 2005). The same perceptual pattern in which spectral discrepancies are less obvious in broadband notes could also occur in songbirds but further empirical evidence is needed.

In such case, receivers could show a preference for narrowband trills as they ensure honesty because low vocal consistency is easily perceived. From the sender's perspective, narrowband trills will be produced only by individuals that sing with high vocal consistency, otherwise the inconsistencies will be highlighted. In line with this idea, previous work has found that common nightingales (*Luscinia megarhynchos*), produce narrowband trills (whistle songs) that are important in mate attraction and vocal consistency within those trills indicates male quality (Bartsch et al., 2016). In the same study, it was also found that birds with higher vocal consistency also produced more narrowband trills (Bartsch et al., 2016). Apart from narrowband trills (whistle songs), common nightingales also produce fast trills of broadband tones and these are meaningful during simulated intrasexual conflicts (Schmidt et al., 2008). This evidence could support the idea that the diversity of song types within individuals may serve to demonstrate neuro-motor skills in relation to different performance constraints (Cardoso, 2017). At the same time, this could also explain the lack of ecological correlates of some performance parameters in some studies that use multiple song types (Cardoso et al., 2012).

Despite possible variation in sensitivity of SPCC due to acoustic properties, previous research has shown that vocal consistency, measured as the similarity scores derived from SPCC is biologically meaningful (Botero and de Kort, 2013; Cardoso, 2017; Sakata and Vehrencamp, 2012). Vocal consistency has been found to increase during song learning (Clark et al., 1987), correlate with social status and reproductive success (Botero et al., 2009), age (de Kort et al., 2009b) or brain structure (Smith et al., 1997a). In blue tits, vocal consistency showed significant seasonal variation, peaking during the fertile period of the female (Sierro et al., in prep) and correlates with seasonal changes in brain song nuclei (Caro et al., 2005), regardless of song type characteristics. Furthermore, I found that females sing with lower vocal consistency (Sierro et al. in prep), that females prefer male songs of higher vocal consistency and that males with higher vocal consistency have higher reproductive success (Sierro et al., in prep).

In conclusion, these results support the use of the SPCC method to measure vocal consistency in tonal sounds of birdsong and possibly in other taxa with similar, tonal sounds. I found that the sensitivity of SPCC varies significantly with bandwidth of the sound, potentially reflecting a similar perceptual bias in acoustic discrimination in human hearing, perhaps also in birds. Despite these results in support of SPCC scores as a biologically meaningful measure of vocal consistency, there are some concerns. The SPCC response was not linear and was significantly affected by the frequency bandwidth of sound. Therefore, absolute values of SPCC scores may not be comparable if the songs analysed have very different spectral structure, or should only be compared after normalizing the measures of acoustic similarity. Finally, my findings indicate that song performance of different song types within individuals may be associated with different performance constraints, highlighting the importance of understanding and validating the methods of measuring song performance in order to provide meaningful measures (Cardoso, 2017).

Chapter 3 - Simultaneous sexual selection for song diversity and song ritualization in birdsong

Manuscript in preparation for the journal *Science*, co-authored by Selvino R. de Kort and Ian R. Hartley

Abstract

A long-held view that song diversity is a signal of quality fails to explain the singing style of many songbird species with small or intermediate repertoire sizes that repeat the same song patterns over and over. New insight into the importance of consistent and ritualized song performance provides an explanation to the high degree of repetition in birdsong. However, no study so far has tried to tease apart the role of both song diversity and repetition in the context of sexual selection. We studied the song of male blue tits during dawn chorus; a song display that takes place during the breeding season, just before mating. We found that consistent repetition of song elements is a signal of fitness: clutch size increased within increasing male vocal consistency. In addition, we found that males increased the consistency of their song performance during the fertile period of the female. Furthermore, a field playback experiment showed that fertile females were more likely to vocalise (mostly with copulation-solicitation calls) to higher consistency songs. Despite preferring consistent songs, females quickly habituated to playback of a repeated song type. However, a switch to a different song type in the playback led to partial recovery of the female vocal response, suggesting a role of song diversity in dishabituation during long courtship displays. Our experimental and observational results are strong evidence that 1) consistent repetition is selected as a signal of quality that is the target of mate choice and 2) song diversity is key during long, courtship displays to avoid habituation. For the first time, we provide evidence that separate behavioural mechanisms, within sexual selection process, promote song ritualization and song diversity simultaneously within the same communication system. These two conflicting evolutionary forces may explain the evolution of song repertoires of intermediate size, typical of many bird species.

Introduction

Birdsong is a classic example of an signal resulting from sexual selection (Andersson, 1994). The sophisticated performance and structure of birdsong is thought to be an honest signal of the quality of the singer. As a communication system, birdsong presents diversity (different song structures share the same function) and repetition (the same structure is produced multiple times) (Collins, 2004; Hebets et al., 2016). Song diversity within the repertoire of individuals, is thought to be under direct sexual selection (Catchpole, 1996) but this hypothesis has mixed support and it is unclear how it might provide an honest signal of quality (Byers and Kroodsma 2009; Gil and Gahr 2002; but see also Soma and Garamszegi 2011) . The problem is highlighted by the fact that birds never produce song with continuous variation (diversity) however, as other mating displays, it is characterise by its repetitive, ritualized nature (Byers et al., 2010; Price, 2013). Although some authors suggested that repetition and diversity respond to the conflicting selective pressures of intra- and intersexual interactions respectively (Searcy and Yasukawa, 1990), this hypothesis has not found wide support across species (Collins, 2004). This apparent paradox between diversity and repetition is still puzzling (Price, 2013) and it suggests there are conflicting evolutionary forces at play.

One suggested mechanism explaining repetitive patterns in birdsong (Botero and de Kort, 2013; Price, 2013) is mate preference for song performance skills (Barske et al., 2011; Byers et al., 2010). As for other ritualized displays (Fusani et al., 2014), theory suggests that song performance, through the physiological demand imposed by singing, indicates the general quality of the neuro-motor system of an individual (Botero and de Kort, 2013; Byers and Kroodsma, 2009; Dubbeldam, 1997). In birds, several studies show that higher song performance is correlated with higher social status (Botero et al., 2009; Christie et al., 2004), is relevant during simulated territorial contests (de Kort et al., 2009b; Rivera-Gutierrez et al., 2011) increased reproductive output (Byers, 2007; Kempenaers et al., 1997; Poesel et al., 2001), longevity (de Kort et al., 2009b; Rivera-Gutierrez et al., 2012) and sexual attractiveness (Ballentine et al., 2004; Wasserman and Cigliano, 1991). Repetition facilitates the assessment of song performance by providing the receiver with the pattern to compare several renditions of the same behaviour and ensuring the display is not a 'lucky move' but that it can be performed consistently at a high level. Furthermore, if the task is physically demanding, it is expected that only the highest quality individuals will be capable of maintaining high performance levels throughout multiple repetitions, so ensuring the honesty of the signal. A fundamental aspect of song performance would be vocal consistency (de Kort et al., 2009b; Zahavi, 1980) which is the ability to repeat the same song elements multiple times with precision, typically associated with trills or rattles.

Many bird species produce trilled songs where the same note is repeated in quick succession, or they deliver multiple repeats of the same song elements many times before switching to a different song type (Catchpole and Slater, 2008). Even if the repetition is not consecutive, most species still repeat the same elements after a short period of time. But repetition might carry a cost in the form of receiver habituation, where the audience loses interest over time (Dong and Clayton, 2009). The habituation hypothesis, also known as the monotony-threshold hypothesis, (Hartshorne, 1956; Kroodsma, 1978), states that birds should reduce habituation by 1) increasing the silent pauses between repetitions or 2) switch more often between different song types in their repertoire. While multi-species analyses have failed to demonstrate a correlation between lower diversity and longer silent pauses (Dobson and Lemon, 1975; Weary and Lemon, 1988), there is some empirical support for the habituation hypothesis; song type switching reduces habituation (Krebs, 1976; Kreutzer et al., 1994; Petrinovich and Patterson, 1981) and it seems meaningful during vocal interactions (Baker et al., 1986; Deoniziak and Osiejuk, 2020; Langmore, 1997; Searcy, 1992b). Often, playback studies simulate short, territorial interactions between males (Marler and Slabbekoorn, 2004) where habituation may be negligible compared to the long vocal displays such as dawn song, associated with mating (Gil and Llusia, 2020).

We studied the song and signal value of a well-established model species to tease apart the seemingly opposing evolutionary forces that select simultaneously for diversity and repetition in the same signal. We directly tested the effect of vocal consistency and song diversity on female choice using playback, under the reasonable assumption that vocal consistency is a signal of quality in birdsong and plays a fundamental role in mate choice. Such an assumption implies that males should perform their best singing during the fertile period, when females make the choice of whom to mate with, and that male vocal consistency should be positively correlated with reproductive success. In the female choice playback, we predicted that females should show a preference towards songs of higher consistency and that switching between song types would reduce female habituation to repeated exposure of playback song (Dong and Clayton, 2009; Petrinovich and Patterson, 1981).

Methods

Study population and model species

The study was conducted using a nest-box breeding population of wild blue tits near Lancaster University in the northwest of England (54.01’ N, 2.78’ W; (Mainwaring and Hartley, 2009). Breeding birds were individually colour ringed after capture at winter feeding-stations or at the nest. Birds were measured (flattened, straightened wing length to nearest mm, tarsus length with foot bent down, to nearest 0.1mm and head-bill length to nearest 0.1mm), weighed (to nearest 0.1g) (Redfern and Clark, 2001). During the breeding season, individuals were sexed and aged in the hand based on plumage features and the presence of a brood patch, present only in females, or cloacal protuberance, present only in males (Svensson, 1992).

Male song data

Blue tits are discontinuous singers which repeat the same song structure with a fixed syntax, known as a song type, to make a song bout, before they switch to another song type (Poesel and Kempenaers, 2000). Individuals use several song types and these are shared between individuals within a population. In general, songs are composed of a few introductory, high-frequency notes followed by a trill, slightly lower in frequency (Bijnens and Dhondt, 1984; Latimer, 1977). We defined a note as a continuous trace in the spectrogram. Between years 2018 and 2020, we made audio recordings of individually identifiable males in the field using a Marantz PMD661 recorder (48kHz sampling rate and 24-bit depth) and a Sennheiser ME67 microphone. From January to May each year, we sampled song of blue tits by making linear transects through the study site, from 1h before sunrise until 12:00 h. Individuals were identified from leg rings using binoculars (Nikon Prostaff 8x42). In the lab, we created a song database by manually selecting ten songs per song type, per date, per male. We focused on the trill section of the songs by selecting individual notes (Mazzoni and Dannenberg,

2014) and then conducting acoustic analysis in R software (Ligges, 2013; R Development Core Team, 2016; Sueur et al., 2006). We used vocal consistency (de Kort et al., 2009b), also known as song stereotypy (Smith et al., 1997a), as our measure of song performance (Cardoso, 2017; Sakata and Vehrencamp, 2012). Song diversity was measured as the number of different song types performed by each male during a complete dawn chorus (Doutrelant et al., 2000a; Poesel et al., 2004). Table 3.1 shows operational definitions of all variables measured.

Breeding data

During the breeding season, nest boxes were monitored every few days to determine occupancy, state of nest building, first egg date, clutch size and fledging success. We defined the onset of the breeding period as the start of nest building, usually two or three weeks before the first egg was laid. Female blue tits typically lay one egg per day and start incubation with the penultimate egg (Tomás, 2015). We defined the female receptive period, (when copulations take place), starting 5 days before egg laying through to the completion of the clutch (Birkhead and Møller, 1993; Kempenaers et al., 1992). During this period, females roost inside the nest box and, at dawn, they interact vocally with their partner that is singing its dawn song outside but near the box (Gorissen and Eens, 2005; Halfwerk et al., 2011). Dawn song is defined as a long, sustained display of song that begins around 30-90 minutes before sunrise and can last for more than 30 min, often in close proximity to the nest (Gil and Llusia, 2020; Poesel et al., 2004). Such a vocal interaction between male and female partners seems to indicate female stimulation, as the most common call produced by the female in the nest box is the copulation-soliciciation call (see Results). Hence, we used this behaviour to test female preferences for song under the premise that females would show a stronger vocal response to male song that they found more attractive (Amy et al., 2015; Dunning et al., 2014). We suspected that female vocal response would decrease after the completion of the clutch, and therefore most of the trials were carried out during the egg laying stage, except two trials that were conducted during the first and second days of incubation.

Experimental design and playback stimuli

The experimental design consisted of the presentation of manipulated male song to females while they were inside the nest box during the dawn chorus. We attached a speaker (MIFA A1 Bluetooth speaker, 5 W, frequency response 80 Hz – 18 kHz) to the outside bottom of the nest box facing towards the box (Figure 3.1). Peak sound level inside the box was set to 60 dB and the speaker was insulated to minimize sound leaking into the territory area, to reduce the likelihood of the male interacting with the playback. We considered that the male reacted to the playback if it interrupted its own dawn song to look inside the nest box while producing mobbing calls. We recorded audio inside the nest box using an AudioMoth^{TM}, (AudioMoth home page, 2017) attached to the underside of the lid, which is an Autonomous Recording Unit (ARU) programmed to record from 1h before to 1h after sunrise (Figure 3.2). The equipment was installed on the nest box the day before the test took place when the birds were not inside the nest, to avoid disturbance. On the morning of testing, we placed a video camera on a tripod 10 m from the nest box to record the entrance hole. The test began 30 minutes before sunrise and lasted 1h or until the female exited the box (see Behavioural analysis). To ensure there was no alteration of the normal breeding behaviour, we compared the clutch size, brood size and fledging success of the females involved in the trial to the rest of the population after the breeding season finished. Previous trials suggested that females were likely to interact vocally with their own mate's song but not with synthetic or a different male's song (personal observation, see also Woolley and Doupe 2008) . Therefore, in order to test female preferences we selected songs of their own mate that varied in acoustic variables, including vocal consistency, song rate, trill rate, trill length and peak frequency (Table 3.2). We chose two songs types from each male recorded during dawn chorus to simulate a complete artificial dawn song for playback. Across males, we used a total of nine song types across tests from 14 identified song types in our population, avoiding song types with trills consisting of only two notes. For two males we could only use one song type to build the simulated dawn song.

For each song type of each male (Figure 3.3), we selected one song in Audacity and then used R software to build a song type bout by repeating it several times for a total time of 1.25 minute. Song rate delivery was measured in ten songs of original recordings of each focal male, and this was the song rate used to build each playback stimulus. Two song type bouts were placed in succession to make a string of playback song of 2.5 minutes, including two bouts for each song types. We then added a silence interval of 2.5 min to make a playback round of 5.0 min. The entire playback stimulus lasted for 1h and consisted of 12 playback rounds, randomizing the order in which the song and the silence bout appeared within each round (see definitions in Table 3.1 & Figure 3.2). In all cases, we introduced an interval of at least 5 minutes of silence at the beginning of the playback file to leave some time for the female to relax after we approached the nest to start the playback. For each test and song type, we carried out a detailed acoustic analysis of the playback songs following the same methodology explained for male song analysis.

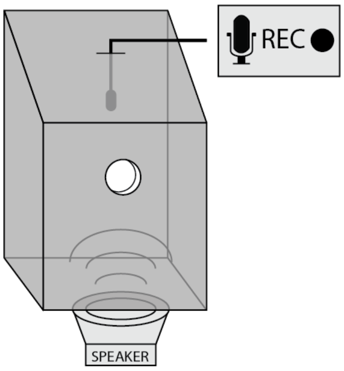


Figure 3.1 – Schematic representation of the equipment deployed in the nest box to conduct the playback experiment and monitor female vocal response.

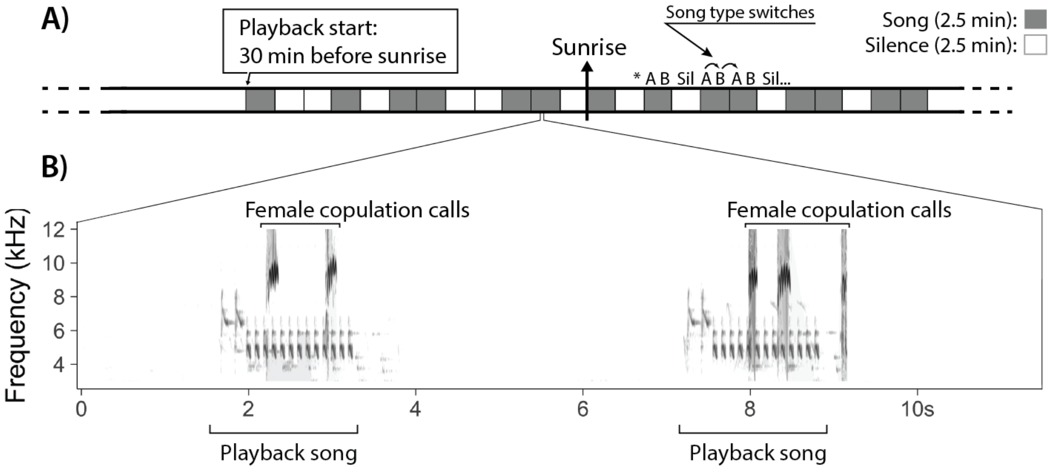


Figure 3.2 – Timeline of the experimental design and example of female vocal response to playback song. Figure 3.2A shows the structure of a playback stimulus, alternating song and silence bouts. Within each song bout, there were two separate, consecutive bouts of two different song types, usually variable depending on each female's partner. The star symbol marks the possible song type switches. Note that switches could be in both directions (from A to B and B to A). Spectrogram in Figure 3.2B is a snapshot of the audio recorded inside the nest box during a playback trial, showing the vocal interaction of the female with the playback song from the speaker. We can see how the vocal response

of the female is tightly overlapping the playback song, suggesting a genuine interaction with the stimulus.

Table 3.1 - Operational definitions

Bio-acoustic variables	**Definition**
Vocal consistency	Mean acoustic similarity between each note to the consecutive note in the same trill using a spectrogram cross-correlation algorithm (SPCC). We calculated the maximum correlation of every pairwise SPCC with a maximum temporal offset of 20 ms and a temporal resolution of 1 ms. The spectrogram matrices were computed using an FFT algorithm with a window size of 512 samples and 90 % overlap between successive windows and 'Hanning' window type.
Song diversity	Number of song types sang by a single individual during one full dawn song
Song rate	Number of songs per minute
Peak frequency	The highest amplitude frequency measured in the mean power-spectrum of each note within the trill [window size = 2048, window type = 'hanning']
Trill length	Number of notes in the trill
Trill rate	Number of notes per second always excluding the last note of the trill to avoid bias of higher rates toward shorter trills.
Playback structure	**Definition**
Song type bout	Portion of the playback stimulus that lasts 1.25 and presents several repetitions of a single song type. The number of songs per minute is established based on the song rate measured on 10 songs of the original recording.
Song bout	A complete song bout presents two song type bouts of different song types chosen for each trial. This lasts 2.5 minutes
Silence bout	A WAVE file of 2.5 minutes of silence
Playback round	This presents a song bout and a silence bout together and it is repeated 12 times in a row through the playback stimulus. The order of presentation of song and silence was randomized along the entire playback stimulus
Playback stimulus	This is the full stimulus presented in one trial lasting 1h and including 12 playback rounds.

Table 3.2 – Variation of acoustic variables in playback song due to natural variation among song types

Song rate	13.8 ± 3.9 songs per min
Peak frequency	5.1 ± 0.6 kHz
Trill rate	8.4 ± 1.9 notes per sec
Trill length	7.4 ± 3.3 notes
Song consistency	0.80 ± 0.07 SPCC similarity score

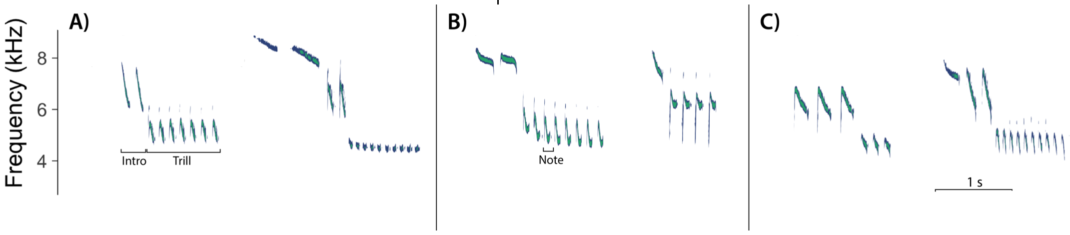


Figure 3.3 – Spectrogram of typical blue tit songs, with the time along the x-axis, shown in the one second scale in Figure 3.3C, and the frequency spectrum in the y-axis. A colour gradient of blue-green tones indicate the amplitude at each time-frequency point. The song of blue tits is usually divided into introductory notes and trill notes as indicated Figure 3.3A, where a note is defined as a continuous trace in the spectrogram, Figure 3.3B. Only the trill part was used to measure vocal consistency. Figure 3.3A-C show three different pairs of song types collected from three different males and each pair was used to build one playback stimulus to test one female

Behavioural analysis of female choice experiment

We measured the vocal response of females in spectrograms (Figure 3.2 & 3.4) produced in in Audacity software (Mazzoni and Dannenberg, 2014) (window type: ‘Hanning’, window length 1024 samples, 90% overlap and -80 dB range). Each vocalization was classified using structural and functional criteria following Bijnens and Dhondt (1984) and Gorissen and Eens (2005) (see Results). Copulation solicitation calls of blue tits are well defined in the literature (Bijnens and Dhondt, 1984; Cramp and Perrins, 1993) and have been recorded previously inside the nest box (Gorissen and Eens, 2005). In blue tits, copulation calls show a very characteristic acoustic structure beginning with an upward frequency sweep followed by quick frequency modulations between 8 and 12 kHz. Female vocal response was used to assess female preferences for song (Amy et al., 2015; Dunning et al., 2014; Salvin et al., 2018; Searcy, 1992a) and it was measured as a binomial variable if there was at least one vocalization during each treatment bout of 1.25 minutes (female response call, yes/no). The complete presentation of the playback lasted for one hour or until the female exited the box, which was easily identified in the audio recordings (Halfwerk et al., 2011) and confirmed with the video recordings. If the female entered the box again before the test was over we continued to take measurements of female vocal behaviour.

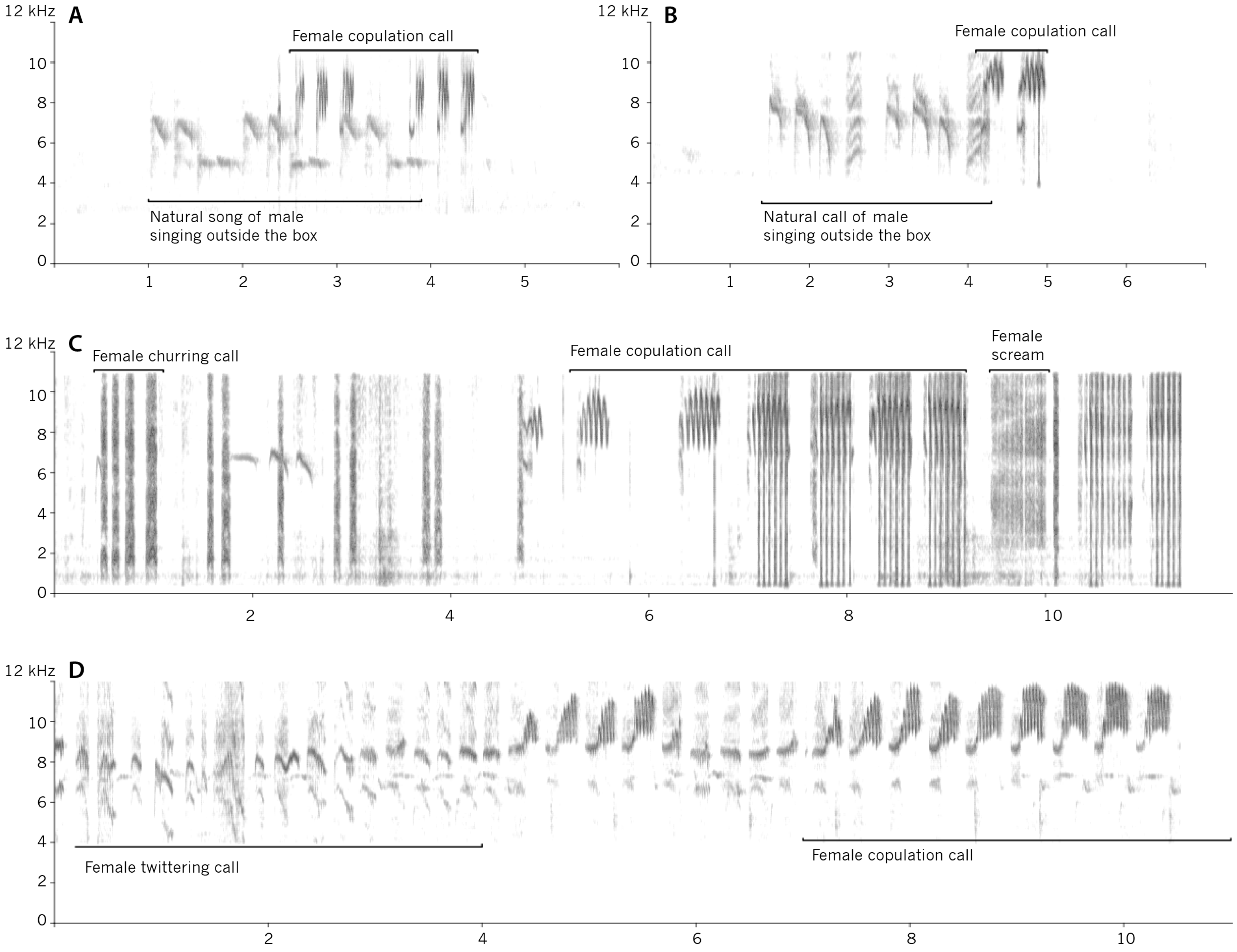

A
12 kHz
10
8
6
4
2
0
1
2
3
4
5
Female copulation call
Natural song of male
singing outside the box
B
12 kHz
10
8
6
4
2
0
1
2
3
4
5
6
Female copulation call
Natural call of male
singing outside the box
C
12 kHz
10
8
6
4
2
0
2
4
6
8
10
Female churring call
Female copulation call
Female
scream
D
12 kHz
10
8
6
4
2
0
2
4
6
8
10
Female twittering call
Female copulation call

Figure 3.4 – Vocal behaviour of female inside the nest, recorded during natural conditions, without experimental playback exposure. Figure 3.4A-B show the natural vocal interaction between the female, inside the nest, and the male, outside the nest. Figure 3.4C and 3.4B show two copulation solicitation displays recorded in two different boxes after the playback trial was over. Different types of vocalizations are labelled inside the spectrogram.

Statistical analyses

All measures are presented as mean ± one standard deviation (SD) and statistical analyses where carried out in R 3.5.1 (R Development Core Team, 2016). In the male song data, we investigated the seasonal variation of male vocal consistency during the breeding season by fitting a Generalized Additive Mixed Model (GAMM) using 'gamm4' package (Wood et al., 2017). The response variable was the vocal consistency of each male, averaged per day of recording, in function of the days to first egg. The days to first egg were measured for each individual male in relation to their own nest and female partner. We also included the male age as a fixed effect outside of the smoothing function and the individual identity as a random effect to control for pseudo-replication or repeated measurements. This model is essentially a Generalized Linear Mixed effects Model (GLMM) that applies unknown smooth functions to the linear relationship between vocal consistency and days to first egg (Schlicht and Kempenaers, 2020). This model included a group of 90 individual males with a total of 484 data points (dates of recording), with in 2.3 ± 1.2 dates per individual from 21 before until 21 after the first egg was laid and 3378 songs analysed in total (37.5 ± 25.4 songs per individual). To investigate the impact of male song on reproductive success, we fitted a Linear Mixed effects Model (LMM) using the clutch size as the response variable as a function of male vocal consistency (SPCC score), male repertoire (number of song types), the head bill length (mm) of the male and of the female partner, as a proxy for body size, the breeding year as a categorical variable and also the individual identity as a random effect to control for repeated measurements within individuals. This model included 142 data points (nests monitored) from a group of 98 individuals from which we analysed 7559 songs (50.1 ± 35.6 songs per individual) to measure vocal consistency and song type repertoire was estimated in one full dawn song for each male. We chose clutch size as our proxy for reproductive success because other measurements (i.e., fledging success) could be affected by chaotic events (weather), desertions, or predation, reducing the sample size and increasing potential noise in the data. Furthermore, we have evidence that high consistency males provide less parental care, which could be a confounding factor (Sierro in prep).

In the female choice experiment, we first compared the vocal response between the song treatments and the silence treatments using a Wilcox Signed Rank test. Female vocal response was quantified as a binomial measure for each song bout, whether or not there was at least one vocalization for every 1.25 min of each song type bout or each 2.5 min of silence bout. To explore how the female vocal response varied in relation to the acoustic variables of playback song, we fitted a binomial GLMM. The female vocal response was the response variable measured as the number of bouts for each song type where female produced one or more vocalizations (successes) vs. the number of bouts without female vocalisations (failures) as a function of vocal consistency, song rate, trill rate, trill duration and peak frequency of each song type within the playback stimulus. The model was fitted with a total of 32 observations, 30 for 15 individuals that were presented with a stimulus formed by two song types and two observations regarding two individuals that were presented with a playback stimulus formed by one song type. This adds up to 17 individuals that were exposed to 398 playback song type bouts (13.3 ± 4.33 song type bouts per individual), each bout lasting 1:15 minutes. The number of bouts that each female was exposed to varied depending on the exit time of the female. For each test, there were two song types that were different in acoustic variables along a continuous gradient, allowing us to test the response within the same female to this variation, thus increasing the statistical power. Even though each female received only two song types, there were nine different song types in 17 playback tests and hence, song type should not be a confounding factor in this analysis (Figure 3.2).

We tested the habituation effect within each playback song bout by modelling the change of female response within each song bout of 2.5 minutes, excluding the silence bouts. For this we fitted a LMM with the female vocal response as a function of song position in the song bout plus its interaction with song type switch. Song position was log transformed since it was clear that decline in vocal response was not linear along the song bout. Song type switch was a categorical variable with two levels: ‘preceded by silence’, for those song type bouts coming right after the silence treatment, or ‘preceded by song type switch’, for those song type bouts that played after a different song type. The latter category includes switches between nine different song types across tests and in all possible directions within test; i.e., from song type A to song type B but also from song type B to song type A within test (Figure 3.2).

In all models, we scaled and centred the predictor variables (Gelman, 2008) and included the individual identity as a random effect to avoid pseudoreplication and group together observations within subjects. To validate all models, we confirmed that the residuals were homoscedastic and close to a normal distribution using diagnostic plots. To find which factors were important we used an information theoretic approach computing all possible model combinations and ranking them by the Akaike information criterion corrected for small sample size (AICc) (Symonds and Moussalli, 2011). We selected all models that were within a ΔAICc < 2 to compute the full average model as the final model (Burnham and Anderson, 2002; Burnham et al., 2011). We used the relative importance of each factor in the final model together with the coefficients and estimated confident intervals (CI) with a threshold of 95% (Burnham et al., 2011; Nakagawa and Cuthill, 2007) to determine which variables had a significant effect in the response variable.

Results

We recorded 87 individual males (age between 1 and 7 years old) and analysed 30,395 notes in total (349 ± 293 notes per individual) and included breeding data of 142 nests. We found that male vocal consistency had a positive, significant effect on clutch size, after controlling for the effect of male body size, female body size and breeding year (Table 3.1 Figure 3.5). In contrast, male song repertoire was not correlated with reproductive success. We also found that vocal consistency increased significantly with age (Figure 3.6) and varied significantly along the breeding season, peaking nearly in synchrony with the onset of the receptive period of individual females, specifically 4 days before the first egg was laid (Table 3.2 Figure 3.6).

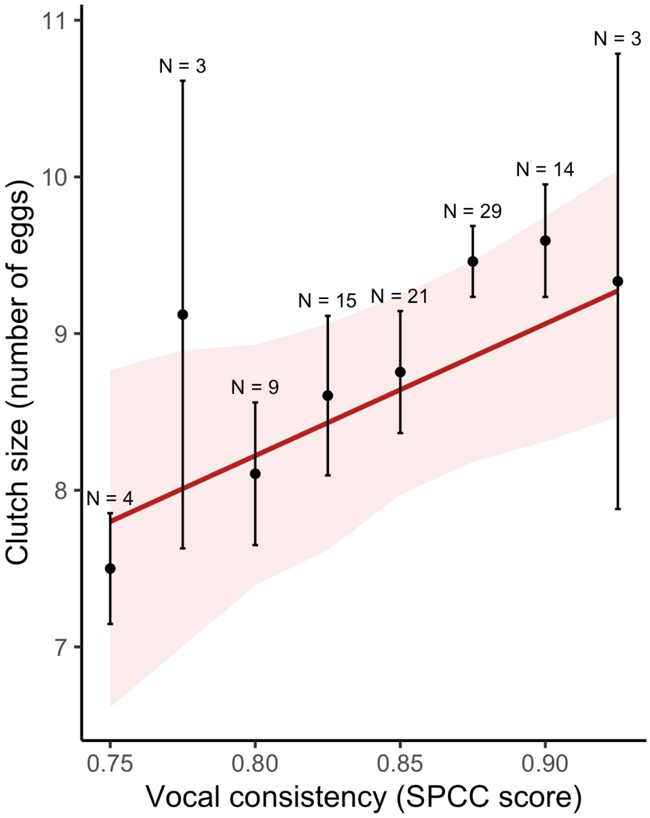


Figure 3.5 – Variation of clutch size in relation to male vocal consistency.

Points show the mean clutch size for every 0.025 interval of vocal consistency for a better visualization of the pattern. Error bars indicate the associated standard error (SE) in clutch size for each interval. Red line traces the model predicted values through the breeding season, with the associated 95% CI in light pink. The model shows that males with higher vocal consistency had significantly larger clutches, after controlling for the effects of male body size, female body size and breeding year. The number above each point shows the sample size of males included in each interval.

Table 3.3 – Results of a Linear Mixed effects Model (LMM) fitted on the reproductive success of each nest (clutch size) as a function of male vocal consistency, male song diversity, the head-bill length of the male and of the female partner, as a proxy for body size, and the breeding year as a categorical variable. For each fixed effect we show the model estimate, which represents the size and direction of the effect. We then proceed to the model selection where unimportant variables were removed, as was the case for song diversity (Supplementary material). Vocal consistency has a positive effect on clutch size, as shown by the positive parameter estimate. Then, the upper and lower 95% confidence intervals (CI) around this estimate do not overlap with zero, indicating that the effect is significantly higher. The significant effect of the year 2019 means that reproductive success was significantly higher in 2019 with respect to 2018. The table also shows the corresponding standard error (SE), the Z statistic derived from Wald tests and the relative importance of that factor in the final model. The last two columns show the marginal R squared represented as R^2_m and the conditional R-squared represented as R^2_c of the full model. These indicate the goodness of fit of the model considering just the fixed effects (R^2_m) and including both fixed and random effects (R^2_c). The values indicate a high goodness-of-fit suggesting the model is biologically meaningful.

Fixed effect	**Estimate**	**SE**	**t**	**2.5% CI**	**97.5% CI**	**R^2_m**	**R^2_c**
(Intercept)	8.70	0.247	35.27	8.22	9.18	0.30	0.60
Vocal consistency	0.440	0.162	2.70	0.122	0.755		
Year: 2019	0.785	0.318	2.468	0.156	1.40		
Year: 2020	-0.210	0.393	-0.530	-0.974	0.556		

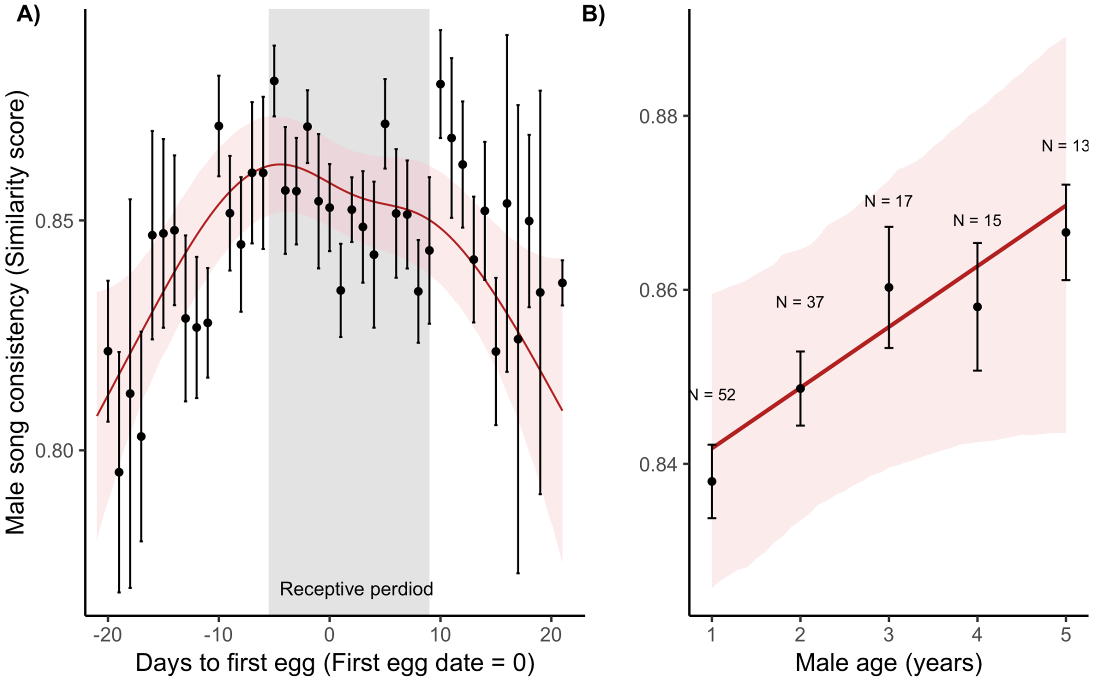


Figure 3.6 – Variation of male vocal consistency along the season (3.6A) and with age (3.6B). Points indicate the mean vocal consistency of all males recorded (per day or per age class) with the associated SE. Red lines trace the model predicted values, with the associated 95% CI in light pink. It shows that male vocal consistency peaked 4 days before egg laying starts and increased significantly with age. The greyed area in the centre of the plot covers the time period of female receptive period from 5 days before egg laying until the completion of the clutch 9 days after first egg. The oldest males recorded were 7 years old but the sample size of males over 5 years old was relatively small for each age class. Hence, we grouped all males that were 5 years or older in the same group. Note that many individuals were recorded in successive years so, in that case, they are included in more than one age class.

Table 3.4 – Output of a GAM model investigating seasonal and age variation of vocal consistency. The estimate and its associated CI show that there is a significant temporal variation of vocal consistency peaking at the start of the female receptive period. The output for the estimates of linear parameters show there is a positive, direct relationship between vocal consistency and age. The table also shows the standard error (SE), the Z statistic derived from Wald tests and the relative importance of that factor in the final model. The last two columns show the marginal R squared represented as R^2_m and the conditional R-squared represented as R^2_c of the full model.

Effects included in the smoothing function

Fixed effect	**Estimate**	**SE**	**F**	**2.5% CI**	**97.5% CI**
(Intercept)	0.458	0.04	11.355	0.379	0.537
Days to first egg	-0.24	0.04	-5.943	-0.32	-0.161

Effects included in the linear model

Fixed effect	**Estimate**	**SE**	**t**	**2.5% CI**	**97.5% CI**
(Intercept)	0.85	0.004	212.118	0.844	0.859
Age	0.01	0.004	3.02	0.004	0.018
Days to first egg	0.001	0.016	0.042	-0.038	0.033

Female choice experiment

We completed trials on 18 individuals. One trial was removed from the analysis due to a reaction from the male occupier of the box to the playback. Of the 17 remaining trials, all were performed during the egg-laying stage except 2 that were carried out in the first and second day of incubation from the 22nd of April to the 4th of May 2020. We found no evidence to suggest that the experimental treatment influenced the breeding success of the females: clutch size, brood size and fledging success were not significantly different between experimental and other females (Table 3.5). In seven trials, we recorded anti-predator hissing displays when approaching the nest to start the playback but this always stopped a few seconds after we walked away and well before the playback trial started. In our trials, 16 of 17 females responded to at least one playback round and, overall, they vocalized during 45% of the playback rounds. Of all calls produced, 85% were copulation calls, the remaining 12% were twittering, churring, chattering and screaming calls; 3% could not be categorized (Gorissen and Eens, 2005).

Table 3.5 – Comparison of breeding behaviour between females included in the experiment and females excluded from the experiment within the same population

Breeding data	Females included the experiment	Females excluded from the experiment	W	P
Clutch size	7.9 ± 1.5 eggs per nest	8.6 ± 1.7 eggs per nest	725	0.13
Brood size	4.8 ± 1.5 chicks per nest	5.8 ± 1.7 chicks per nest	560	0.17
Fledging success	4.4 ± 3.2 fledglings per nest	5.1 ± 3.4 fledglings per nest	520	0.40

We found that female vocal response during the song treatment was significantly higher than during the silence treatment (Song treatment = 0.36 ± 0.24 proportion of bouts with vocalizations, Silence treatment = 0.06 ± 0.10 proportion of bouts with vocalizations, W = 5, P = 0.001, 5% CI = -0.47, 95% CI= -0.19, Figure 3.7A). Within the song treatment, we found that vocal response was significantly positively correlated with vocal consistency but not with song rate, trill rate, trill length or peak frequency (Figure 3.7B, Table 3.3). We obtained the same results when selecting only the copulation calls from the female vocal response but with a smaller sample size. Analysis investigating habituation to playback stimuli shows a significant decline of female vocal response within a song bout (Table 3.4, Figure 3.8). Importantly, such a habituation effect was disrupted if there was a song type switch (Table 3.4, Figure 3.8). This means that female vocal response faded along the song bout but recovered partially if there was a switch in song type. In general, vocal response was significantly higher in song bouts preceded by silence than song bouts preceded by song (Table 3.4).

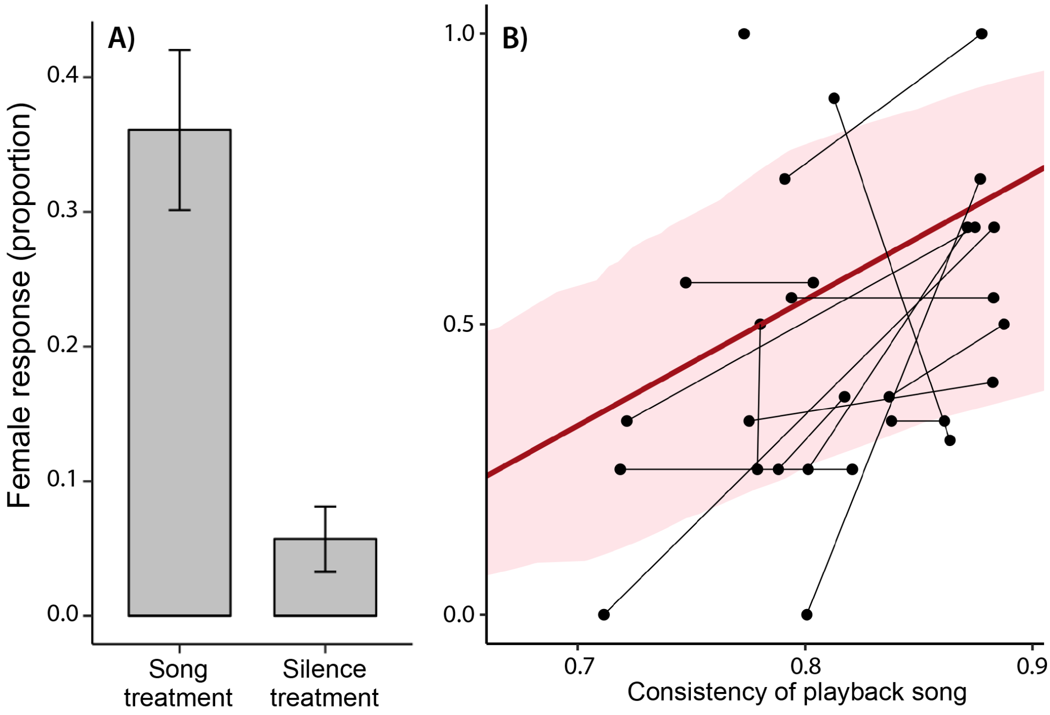


Figure 3.7 – Effect of vocal consistency on female vocal response measured as a proportion of bouts with female response out of the total. Figure 3.7A shows that female vocal response was significantly higher during the song treatment than during the silence treatment, using a Wilcox Signed Rank test (Song treatment = 0.36 ± 0.24 proportion of bouts with vocalizations, Silence treatment = 0.06 ± 0.10 proportion of bouts with vocalizations, W = 5, P = 0.001, 5% CI = -0.47, 95% CI= -0.19). Figure 3.7B shows the significant, positive correlation between female vocal response and vocal consistency of playback, measured as acoustic similarity between trill notes. Points represent raw data with black lines connecting the response of the same female to each song type within trial. The red line represents the predictions from the model together with the upper and lower 95% (CI) in red, dashed lines, keeping all other variables fixed.

Table 3.6 – Output of a binomial GLMM exploring the variation in female vocal response as a function of the acoustic variables of playback song. The effect of vocal consistency was significantly higher than zero. For each factor we present the model estimate, its corresponding standard error (SE), the lower and higher 95% CI of the estimate, the Z statistic derived from Wald tests and the relative importance of that factor in the final model. The last two columns show the marginal R squared represented as R^2_m and the conditional R-squared represented as R^2_c of the full model.

Fixed effect	Estimate	SE	Z	2.5% CI	97.5% CI	Relative importance	R^2_m	R^2_c
(Intercept)	-0.14	0.174	0.764	-0.498	0.219	-	0.39	0.51
Vocal consistency	0.505	0.178	2.713	0.14	0.87	1		
Song rate	-0.335	0.16	1.995	-0.665	0.006	0.459		
Trill length	0.269	0.163	1.57	-0.067	0.605	0.256		

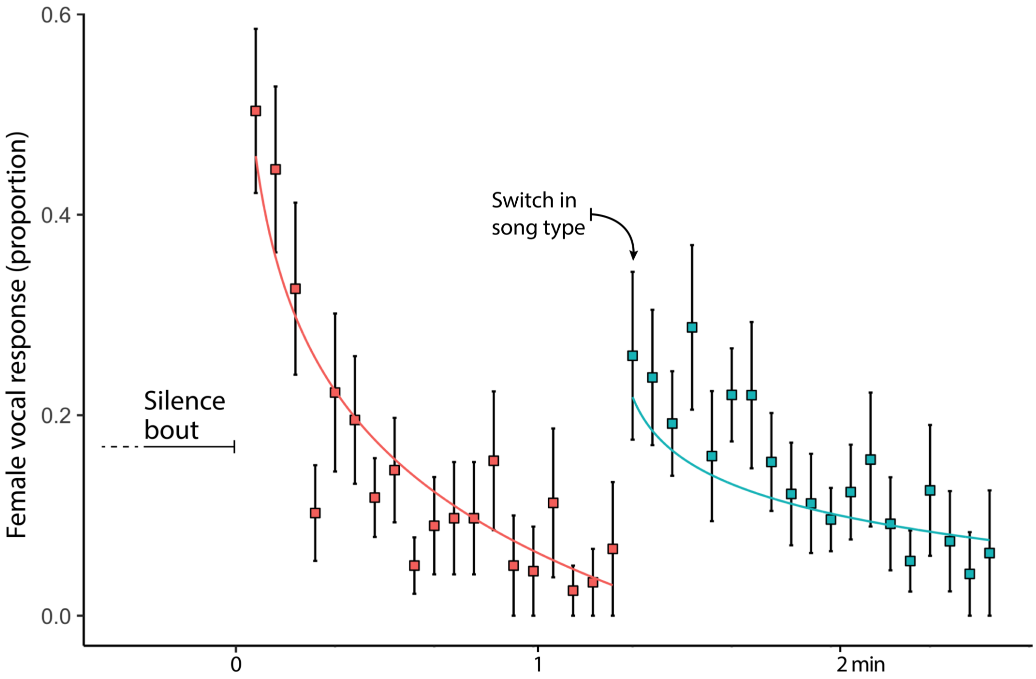


Figure 3.8 – Schematic view of the female response within song bout of 2.5 minutes that includes two song types of 1.25 min each. The y-axis shows the proportion of playback songs with female vocalizations and the x-axis shows position of each song within a song bout. Each point represents the mean female vocal response for all songs recorded in that position within the bout and its associated SE, summarizing the response of 17 females to 41 song type bouts preceded by silence (in red) and 72 preceded by a different song type (in blue). The decrease in response as the song is repeated indicates significant behavioural habituation in the female while the switch to a different song type increases dishabituation.

Table 3.7 – Summary of the results from modelling female vocal response over time for each bout of playback song that lasts 2.5 min. We fitted a LMM on the female response as a function of song position within the bout and its interaction with a song type switch. The results show that female vocal response decreased significantly with song position indicating behavioural habituation but that female response increased significantly after a song type switch. The significant effect of the interaction between song type switch and song position shows that female vocal response to a song bout that was preceded by another song type was significantly lower than to song bout that was preceded by silence. The last two columns show the marginal R squared represented as R^2_m and the conditional R squared represented as R^2_c of the full model.

Fixed effect	**Estimate**	**SE**	**t**	**2.5% CI**	**97.5% CI**	**R^2_m**	**R^2_c**
(Intercept)	0.458	0.040	11.355	0.379	0.537	0.176	0.414
Song position	-0.145	0.014	-10.689	-0.172	-0.119		
Song type switch	-0.240	0.041	-5.943	-0.32	-0.161		
Song position: song type switch	0.097	0.019	5.186	0.061	0.134		

Discussion

We show that female choice could select, simultaneously, for both consistent repetition and increased diversity in the song of male blue tits. Vocal consistency, as a signal of quality that reflects fitness, is preferred by females while song diversity prevents habituation during long, sexual displays prior to mating. These results can thus provide a key to the solution of the long-standing puzzle about the simultaneous selection for ritualized, consistent performance and increased diversity in birdsong. We also present a novel experimental design to assess female preference in the field, opening new possibilities to study intersexual selection in natural conditions.

Our results show that vocal consistency is a signal of quality under positive sexual selection in the blue tit. The relationship between male vocal consistency and clutch size could be explained by differential investment by females in response to the quality of their partner (Cunningham and Russell, 2000; Gil et al., 1999) and/or assortative mating between males with high vocal consistency and high quality females (Andersson et al., 1998). Another possibility is that high consistency males are also better parents and can raise a larger number of offspring. Our findings that male vocal consistency increases during late winter is in line with previous findings (Nottebohm et al., 1986; Smith et al., 1997a). Furthermore, the seasonal maximum of male vocal consistency coincides with the start of the female receptive. This period, before the start of egg laying, is when most extra-pair copulations take place (Magrath et al., 2009; Santema et al., 2020). Performing the highest quality song at this time of the breeding cycle indicates that vocal consistency may play an important role in (extra-pair) mate choice and mate guarding (Halfwerk et al., 2011; Kempenaers et al., 1992).

So while the data suggest that females prefer consistent, well-repeated male songs, we also found that females habituated to male songs in a matter of minutes (Dong and Clayton, 2009). Importantly, a song type switch increased female response (Krebs, 1976; Petrinovich and Patterson, 1981) which may cause increased sexual arousal, offering strong support for the habituation hypothesis as a behavioural mechanism that may promote, or maintain, the evolution of song diversity (Hartshorne, 1956; Krebs, 1976). The design of our playback method allowed us to quantify repeated habituation and dishabituation of female sexual response reporting results highly similar to previous laboratory studies on female song preference in domesticated canaries (*Serinus canaria*) (Kreutzer et al., 1994). These results suggest that there is an evolutionary conflict between diversity and repetition during song displays, as females prefer consistent songs but habituate quickly, while a switch in song type elicits an immediate response recovery. The conflicting action of female preference and habituation over repetition and diversity respectively, may explain the song structure of many bird species with small and intermediate song repertoires. We speculate, based on previous claims, that song diversity could have evolved secondarily as an ornament to enhance the courtship display of motor performance skills (Byers et al., 2010). Many birdsong experiments focus on short, male-male interactions but here, we studied long, courtship displays where the importance of habituation is likely accentuated.

The balance between repetition (ritualization) and diversity in birdsong likely depends on several factors. First, selection for male neuro-motor control, despite the need for repetition, may promote song diversity if the production of distinct elements is necessary to demonstrate neuro-motor skills, i.e., frequency jumps (Christie et al., 2004; Zollinger et al., 2008). Second, the penalty on repetition in the form of habituation will depend on the duration of sexual display as well as the time that the audience is listening (Weary and Lemon, 1988). Third, the anti-exhaustion hypothesis suggests that song diversity can arise from the need to avoid muscle fatigue after many renditions of the same task (Brumm et al., 2009; Lambrechts and Dhondt, 1988a). It is possible that a combination of different evolutionary mechanisms are responsible for the variety of singing styles (and repertoire sizes) in birdsong (Byers et al., 2010).

Conclusion

The apparent evolutionary conflict between repetition and diversity in birdsong has puzzled scientists for decades (Catchpole, 1996; Price, 2013). For the first time, within the same communication system, we show how sexual selection on song diversity and song performance might act antagonistically to produce song repertoires of intermediate size. Our empirical and observational evidence shows that vocal consistency signals male quality under targeted sexual selection, while song diversity may be under selection to reduce habituation during long sexual displays. These findings provide new evidence to reconcile two separate theoretical frameworks and shed new light into the co-evolution of multiple facets of birdsong

Chapter 4 - Trade-off between vocal performance and song length ensures honesty and has communicative value in birdsong

Manuscript in preparation for the journal *Proceedings of the Royal Society B: Biological Sciences*, co-authored by Selvino R. de Kort and Ian R. Hartley

Abstract

Birdsong is a crucial signal involved in the access to breeding resources. But not all songs are similar and song performance plays an important role during intra- and intersexual interactions. Increased vocal performance has been associated with increased fitness but it is possible that sustained production of high-performance vocalizations is limited due to exhaustion, especially in trilled songs that have many repetitions of the same sound. Based on this premise, we present the stable performance hypothesis, which predicts that birds adjust the length of song to avoid a decline in vocal performance within songs. Across 31 songbird species, we found that vocal performance, including vocal consistency, vocal power and sound density, was stable (or increased) along the duration of the trill to then decline abruptly towards the end, irrespective of trill length. The sharp decline in vocal power (lower sound amplitude) and sound density (longer inter-note gaps) can be indicator of exhaustion. Based on these findings, we designed a double playback experiment to investigate the communicative value conveyed in this performance trade-off using a wild population of blue tits (*Cyanistes caeruleus*). We found that long songs elicited a stronger response than short songs during simulated territorial interactions. But, when presented with long songs with a decline in performance, male blue tits showed a stronger response to short songs of high, stable performance. This penalty on the decline in vocal performance supports the fundamental role of this feature in birdsong and provides a mechanism that makes song length an honest signal of quality. The study highlights the multifaceted nature of birdsong, where multiple aspects can interact causing performance constraints that are meaningful during communication.

Introduction

Birdsong is famous for the diversity, variety and complexity of sounds within and between species, but the overwhelming majority of species have one thing in common; repetition of songs or parts thereof (Botero and de Kort, 2013; Price, 2013). Even though each species has its own, specific singing style, repetition is common at various levels. Many species repeat one note in quick succession commonly referred to as a trill, or they repeat phrases or the entire song. It is still unclear why birds repeat but it has been proposed as a way to demonstrate motor performance skills that honestly signal quality differences between individuals (Botero and de Kort, 2013; Price, 2013).

As for other ritualized displays (Fusani et al., 2014), theory suggests the physiological demand required for singing indicates the general quality of the neuro-motor system of the individual (Botero and de Kort, 2013; Byers and Kroodsma, 2009; Cardoso, 2017; Dubbeldam, 1997). In birds, higher vocal performance correlates with higher social status (Botero et al., 2009; Christie et al., 2004), increased reproductive output (Byers 2007; Kempenaers et al. 1997; Poesel et al. 2001; Sierro et al. in prep) , longevity (de Kort et al., 2009b) and sexual attractiveness (Ballentine et al. 2004; Wasserman and Cigliano 1991; Sierro et al. in prep) . The multidimensionality of song displays means that various aspects of performance can have interacting effects, which may cause performance constraints (Podos, 1997). In many species that produce trills, there seems to be a trade-off between the length of gap between notes (i.e., rate of notes per second) and the extent of mechanical reconfiguration in the vocal apparatus between notes, (i.e., frequency bandwidth of trill notes) (Podos and Nowicki, 2004). This trade-off implies a physiological limitation, as fast rates restrict the bandwidth and vice versa, measured by the vocal deviation (Podos, 1997; Podos et al., 2016).

Another major feature of vocal performance is vocal consistency, which is the ability to produce highly stereotyped repetitions of the same sound. Consistent repetition of song components within a display may provide an honest signal of neuro-motor skills (de Kort et al., 2009b; Zahavi, 1980). Increased vocal consistency has been associated with age (de Kort et al., 2009b; Rivera-Gutierrez et al., 2012; Vehrencamp et al., 2013), higher social status and increased fitness (Botero et al., 2009). During territorial interactions, several studies showed that birds can discriminate between high and low consistency songs (de Kort et al. 2009, Porcedda unpublished, Riviera-Gutierrez et al. 2011). Even if vocal consistency has been demonstrated to be biologically meaningful in itself (Sakata and Vehrencamp, 2012), it can be affected by other performance parameters such as the trade-off between vocal deviation and vocal consistency (Vehrencamp et al., 2013).

When birds repeat, it might be expected the number of repetitions to impose a trade-off on general aspects of vocal performance within song. Some field and laboratory studies have shown that singing is costly (Eberhardt, 1994; Thomas, 2002), although others have shown that the metabolic cost of the act of singing is small (Oberweger and Goller, 2001; Ward et al., 2003). Nonetheless, singing long songs may be challenging. During trills, songbirds are able to take short breaths in between notes, i.e., mini-breaths, through a precise coordination of syringeal and respiratory muscles (Hartley and Suthers, 1989). Such mini-breaths have been shown to replenish air pressure during song but oxygen debt might occur, as the mini-breaths might not be sufficient to renew the air in all areas where gas exchange takes place (Suthers and Goller, 1997). Therefore, producing long trills can cause oxygen deprivation (Suthers and Goller, 1997) and the repetition of the same motor pattern can cause muscle fatigue (Lambrechts and Dhondt, 1988b). The duration of the song has been shown to be informative on the quality and physical state of an individual (Galeotti et al., 1997; Lambrechts and Dhondt, 1986) and it is under positive sexual selection in several species (Gentner and Hulse, 2000; Gentner et al., 2001; Kempenaers et al., 1997; Mennill et al., 2006; Nolan and Hill, 2004). Furthermore, there is evidence that longer songs elicit a stronger response during simulated territorial interactions (Lattin and Ritchison, 2009; Linhart et al., 2012). But producing long trills could cause a decline in the quality of the performance due to exhaustion and this would mean a trade-off between high vocal performance and trill length. In fact, vocal drift, the increase in silent pauses between songs, has been shown to increase after many repetitions of the same song structure, possibly as a result of muscle exhaustion from repetition (Lambrechts and Dhondt, 1988b).

In this hypothetical scenario, a songbird faces two options while singing: produce long trills despite the decrease in vocal performance or stop the trill before the decline in performance is apparent and, therefore, produce shorter trills. Here, we propose the stable performance hypothesis that predicts a trade-off between trill-length and quality of vocal performance. Given such a trade-off, it also predicts that songbirds should keep vocal performance stable throughout the trill and avoid a decline in performance by stopping the song once vocal performance starts to drop. This hypothesis takes into account that: 1) vocal performance is crucial in bird communication and 2) producing long songs has communicative value associated with increased sexual attractiveness or intimidation, but 3) increased song length at the expense of lowered vocal performance is detrimental because vocal performance is an honest signal of quality.

The stable performance hypothesis predicts that birds adjust the length of their trills to retain high vocal performance throughout the trill; a drop in vocal performance therefore ends the trill regardless of trill length. Alternatively, birds might emphasize trill length over performance quality, so we might see a gradual decrease in vocal performance as the trill progresses. To test this, we analysed the trilled songs of 31 bird species from a wide range of distinct taxa to investigate the change in vocal performance along the trill. These observations suggest there is a performance constraint between trill length and vocal performance. We then conducted two playback experiments to assess the communicative value of such performance trade-off during simulated agonistic interactions using the blue tit as model species.

Methods

Multi-species analysis of vocal performance within trills

We collected audio recordings from Xeno-Canto repository (www.xeno-canto.org, recording category: ‘song’, recording quality = ‘A’), for 31 bird species that have been the subject of birdsong research; this included 16 different families (see Table 4.1 for the selection criteria). Choosing locations that were far away (i.e., different countries), we chose three to five recordings of each species assuming these were separate individuals that could be considered independent data points. For each individual we selected a maximum of five trills per individual to analyse the change in vocal performance along the trill. A trill was defined as any repetition of the same note that was separated from other notes by 1) a change in the note type or 2) a silence pause between notes (inter-note gap) of more than 2 times the inter-note gap of adjacent trill notes. We marked the start and end points of each note using the cursor on the spectrogram in Audacity (Mazzoni and Dannenberg, 2014) (window type: 'Hanning', window length 1024 samples, 90% overlap and -80 dB range). We then proceeded with the bioacoustics analysis conducted in R software (Ligges, 2013; R Development Core Team, 2016; Sueur et al., 2006). As three main aspects of vocal performance, we measured vocal consistency (*sensu* de Kort et al., 2009), sound density (*sensu* Holveck et al., 2008) and vocal power (see also 'percentage peak performance' for a related measure, Forstmeier et al., 2002). Vocal consistency was calculated as the acoustic similarity of each rendition to all other renditions within the same trill, using a spectrogram cross-correlation algorithm (SPCC) (Botero et al., 2009; Clark et al., 1987; Coleman et al., 2007). We calculated the maximum correlation of every pairwise SPCC with a maximum temporal offset of 20 ms and a temporal resolution of 1 ms. The spectrogram matrices were computed using a Fast Fourier Transformation algorithm with a window size of 512 samples and 80% overlap between successive windows and a 'Hanning' window type. Sound density was measured as the ratio between note length from the sum of note length plus the preceding inter-note gap (Figure 4.1). Vocal power was measured as the maximum amplitude of each note relative to the maximum amplitude of the trill (Figure 4.1). For this, we calculated the amplitude envelope of each trill using a smoothing function (window length: 128 samples, overlap:10 samples). In this case, since the amplitude of the note directly influenced manual measures of the start and end of the note (Brumm et al., 2017), we split the trill in the middle points of the inter-note gaps and then measured the maximum amplitude of each note, thus obtaining a proxy of vocal power for each note that was independent of measures of note length and note gap, used to estimate sound density. We measured vocal power and sound density as important aspects of vocal performance that may also provide an indicator of exhaustion along the trill. In all trills, we identified the position of the note in relation to the last note of the trill.

Table 4.1 – List of species selected for the analysis of change in vocal performance within trills. The family, number of individuals analysed* and the file names of the recordings downloaded from Xeno-Canto are shown. We searched for bird species with a singing style that included trilled songs, where the same note is repeated within the trill with similar spectro-temporal characteristics. For this we focused on several families with species that are common across the European continent. We also selected five non-european species that have been subjects of bioacoustic research. For each family, we aimed to select at least two or more species when possible. In some cases, we chose songs that repeat two notes in alternation (i.e., the great tit, *Parus major*). In that case, we measured the vocal performance of each note type within the trill, as if there were two trills in alternation.

Species name	Family	No. ID*	File name ID
Acrocephalus paludicola	*Acrocephalidae*	5	XC105381, XC105381, XC113298, XC183177, XC183177, XC327982, XC327982, XC507639, XC507639, XC561499, XC561499
Acrocephalus scirpaceus	*Acrocephalidae*	3	XC574161, XC574161, XC611025, XC611025
Anthus spinoletta	*Motacillidae*	3	XC429383, XC429383, XC476806, XC569373, XC569373, XC572360, XC608236, XC608236
Anthus trivialis	*Motacillidae*	5	XC105562, XC105562, XC551872, XC551872, XC553209, XC553209, XC567039, XC567039, XC632750, XC632750
Carduelis carduelis	*Fringillidae*	5	XC359018, XC359018, XC547421, XC547421, XC572995, XC572995, XC613631, XC613631, XC627097, XC627097
Carduelis chloris	*Fringillidae*	3	XC637320, XC637320, XC637874, XC637874, XC642456, XC642456
Cettia cetti	*Cettiidae*	5	XC463515, XC463515, XC502818, XC502818, XC614853, XC614853, XC622408, XC622408, XC623348, XC623348
Cyanistes caeruleus	*Paridae*	5	Own recordings from our study population
Emberiza cirlus	*Emberizidae*	5	XC131708, XC131708, XC132000, XC132000, XC367918, XC367918, XC470752, XC470752, XC632510, XC632510
Emberiza citrinella	*Emberizidae*	2	XC304018, XC304018, XC321553, XC321553
Emberiza hortulana	*Emberizidae*	1	XC487269, XC487269
Fringilla coelebs	*Fringillidae*	5	XC291601, XC291601, XC323914, XC323914, XC565441, XC565441, XC603767, XC603767, XC637055, XC637055
Hippolais polyglotta	*Acrocephalidae*	1	XC251956, XC251956
Luscinia megarhynchos	*Muscicapidae*	5	XC379584, XC379584, XC563326, XC563326, XC602143, XC602143, XC635145, XC635145, XC639587, XC639587
Luscinia svecica	*Muscicapidae*	1	XC402262, XC402262, XC402262

Malurus cyaneus	*Maluridae*	5	XC172263, XC172263, XC183569, XC183569, XC340990, XC340990, XC571184, XC571184, XC98477, XC98477
Melospiza melodia	*Emberizidae*	1	XC638312, XC638312
Mimus polyglottos	*Mimidae*	4	XC608486, XC608486, XC608489, XC608489, XC627759, XC627759, XC629346, XC629346
Motacilla cinerea	*Motacillidae*	5	XC141165, XC141165, XC343363, XC343363, XC483657, XC483657, XC608325, XC608325, XC637655, XC637655
Parus major	*Paridae*	5	XC621072, XC621072, XC621329, XC621329, XC624886, XC624886, XC635972, XC635972, XC72819, XC72819
Phoenicurus ochruros	*Muscicapidae*	3	XC367673, XC367673, XC516352, XC516352, XC633412, XC633412
Phylloscopus bonelli	*Phylloscopidae*	5	BR_JS021.txt, BR_JS022.WAV, XC482637, XC482637, XC574344, XC574344
Poecile montanus	*Paridae*	3	XC623386, XC623386, XC640078, XC640078, XC640081, XC640081
Poecile palustris	*Paridae*	5	XC355315, XC355315, XC361421, XC361421, XC618397, XC618397, XC628160, XC628160, XC629240, XC629240
Remiz pendulinus	*Remizidae*	3	XC131669, XC131669, XC35982, XC35982, XC463105, XC463105
Setophaga pensylvanica	*Parulinae*	2	XC253752, XC253752, XC296176, XC296176
Sitta europaea	*Sittidae*	5	XC378852, XC378852, XC535505, XC535505, XC602179, XC602179, XC602184, XC602184, XC637762, XC637762
Sylvia borin	*Sylviidae*	2	XC631570, XC631570, XC633006, XC633006
Thryophilus pleurostictus	*Troglodytidae*	2	XC578741, XC578741, XC578743, XC578743
Troglodytes troglodytes	*Troglodytidae*	5	XC310189, XC310189, XC397554, XC397554, XC49824, XC49824, XC574762, XC574762, XC581790, XC581790
Turdus philomelos	*Turdidae*	5	XC310540, XC310540, XC571714, XC571714, XC573467, XC573467, XC580625, XC580625, XC599057, XC599057

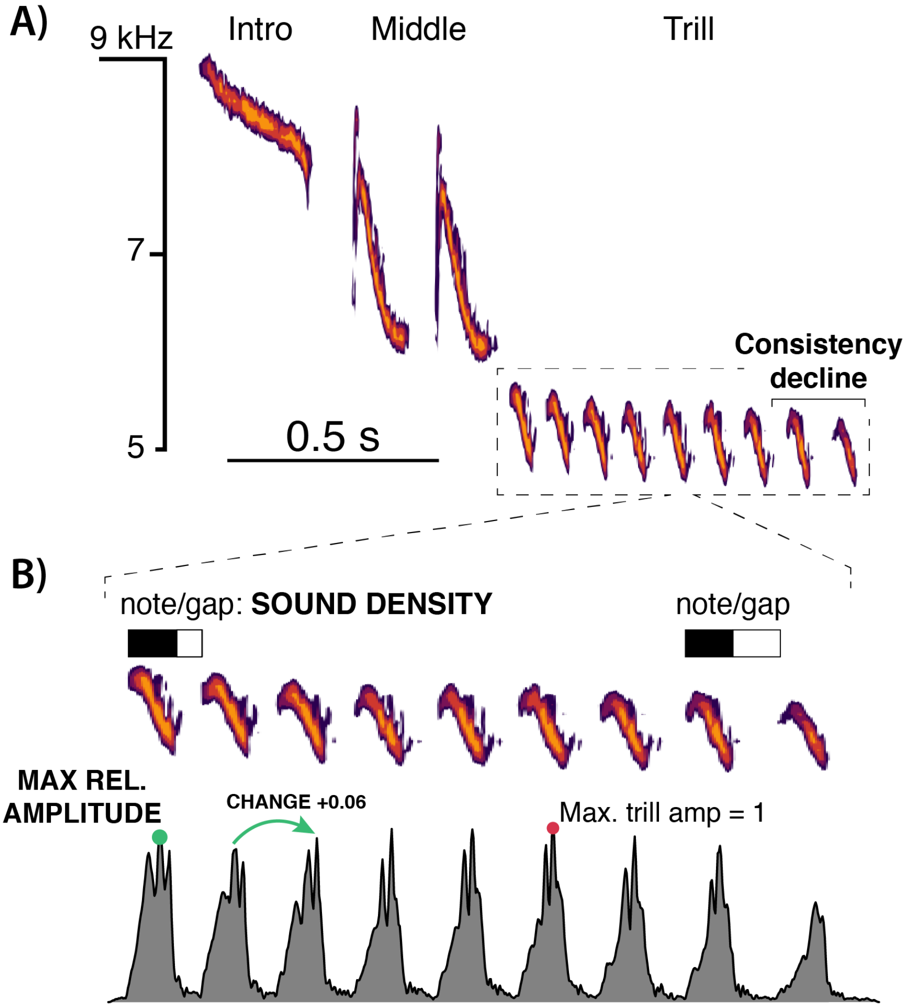


Figure 4.1 – Example of a spectrogram of a natural blue tit song (Figure 4.1A) with annotations of the performance variables as measured for this study. The trill was selected from the song (Figure 4.1B), as in the other 30 species, to conduct the analysis of vocal performance within trills. Both figures show the decline in vocal consistency, note gap ratio and maximum relative amplitude at the end of the trill.

Blue tit population and song analysis

We conducted a field study on a wild population of blue tits breeding in nest-boxes at Lancaster University campus, UK (Mainwaring and Hartley, 2009). Every year, adult breeding pairs were caught and ringed with a unique combination of three coloured rings and one numbered metal ring for identification in the field (Redfern and Clark, 2001). During the breeding season, individuals were sexed in the hand, based on the presence of a brood patch, present only in females, or cloacal protuberance, present only in males (Svensson, 1992). Blue tit song is usually composed of a few introductory notes followed by a trill (Figure 4.1A) (Bijnens and Dhondt, 1984; Cramp and Perrins, 1993; Latimer, 1977). In this species, each individual presents several stereotyped song structures referred to as song types and these are usually comparable between individuals within a population. From January to May, 2018 to 2020 (egg laying starts in April), song recordings were collected daily using a Marantz PMD661 recorder (48kHz sampling rate and 24-bit depth) and a Sennheiser ME67 microphone. We collected song recordings following linear transects since nest boxes were placed in lines within strips of woodland. We then analysed these songs following a similar procedure as for the multi-species song recordings downloaded from Xeno-Canto (see above).

Blue tit playback experiments

We carried out two song playback experiments in the field, to study: 1) the communicative value of song length, hereafter the 'song length experiment', and 2) the impact of a decline in vocal consistency along the trill, hereafter the 'consistency decline experiment'. In the song length experiment, we compared the reaction of wild male blue tits to two treatments: a short-stable treatment that was a song with a trill of six notes of high, stable vocal consistency, and a long-stable treatment, a song with a trill of twelve notes of high, stable vocal consistency (Figure 4.2A). This was within the natural range of trills found in our population (see Results). In the consistency decline experiment we presented wild males with two treatments: a short-stable treatment, a song with a 6-note trill of high, stable vocal consistency and a long-decline treatment, a song with a 12-note trill with the first 6 notes of high, stable vocal consistency and lowered vocal consistency from the 7th rendition to the end of the trill (Figure 4.2B).

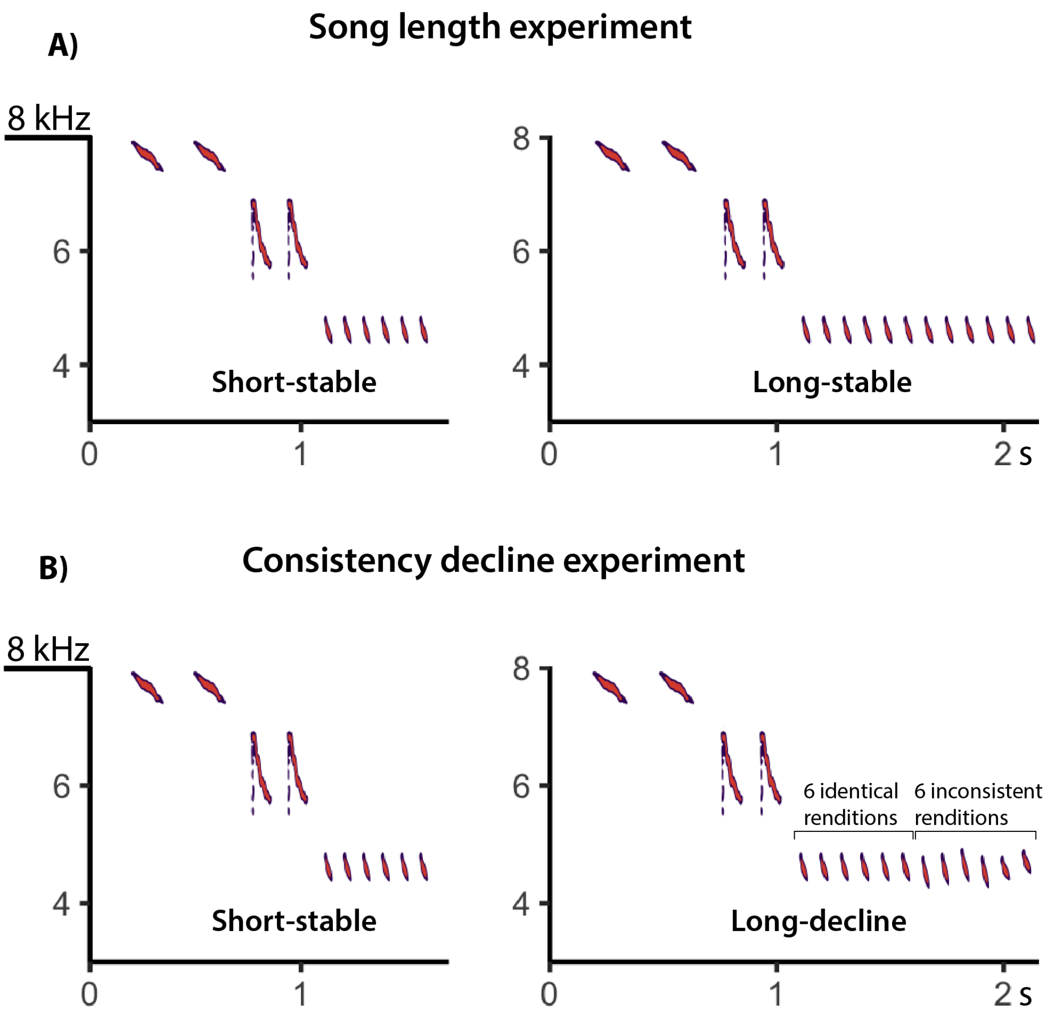


Figure 4.2 – Examples of playback song treatments in both experiments. Figure 4.2A show the short-stable treatment and the long-stable treatment for the song length experiment. Figure 4.2B show the short-stable treatment and the long-decline treatment for the consistency decline experiment. Note that the short-stable treatment is identical in both experiments. The first six renditions of the short-stable treatment and the long-decline treatment were always identical, only the additional six notes at the end were lowered in consistency by introducing small variations in the individual notes, based on natural range of inconsistencies within trill in note duration, note peak frequency and note bandwidth. The long-decline treatment represents a bird that, despite the decline in vocal consistency in the 7th rendition, would keep singing up to 12 renditions.

For both experiments, the playback stimuli were synthesized using the natural song as a template. Synthetic stimuli allowed us to manipulate several acoustic parameters and "avoids many of the design problems arising from the multidimensional nature of natural signals and variations in recording quality" (Gerhardt, 1992). Previously, we had shown that blue tits respond similarly to natural and synthetic song (Sierro unpublished). Other studies have also used synthetic song to simulate territorial interactions in blue tits (Poesel et al., 2004) and other species (Bremond, 1976; Fishbein et al., 2018). We chose a song type in our population that presented long trills and had tonal sounds, produced when upper harmonics are filtered in the vocal tract (Nowicki, 1987; Nowicki et al., 1989) (Figure 4.1). This type of sound can be closely recreated by a synthesizer. This song type had three distinct parts: the introduction, the middle and the trill (Figure 4.1A). To create the synthetic stimuli, we selected 32 different individual male blue tits that were recorded previously and, for each individual, we tracked the fundamental frequency of one introduction, one middle and one trill note to recreate the same shape synthetically for each note. Preserving the spectrographic shape of each note, we shifted each note in the spectrum to match the peak frequency of the mean peak frequency of the population for each note type. Furthermore, we stretched or shrank the note to match the mean note duration of the population and finally we adjusted the trill rate to the mean trill rate of the population (see Results). In this way, all playback stimuli recreated the natural song of 32 different individual males, avoiding pseudo-replication, but they were all re-adjusted to the same note duration, note peak frequency and trill rate, removing these potential confounding factors from the experiment (Cardoso, 2017; Gerhardt, 1992). All songs, regardless of treatment, had two introductory and two middle notes prior to the trill.

Song length experiment

Each trial in the song consistency experiment required two playback stimuli: a short-stable, and a long-stable stimulus. These were paired so that both stimuli within one trial were synthesized based on the same natural recording. In the short-stable treatment, the trill had six notes whereas in the long-stable treatment it had twelve notes (Figure 4.2A) and in both cases the trill had the highest possible consistency since notes were identical within trills. A complete stimulus had ten songs of one of the treatments, delivered at a rate of one song every six seconds, making one minute of playback song.

Consistency decline experiment

In the consistency decline experiment, we created the short-stable treatment using the same method as the short-stable treatment in the song length experiment (i.e., a song with a six-note trill of high, stable vocal consistency). To build the song for the long-decline treatment, we added six renditions to the short-stable treatment and created a song with a trill of 12 notes. The final six notes of the long-decline trill were synthesized including small variations in each rendition respect to the model note, to mimic a decrease in consistency in the second half of the trill. The variation introduced was not arbitrary (Rivera-Gutierrez et al., 2011) but based on the natural variation of duration, peak frequency and bandwidth for notes within each trill recorded, using the range between the 75% and 95% quantiles of the distribution (see Results). Each of the ten songs of a playback stimulus was individually distinct since the variation introduced to create inconsistencies was fully randomized in each song. Again, a complete stimulus had ten songs of one of the treatments, delivered at a rate of one song every six seconds, making one minute of playback song.

Field trials

The two experiments were performed from February to April 2020. A trial consisted of two consecutive tests, one for each treatment of one experiment. Playback stimuli were played using a smartphone (Iphone 5s) and a wireless, bluetooth speaker (Ultimate Ears BOOM 3, 10W, frequency response 90 Hz – 20 kHz). Playback stimuli were stored in the smartphone internal memory with an alias filename hiding the treatment so the field observer was blind to the treatment and experiment type. Before each trial, the speaker was placed within 10 m of the nest box occupied by a focal bird, and a 4 m perimeter was measured around the speaker, establishing visual landmarks in the horizontal and vertical axis. Often, birds approached and mobbed the human observer upon arrival in the territory. To allow the subject to resume baseline behaviour, we waited several minutes until such mobbing behaviour stopped before the trial commenced. We began by locating the breeding male, identifiable by colour rings, and recording the vocal behaviour of the subject using the same audio recording equipment as above. Simultaneously, the approach behaviour was registered by dictating each time that the bird entered and left the 4 m perimeter around the speaker using a separate dictaphone recorder. We recorded one minute of natural baseline behaviour (control phase), followed by the first treatment stimulus (playback phase), and continued recording for another minute after the playback ended (post-playback phase). After the first test ended, we waited for two minutes and then conducted the second test with the other treatment using the same procedure. This completed the trial of one experiment, for example, the presentation of the short-stable and long-stable treatments of the song length experiment. The order of treatment presentation was alternated within each trial to avoid an order effect. The same focal bird was part of two trials, one for each experiment. Thus, one week later we returned to the focal bird's territory and repeated the procedure using the two treatment stimuli of the other experiment. The order of experiments was also alternated to control a possible order bias.

Each test (two tests per trial), was split into three phases of one minute each: control, playback and post-playback. We then divided each phase into 10 intervals of 6 seconds each, based on the song interval of the playback stimuli, to study the behavioural response at a higher temporal resolution (Sierro et al., 2020). For each interval, we used Audacity (Mazzoni and Dannenberg, 2014) to measure the sum of durations of all vocalizations per interval and the time spent within a 4 m radius of the speaker.

Statistical analysis of consistency decline

All measures are presented as mean ± one standard deviation, unless otherwise indicated. Statistical analyses were carried out in R software (R Development Core Team, 2016). We used the multi-species song data to investigate the change in vocal performance along the trill. Production and perception of vocal performance can vary depending on song structure and acoustic properties (Cardoso, 2017). Therefore, we measured the change in vocal performance (consistency, power and sound density) by subtracting vocal performance of each note to the previous note within a trill, to obtain a standardized measure for all species. We fitted three Linear Mixed-Effects models (LMMs), using the change in vocal consistency, vocal power and sound density of each note within a single trill as response variables as a function of the note position relative to the last note of the trill. We included the species as a random effect. We selected trills with at least 5 repetitions. Since the number of individuals and species was much lower for trills longer than 10 notes, we selected only the last 10 notes of all trills. This led to a sample size of 3,600 notes in 432 trills from 112 individuals in 31 different species (3.9 ± 1.9 trills per individual). The note position was a categorical variable of 11 levels, ten levels for each position in the last 10 notes of the trill and one extra level labelled as "control". The control was an extra, 'dummy' note introduced in every trill that was set to zero change in vocal performance, representing a 'fake' note that suffered no change in vocal performance within the trill. By setting the control level as the reference level in the model, we could test whether there was a significant change in vocal performance at each position of the trill, relative to zero (no change).

Statistical analysis of playback experiment

For each playback experiment, we fitted two LMMs. In the first we used the approach behaviour, the time spent within 4 m radius, as the response variable and in the second we used the vocal behaviour, the duration of all vocalizations per each 6-second interval, as the response variables. In both models the explanatory variables were the full interaction between treatment (short-stable vs. long-stable or short-stable vs. long-decline), phase (pre-, playback and post-) and interval, a numerical variable from 1 to 30 intervals of 6 seconds each per test, 10 intervals per phase. The order of treatment presentation was included outside of the interaction to control for order effects and the individual identity was included as a random effect to control for repeated measurements. In the playback variable, we set the playback phase as the reference level to compare the response in relation to the control phase and the post-playback phase. We considered there was a significant effect on the response variable if the 95% Confidence Intervals (CI) did not overlap with zero.

Results

Change in vocal performance along the trill – multi-species analysis

We analysed the change in vocal performance of 3600 notes in 432 trills from 112 individuals in 31 different species (3.9 ± 1.9 trills per individual). We found a significant decline in vocal performance in all three measures but only in the last note of the trill, with the exception of vocal consistency which declined significantly in the second to last note, i.e., negative change in vocal performance dropping significantly below zero (Figure 4.3, Table 4.2-4.4). From the -10th to the -3th rendition, vocal performance was either stable (no change) or increased (positive change) but did not decrease significantly in any position.

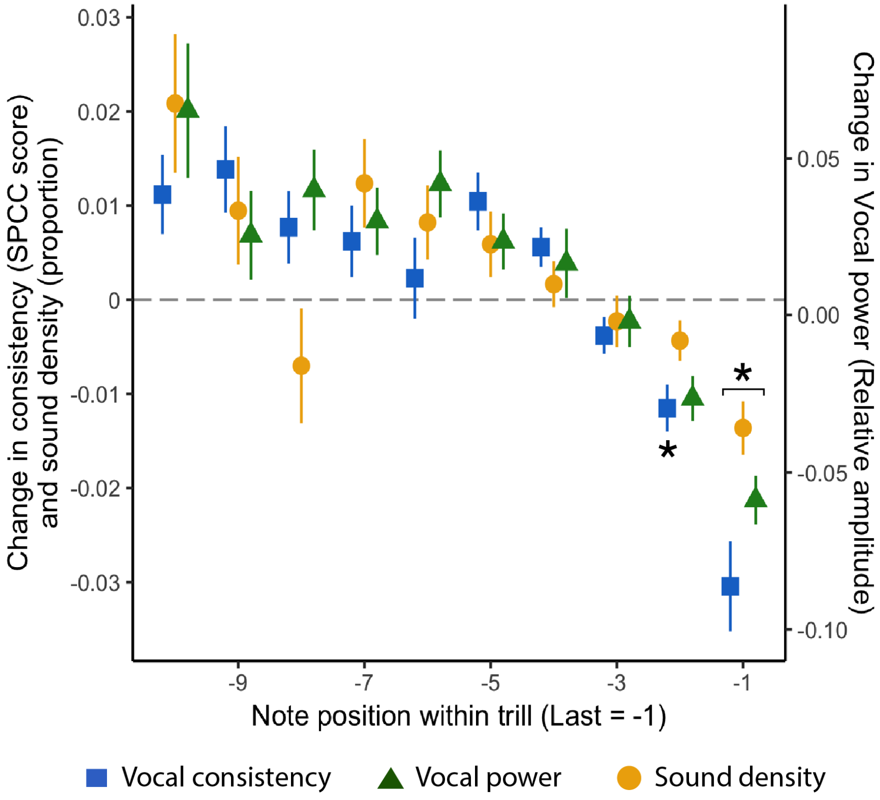


Figure 4.3 – Change in vocal performance along the trill, analysed in 432 trills of 31 different species. Each point is the mean value of all notes in that position, in relation to the last note of the trill, with the associated standard error. The Y-axis represent the change in vocal performance by subtracting the value measured in each note from the preceding note within the trill. The star indicates a significant decline in vocal performance (the change is negative, significantly lower than zero) in the last rendition of the trill. In all three parameters there is also a decline in the last 2 or 3 notes but only vocal consistency showed a significant decline in the second to last note within trills. Note that the change is

either zero or positive in all other notes before the 3rd from the last note, indicating that the vocal performance either stays stable or increases but never declines until the end.

Table 4.2 – Estimated coefficients from the model describing the change in vocal consistency along the trill for all the species. The "control" level in note position represents a "dummy", false note in every trill with no change in vocal consistency (control = 0, no change). All other levels, from the 10th note to the last note of the trill, are compared to the control level and, therefore, the statistical differences are in relation to "no change". Positive values imply an increase in vocal consistency whereas negative values represent a decline in vocal consistency in that position. For every estimate, we present the standard error (SE), the associated t statistic, the high and low 95% confidence interval and the R^2 of the model, both the marginal (without) and the condition (with) the random fixed effects (species).

Fixed effects	**Estimate**	**SE**	**t value**	**5% CI**	**95% CI**	**R^2_m**	**R^2_c**
Note position: Control	0	0.003	0.004	-0.006	0.006	0.141	0.155
Note position: -10	0.013	0.006	2.003	0	0.025		
Note position: -9	0.015	0.006	2.554	0.003	0.026		
Note position: -8	0.008	0.005	1.605	-0.002	0.019		
Note position: -7	0.007	0.005	1.363	-0.003	0.016		
Note position: -6	0.003	0.005	0.553	-0.007	0.012		
Note position: -5	0.011	0.004	2.354	0.002	0.019		
Note position: -4	0.006	0.004	1.283	-0.003	0.014		
Note position: -3	-0.004	0.004	-0.889	-0.012	0.005		
Note position: -2	-0.011	0.004	-2.697	-0.02	-0.003		
Note position: -1 (Last note)	-0.03	0.004	-7.138	-0.039	-0.022		

Table 4.3 – Estimated coefficients from the model describing the change in sound density along the trill for all the species. The “control” level in note position represents a “dummy”, false note in every trill with no change in sound density (control = 0, no change). All other levels, from the 10th note to the last note of the trill, are compared to the control level and, therefore, the statistical differences are in relation to “no change”. Positive values imply an increase in vocal consistency whereas negative values represent a decline in vocal consistency in that position. For every estimate, we present the standard error (SE), the associated t statistic, the high and low 95% confidence interval and the R^2 of the model, both the marginal (without) and the condition (with) the random fixed effects (species).

Fixed effects	**Estimate**	**SE**	**t value**	**5% CI**	**95% CI**	R^2_m	R^2_c
Note position: Control	0.965	3.609	0.267	-6.065	8.031	0.059	0.138
Note position: -10	22.548	6.423	3.511	9.997	35.064		
Note position: -9	10.643	5.992	1.776	-1.063	22.322		
Note position: -8	-5.598	5.458	-1.026	-16.274	5.036		
Note position: -7	12.602	5.054	2.493	2.739	22.461		
Note position: -6	8.397	4.774	1.759	-0.923	17.709		
Note position: -5	6.145	4.63	1.327	-2.895	15.173		
Note position: -4	1.839	4.486	0.41	-6.922	10.585		
Note position: -3	-2.283	4.385	-0.521	-10.839	6.273		
Note position: -2	-4.334	4.385	-0.988	-12.89	4.222		
Note position: -1 (Last note)	-13.612	4.385	-3.104	-22.167	-5.056		

Table 4.4 – Estimated coefficients from the model describing the change in vocal power along the trill for all the species. The “control” level in note position represents a “dummy”, false note in every trill with no change in vocal power (control = 0, no change). All other levels, from the 10th note to the last note of the trill, are compared to the control level and, therefore, the statistical differences are in relation to “no change”. Positive values imply an increase in vocal consistency whereas negative values represent a decline in vocal consistency in that position. For every estimate, we present the standard error (SE), the associated t statistic, the high and low 95% confidence interval and the R^2 of the model, both the marginal (without) and the condition (with) the random fixed effects (species).

Fixed effects	**Estimate**	**SE**	**t value**	**5% CI**	**95% CI**	R^2_m	R^2_c
Note position: Control	0.001	0.011	0.111	-0.02	0.022	0.116	0.236
Note position: -10	0.085	0.017	5.132	0.053	0.117		
Note position: -9	0.045	0.015	3.004	0.016	0.075		
Note position: -8	0.057	0.014	4.006	0.029	0.085		
Note position: -7	0.044	0.013	3.293	0.018	0.07		
Note position: -6	0.055	0.013	4.307	0.03	0.08		
Note position: -5	0.036	0.013	2.896	0.012	0.061		
Note position: -4	0.027	0.012	2.193	0.003	0.05		
Note position: -3	0.008	0.012	0.671	-0.016	0.032		
Note position: -2	-0.016	0.012	-1.354	-0.04	0.007		
Note position: -1 (Last note)	-0.049	0.012	-4.02	-0.072	-0.025		

Playback experiment

The 75%-95% quantile of the within-trill variation was 8.3 – 17.5% for note length, 1.7 – 3.2% for peak frequency and 25.3 – 60.5% for note bandwidth, measured in 708 trills of 35 male blue tits (20.2 ± 14.9 trills per individual). These values were then used as the minimum and maximum variation introduced to create the inconsistent notes within the playback trill. We conducted 52 playback trials for each experiment, testing a total of 30 individual males. In each experiment, 26 different individuals were tested, 22 of which were common to both experiments. The natural trills of the song type used in the playback experiment had 9.1 ± 4.0 notes per trill, measured in 708 trills of 35 individuals.

Blue tit males spent significantly more time vocalizing and within 4 m of the speaker in response to playback of long-stable trills compared to short-stable trills (Figure 4.4, Table 4.5 & 4.6). However, in the consistency decline experiment, the vocal response of males was significantly higher in response to the short-stable treatment than to the long-decline treatment (Figure 4.4, Table 4.5 & 4.6). In the consistency decline experiment, the time spent within 4 m was not significantly different between treatments during the playback phase but, during the post-playback phase, males spent more time within 4 m during the short-stable than during the long-decline treatment. Importantly, the approach and vocal response to the short treatments did not differ significantly between experiments (not shown). We found a significant effect of the order of playback presentation, as the vocal response was significantly higher in the second test within a trial (Table 4.5). The approach response, however, was not significantly different between the first and the second tests within a trial.

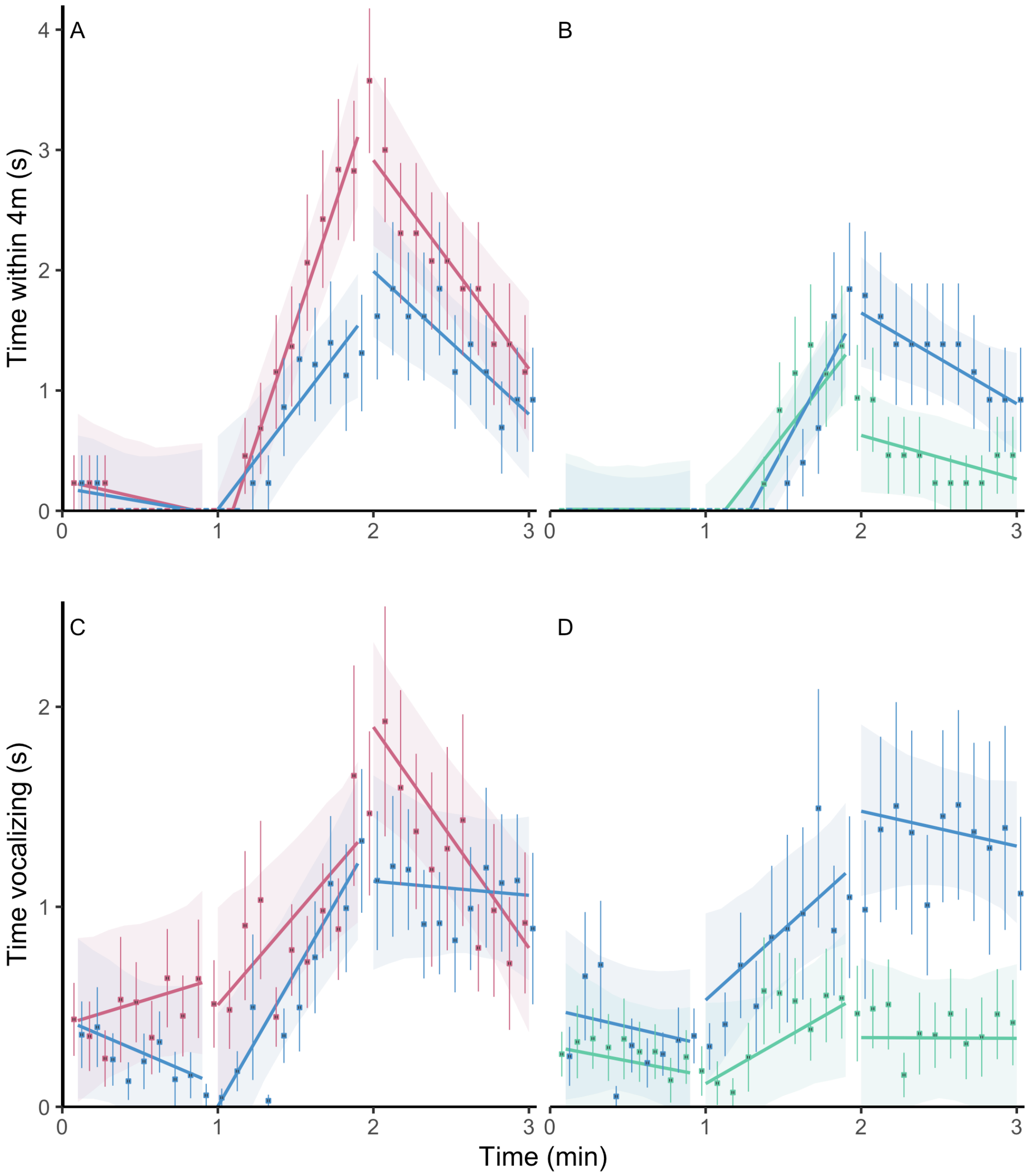


Figure 4.4 – Behavioural response of male blue tits before (0-1) during (1-2) and after (2-3) broadcast of playback song. Figure 4.4A and 4.4C show the approach response (time within 4 m radius of the speaker) and vocal response (sum of durations of vocalizations) per each time interval to the long-stable (red) or short-stable (blue) treatment. Figure 4.4B and 4.4C show the approach and vocal response to the long-decline (green) or short-stable (blue) treatment. In the song length experiment (Figure 4.4A & 4.4C), subjects showed a significantly stronger vocal and approach response towards the long-stable treatment. In the consistency decline experiment, subjects showed a significantly stronger response towards the short-stable than to the long-decline treatment.

Table 4.5 – Estimated coefficients of the model describing the vocal behaviour of male blue tit in response to playback of conspecific song. In the song length experiment, we found that male blue tits showed a stronger response (and steeper increase) towards the long treatment during the playback phase. In the consistency decline experiment, males responded significantly more to the short (stable consistency) treatment than to the long-decline treatment. Estimates are presented with the associated SE, t statistic and the lower and higher 95% CI.

Experiment	**Fixed effects**	**Estimate**	**SE**	**t value**	**5% CI**	**95% CI**
Song length	Intercept	0.332	0.203	1.637	-0.066	0.73
	Treatment (Short vs. Long)	-0.263	0.115	-2.295	-0.486	-0.039
	Time	0.784	0.245	3.204	0.306	1.262
	Phase (Control vs. Playback)	-0.231	0.303	-0.763	-0.823	0.361
	Phase (Post-Playback vs. Playback)	1.391	0.304	4.583	0.798	1.984
	Order	0.449	0.066	6.808	0.32	0.578
	Treatment (Short vs. Long) : Time	0.385	0.347	1.111	-0.292	1.063
	Treatment (Short vs. Long) : Control	-0.586	0.429	-1.367	-1.423	0.251
	Treatment (Short vs. Long): Post-Playback	-0.975	0.429	-2.271	-1.813	-0.136
	Time: Control	-0.578	0.345	-1.676	-1.252	0.096
	Time: Post-Playback	-1.741	0.345	-5.044	-2.415	-1.067
	Treatment (Short vs. Long): Time : Control	-0.879	0.489	-1.799	-1.834	0.076
	Treatment (Short vs. Long): Time : Post-Playback	0.511	0.489	1.045	-0.444	1.466
Consistency decline	Intercept	-0.052	0.191	-0.272	-0.424	0.321
	Treatment (Short vs. Long decline)	0.56	0.118	4.752	0.33	0.79
	Time	0.386	0.246	1.57	-0.094	0.867
	Phase (Control vs. Playback)	-0.288	0.307	-0.938	-0.886	0.311
	Phase (Post-Playback vs. Playback)	-0.013	0.307	-0.042	-0.612	0.586
	Order	0.275	0.071	3.892	0.137	0.413
	Treatment (Short vs. Long decline) : Time	0.224	0.349	0.643	-0.457	0.906
	Treatment (Short vs. Long decline): Control	-0.423	0.434	-0.976	-1.27	0.423
	Treatment (Short vs. Long decline): Post-Playback	0.649	0.434	1.495	-0.198	1.496
	Time: Control	-0.515	0.348	-1.48	-1.195	0.164
	Time: Post-Playback	-0.39	0.348	-1.12	-1.069	0.29
	Treatment (Short vs. Long decline): Time : Control	-0.252	0.493	-0.511	-1.214	0.71
	Treatment (Short vs. Long decline): Time : Post-Playback	-0.372	0.493	-0.755	-1.334	0.59

Table 4.6 – Estimated coefficients of the model describing the approach behaviour of male blue tit in response to playback of conspecific song. In the song length experiment, we found that male blue tits showed a stronger response (and steeper increase) towards the long treatment during the playback phase. In the consistency decline experiment, males responded significantly more to the short (stable consistency) treatment than to the long-decline treatment during the post-playback phase but not during the playback phase. Estimates are presented with its associated SE, t statistic and the lower and higher 95% CI.

Experiment	Fixed effects	Estimate	SE	t value	5% CI	95% CI
Song length	Intercept	1.757	0.268	6.555	1.232	2.283
	Treatment (Short vs. Long)	-0.81	0.159	-5.104	-1.12	-0.5
	Time	3.344	0.339	9.863	2.682	4.006
	Phase (Control vs. Playback)	-1.956	0.42	-4.655	-2.776	-1.135
	Phase (Post-Playback vs. Playback)	1.943	0.421	4.618	1.121	2.765
	Order	-0.003	0.091	-0.028	-0.181	0.176
	Treatment (Short vs. Long):Time	-1.874	0.481	-3.899	-2.813	-0.935
	Treatment (Short vs. Long): Control	0.857	0.594	1.442	-0.304	2.017
	Treatment (Short vs. Long): Post-Playback	-0.359	0.595	-0.604	-1.522	0.803
	Time: Control	-3.598	0.478	-7.523	-4.533	-2.664
	Time: Post-Playback	-4.847	0.478	-10.132	-5.781	-3.912
	Treatment (Short vs. Long): Time : Control	1.935	0.677	2.857	0.612	3.258
	Treatment (Short vs. Long): Time : Post-Playback	2.347	0.677	3.465	1.024	3.67
Consistency decline	Intercept	0.895	0.198	4.527	0.509	1.282
	Treatment (Short vs. Long decline)	-0.075	0.126	-0.592	-0.322	0.172
	Time	1.443	0.267	5.408	0.922	1.965
	Phase (Control vs. Playback)	-0.703	0.333	-2.112	-1.352	-0.053
	Phase (Post-Playback vs. Playback)	0.077	0.333	0.23	-0.574	0.727
	Order	-0.121	0.074	-1.626	-0.266	0.024
	Treatment (Short vs. Long decline):Time	0.619	0.378	1.636	-0.12	1.358
	Treatment (Short vs. Long decline): Control	0.047	0.47	0.1	-0.872	0.966
	Treatment (Short vs. Long decline): Post-Playback	1.269	0.471	2.697	0.35	2.189
	Time: Control	-1.443	0.377	-3.824	-2.181	-0.706
	Time: Post-Playback	-1.758	0.377	-4.658	-2.496	-1.021
	Treatment (Short vs. Long decline) : Time : Control	-0.619	0.535	-1.158	-1.663	0.425
	Treatment (Short vs. Long decline) : Time : Post-Playback	-0.958	0.535	-1.793	-2.002	0.086

Discussion

We proposed the stable performance hypothesis and tested it with comparative and experimental data. Across 31 taxonomically diverse bird species, the vocal performance within trills showed a sudden decline in the last note, irrespective of total trill duration. These results support the premise that there is a general limitation in the sustained production of high-performance vocalizations in birdsong. The two playback experiments showed that longer songs with consistent performance elicited the strongest response during simulated agonistic interactions. However, long songs with a decline in consistency elicited the lowest response, significantly lower than short songs of high consistency, supporting our prediction that birds adjust the length of song to avoid a decline in performance. These empirical results further indicate that there is communicative value in the performance trade-off between continued repetition of high-performance vocalisations and the length of the song. The playback experiments also demonstrate the importance of vocal consistency in birdsong and provide a possible link between song length and individual quality because longer songs of lower consistency seem to be penalized. Although several studies showed the relevance of song length in communication, such link was missing since singing is not costly, respect to resting metabolic rate. The finding of a novel performance constraint that seems common across taxonomically diverse species is an important step in the search for meaningful measures of vocal performance in birdsong (Cardoso, 2017).

The fact that vocal performance declines at the end of a trill in the songs of many bird species is indicative of a general limitation in the sustained production of high-performance vocalisations within song. We hypothesized that the mechanism underlying the decline in performance may be exhaustion due to muscle fatigue caused by the repetition (Lambrechts and Dhondt, 1988b), or to oxygen deprivation within song (Suthers and Goller, 1997; Suthers and Zollinger, 2008). In support of this idea, we found that song became slower and weaker at the end of trills, as vocal power declined and inter-note gaps became larger (lower sound density). Another novel finding was the increase in vocal performance at the beginning of the song, i.e., a warming up effect within the song. Normally, the air volume replenished during a mini-breath between notes is directly related to the duration of the inter-note gap (Hartley and Suthers, 1989). Importantly, vocal power increases in the first notes despite the higher sound density, which presumably also indicate shorter mini-breaths.

This contrasts with the pattern observed in the final notes of the trill, where vocal power declines despite longer inter-note gaps and, presumably, longer mini-breaths. This suggests that loss of vocal power occurs even though air pressure should be stable, pointing towards muscle exhaustion or oxygen debt as possible explanations of lowered vocal performance. Nevertheless, the same mechanism that leads to slower and weaker vocalizations at the end of songs seems to affect the fine coordination skills needed to repeat the same spectral shape consistently, as we found a decline in vocal consistency.

Our finding that sound density declines at the end of songs is similar to that of vocal drift found in blue tits and great tits (*Parus major*) where males lowered the proportion of song versus inter-song gaps after many repetitions of the same song type (Lambrechts and Dhondt, 1988b; Poesel and Kempenaers, 2000). Lambrechts and Dhondt (1988) also found that males with higher drift had lower life-time reproductive success. Similarly, higher vocal power in the dusky warbler (*Phylloscopus fuscatus*) is associated with increased reproductive success and extra-pair copulations (Forstmeier et al., 2002). In that case, the measure of vocal power was the proportion of time that birds produced sound above a certain threshold of relative amplitude. We did not use this measure as it can be associated with note length and therefore to sound density. Given that vocal consistency, vocal power, sound density and song length play an important role in bird communication and the observational data strongly indicates a performance constraint between a stable, high performance and song length, our experiments provided an opportunity to disentangle the communicative value of the trade-off and how it might be resolved by birds.

During the simulated territorial interactions in the experimental playback, we found that male blue tits showed a lower response to long songs with a decline in consistency, than to short songs with high consistency. Given the fact that long, high consistency songs elicited the strongest response and that singing longer songs increases the rates of extra-pair mating in male blue tits (Kempenaers et al., 1997), we suggest that song length plays an important role in both inter- and intrasexual interactions in the blue tit. However, the empirical results strongly indicate that maintaining consistency throughout the song is fundamentally important, as the communicative value of song length disappears if vocal consistency drops. This penalty on decline in vocal performance supports the fundamental role of vocal performance in birdsong (Botero and de Kort, 2013; Byers et al., 2010; Sakata and Vehrencamp, 2012) and suggests that birds adjust their singing behaviour to minimize the decrease in vocal performance rather than to maximise song length.

The presence of a limitation in the sustained production of high performance vocalisations, and the penalty on elongating the song at the expense of low performance quality, reinforces the honesty of song length as a signal that cannot be cheated. In this way, the stable performance hypothesis provides insight into the mechanism that maintains an honest signal of individual quality during communication. The production of relatively long songs without a decline in performance provides information to the receiver, as it is expected that neuro-motor patterns performed close to a physiological limit are meaningful in communication (Byers et al., 2010). During agonistic contests, song length can also be important in those species where song overlapping is a signal of aggression (Naguib and Mennill, 2010). As songbirds counter-sing during agonistic interactions, they convey both individual qualities and motivation in their vocalisations. Overlapping the song of the opponent is considered a motivational signal that leads to escalate the conflict (Naguib and Mennill, 2010). While alternating, song overlapping can be achieved by one bird elongating its own song to overlap the next song of the opponent (personal observation). Elongating the song while avoiding a decline in vocal performance within song may be possible only for individuals with enhanced motor skills, therefore song overlapping could be both a motivational and a quality signal during agonistic interactions.

Conclusion

We found a trade-off between vocal performance and song length, as three measures of vocal performance declined abruptly at the end of the trill in a range of species. Furthermore, our playback experiments indicate that elongating the song at the expense of lowered vocal performance is highly penalized during territorial contests. This study reinforces the idea that vocal performance is a crucial aspect of birdsong (Byers et al., 2010; Sakata and Vehrencamp, 2012) and that multiple aspects of song often interact, causing performance constraints (Cardoso, 2017; Podos, 1997). Finally, although many studies focus on the diversity of birdsong, the study of repetition in song is proving fundamental as it provides a common ground to find generalizations, despite the great differences in singing style between species

Chapter 5 - Frequent female song in blue tits: sexual differences in behavioural context and performance of song

Manuscript submitted to the journal *Behavioural Ecology*, co-authored by Selvino R. de Kort, Katharina Riebel and Ian R. Hartley

Abstract

The blue tit (*Cyanistes caeruleus*) is an important avian model in evolutionary ecology, mentioned in over 30,000 published scientific studies. Song, like in other songbird species, is generally described as a male trait and plays an important role in mate attraction and territory defence. Over the decades, there have been recurring reports of anecdotal female song but these have not led to any quantitative study of female song in blue tits. Our systematic sampling over three years revealed prolific female singing in a northern population of individually colour-ringed blue tits. Daytime singing of females occurred throughout the breeding season during agonistic interactions, solo songs and alarm situations, and these contexts are similar to male song. Notably, female song was absent during the dawn chorus; the period around sunrise when males sing intensively just before copulations take place. Female and male song overlapped substantially in acoustic structure (i.e., same song types, peak frequency or trill rates) but there were also significant differences in that females had smaller repertoires, shorter trills and lower vocal consistency. Differential selections pressures related with contextual (functional) differences in the role of male and female song could explain the observed differences in acoustic structure. The new finding of prolific female singing in such a well-studied species suggests we ought to revise our understanding of male and female vocal communication in this (and probably other) species. Identifying the selection pressures associated with the convergence versus divergence of male and female song may provide important insight in understanding birdsong evolution.

Introduction

Birdsong plays an important role in the acquisition of breeding resources, to mediate social conflicts and to attract mates (Catchpole and Slater, 2008; Marler and Slabbekoorn, 2004) but also for pair coordination and in alarm situations (Cresswell, 1994; Halkin, 1997). Birdsong is, therefore, under strong sexual selection (but see Tobias et al. 2011) and it has been assumed to be, predominantly, a male trait (Collins, 2004; Searcy and Andersson, 1986; Searcy and Yasukawa, 1990). Consequently, its function has been mostly studied in males (Austin et al., 2021; Langmore et al., 1996; Odom et al., 2014; Riebel et al., 2005) regardless of many reports of female song right from the early days of modern birdsong research (e.g. Nice 1943; Robinson 1949, Hinde 1952; Hoelzel 1986; Ritchison 1986).

It was not until this century that the first systematic worldwide survey was conducted showing that female song is common, particularly in the basal clades of passerines, making concurrent male and female song the most likely ancestral state (Odom et al., 2014). With this shifting view, new questions arise regarding the function of female song and the selection pressures underlying sexual differences. Even though the study of female song lacks the literature background of male song, the evidence indicates that it can serve similar functions such as territory advertisement (Cain et al., 2015; Cooney and Cockburn, 1995), mate attraction (Langmore et al., 1996), mate guarding (Reichard et al., 2018) or resource defence (Tobias and Seddon, 2009). One of the most common roles of female song, regarding non-duetting species, is related to the competition for breeding resources (and mates) between females (Austin et al., 2021; Langmore, 1998). Systematic research is needed to gain a complete picture of shared versus sex-specific functions of song in passerines (Austin et al., 2021; Riebel et al., 2019).

One of the unresolved issues is why female song is common in the (sub)tropics and in ancestral clades, while it seems rare in temperate zones and in Passerida. A current working hypotheses is that short breeding seasons, seasonal territoriality and migration might be associated with the loss of song (Benedict, 2008; Odom et al., 2014; Price, 2009). However, the rare documented cases of functional song in Passerida could have partly arisen from sampling biases (sexing a singing bird as male) (Odom and Benedict, 2018). Furthermore, there have been especially few systematic studies in Northern Temperate regions to quantify female song and its function(s) (Riebel et al., 2019). Important first steps are the documentation of female song across and within species, finding ecological correlates, describing behavioural contexts and sex specific structure of song.

The blue tit might be a case in point: this common, widespread and non-migratory passerine breeds in the temperate regions of Europe and western Asia. It is a model species for studies of birdsong, mating systems and other aspects of behavioural ecology (reviewed in Mainwaring and Hartley 2019), with more than 20,000 scientific publications. To human observers, males and females show only minor sexual plumage and size dimorphism and there is much overlap in colour intensity and size between the sexes. While there are recurring anecdotal reports of female song (Bijnens and Dhondt, 1984; Cramp and Perrins, 1993; Hinde, 1952; Mahr et al., 2016), there are no detailed, quantitative descriptions of female singing behaviour, song structure or context, despite the extensive literature of song research in this species (Doutrelant et al., 1998; Doutrelant et al., 2000a; Doutrelant et al., 1999; Doutrelant et al., 2000b; Gorissen and Eens, 2005; Gorissen et al., 2002; Hinde, 1952; Latimer, 1977; Poesel and Dabelsteen, 2005; Poesel et al., 2004; Poesel et al., 2001; Poesel and Kempenaers, 2000; Stadler, 1951)

Overall, from the literature one gains the impression that female song is rare in blue tits. In contrast, during a systematic song recording effort in an individually colour-ringed population, we encountered frequent female song throughout the breeding season. Here, we present a quantitative analysis of the context, occurrence and acoustic structure of female song, in comparison with the song of their male partners. Blue tits generally breed in monogamous pairs but are occasionally socially polygynous and frequently genetically polyandrous (Leech et al. 2001; Schlicht and Kempenaers 2021). In other facultative polygynous species, females use song as a mate-guarding signal during female-female competition (Austin et al., 2021; Langmore, 1998). If this is the same for blue tits, we would predict that females should sing during agonistic interaction with other females. If the function of female song in blue tits is associated with territory defence, we would expect to find females producing solo songs, as a territory advertisement, and during agonistic interactions with both sexes. Solo singing by females would also be consistent with a potential function in mate attraction. Furthermore, female song during the dawn chorus, a time when copulations take place in blue tits, could also be interpreted as providing a function in mate attraction or seeking within- or extra-pair copulations. Finally, we discuss the sexual similarities and differences in song structure in relation to possible functional differences based on the context where we find females and males singing. Results derived from this study might provide clues with respect to potential sex-specific functions of female song and help with developing testable hypothesis for future experimental work.

Methods

Study species and sampling methods

Blue tit song is usually composed of a few introductory, high-frequency notes followed by a trill, defined as the last part of the song where a note is repeated in succession (Figure 5.1), (Bijnens and Dhondt, 1984; Cramp and Perrins, 1993), where a note is defined as a continuous trace in the spectrogram, separated from other notes by silent gaps. Each individual presents several stereotyped song structures referred to as song types (see also ‘phrase’ in Poesel and Kempenaers 2000) and these are usually comparable between individuals within the population (Figure 5.1). During singing, blue tits repeat the same song type many times, alternated with silent pauses (i.e., discontinuous singers), before switching to a different song type which results in so-called song type bouts

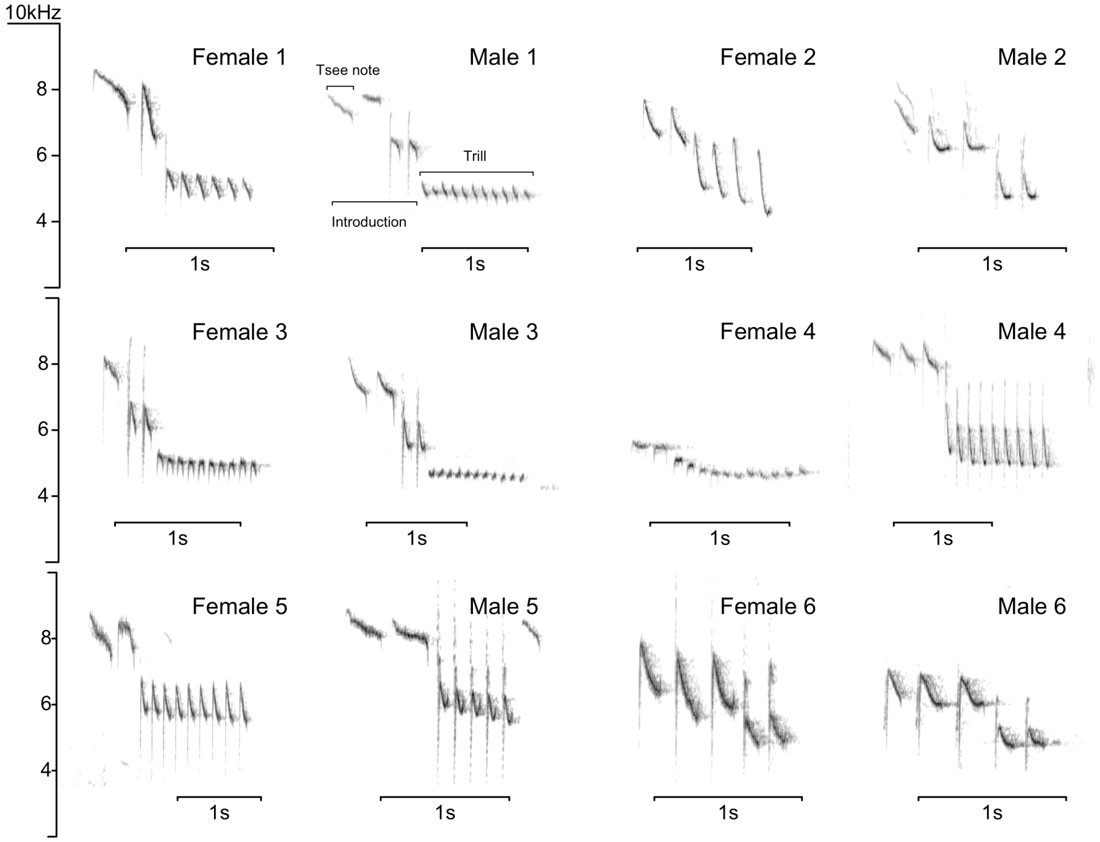


Figure 5.1 – Example of female and male songs. For each female, we selected the same, or similar, song type of her mate for display. The basic features of blue tit song structure are indicated in the first example. The songs were selected to provide a good visualization of song structure, not necessarily a statistical representation of sexual differences.

All birds observed in this study were part of a breeding population occupying nest boxes in deciduous and mixed woodland at Lancaster University campus (54.01° N, 2.78° W) and form part of long-term monitoring study based on approximately 120 available nest boxes (Mainwaring and Hartley, 2009). Each year, adult breeding pairs are caught, measured (straightened, flattened, wing length to nearest mm, tarsus length with foot bent down, to nearest 0.1 mm and head-bill length to nearest 0.1 mm), weighed (to nearest 0.1 g) and ringed with a unique combination of three coloured rings and one numbered metal ring (Redfern and Clark, 2001). During the breeding season, individuals were sexed in the hand based on the presence of a brood patch, present only in females, or cloacal protuberance, present only in males (Svensson, 1992). Birds were aged as first-year or older than first-year, based on plumage characteristics described in (Svensson, 1992).

From Jan-May 2018 to 2020, we conducted daily walked transects to collect song recordings using a Marantz PMD661 recorder (48kHz sampling rate and 24-bit depth) and a Sennheiser ME67 microphone. We followed linear transects since nest boxes were placed in lines within strips of woodland (Figure 5.2). The blue tits found along the transect were identified from their colour rings. The study was initially designed to investigate variation in song performance of males, but when collecting male recordings, we regularly encountered singing females and decided to concurrently establish a data base of female song (i.e., Wilkins et al. 2020) . When a blue tit was encountered singing, its behaviour and song were recorded simultaneously. The observation ended whenever the focal bird stopped singing for more than a few minutes or went out of sight. In the field, the sex of the singing bird was generally unknown to the observer, but using the ring details recorded, individuals could be sexed after cross-checking with the database. We used a dictaphone to take voice notes in a continuous recording for each day of fieldwork. The dictaphone also recorded the ambient sounds, including songs of blue tits that were nearby and this was useful to record the first songs of a bout if song was unpredictable, just before we started the high-quality audio recorder. During the pre-breeding period, from January to the beginning of April, we collected song during daytime singing, after sunrise throughout the morning since singing activity was still very low at dawn (Hinde, 1952). In April and May, during the egg-laying (fertile) and incubation periods, when dawn singing behaviour was more predictable (Hinde, 1952), we began sampling one hour before sunrise until one hour after sunrise. At dawn, visibility was poor and therefore identification of birds using colour rings was difficult. To confirm the identity of a bird singing at dawn, we kept following and recording it until light levels were high enough for identification.

Figure 5.2 – Study area where data was collected (Google Maps, 2021). The nest boxes, shown as grey symbols of a nest box, are placed in lines along the strips of forest that surround the campus of Lancaster University (54.01’ N, 2.78’ W)

Categorizing the behavioural context of blue tit song

After a first revision of song recordings, we chose four distinct behavioural contexts to categorize all singing observations: alarm, agonistic interactions, solo song and dawn song (see Table 5.1 for operational definitions). For each bird we categorized every singing observation recorded into one of the four behavioural contexts. We then selected a set of males that were partners of the recorded females and conducted a similar categorization of behavioural context for all singing observations recorded.

Table 5.1 - Operational definition of behavioural and acoustic terms

Term	Definition
Alarm display	Characterised by the production of 1) isolated, high-frequency 'tsee' calls or scolding calls, both typically produced in the presence of a predator (Bijnens and Dhondt, 1984; Hinde, 1952; Latimer, 1977; Mahr et al., 2016) or 2) mobbing behaviour such as approaching closely and circling a potential predator (Cramp and Perrins, 1993; Shirihai and Svensson, 2018).
Solo songs	Characterised by an individual bird singing alone without any active social interactions with conspecific or alarm responses.
Agonistic interactions	Defined as any interaction where two or more blue tits actively chased, displaced or attacked each other (Irschick et al., 2007). In some cases, the identity and therefore sex of the non-focal individual(s) was unknown and then we did not further classify these interactions into intra- or inter- sexual interactions.
Dawn song	A long vocal display that starts 90-30 min before sunrise, in line with literature on this and other species (Gil and Llusia, 2020; McGregor, 2005; Poesel et al., 2001).
Song repertoire	Total number of song types found across all recordings of a single individual
Peak frequency	The maximum amplitude frequency in the mean power-spectrum (window size = 2048, window type = 'hanning')
Trill length	Duration of the trill in seconds
Trill rate	Number of notes per second always excluding the last rendition of the trill to avoid bias of higher rates toward shorter trills
Vocal deviation	Using song of both sexes, we first calculated the upper bound regression between trill bandwidth and trill rate using bins of 100 Hz (Blackburn et al., 1992; Podos, 1997). The vocal deviation is the orthogonal distance of each trill to the estimated upper bound limit (Figure 5.5, see Podos et al., 1997).
Vocal consistency	Acoustic similarity between each note to the consecutive rendition within the trill using a spectrogram cross-correlation algorithm (SPCC) (Botero et al., 2009; Clark et al., 1987; Coleman et al., 2007). We calculated the maximum correlation of every pairwise SPCC with a maximum temporal offset of 20 ms and a temporal resolution of 1 ms. The spectrogram matrices were computed using an FFT algorithm with a window size of 512 samples and 90 % overlap between successive windows and 'Hanning' window type.

Song type repertoire analysis

Blue tits have relatively small individual repertoires, ranging from 3-8 different song types (Bijnens and Dhondt, 1984). Since blue tits repeat the same song many times before switching, repertoire size is often estimated on the basis of the long, sustained singing displayed during dawn song (Doutrelant et al., 2000a; Poesel et al., 2004). We did not record cases of females singing dawn song, and female daytime singing, like that of males, consisted of much shorter singing bouts. Hence, to compare repertoire usage between males and females, we focussed on daytime singing and counted the number of distinct song types across several recording days of the same individual. Following this method, we selected all those females recorded on two or more dates, and called this the repertoire data set. We categorized song types by visual inspection of spectrograms (Bijnens and Dhondt, 1984; Doutrelant et al., 1998) in Audacity (Mazzoni and Dannenberg, 2014) (window type: ‘Hanning’, window length 1024 samples, 90% overlap and -80 dB range). Based on the song delivery mode of blue tits, that repeat the same song type many times, we focused on the switching points between song types within individuals as the main criteria for identifying song types. The switching point between song types was easily identifiable even if song types were similar. Some of the main features of song that helped in categorizing song types were trill rate, frequency modulation of trill notes, the trill length and the structure of the introductory part (Figure 5.1) but it was nevertheless a visual categorization. For every female, we selected her male partner to compare song repertoires between sexes. Matching female and male recording dates, we selected the same (or similar) number of songs from the male recording, starting always in the first song recorded. In this case, we also excluded male dawn chorus recordings to avoid a contextual bias on the comparison of song repertoires between sexes.

Acoustic analysis

For detailed analysis of acoustic parameters of female song, we chose a subset of females with song recordings of high signal-to-noise ratio, hereafter the acoustic data set (Table 5.2). We then selected individual males that were breeding partners of these females in at least one breeding season. For each male partner, we selected a random sample of songs, recorded in similar days in the season and excluding songs recorded during the dawn chorus (Table 5.2).

Table 5.2 - Sample collected for each individual female, specifying the number of days and years recorded. For each female, the corresponding male partner is shown in the same row. Individuals included in the song repertoire analysis are shaded in light grey. Individuals included in the acoustic analysis with high-quality recordings (HQ rec) are shaded in light and dark grey. The entire data set was used in the behavioural context analyses.

Data set	Female ID	Sample collected in:		No. songs recorded	Song types	No. HQ rec songs	Male ID	Sample collected in:		No. songs selected	Song types	No. HQ rec songs
		No. days	No. years					No. days	No. years			
Complete data set: behavioural context analysis Acoustic analysis Song repertoire analysis	BGOM	5	1	163	2	10	VMBY	6	2	125	3	10
	BRW3OM	4	1	127	2	20	LMOG	4	2	140	3	20
	LGRM	5	1	108	1	26	VMOW	4	1	115	3	26
	GMLG	6	2	98	1	10	VGRM	3	1	65	2	13
	GYRW3M	3	2	95	3	31	WMGR	3	3	100	3	29
	BYLM	8	2	93	1	46	OMRW3G	8	2	110	4	46
	YBYM	5	2	70	3	40	GMLO	5	3	85	3	32
	BBWM	7	3	67	2	43	GBRW3M	7	2	75	5	43
	GBNM	6	1	58	1	22	RW3MOP	4	2	65	2	21
	GMOB	2	1	55	1	20	GRW3VM	2	1	55	2	3
	LMBW	4	2	37	2	17	RW3GRW3M	5	2	45	3	14
	BYVM	4	1	26	1	4	RGYM	2	1	25	2	4
	LMGL	5	2	25	1	12	VYLM	3	1	25	3	4
	YMBR	2	1	20	2	18	VMYW	2	1	25	2	5
	RMRW3V	3	2	18	1	5	NVBM	2	1	20	2	3
	VMGB	2	1	17	1	14	RW3MOP	2	2	15	2	14
	PMVV	3	2	12	1	11	PGRW3M	3	2	14	3	11
	GYPM	2	1	12	2	11	OMRW3Y	2	1	15	2	11
	LMBR	2	1	12	2	8	BMYP	2	2	8	2	8
	RPGM	1	1	27	-	0	-	2	1	20	-	0
	BPNM	1	1	25	-	12	OGRM	1	1	25	-	12
	YYGM	1	1	17	-	8	BYNM	1	1	20	-	8
	GMBG	2	1	14	-	5	RGNM	2	1	5	-	5
	LMWB	1	1	14	-	0	-	-	-	-	-	0
	BRW3VM	1	1	9	-	4	PBNM	1	1	4	-	4
	WMBW	1	1	9	-	9	BMYP	2	2	8	2	8
	YMGB	1	1	9	-	9	RW3MBG	1	1	2	-	2
	WBRM	1	1	7	-	4	-	-	-	-	-	0
	WMBL	1	1	6	-	5	OMOB	1	1	5	-	5

OMRG	2	1	5	-	3	PMVL	2	1	3	-	3
GMVY	1	1	5	-	4	YBLM	1	1	4	-	4
LGPM	1	1	4	-	3	YVGM	1	1	3	-	3
LMRW3G	1	1	3	-	2	LMRW3V	1	1	2	-	2
RLGM	1	1	3	-	3	-	-	-	-	-	0
OMBY	2	-	2	-	0	-	-	-	-	-	0
WMYG	1	-	2	-	0	-	-	-	-	-	0

We conducted all acoustic analysis in R (*tuneR*: Ligges 2013, *seewave*: Sueur et al. 2006; R Development Core Team 2019) . For each individual, we analysed a maximum of ten songs for each song type and each date recorded. Acoustic measurements were made in the terminal trill only since the introductory notes of blue tit songs are more variable and may be absent in some song types (Figure 5.1). Despite having selected high quality recordings, we still had to exclude some notes that were masked by extraneous sounds but included the rest of the song in the analysis (7.04 ± 5.3% of notes in females and 6.4 ± 5.5% of notes in males). We did this to avoid biasing the sample towards shorter songs, as longer songs were more likely to be partly masked.

Every note was manually labelled in the spectrogram, using the cursor to mark the start and end times. Each note was then cut out of the recording and saved as a single, normalized *wav* file. Following Podos (1997), we measured the peak, maximum and minimum frequency of each note in the mean power-spectrum (window size: 1024 samples; amplitude threshold; -20 dB) and from this we derived the bandwidth of the note. Apart from spectral features, we also took measurements of song performance including vocal consistency (*sensu* de Kort et al. 2009), trill length, trill rate and vocal deviation (*sensu* Podos 1997). Table 5.1 shows detailed operational definitions of each acoustic variable. All acoustic variables were measured in each note and summarised as mean values per song unit for statistical analysis.

Statistical analysis

We created three data sets from the same group of females, one to investigate behavioural context of song, one to investigate song type repertoire and one to investigate acoustic structure in female song in relation to male song. The same individuals could be part of one or all three data sets. For a summary of the three data sets for each analysis see Table 5.2. All measures are presented as mean ± one standard deviation (SD), unless otherwise indicated. Statistical analyses were carried out in R software (R Development Core Team, 2016).

We explored the sexual differences in singing context by fitting a Linear Mixed-effect Model (LMM) with the number of singing observations in each behavioural context per individual as the response variable, excluding those individuals with less than two observations. We also excluded all the observations of males during the dawn chorus.

The model included 86 data points, from 23 females and 20 males. The exploratory variables were the sex and its interaction with the behavioural context. We also included the total number of observations per individual as an offset term in the model. To account for repeated measurements, observations were nested within individual and within pair as random effects (23 pairs in total). In this behavioural data set, we had different numbers of males and females because some males were not recorded and, in other cases, different females had the same partner in different years (see *Results*). Using LMM allowed us to cope with this unbalanced design. All males included in the analysis were partners of at least one of the females.

For the song repertoire analysis, we compared females and males for the number of song types within the sample of each individual matched by pairs using Wilcoxon signed-rank tests. Considering the day of the year of each recording, male and female partners of this data set were recorded a mean of 19.8 ± 12.5 days apart and the difference in the total number of songs collected within pairs was 7.9 ± 10.7, from a total of 2,042 songs analysed for both sexes combined. While this means that we might not have included the complete repertoire of each individual, this sampling method allowed us to conduct an unbiased comparison of repertoire usage between the sexes.

For the acoustic analysis, we built five LMMs to investigate sex differences using, in each model, one of the following five parameters as the response variable: peak frequency, vocal consistency, trill length, trill rate and vocal deviation (Table 5.3). These models were fitted on a sample of 402 songs from 32 females (12.6 ± 12.3, songs per individual) and 345 songs from 28 male partners (12.3 ± 12.2, songs per individual, Table 5.1). To define the spectral features of song, we selected only the mean peak frequency of each song, because it is a robust measurement as it is little affected by recording quality (Linhart et al., 2012) and it was strongly correlated with the maximum frequency (r (798) = 0.90, $P < 0.001$) and the minimum frequency (r (798) = 0. 89, $P < 0.001$). As explanatory variables in all models, we used sex (male, female), age (first-year or older than first-year) and weeks in relation to first egg date to account for seasonal variation (week of first egg = 0) (*sensu* Schlicht and Kempenaers 2020). If breeding data were missing for a particular individual, we used the mean date of the first egg in our study site for that year to estimate weeks to first egg. We included sex-specific interactions with age and season effects (in weeks in relation to first egg) to investigate their potential effects on each sex independently. To model peak frequency, we also included the tarsus length and its interaction with sex, since this song feature could be affected by sexual dimorphism in body size. To account for repeated measurements, observations were nested within individual and within pair, using both variables as random effects.

To validate all models, we confirmed that the residuals were homoscedastic and close to a normal distribution using diagnostic plots (Knief and Forstmeier, 2021; Zuur et al., 2009). To find which factors were important in explaining variation in song we computed all possible model combinations ranking them by the Akaike Information Criterion for small samples (AICc). This procedure compares the fit of all possible models while penalizing the complexity, in terms of the number of explanatory variables included. We selected all models that had ΔAICc < 2, in relation to the model with the lowest AICc score, to compute the full average model as the final model (Burnham and Anderson, 2002; Burnham et al., 2011). We used the relative importance of each factor in the final model together with the coefficients and estimated confident intervals (CI) with a threshold of 95% (Burnham et al., 2011; Nakagawa and Cuthill, 2007), considering there was a significant effect if the CI did not overlap with zero. We considered there was a non-significant trend when the 90% CI did not overlap with zero, and in those cases, it is embedded in the text. Finally, we calculated the R^2_{GLMM} of the full models to measure the goodness of fit (Nakagawa and Schielzeth, 2013). All numerical variables are scaled and centred so model estimates are standardized (Gelman, 2008). All the model coefficients are reported in Tables 5.2 & 5.3.

Results

Behavioural context and singing activity

We recorded one individual female singing every 110 min in the field, at 8.2 ± 6.6 % of the boxes visited. For the same period, we recorded a male singing every 33 minutes of fieldwork, at 28.2 ± 21.5 % of the boxes visited. Females were observed and recorded singing in all pre-breeding and breeding stages, from January to the end of May (mean first egg date during the three years was the 22nd of April).

In total, we recorded 99 singing observations from 36 different females and 192 singing instances from 31 selected male partners (Table S1). From all observations of female singing, 42% were during alarm displays, 42% in agonistic interactions and 16% during solo song During agonistic interactions, we recorded two cases where females produced song along with direct physical aggression directed towards other females. No female was observed singing during the dawn chorus (i.e., concurrent with the long singing displays, starting 90-30 min before sunrise (Gil and Llusia, 2020; McGregor, 2005; Poesel et al., 2001). In males, 34% of the total of 192 singing instances were observed during the dawn chorus. After excluding dawn chorus observations, we were left with 127 observations, 61% during solo song, 19% during alarm displays and 20% during agonistic interactions. To compare the behavioural context of song between sexes, we selected 86 observations from 23 females and 18 males that were recorded at least in two different occasions. Our results show that females use song in alarm behaviour significantly more often than males in relation to other contexts (Table 5.3, Figure 5.3). On the contrary, the use of song during agonistic interactions or solo song was not significantly different between sexes, in relation to other contexts (Table 5.3, Figure 5.3).

Table 5.3. Model output comparing behavioural context of song between sexes. For each fixed effect we present the model estimate, the standard error around the estimate (SE), the lower and higher CI of the estimate, the Z statistic derived from Wald tests and the relative importance of that factor in the final model. The last two columns show the marginal R squared represented as R^2_m and the conditional R-squared represented as R^2_c of the full model. This output shows that proportion of song observations in solo songs, agonistic context is similar between sexes, whereas in the alarm context, females were observed significantly more often in proportion.

Parameter	**Fixed effect**	**Estimate**	**SE**	**2.5 % CI**	**97.5 % CI**	**T value**	**R^2_m**	**R^2_c**
Behavioural context of song	Intercept	-1.846	0.466	-2.743	-0.941	-3.958		
	Context (Alarm)	0.604	0.503	-0.368	1.566	1.201	0.137	0.663
	Context (Agonistic)	-1.027	0.568	-2.115	0.06	-1.807		
	Sex (male)	-0.207	0.557	-1.281	0.874	-0.371		
	Context (Alarm) : Sex (male)	-1.894	0.741	-3.312	-0.478	-2.557		
	Context (Agonistic) : Sex (male)	-0.586	0.754	-2.033	0.855	-0.778		

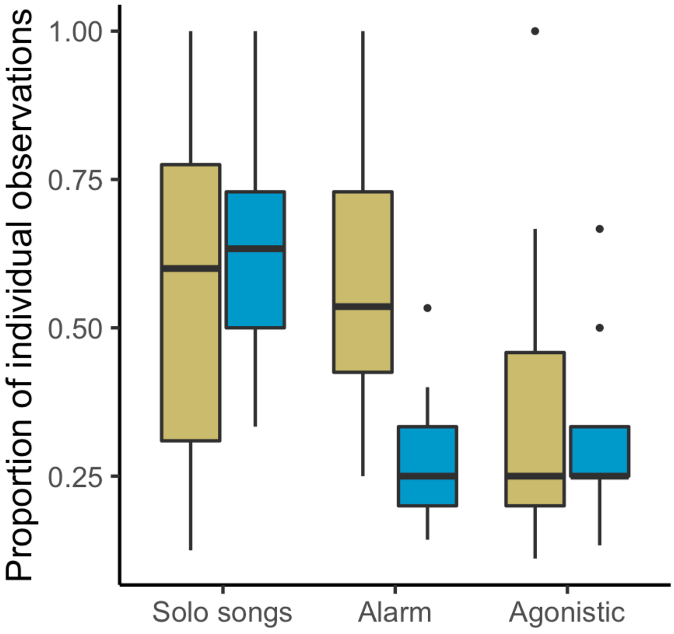


Figure 5.3 – Behavioural context of song production (N = 86 observations) in 23 females and 18 males with more than one observation, after removing observations made during dawn chorus. Females are shown in yellow and males in blue. Even though the response variable in the model was the number of observations per individual per context with the total observations per individual as an offset term, here we show the proportion of observations per individual per context as it is easier to understand visually. The model shows that females sing during the alarm context significantly more often than males, in relation to other contexts.

Song repertoire

Song repertoires were assessed for 19 females that were recorded at least on two different dates (60.5 ± 43.5 songs and 3.8 ± 1.8 dates per female) and their 19 male partners (58.4 ± 40.2 songs and 3.8 ± 1.8 dates per male, Table 5.2). Following visual inspection of songs, we found that males and females used the same song type categories but the number of song types per individual sample was significantly lower in females than in males, with similar sampling effort within pairs (Females = 1.50 ± 0.70 versus males = 2.65 ± 0.96 song types; N=19 pairs, W = 63, $P < 0.001$, 5% CI = -1.00, 95% CI= -2.00; Figure 5.4). For 13 of 19 females we only recorded one song type, in contrast, none of males had fewer than two song types in this or other populations (Bijnens and Dhondt, 1984; Poesel et al., 2001). Three of the most recorded females sang only one song type, with more than 5 dates of recording, in two (or more) years and approximately 100 songs sampled, (Table 5.2).

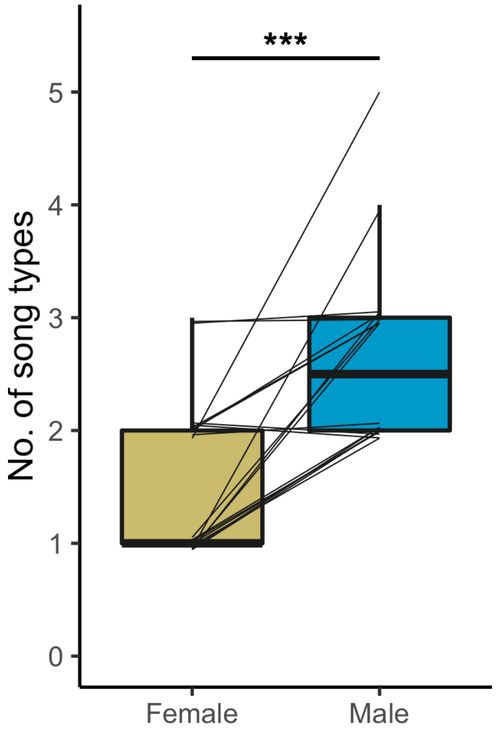


Figure 5.4 – Comparative analysis of song repertoire between sexes, measured as number of song types per individual across all recordings collected during the sampling period (only individuals recorded at least in two dates). Females are shown in gold and males in blue with lines connecting female-male partners with similar sample size in terms of number of songs and recording dates. *** = $P < 0.001$, Wilcox Signed-Rank test (Females = 1.50 ± 0.70 versus males = 2.65 ± 0.96 song types; N=19 pairs, W = 63, $P < 0.001$, 5% CI = -1.00, 95% CI= -2.00).

Acoustic analysis: Spectral and performance variables

Female blue tits sang with significantly lower vocal consistency and produced shorter trills than males (Figure 5.6, Table 5.4), but they did not differ in peak frequency, trill rate or vocal deviation (Figure 5.6, Table 5.4). Upper bound regression of vocal deviation including male and female song rendered a significantly negative slope (r(35) = -0.90, $P < 0.001$, Figure 5.5). In the case of vocal consistency, both males and females increased vocal consistency during winter towards spring (fertile period), independently of age (Figure 5.7, Table 5.4). In both males and females, peak frequency of song decreased significantly during winter towards the spring (fertile period) and older birds showed significantly lower peak frequency than first-year birds (Figure 5.7, Table 5.4). Even though we did not test it specifically, we have evidence that the change in peak frequency is due to spectral flexibility, shifting the entire song in the spectrum (Figure 5.8).

Peak frequency also showed a marginally non-significant, negative trend in relation to tarsus length in both sexes (5% CI = -0.19, 90% CI = -0.016, Figure 5.7A). Note that the effects of age, season and tarsus length were estimated separately for males and females.

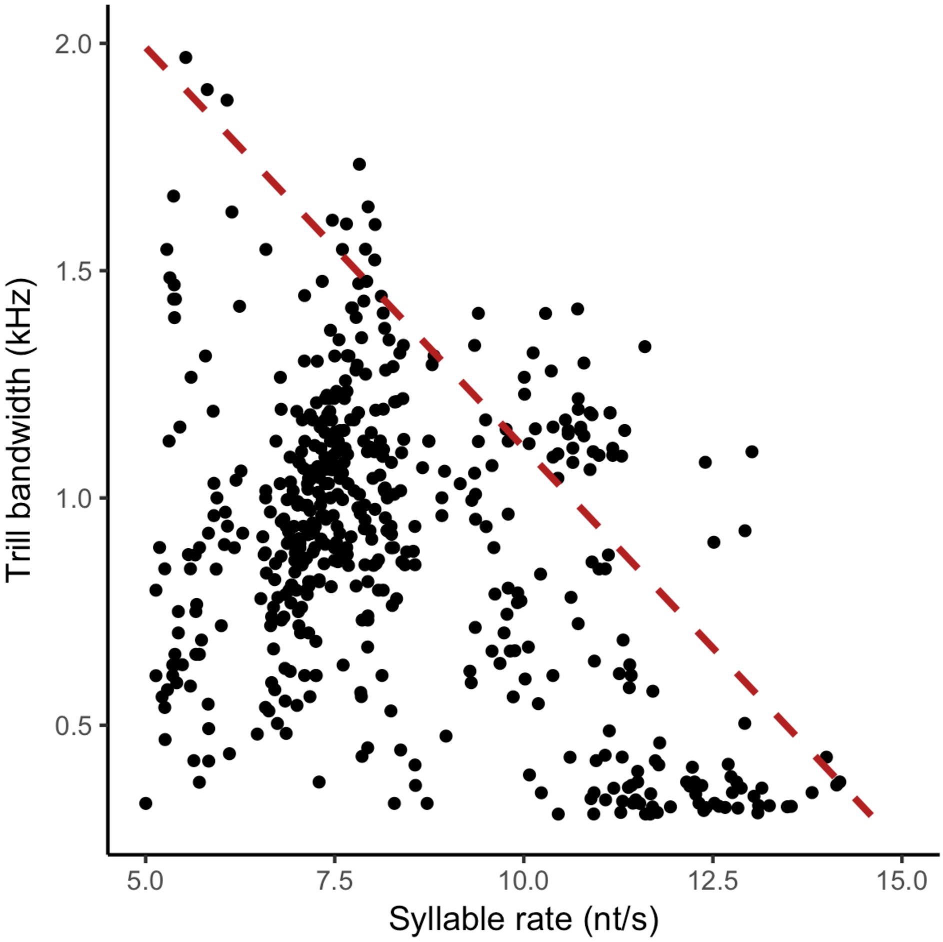


Figure 5.5 – 'Triangular' distribution of trill bandwidth in relation with trill rate, red dashed line showing the upper-bound regression (*r* (35) = -0.90, $P < 0.001$). Degrees of freedom in the upper-bound regression are related to the number of bins selected to estimate local maxima (Podos, 1997). Each dot represents a song, with a total of 747 songs from males and females.

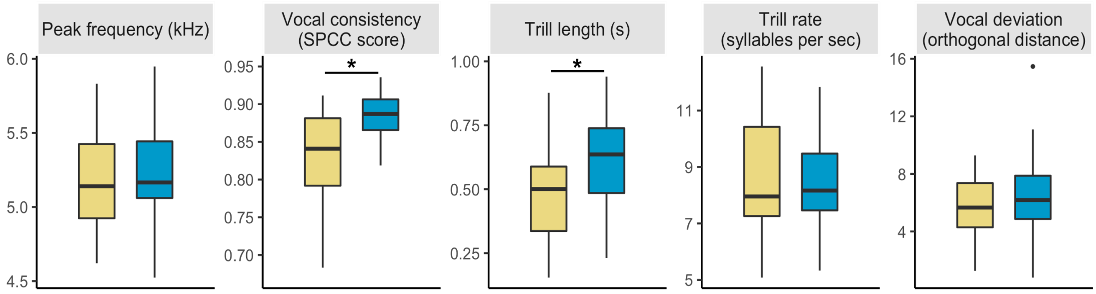


Figure 5.6 – Results of sex differences in acoustic variables. Females sing with lower vocal consistency and shorter trill length than males. Peak frequency, trill rate and vocal deviation are not statistically different between sexes. Significance code: * means that the CI for the effect of sex in the model does not overlap with zero.

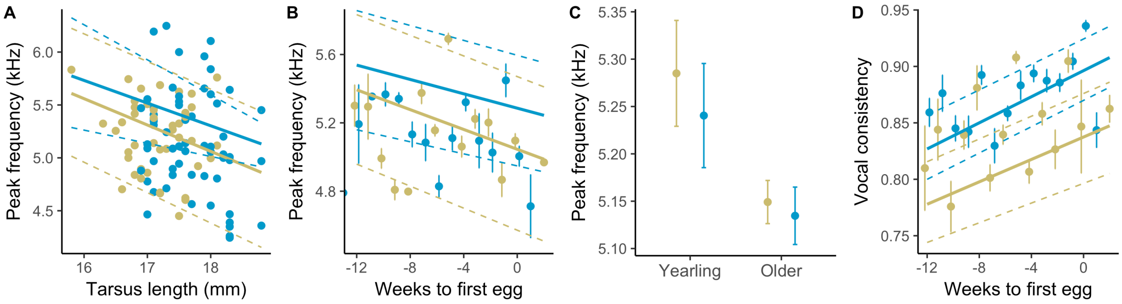


Figure 5.7 – Scatter plots showing the significant effects of tarsus length, age and season on various song parameters. These effects are estimated separately and specifically for males and females in the model. Points represent raw data and lines represent predicted values from the model, with the associated confidence interval in dashed lines. Females are represented in gold and males in blue.

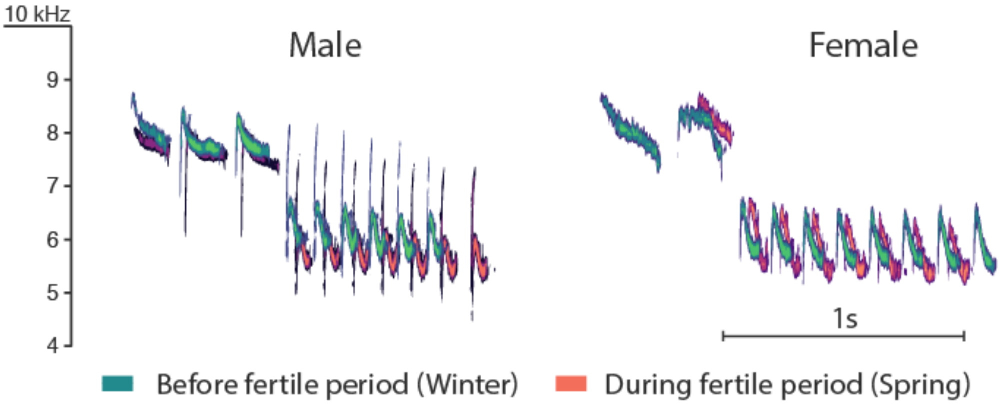


Figure 5.8 – Two examples of spectral flexibility. In the left, a song recorded from one male, 68 days before her female partner began to lay eggs (green spectrogram), overlaid onto a recording of the same song type during the egg laying period of her female partner (red spectrogram). On the right, two song recordings of the same female 40 days before she began to lay eggs (green) and while she was laying eggs (red). Red spectrogram is slightly delayed respect to green spectrogram, for a better visualisation.

In both cases, we can see how the entire song structure is shifted downwards in the frequency spectrum rendering a lower minimum peak frequency.

Table 5.4 - Model output of the full average model for each song trait, comparing male and female song. For each fixed effect we present the model estimate, the lower and higher CI of the estimate, the Z statistic derived from Wald tests and the relative importance of that factor in the final model. The last two columns show the marginal R squared represented as R^2_m and the conditional R-squared represented as R^2_c of the full model. Overall, we found a significant difference between sexes in vocal consistency and trill length.

Parameter	Fixed effect	Estimate	2.5 % CI	97.5 % CI	Z value	Relative importance	R^2_m	R^2_c
Peak frequency (kHz)	Intercept	5.073	4.956	5.191	84.541	-		
	Weeks to first egg	-0.136	-0.199	-0.072	4.191	1	0.08	0.644
	Age	-0.055	-0.101	-0.009	2.328	0.201		
	Tarsus length	-0.096	-0.194	0.001	1.93	0.175		
Trill length (s)	Intercept	0.451	0.367	0.535	10.531	-	0.126	0.604
	Sex (male)	0.224	0.103	0.347	3.609	1		
Vocal deviation (orthogonal distance)	Intercept	6.058	5.244	6.871	14.59	-	0.017	0.286
	Sex (male)	0.702	-0.514	1.918	1.132	0.517		
	Age (older)	0.303	-0.1	0.706	1.475	0.357		
Vocal consistency (SPCC score)	Intercept	0.869	0.841	0.898	59.574	-	0.125	0.416
	Weeks to first egg	0.038	0.024	0.051	5.541	1		
	Sex (male)	0.042	0.013	0.071	2.828	0.37		
Trill rate (notes per sec)	Intercept	8.52	7.926	9.114	28.115	-	0.04	0.732
	Sex (male)	-0.125	-1.163	0.913	0.236	0.33		

Discussion

In blue tits, female song was common across a variety of behavioural contexts. This contrasts with previous anecdotal reports of female song in this species, associated mostly with alarm contexts. Based on an *ad hoc* sampling scheme, registering whether each identified individual was singing during the daytime, we observed a high incidence of females singing (36%) relative to males (53%), normally considered the "singing sex". In contrast, we did not collect any observations of females singing dawn song (but see Gorissen and Eens (2005) for female vocal activity inside the nest box). As there is little evidence that ecological or environmental factors are unusual in our population, we assume that females in other populations might sing as much. Hence, the most parsimonious explanation for the under-reporting of female song in the literature is that singing in female blue tits has been overlooked. Apart from the general research bias towards male song, two more specific reasons may account for these oversights. One, a focus of many studies on dawn song, a context when female blue tits indeed do not seem to sing, and second, the low dimorphism (to the human eye) that makes it difficult to sex birds from a distance in the field. As song is often used to sex birds, singing females might have been registered as singing males. Importantly, female song, combined with physical aggression, was observed only during female-female interactions, which is in line with a potential function in intrasexual competition, supporting previous reports of female singing and fighting (Hinde, 1952; Kempenaers et al., 1995). Female, like male song, showed seasonal variation in relation to the fertile period in two key parameters: peak frequency and vocal consistency. Previous studies suggested that lower peak frequency was a result of different song type usage (Halfwerk et al., 2011) but we have evidence that there is spectral flexibility within song type, although these two mechanisms are not mutually exclusive. Moreover, peak frequency was related to age and weakly to body size in both sexes, indicating that female song could potentially inform receivers about the qualities of an individual singer. By comparing male and female partners, we found that individual females sang fewer song types than males and their songs were significantly shorter and lower in vocal consistency. The three most recorded females had a repertoire of one song type and this contrasts with males, that have not been reported with fewer than two song types (Bijnens and Dhondt, 1984; Poesel et al., 2001). Finally, female song was similar to male song in peak frequency, trill rate and vocal deviation, raising the question as to whether some function of song is shared in both sexes. Additional work and experimental approaches are necessary to fully characterise all functions of female song, but our observational data allow us to formulate some hypotheses for future work.

First, it is important to realise that functions of female song might partly overlap with male song, but also have different functions (Austin et al., 2021; Riebel et al., 2019). Importantly, if females sing less often than males this does not mean that female song has lower ecological importance (Austin et al., 2021). In our population, female song occurred throughout the breeding season in several behavioural contexts, most often during alarm situations. In a comparable experimental context, Mahr et al. (2016) observed at least 3 different females and two different males singing when they positioned a taxidermy mount of a sparrowhawk, *Accipiter nisus*, near nests with fledglings. Our extensive observations of song during mobbing events and predator presence support those findings but show that such behaviour occurs throughout the breeding season and not only during the fledging period. Blue tits are thus among a growing number of species known to produce song in the presence of predators (Cresswell, 1994; Laiolo et al., 2004; Langmore and Mulder, 1992). However, anti-predator behaviour was but one context where we observed females singing. Females, like males, were regularly observed producing solo songs and also singing during agonistic interactions. Both contexts are in line with a possible territorial function of female song (Langmore, 1998) that has been demonstrated in several species using playback experiments (Cooney and Cockburn, 1995; Hoelzel, 1986; Krieg and Getty, 2016; Magoolagan and Sharp, 2018). The fact that both sexes use the same song types can also facilitate joint territory defence since song type matching is an aggressive signal in this and related species (Krebs et al., 1981; Langemann et al., 2000; Poesel and Dabelsteen, 2005).

Although further empirical work is needed to establish the functional value of female song in blue tits, there are three lines of evidence from our observational study indicating that female song plays a role during intrasexual competition for breeding resources and mates in the blue tit (Clutton-Brock, 2009; Langmore, 1998) and supporting previous reports in the blue tit (i.e., female song in ‘reproductive fighting’ Hinde 1952). First, females sing throughout the breeding season, a time of the year when blue tits are strongly territorial. Second, female song varies seasonally in acoustic structure towards the fertile period and third, females produce song during aggressive encounters with other females. A similar role of female song during intrasexual conflicts has been described in other species. In dunnocks (*Prunella modularis*) and great reed warblers (*Acrocephalus arundinaceus*), artificially high female-female competition increased the incidence of female song during intra-sexual conflicts (Kluyver, 1955; Langmore and Davies, 1997). Female European starlings (*Sturnus vulgaris*) produced song during aggression directed towards caged females placed in their territory (Sandell and Smith, 1997). Dark-eyed junco (*Junco hyemalis*) females sang while showing aggressive behaviour towards the presentation of caged females (Reichard et al., 2018). In barn swallows (*Hirundo rustica*) females used song to interrupt male song, potentially disrupting mate attraction by the male (Wilkins et al., 2020), and in eastern whipbirds (*Psophodes olivaceus*) females approached more closely to playback of female solo song than to male or duet songs (Rogers et al., 2007).

Many of these species, where female song has a function in intrasexual competition, are opportunistically polygynous, just like blue tits (Langmore, 1998). Blue tit males try to attract secondary females after the primary female starts to lay eggs (Kempenaers, 1994, 1995) and polygyny can affect female fitness, since male parental care will be reduced in their nest (Hartley unpublished data). This may explain why female blue tits show strong aggressive behaviour towards intruding females during the breeding period (Gorissen et al., 2002; Kempenaers, 1994, 1995; Midamegbe et al., 2011) and we observed female song during female-female fights. Female song may, therefore, play a role in intrasexual competition over mates and territories to actively ward off polygyny. To establish whether female song in blue tits functions in mate competition (mate guarding), one possibility is to test for links between the incidence of female song and the proportion of polygynous partnerships across and within populations. Playback studies are also fundamental for testing the response of female (and male) blue tits to female song within (and outside) the breeding period.

In relation to acoustic variables of song, we found that peak frequency was related to age and (weakly) to body size in both females and males. This could play an important role during female vocal interactions as it does for female Mexican antthrushes (*Formicarius moniliger*) (Kirschel et al., 2020). Female song, like male song, varied in acoustic structure during the breeding period; decreasing in peak frequency and increasing in vocal consistency towards the fertile period. In males, such seasonal variation has been associated with specific functions of song during reproduction (Ballentine et al., 2003; Halfwerk et al., 2011). The seasonal change occurred similarly in all three years, indicating that song parameters must “return to winter values” after the breeding period (summer-autumn), completing the seasonal cycle. To our knowledge, this is the first study to document seasonal changes in acoustic structure of female song.

Female song could potentially play a role in mate choice and pair formation as females display solo song during daytime singing from winter to spring, and winter associations predict pair formation (Beck et al., 2020). But importantly, the truly qualitative difference in song between sexes was the absence of female dawn song, a song display that is associated with seeking within- and extra-pair copulations by males (Kempenaers et al., 1997; Parker et al., 2006; Poesel et al., 2004; Welling et al., 1995). Given that copulations are mostly under female control, this context implies a particular selection pressure over male song that seems absent in female song. At this point, even though we need to know more about the function of female song in blue tits, we can already deduce that selection pressures shaping this signal are not congruent between the sexes (Austin et al., 2021). Theoretically, such differential pressures could imply differences in song parameters associated with female choice during copulations. In line with this, we found that song repertoire, song length and vocal consistency were different between males and females. Previous studies in blue tits showed that longer songs of blue tit males are associated with higher rates of extra-pair copulations (Kempenaers et al., 1997) and a playback study on female choice showed that females had a preference for higher consistency songs in blue tits (Sierro et al., in prep). Future empirical work will have to address which sexual differences, and similarities, in song are meaningful in communication and which selection pressures favour the observed differences.

Our observations of frequent singing in female blue tits are in line with recent comparative analyses that found a higher incidence of female song in non-migratory species with low sexual dichromatism (Webb et al., 2016) that hold year round territories (Benedict, 2008; Logue and Hall, 2014; Odom et al., 2014; Price, 2009; Riebel et al., 2019). Our comprehensive report provides an example of scarce descriptions of female song in the Northern Temperate regions (reviewed in Odom and Benedict 2018; Odom et al. 2014; Riebel et al. 2019) despite a large number of species with anecdotal reports of female song (e.g. Garamszegi et al. 2007). The behavioural contexts and seasonal variation of female song in blue tits suggest a territorial and/or a mate guarding function to avoid polygyny (Langmore, 1998). Furthermore, our finding that males, but not females, produce dawn song likely implies differential selection pressures between sexes since this context is highly associated with attracting females for copulations (Austin et al., 2021). For now, our observations of prolific female singing and its associations with specific contexts provide robust foundations for further hypothesis development and testing. It is crucial to design future playback studies presenting male and female blue tits with female song, as empirical evidence is key to understand its function. Finally, increasing documentation of female song in all biogeographic regions is crucial for our understanding of the evolution of birdsong.

Chapter 6 – General discussion

Communicative value of vocal performance

Throughout the thesis, I have covered several aspects of vocal performance and I show their communicative value. The main finding of this thesis is that higher vocal consistency in males is correlated with higher reproductive success, in terms of number of eggs per clutch. This is crucial evidence that puts the focus on vocal consistency as a signal of male fitness. Furthermore, our novel playback experimental design with females inside the nest box revealed a female preference for high consistency songs by showing a stronger sexual response. In this experiment I also found that song type switching was fundamental in keeping a high sexual response from females through dishabituation, a possible behavioural mechanism explaining the evolution of small or intermediate repertoires in songbirds. Theoretical and empirical research have shown how song complexity (song repertoires) and song ritualization (vocal performance) have communicative value and play a role during inter- and intrasexual interaction. In this thesis I provide a possible solution to this apparent paradox (Byers and Kroodsma, 2009; Price, 2013), by showing a behavioural mechanisms that may lead to simultaneous positive selection pressures over vocal consistency (repetition) and song diversity (variation) within the same communication system.

Birdsong is a multidimensional acoustic signal where many parameters interact to shape each individual' and species' song (Gil and Gahr, 2002). To understand the interaction, and possible trade-off, between several aspects of vocal performance within song, I created a multi-species data set of 31 different species. The results suggest there is a limitation to the sustained production of high-performance vocalisations as vocal performance drops at the end of trilled songs. I investigated the communicative value found in such a trade-off, between song length and vocal performance, by conducting two playback experiments in wild male blue tits. Our findings suggest that longer songs are more intimidating during male-male interactions, if vocal performance is high and stable. However, when long songs showed a decline in vocal performance (consistency), short songs of high, stable consistency elicited a stronger response in territorial males. This result indicates that vocal performance is a *high priority* feature of blue tit song, while song length is also important, but the latter will be compromised if vocal performance starts to decline. At the same time, the findings suggest a possible mechanism that ensures honesty in song length, as only those individuals with high performance skills are able to produce long songs without a decrease in vocal performance. This is because long song, despite being attractive, during mate choice (Gentner and Hulse, 2000; Gentner et al., 2001; Kempenaers et al., 1997; Mennill et al., 2006), and intimidating, during territorial contests (Lattin and Ritchison, 2009; Linhart et al., 2012) are penalized if vocal performance drops.

Altogether, these findings support previous theoretical and empirical work that proposed vocal performance as a fundamental aspect of vocal communication in birds (Botero and de Kort, 2013; Byers et al., 2010; Podos, 1997) and are further supported by our validation of the methods to measure vocal consistency in song; the SPCC algorithm. I found that the SPCC, as a method to measure acoustic similarity, is sensitive to small discrepancies in bird-like sounds. This is fundamental to assess vocal consistency since the discrepancies in the sounds recorded are ultimately related to variation in the execution of underlying neuro-motor patterns (Allan and Suthers, 1994; Zollinger and Suthers, 2004).

Finally, our finding that female blue tits sing prolifically opens a new window into birdsong research in this species. Our extensive recording effort on individually colour ringed females provides a solid foundation to investigate sexual differences in song in the blue tit and other species. I found that the context of song was similar between sexes when it came to territorial behaviour and alarm situations. However, females did not sing during the dawn chorus, a song display that is tightly associated with copulations. This contextual (functional) difference could explain the observed sexual differences in song repertoires, song length and vocal consistency. The study of female song in blue tits has further relevance as one of the few northern temperate species of songbirds with extensive data on female song. Similar to other sexually dimorphic traits (i.e., body size or plumage colouration), sexual differences in song can provide a powerful tool in understanding the evolution of sexually divergent traits in relation to certain ecological and social conditions

Vocal consistency and reproductive success in the blue tit

One major finding was the significant association between higher male vocal consistency and higher reproductive success. However, our data cannot shed light onto the exact mechanism underlying the association between these two variables. If vocal consistency is associated with male viability, it is possible that females paired with males of higher vocal consistency invest more in reproduction, by laying more eggs, the ‘differential-allocation hypothesis’ (Burley, 1986, 1988). Females would benefit from the higher resource allocation by having offspring that were more viable, including sons that were themselves more attractive as breeders (Weatherhead and Robertson, 1979). Another way to explain the link between higher male vocal consistency and larger clutches would be through assortative mating, where high consistency males breed with high quality females that can invest more in reproduction (Härdling and Kokko, 2005; Real, 1991), although these two mechanisms are not mutually exclusive (Holveck and Riebel, 2010).

I have evidence of assortative mating in our population based on another song feature: minimum song frequency (Figure 6.1). This feature of song is normally correlated with body size (Bradbury and Vehrencamp, 1998) but, importantly, the correlation of minimum song frequency between male and female breeding partners is stronger than the correlation of tarsus length, a biometric parameter often used as a proxy for body size and relative quality (Figure 6.1). Previous research has shown there is assortative mating in blue tits based on UV-reflectance of the crown feathers, a physical attribute that is associated with individual quality (Andersson et al., 1998). Even though I can only speculate, the data suggest that the link between male vocal consistency and clutch size can be through assortative mating of high-quality females with males of higher vocal consistency.

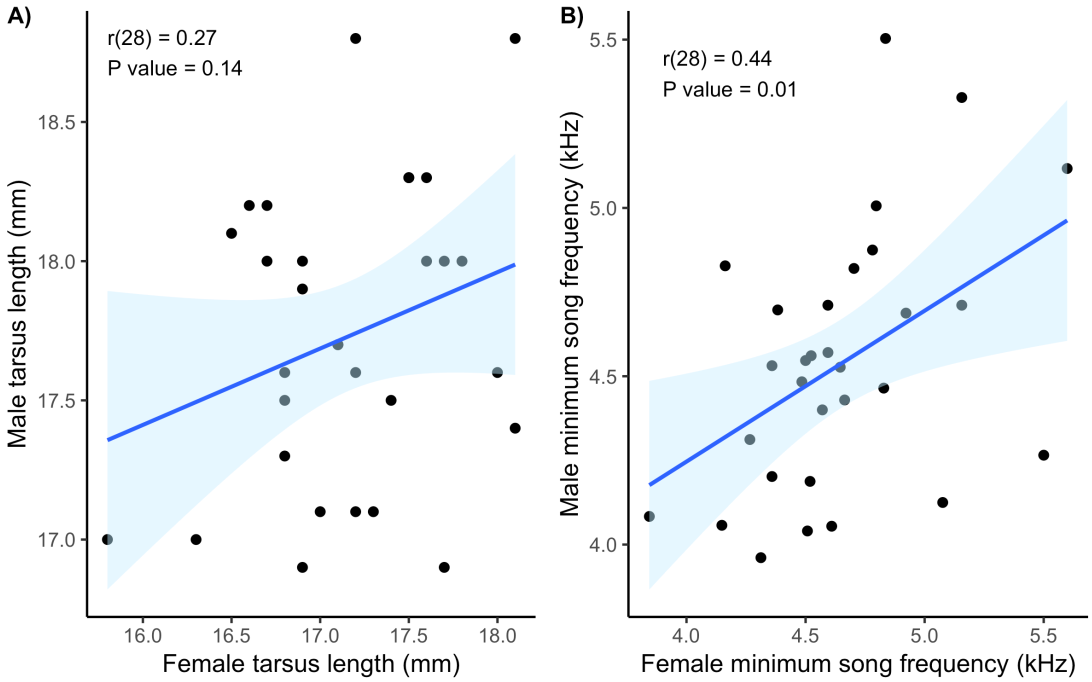


Figure 6.1 – Correlation of tarsus length (6.1A) and minimum song frequency (6.1B) between male and female partners. Each point represents two partners. Minimum song frequency was measured as the lowest peak frequency in the trill, as it was always the lowest part of the song and peak frequency is the most robust measure. Figure 6.1B shows the significant and strong correlation between partners in their minimum song frequency.

Another important behaviour that determines lifetime reproductive success is that of parental care. In blue tits, as in most songbirds, both sexes care for the offspring (Kendeigh, 1952; Wesołowski, 2004). I expected males of higher consistency, as an indicator of whole-organism quality, to invest in higher parental effort, assessed as feeds per hour per nestling at the age of 10-12 days old. This would translate into a positive correlation between parental effort and vocal consistency of the males. However, our data showed that such a correlation is indeed significant, but negative, as males of higher vocal consistency showed fewer feeds per chick per hour (Figure 6.2).

Results from a meta-analysis have shown that empirically increasing parental effort significantly decreased the chance of survival in males (Santos and Nakagawa, 2012). Furthermore, several empirical studies have shown that more attractive males show lower parental effort but overall higher long term reproductive success (Burley, 1988; Santos and Nakagawa, 2012). I have shown that high vocal consistency is attractive to females and that it signals male fitness, therefore our findings that high consistency males show lower parental effort are consistent with previous results. If male blue tits with high vocal consistency have larger clutches, survive longer or sire more extra-pair offspring, it is possible that reducing parental effort to maximize survival is adaptive. Unfortunately I do not currently have the data of a paternity analysis of the birds in this study but our playback experiment, showing that females prefer high consistency song during dawn chorus, suggests that high consistency males may have higher EPC rates (Holveck and Riebel, 2007). Alternatively, it could also be that high consistency males bring higher quality prey items if, for instance, they occupy high quality territories, explaining the lower number of visits per hour. Another possible mechanism explaining this pattern is that males with low consistency songs increase parental effort in their own nest to maximize the chances of success in that particular brood, if they have fewer extra-pair offspring or lower prospects for survival to the following breeding season (Clutton-Brock, 1984; Williams, 1966).

There are anecdotal reports of extra-pair males providing parental care or predator defence in other nests where they sired extra-pair offspring (Kempenaers, 1993) but this is extremely rare in our population (Hartley unpublished) and therefore would not explain the observed pattern. An empirical study in blue tits showed that females increased the parental effort when paired with more attractive males (Limbourg et al., 2004). This could partially compensate for the lowered parental effort from males of higher vocal consistency but I did not find such correlation between higher male vocal consistency and higher female parental investment.

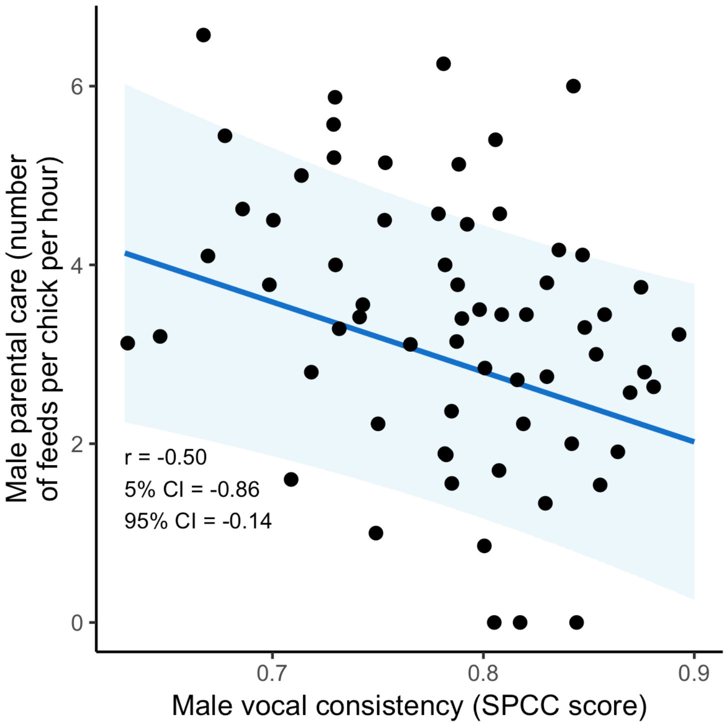


Figure 6.2 – Relationship between male parental effort, in number of feeds per hour per nestling, and male vocal consistency. The linear model shows a significant, negative relationship after controlling for male body size (head-bill length), male age (in years) and female parental effort (female feeds per hour per chick).

Functional value of seasonal variation in vocal performance

Another major finding of this thesis is the significant seasonal variation in vocal consistency (Gahr, 2020). The striking synchrony between the seasonal peak in vocal consistency and the start of the receptive period of the female, considering that song plays is an important role in within- and extra-pair copulations (Kempenaers et al., 1997; Mace, 1987; Welling et al., 1995), is another line of evidence in support of the importance of vocal consistency in mate choice. Seasonal changes in the singing activity have been known for decades (Avey et al., 2011; Hinde, 1952) but very few studies provide a detailed analysis of changes in song structure throughout the season with high enough resolution to detect such close-fitting synchronization (reviewed in Gahr, 2020) . Some studies have shown intra-annual variation in song structure, normally comparing song within and outside the breeding season in a two-level comparison (Ballentine et al., 2003; Hill et al., 2015). But, to my knowledge, this is the first study to show a daily, seasonal change in song with high temporal resolution (but see similar approach in Halfwerk et al., 2011) .

Seasonal changes in vocal consistency have been associated with seasonal changes in brain song-control nuclei in the song sparrow (Smith et al., 1997a) and these changes in brain structure related with song nuclei are common (Nottebohm et al., 1986; Smith, 1996; Smith et al., 1997b). Caro et al., (2005) showed a similar increase in the size of song control nuclei in the blue tit. As well as a seasonal increase in vocal consistency, I found a similar change in trill length and peak frequency and, in all cases, these changes match with the reported increase in brain song nuclei (Figure 6.3). But, what is the functional value of such strong seasonal variation?

At the end of winter (January - March), winter flocks break up and blue tit breeding pairs become territorial (Cramp and Perrins, 1993). Territoriality often leads to intrasexual conflicts for territories, which are common in late winter (personal observation). Following our own and previous findings that vocal consistency, trill length or song frequency are relevant features during such agonistic interactions in several species (Sierro in prep, de Kort el al. 2009; Linhart et al. 2012; Rivera-Gutierrez et al. 2011) , the question arises of why these features of song are so low (or high in the case of frequency) during this period.

Even though our data cannot shed much light on this, the seasonal association between song and brain structure indicates that increased size in brain song control nuclei is necessary to produce high performance, low frequency songs. It is possible that such an increase in brain song control centres implies a metabolic cost that may be unaffordable for small birds during the winter months, as has been suggested for other small vertebrates with seasonal changes in brain size (Lázaro et al., 2019; Yaskin, 1984). If this is true, it implies that song cannot be of high performance in late winter despite positive selection on vocal performance because the growth in brain structures necessary to sing high performance song is only possible when food supplies increase at the beginning of spring. This could be reflected in a latitudinal pattern where southern populations of blue tit, that experience milder winter conditions, show lower seasonal variation in song structure. Alternatively, it could be the development of different brain structures during winter that determines the size of song control centres. Pre-winter growth of hippocampal centres has been observed in many bird species (reviewed in Yaskin 2001) , associated with food caching behaviour and the spatial cognition required to relocate those caches at a later stage. A latitudinal variation in brain structure was found in the hippocampal centres of black cap chickadees (*Poecile atricapilla*) as northern populations, that experience harsh winter conditions, have larger hippocampal centres than southern populations (Pravosudov and Clayton, 2002). Even if the blue tit is not considered a food-storing species, seasonal changes in hippocampus have been shown in birds and mammals associated with changes in spatial behaviour (Pravosudov and Clayton, 2002; Yaskin, 2011) and these changes could also impact the resources allocated to other neural centres such as the song nuclei centres.

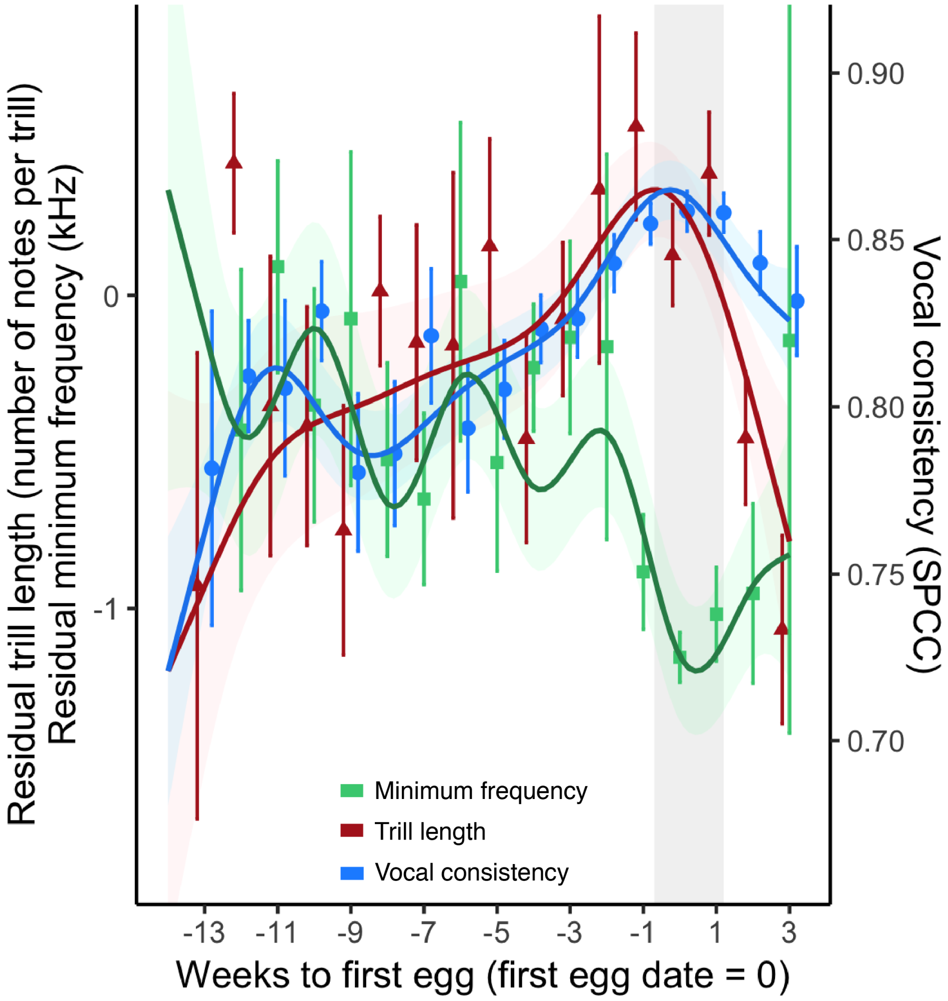


Figure 6.3 – Seasonal change in residual trill length (red), residual minimum song frequency (green) and vocal consistency (blue) including data from 3 different years. The residual trill length was measured as the number of notes per trill in relation to the mean number of notes per trill for each song type within individual. Similarly, residual minimum song frequency was measured in kHz as the differences between each song and the mean minimum song frequency of that song type within individual. Points show the mean values per week, for all three years, with the associated standard error. Lines show the predicted values from two GAM models with the associated 95% confidence interval. Trill length and vocal consistency increased while minimum song frequency decreased during

the weeks of late winter and early spring and showed their maximum, or minimum, values during the receptive period of the female. This period was defined as the period where copulation take place from 5 days before the first egg was laid until the completion of the clutch ~9 days after the first egg. In all cases, there seemed to be a decline (or increase) after the incubation (fertile) period is over.

Performance constraints

In Chapter 4, I found a trade-off between vocal performance and trill length as vocal power, sound density and vocal consistency declined at the end of song. In Chapter 3, I found that, within the same individuals, two song types emphasized opposite extremes of one parameter; the wide and narrow bandwidth trills in Figure 3.1. In that case, I suggest that both extremes imply a challenging performance in relation to different constraints. On one hand, broadband trills of fast note rate are challenging because of the need for fast, mechanical rearrangement of the vocal apparatus in between trill notes (i.e., low vocal deviation, Podos 1997) . On the other hand, narrowband trills may be challenging because they highlight possible errors in motor execution, as inconsistencies in frequency are easier to detect in pure tones than in frequency modulated tones. If vocal performance is meaningful in relation to physiological constraints, it is possible that different song types or singing styles exploit different performance limits (Cardoso, 2017). This, together with the multiple interactions between various aspects of vocal performance, highlights the complexity of understanding and measuring song performance in birdsong. The interaction between various aspects of song performance implies that the communicative value of each parameter must be seen in relation to the rest as the execution of motor patterns near a physiological limit is challenging and thus an honest signal of neuro-motor skills.

Conclusions

From this thesis it can be concluded that vocal performance, with emphasis on vocal consistency, is a fundamental aspect of birdsong and has an important signal value during intra- and intersexual interactions. I found that vocal consistency 1) relates to individual male fitness, 2) is preferred by females during courtship displays, 3) is an important aspect of sexual differences in song, 4) is meaningful during male-male territorial contests, 5) shows a large seasonal variation with the highest performance observed during the fertile period, 6) is positively correlated with age and 7) correlates negatively with parental effort in males. Another important finding was that of frequent female song in blue tits, which highlights the value of observation of individually marked animals in understanding natural phenomena, complemented by theory-based research. The thorough, inquisitive and detailed observations have proven fundamentally important throughout this thesis, from the analytical approach to validate the SPCC algorithm, to the detailed description of change in vocal performance within song. To end, my findings stress the multifaceted nature of birdsong, where many parameters interact to shape the particular singing style of each species. It is important that we take into consideration such complexity if we are to continue to disentangle the processes underlying the evolution of birdsong.

References

Allan SE, Suthers RA, 1994. Lateralization and motor stereotype of song production in the brown‐headed cowbird. Journal of neurobiology 25:1154-1166.

Amy M, Salvin P, Naguib M, Leboucher G, 2015. Female signalling to male song in the domestic canary, Serinus canaria. Royal Society Open Science 2:140196.

Andersson M, 1994. Sexual selection: Princeton University Press.

Andersson S, Örnborg J, Andersson M, 1998. Ultraviolet sexual dimorphism and assortative mating in blue tits. Philosophical Transactions of the Royal Society B: Biological Sciences 265:445-450.

AudioMoth home page, 2017. https://www.openacousticdevices.info.

Austin VI, Dalziell AH, Langmore NE, Welbergen JA, 2021. Avian vocalisations: the female perspective. Biological Reviews.

Avey MT, Rodriguez A, Sturdy CB, 2011. Seasonal variation of vocal behaviour in a temperate songbird: Assessing the effects of laboratory housing on wild-caught, seasonally breeding birds. Behavioural processes 88:177-183.

Baker MC, Bjerke TK, Lampe H, Espmark Y, 1986. Sexual response of female great tits to variation in size of males' song repertoires. The American Naturalist 128:491-498.

Ballentine B, Badyaev A, Hill GE, 2003. Changes in song complexity correspond to periods of female fertility in blue grosbeaks (*Guiraca caerulea*). Ethology 109:55-66.

Ballentine B, Hyman J, Nowicki S, 2004. Vocal performance influences female response to male bird song: an experimental test. Behav Ecol 15:163-168.

Barske J, Schlinger BA, Wikelski M, Fusani L, 2011. Female choice for male motor skills. Philosophical Transactions of the Royal Society B: Biological Sciences 278:3523-3528.

Bartsch C, Hultsch H, Scharff C, Kipper S, 2016. What is the whistle all about? A study on whistle songs, related male characteristics, and female song preferences in common nightingales. Journal of Ornithology 157:49-60.

Beck KB, Farine DR, Kempenaers B, 2020. Winter associations predict social and extra-pair mating patterns in a wild songbird. Philosophical Transactions of the Royal Society B: Biological Sciences 287:20192606.

Benedict L, 2008. Occurrence and life history correlates of vocal duetting in North American passerines. Journal of Avian Biology 39:57-65.

Berwick RC, Chomsky N, 2013. Birdsong, speech, and language: exploring the evolution of mind and brain: MIT press.

Bijnens L, Dhondt A, 1984. Vocalizations in a Belgian blue tit, *Parus c. caeruleus*, population. Le Gerfaut 74:243-269.

Birkhead TR, Møller AP, 1993. Why do male birds stop copulating while their partners are still fertile? Animal Behaviour 45:105-118.

Blackburn TM, Lawton JH, Perry JN, 1992. A method of estimating the slope of upper bounds of plots of body size and abundance in natural animal assemblages. Oikos:107-112.

Bostwick KS, Prum RO, 2003. High-speed video analysis of wing-snapping in two manakin clades (Pipridae: Aves). Journal of Experimental Biology 206:3693-3706.

Botero CA, de Kort SR, 2013. Learned signals and consistency of delivery: a case against receiver manipulation in animal communication. Animal communication theory: information and influence. p. 281-296.

Botero CA, Rossman RJ, Caro LM, Stenzler LM, Lovette IJ, de Kort SR, Vehrencamp SL, 2009. Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. Animal Behaviour 77:701-706.

Bradbury JW, Vehrencamp SL, 1998. Principles of animal communication. Sunderland, Massachusetts: Sinauer Associates.

Breiehagen T, Slagsvold T, 1988. Male polyterritoriality and female-female aggression in pied flycatchers *Ficedula hypoleuca*. Animal Behaviour 36:604-606.

Bremond J-C, 1976. Specific recognition in the song of Bonelli's warbler (*Phylloscopus bonelli*). Behaviour 58:99-116.

Brumm H, Lachlan RF, Riebel K, Slater PJ, 2009. On the function of song type repertoires: testing the ‘antiexhaustion hypothesis’ in chaffinches. Animal Behaviour 77:37-42.

Brumm H, Todt D, 2004. Male–male vocal interactions and the adjustment of song amplitude in a territorial bird. Animal Behaviour 67:281-286.

Brumm H, Zollinger SA, Niemelä PT, Sprau P, 2017. Measurement artefacts lead to false positives in the study of birdsong in noise. Methods in Ecology and Evolution 8:1617-1625.

Burley N, 1986. Sexual selection for aesthetic traits in species with biparental care. The American Naturalist 127:415-445.

Burley N, 1988. The differential-allocation hypothesis: an experimental test. The American Naturalist 132:611-628.

Burnham KP, Anderson DR, 2002. Model selection and multimodel inference: a practical information-theoretic approach: Springer, New York.

Burnham KP, Anderson DR, Huyvaert KP, 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behavioral Ecology and Sociobiology 65:23-35.

Byers BE, 2007. Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance. Behav Ecol 18:130-136.

Byers BE, Kroodsma DE, 2009. Female mate choice and songbird song repertoires. Animal Behaviour 77:13-22.

Byers J, Hebets E, Podos J, 2010. Female mate choice based upon male motor performance. Animal Behaviour 79:771-778.

Cain KE, Cockburn A, Langmore NE, 2015. Female song rates in response to simulated intruder are positively related to reproductive success. Frontiers in Ecology and Evolution 3:119.

Cardoso GC, 2017. Advancing the inference of performance in birdsong. Anim Behav 125:e29-e32.

Cardoso GC, Atwell JW, Hu Y, Ketterson ED, Price TD, 2012. No Correlation Between Three Selected Trade‐Offs in Birdsong Performance and Male Quality for a Species With Song Repertoires. Ethology 118:584-593.

Caro SP, Lambrechts MM, Balthazart J, 2005. Early seasonal development of brain song control nuclei in male blue tits. Neuroscience letters 386:139-144.

Catchpole CK, 1996. Song and female choice: good genes and big brains? Trends in Ecology & Evolution 11:358-360. doi: http://dx.doi.org/10.1016/0169-5347(96)30042-6.

Catchpole CK, Slater PJB, 2008. Bird song: biological themes and variations. Cambridge, UK: Cambridge University Press.

Charmantier A, Blondel J, 2003. A contrast in extra - pair paternity levels on mainland and island populations of Mediterranean blue tits. Ethology 109:351-363.

Christie PJ, Mennill DJ, Ratcliffe LM, 2004. Pitch shifts and song structure indicate male quality in the dawn chorus of black-capped chickadees. Behavioral Ecology and Sociobiology 55:341-348.

Clark CW, Marler P, Beeman K, 1987. Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. Ethology 76:101-115.

Clutton-Brock T, 2009. Sexual selection in females. Animal behaviour 77:3-11.

Clutton-Brock TH, 1984. Reproductive effort and terminal investment in iteroparous animals. The American Naturalist 123:212-229.

Coleman SW, Patricelli GL, Coyle B, Siani J, Borgia G, 2007. Female preferences drive the evolution of mimetic accuracy in male sexual displays. Biology letters 3:463-466.

Collins S, 2004. Vocal fighting and flirting: The functions of birdsong. In: Slabbekoorn PM, Hans, editor. Nature's music: the science of birdsong San Diego: Elsevier Academic Press. p. 39-79.

Cooney R, Cockburn A, 1995. Territorial defence is the major function of female song in the superb fairy-wren, *Malurus cyaneus*. Animal Behaviour 49:1635-1647.

Cortopassi KA, Bradbury JW, 2000. The comparison of harmonically rich sounds using spectrographic cross-correlation and principal coordinates analysis. Bioacoustics 11:89-127.

Cramer ERA, 2013a. Measuring consistency: spectrogram cross-correlation versus targeted acoustic parameters. Bioacoustics-the International Journal of Animal Sound and Its Recording 22:247-257.

Cramer ERA, 2013b. Vocal deviation and trill consistency do not affect male response to playback in house wrens. Behav Ecol 24:412-420.

Cramp S, Perrins C, 1993. Handbook of the birds of Europe, the Middle East and North Africa. The birds of the western Palearctic. Volume VII. United Kingdom: Oxford University Press, Oxford.

Cresswell W, 1994. Song as a pursuit-deterrent signal, and its occurrence relative to other anti-predation behaviours of skylark (*Alauda arvensis*) on attack by merlins (*Falco columbarius*). Behavioral Ecology and Sociobiology 34:217-223.

Cunningham EJ, Russell AF, 2000. Egg investment is influenced by male attractiveness in the mallard. Nature 404:74-77.

Darwin C, 1872. The descent of man, and selection in relation to sex: D. Appleton.

De Cheveigne A, 2005. Pitch perception models. Pitch: Springer. p. 169-233.

de Kort SR, Eldermire ER, Cramer ER, Vehrencamp SL, 2009a. The deterrent effect of bird song in territory defense. Behav Ecol 20:200-206.

de Kort SR, Eldermire ERB, Valderrama S, Botero CA, Vehrencamp SL, 2009b. Trill consistency is an age-related assessment signal in banded wrens. Philosophical Transactions of the Royal Society B: Biological Sciences 276:2315-2321.

Deoniziak K, Osiejuk TS, 2020. Song-type switching rate in the chaffinch carries a message during simulated intrusion. Behavioral Ecology and Sociobiology 74:1-11.

Dhondt AA, Adriaensen F, 1994. Causes and effects of divorce in the blue tit Parus caeruleus. Journal of Animal Ecology:979-987.

Dobson CW, Lemon RE, 1975. Re-examination of monotony threshold hypothesis in bird song. Nature 257:126-128.

Dong S, Clayton DF, 2009. Habituation in songbirds. Neurobiology of learning and memory 92:183-188.

Dooley GJ, Moore BC, 1988. Detection of linear frequency glides as a function of frequency and duration. The Journal of the Acoustical Society of America 84:2045-2057.

Doutrelant C, Aubin T, Hitier S, Lambrechts MM, 1998. Two distinct song populations of blue tit (*Parus caeruleus*) in the French Mediterranean. Bioacoustics 9:1-16.

Doutrelant C, Blondel J, Perret P, Lambrechts MM, 2000a. Blue tit song repertoire size, male quality and interspecific competition. Journal of Avian Biology:360-366.

Doutrelant C, Leitao A, Giorgi H, Lambrechts MM, 1999. Geographical variation in blue tit song, the result of an adjustment to vegetation type? Behaviour 136:481-493.

Doutrelant C, Leitao A, Otter K, Lambrechts MM, 2000b. Effect of blue tit song syntax on great tit territorial responsiveness–an experimental test of the character shift hypothesis. Behavioral Ecology and Sociobiology 48:119-124.

Dubbeldam J, 1997. Intratelencephalic sensorimotor circuits in birds-What have feeding and vocalization in common? Netherlands journal of zoology 48:199-212.

Dunning JL, Pant S, Bass A, Coburn Z, Prather JF, 2014. Mate choice in adult female Bengalese finches: females express consistent preferences for individual males and prefer female-directed song performances. PLoS One 9:e89438.

Düring DN, Ziegler A, Thompson CK, Ziegler A, Faber C, Müller J, Scharff C, Elemans CP, 2013. The songbird syrinx morphome: a three-dimensional, high-resolution, interactive morphological map of the zebra finch vocal organ. BMC biology 11:1-27.

Eberhardt LS, 1994. Oxygen consumption during singing by male Carolina wrens (Thryothorus ludovicianus). The Auk 111:124-130.

Elemans CP, Spierts IL, Müller UK, Van Leeuwen JL, Goller F, 2004. Superfast muscles control dove's trill. Nature 431:146-146.

Fastl H, 1978. Frequency discrimination for pulsed versus modulated tones. The Journal of the Acoustical Society of America 63:275-277.

Fishbein AR, Idsardi WJ, Ball GF, Dooling RJ, 2019. Sound sequences in birdsong: how much do birds really care? Philosophical Transactions of the Royal Society B: Biological Sciences 375:20190044.

Fishbein AR, Löschner J, Mallon JM, Wilkinson GS, 2018. Dynamic sex-specific responses to synthetic songs in a duetting suboscine passerine. PloS one 13:e0202353.

Fletcher NH, 2004. A simple frequency-scaling rule for animal communication. The Journal of the Acoustical Society of America 115:2334-2338.

Forstmeier W, Kempenaers B, Meyer A, Leisler B, 2002. A novel song parameter correlates with extra-pair paternity and reflects male longevity. Philosophical Transactions of the Royal Society B: Biological Sciences 269:1479-1485.

Fusani L, Barske J, Day LD, Fuxjager MJ, Schlinger BA, 2014. Physiological control of elaborate male courtship: female choice for neuromuscular systems. Neuroscience & Biobehavioral Reviews 46:534-546.

Fusani L, Giordano M, Day LB, Schlinger BA, 2007. High‐speed video analysis reveals individual variability in the courtship displays of male golden‐collared manakins. Ethology 113:964-972.

Gahr M, 2020. Seasonal hormone fluctuations and song structure of birds. Coding strategies in vertebrate acoustic communication:163-201.

Galeotti P, Saino N, Sacchi R, Møller AP, 1997. Song correlates with social context, testosterone and body condition in male barn swallows. Animal Behaviour 53:687-700.

Gelman A, 2008. Scaling regression inputs by dividing by two standard deviations. Statistics in medicine 27:2865-2873.

Gentner TQ, Hulse SH, 2000. Female European starling preference and choice for variation in conspecific male song. Animal behaviour 59:443-458.

Gentner TQ, Hulse SH, Duffy D, Ball GF, 2001. Response biases in auditory forebrain regions of female songbirds following exposure to sexually relevant variation in male song. Journal of neurobiology 46:48-58.

Gerhardt HC, 1992. Conducting playback experiments and interpreting their results. Playback and studies of animal communication: Springer. p. 59-77.

Gil D, Gahr M, 2002. The honesty of bird song: multiple constraints for multiple traits. Trends in Ecology & Evolution 17:133-141.

Gil D, Graves J, Hazon N, Wells A, 1999. Male attractiveness and differential testosterone investment in zebra finch eggs. Science 286:126-128.

Gil D, Llusia D, 2020. The bird dawn chorus revisited. Coding strategies in vertebrate acoustic communication: Springer. p. 45-90.

Gorissen L, Eens M, 2005. Complex female vocal behaviour of great and blue tits inside the nesting cavity. Behaviour 142:489-506.

Gorissen L, Janssens E, Pinxten R, Eens M, 2002. Differences in song repertoire size and composition between two populations of blue tits (*Parus caeruleus*). Avian Science 2:39-48.

Halfwerk W, Bot S, Buikx J, van der Velde M, Komdeur J, ten Cate C, Slabbekoorn H, 2011. Low-frequency songs lose their potency in noisy urban conditions. Proceedings of the National Academy of Sciences 108:14549-14554.

Halkin SL, 1997. Nest-vicinity song exchanges may coordinate biparental care of northern cardinals. Animal Behaviour 54:189-198.

Härdling R, Kokko H, 2005. The evolution of prudent choice. Evolutionary Ecology Research 7:697-715.

Hartley RS, Suthers RA, 1989. Airflow and pressure during canary song: direct evidence for mini-breaths. Journal of Comparative Physiology A 165:15-26.

Hartshorne C, 1956. The monotony-threshold in singing birds. The Auk 73:176-192.

Hebets EA, Barron AB, Balakrishnan CN, Hauber ME, Mason PH, Hoke KL, A systems approach to animal communication. Philosophical Transactions of the Royal Society B: Biological Sciences2016. The Royal Society. p. 20152889.

Hill CE, Akçay Ç, Campbell SE, Beecher MD, 2011. Extrapair paternity, song, and genetic quality in song sparrows. Behav Ecol 22:73-81.

Hinde RA, 1952. The behaviour of the great tit (*Parus major*) and some other related species. Behaviour Supplement No 2:III, V-X, 1-201.

Hoelzel AR, 1986. Song characteristics and response to playback of male and female robins (*Erithacus rubecula*). Ibis 128:115-127.

Holveck M-J, Riebel K, 2007. Preferred songs predict preferred males: consistency and repeatability of zebra finch females across three test contexts. Animal Behaviour 74:297-309.

Holveck M-J, Riebel K, 2010. Low-quality females prefer low-quality males when choosing a mate. Philosophical Transactions of the Royal Society B: Biological Sciences 277:153-160.

Holveck MJ, de Castro ACV, Lachlan RF, ten Cate C, Riebel K, 2008. Accuracy of song syntax learning and singing consistency signal early condition in zebra finches. Behav Ecol 19:1267-1281. doi: 10.1093/beheco/arn078.

Hunt S, Bennett ATD, Cuthill IC, Griffiths R, 1998. Blue tits are ultraviolet tits. Philosophical Transactions of the Royal Society B: Biological Sciences 265:451-455.

Hunt S, Cuthill IC, Bennett AT, Griffiths R, 1999. Preferences for ultraviolet partners in the blue tit. Animal Behaviour 58:809-815.

Irschick DJ, Herrel A, Vanhooydonck B, van Damme R, 2007. A functional approach to sexual selection. Functional Ecology 21:621-626.

Kempenaers B, 1993. A case of polyandry in the blue tit: female extra-pair behaviour results in extra male help. Ornis Scandinavica:246-249.

Kempenaers B, 1994. Polygyny in the blue tit: unbalanced sex ratio and female aggression restrict mate choice. Animal behaviour 47:943-957.

Kempenaers B, 1995. Polygyny in the blue tit: intra-and inter-sexual conflicts. Animal Behaviour 49:1047-1064.

Kempenaers B, Verheyen GR, Dhondi AA, 1997. Extrapair paternity in the blue tit (*Parus caeruleus*): female choice, male characteristics, and offspring quality. Behav Ecol 8:481-492.

Kempenaers B, Verheyen GR, Dhondt AA, 1995. Mate guarding and copulation behaviour in monogamous and polygynous blue tits: do males follow a best-of-a-bad-job strategy? Behavioral Ecology and Sociobiology 36:33-42.

Kempenaers B, Verheyen GR, Van den Broeck M, Burke T, Van Broeckhoven C, Dhondt A, 1992. Extra-pair paternity results from female preference for high-quality males in the blue tit. Nature 357:494-496.

Kendeigh SC, 1952. Parental care and its evolution in birds. Urbana: University of Illinois Press.

Kirschel AN, Zanti Z, Harlow ZT, Vallejo EE, Cody ML, Taylor CE, 2020. Females don't always sing in response to male song, but when they do, they sing to males with higher-pitched songs. Animal Behaviour 166:129-138.

Kluyver H, 1955. The behaviour of the great reed warbler, *Acrocephalus arundinaceus*; with special reference to nest building and territorial defence. Ardea 43:1-50.

Knief U, Forstmeier W, 2021. Violating the normality assumption may be the lesser of two evils. Behavior Research Methods:1-15.

Kogan JA, Margoliash D, 1998. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. The Journal of the Acoustical Society of America 103:2185-2196.

Krebs J, 1976. Habituation and song repertoires in the great tit. Behavioral Ecology and Sociobiology 1:215-227.

Krebs JR, Ashcroft R, Van Orsdol K, 1981. Song matching in the Great Tit, *Parus major* L. Animal Behaviour 29:918-923.

Kreutzer M, Nagle L, Vallet E, 1994. Sexual responsiveness of female canaries to song bout organization. Behaviour 129:293-305.

Krieg CA, Getty T, 2016. Not just for males: females use song against male and female rivals in a temperate zone songbird. Animal Behaviour 113:39-47.

Kroodsma DE, 1978. Continuity and versatility in bird song - support for the monotony threshold-hypothesis. Nature 274:681-683. doi: 10.1038/274681a0.

Lachlan R, Verhagen L, Peters S, Cate Ct, 2010. Are there species-universal categories in bird song phonology and syntax? A comparative study of chaffinches (*Fringilla coelebs*), zebra finches (*Taenopygia guttata*), and swamp sparrows (*Melospiza georgiana*). Journal of Comparative Psychology 124:92.

Laiolo P, Tella JL, Carrete M, Serrano D, López G, 2004. Distress calls may honestly signal bird quality to predators. Philosophical Transactions of the Royal Society B: Biological Sciences 271:S513-S515.

Lambrechts M, Dhondt AA, 1986. Male quality, reproduction, and survival in the great tit (Parus major). Behavioral Ecology and Sociobiology 19:57-63.

Lambrechts M, Dhondt AA, 1988a. The anti-exhaustion hypothesis: a new hypothesis to explain song performance and song switching in the great tit. Animal Behaviour 36:327-334. doi: 10.1016/s0003-3472(88)80002-2.

Lambrechts M, Dhondt AA, 1988b. The anti-exhaustion hypothesis: a new hypothesis to explain song performance and song switching in the great tit. Animal Behaviour 36:327-334.

Langemann U, Tavares J, Peake TM, McGregor PK, 2000. Response of great tits to escalating patterns of playback. Behaviour 137:451-472.

Langmore N, Davies N, 1997. Female dunnocks use vocalizations to compete for males. Animal Behaviour 53:881-890.

Langmore NE, 1997. Song switching in monandrous and polyandrous dunnocks, *Prunella modularis*. Animal Behaviour 53:757-766. doi: 10.1006/anbe.1996.0312.

Langmore NE, 1998. Functions of duet and solo songs of female birds. Trends in ecology & evolution 13:136-140.

Langmore NE, Davies N, Hatchwell B, Hartley IR, 1996. Female song attracts males in the alpine accentor (*Prunella collaris*). Philosophical Transactions of the Royal Society B: Biological Sciences 263:141-146.

Langmore NE, Mulder RA, 1992. A novel context for bird song: predator calls prompt male singing in the kleptogamous superb fairy ‐ wren, *Malurus cyaneus*. Ethology 90:143-153.

Latimer W, 1977. A comparative study of the songs and alarm calls of some *Parus* species. Zeitschrift für Tierpsychologie 45:414-433.

Lattin C, Ritchison G, 2009. Intra-and intersexual functions of singing by male blue grosbeaks: the role of within-song variation. The Wilson Journal of Ornithology 121:714-721.

Lawson SL, Fishbein AR, Prior NH, Ball GF, Dooling RJ, 2018. Relative salience of syllable structure and syllable order in zebra finch song. Animal cognition 21:467-480.

Lázaro J, Hertel M, Muturi M, Dechmann DK, 2019. Seasonal reversible size changes in the braincase and mass of common shrews are flexibly modified by environmental conditions. Scientific reports 9:1-10.

Leadbeater E, Goller F, Riebel K, 2005. Unusual phonation, covarying song characteristics and song preferences in female zebra finches. Animal Behaviour 70:909-919.

Leech DI, Hartley IR, Stewart IR, Griffith SC, Burke T, 2001. No effect of parental quality or extrapair paternity on brood sex ratio in the blue tit (*Parus caeruleus*). Behav Ecol 12:674-680.

Author. 2013. tuneR–analysis of music.

Limbourg T, Mateman AC, Andersson S, Lessells CM, 2004. Female blue tits adjust parental effort to manipulated male UV attractiveness. Philosophical Transactions of the Royal Society B: Biological Sciences 271:1903-1908.

Linhart P, Slabbekoorn H, Fuchs R, 2012. The communicative significance of song frequency and song length in territorial chiffchaffs. Behav Ecol 23:1338-1347.

Logue DM, Hall ML, 2014. Migration and the evolution of duetting in songbirds. Philosophical Transactions of the Royal Society B: Biological Sciences 281:20140103.

Mace R, 1987. The dawn chorus in the great tit *Parus major* is directly related to female fertility. Nature 330:745-746. doi: 10.1038/330745a0.

Magoolagan L, Sharp SP, 2018. Song function and territoriality in male and female White-throated Dippers (*Cinclus cinclus*). Bird Study 65:396-403.

Magrath MJ, Vedder O, Van der Velde M, Komdeur J, 2009. Maternal effects contribute to the superior performance of extra-pair offspring. Current Biology 19:792-797.

Mahr K, Seifert CL, Hoi H, 2016. Female and male Blue Tits (*Cyanistes caeruleus*) sing in response to experimental predator exposition. Journal of Ornithology 157:907-911.

Mainwaring MC, Hartley IR, 2009. Experimental evidence for state-dependent nest weight in the blue tit, *Cyanistes caeruleus*. Behavioural Processes 81:144-146.

Mainwaring MC, Hartley IR, 2019. Blue tits. Encyclopedia of Animal Behavior (Second Edition) ed. p. 11-22.

Author. 2021. Lancaster University Campus

Margoliash D, 1983. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. Journal of Neuroscience 3:1039-1057.

Marler PR, Slabbekoorn H, 2004. Nature's music: the science of birdsong. San Diego: Elsevier Academic Press.

Author. 2014. Audacity.

McGregor PK, 2005. Animal communication networks: Cambridge University Press.

Mennill DJ, Badyaev AV, Jonart LM, Hill GE, 2006. Male house finches with elaborate songs have higher reproductive performance. Ethology 112:174-180. doi: 10.1111/j.1439-0310.2006.01145.x.

Midamegbe A, Grégoire A, Perret P, Doutrelant C, 2011. Female–female aggressiveness is influenced by female coloration in blue tits. Animal Behaviour 82:245-253.

Miskimen M, 1951. Sound production in passerine birds. The Auk 68:493-504.

Naguib M, Mennill DJ, 2010. The signal value of birdsong: empirical evidence suggests song overlapping is a signal. Animal Behaviour 80:e11-e15.

Nakagawa S, Cuthill IC, 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biological Reviews 82:591-605.

Nakagawa S, Schielzeth H, 2013. A general and simple method for obtaining R2 from generalized linear mixed - effects models. Methods in Ecology and Evolution 4:133-142.

Nolan PM, Hill GE, 2004. Female choice for song characteristics in the house finch. Animal Behaviour 67:403-410.

Nottebohm F, 1971. Neural lateralization of vocal control in a passerine bird. I. Song. Journal of Experimental Zoology 177:229-261. doi: 10.1002/jez.1401770210.

Nottebohm F, 1972. Neural lateralization of vocal control in a passerine bird. II. Subsong, calls, and a theory of vocal learning. Journal of Experimental Zoology 179:35-49. doi: 10.1002/jez.1401790104.

Nottebohm F, Nottebohm ME, Crane L, 1986. Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei. Behavioral and neural biology 46:445-471.

Nowicki S, 1987. Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere. Nature 325:53-55.

Nowicki S, Capranica RR, 1986. Bilateral syringeal coupling during phonation of a songbird. Journal of Neuroscience 6:3595-3610.

Nowicki S, Mitani JC, Nelson DA, Marler P, 1989. The communicative significance of tonality in birdsong: responses to songs produced in helium. Bioacoustics 2:35-46.

Nowicki S, Searcy W, Peters S, 2002. Brain development, song learning and mate choice in birds: a review and experimental test of the "nutritional stress hypothesis". Journal of Comparative Physiology A 188:1003-1014.

Oberweger K, Goller F, 2001. The metabolic cost of birdsong production. Journal of Experimental Biology 204:3379-3388.

Odom KJ, Benedict L, 2018. A call to document female bird songs: applications for diverse fields. The Auk: Ornithological Advances 135:314-325.

Odom KJ, Hall ML, Riebel K, Omland KE, Langmore NE, 2014. Female song is widespread and ancestral in songbirds. Nature Communications 5:1-6.

Parker TH, Barr IR, Griffith SC, 2006. The blue tit's song is an inconsistent signal of male condition. Behav Ecol 17:1029-1040.

Petrinovich L, Patterson TL, 1981. Field studies of habituation: IV. Sensitization as a function of the distribution and novelty of song playback to white-crowned sparrows. Journal of Comparative and Physiological Psychology 95:805.

Podos J, 1997. A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: *Emberizidae*). Evolution 51:537-551.

Podos J, 2001. Correlated evolution of morphology and vocal signal structure in Darwin's finches. Nature 409:185-188.

Podos J, Moseley DL, Goodwin SE, McClure J, Taft BN, Strauss AV, Rega-Brodsky C, Lahti DC, 2016. A fine-scale, broadly applicable index of vocal performance: frequency excursion. Animal behaviour 116:203-212.

Podos J, Nowicki S, 2004. Performance limits on birdsong. In: Slabbekoorn PM, Hans, editor. Nature's music: the science of birdsong San Diego: Elsevier Academic Press. p. 318-342.

Poesel A, Dabelsteen T, 2005. Territorial responses of male blue tits to simulated dynamic intrusions: effects of song overlap and intruder location. Animal Behaviour 70:1419-1427.

Poesel A, Dabelsteen T, Pedersen SB, 2004. Dawn song of male blue tits as a predictor of competitiveness in midmorning singing interactions. Acta Ethologica 6:65-71.

Poesel A, Foerster K, Kempenaers B, 2001. The dawn song of the blue tit (*Parus caeruleus*) nd its role in sexual selection. Ethology 107:521-531.

Poesel A, Kempenaers B, 2000. When a bird is tired from singing: a study of drift during the dawn chorus. Etología 8:1-7.

Prather J, Peters S, Mooney R, Nowicki S, 2012. Sensory constraints on birdsong syntax: neural responses to swamp sparrow songs with accelerated trill rates. Animal behaviour 83:1411-1420.

Pravosudov VV, Clayton NS, 2002. A test of the adaptive specialization hypothesis: population differences in caching, memory, and the hippocampus in black-capped chickadees (*Poecile atricapilla*). Behavioral neuroscience 116:515.

Price JJ, 2009. Evolution and life-history correlates of female song in the New World blackbirds. Behav Ecol 20:967-977.

Price JJ, 2013. Why is birdsong so repetitive? Signal detection and the evolution of avian singing modes. Behaviour 150:995-1013.

Author. 2016. R: A language and environment for statistical computing. Vienna.

Real LA, 1991. Search theory and mate choice. II. Mutual interaction, assortative mating, and equilibrium variation in male and female fitness. The American Naturalist 138:901-917.

Redfern CPF, Clark JA, 2001. Ringers' Manual. BTO, Thetford, UK.

Reichard DG, Brothers DE, George SE, Atwell JW, Ketterson ED, 2018. Female Dark‐eyed Juncos (*Junco hyemalis*) thurberi produce male‐like song in a territorial context during the early breeding season. Journal of Avian Biology 49:jav-01566.

Riebel K, Hall ML, Langmore NE, 2005. Female songbirds still struggling to be heard. Trends in Ecology & Evolution 20:419-420.

Riebel K, Odom KJ, Langmore NE, Hall ML, 2019. New insights from female bird song: towards an integrated approach to studying male and female communication roles. Biology letters 15:20190059.

Ritchison G, 1986. The singing behavior of female northern cardinals. The Condor 88:156-159.

Rivera-Gutierrez HF, Pinxten R, Eens M, 2011. Songs differing in consistency elicit differential aggressive response in territorial birds. Biology Letters 7:339-342. doi: 10.1098/rsbl.2010.0962.

Rivera-Gutierrez HF, Pinxten R, Eens M, 2012. Tuning and fading voices in songbirds: age-dependent changes in two acoustic traits across the life span. Animal Behaviour 83:1279-1283. doi: 10.1016/j.anbehav.2012.03.001.

Rogers AC, Langmore NE, Mulder RA, 2007. Function of pair duets in the eastern whipbird: cooperative defense or sexual conflict? Behav Ecol 18:182-188.

Rome LC, 2006. Design and function of superfast muscles: new insights into the physiology of skeletal muscle. Annual Review of Physiology 68:193-221.

Rome LC, Syme DA, Hollingworth S, Lindstedt SL, Baylor SM, 1996. The whistle and the rattle: the design of sound producing muscles. Proceedings of the National Academy of Sciences 93:8095-8100.

Ryan MJ, Brenowitz EA, 1985. The role of body size, phylogeny, and ambient noise in the evolution of bird song. The American Naturalist 126:87-100.

Sakata JT, Vehrencamp SL, 2012. Integrating perspectives on vocal performance and consistency. Journal of Experimental Biology 215:201-209.

Salvin P, Derégnaucourt S, Leboucher G, Amy M, 2018. Consistency of female preference for male song in the domestic canary using two measures: Operant conditioning and vocal response. Behavioural processes 157:238-243.

Sandell MI, Smith HG, 1997. Female aggression in the European starling during the breeding season. Animal Behaviour 53:13-23.

Santema P, Teltscher K, Kempenaers B, 2020. Extra‐pair copulations can insure female blue tits against male infertility. Journal of Avian Biology 51:1-6.

Santos E, Nakagawa S, 2012. The costs of parental care: A meta‐analysis of the trade‐off between parental effort and survival in birds. Journal of evolutionary biology 25:1911-1917.

Schlicht L, Kempenaers B, 2020. The effects of season, sex, age and weather on population‐level variation in the timing of activity in Eurasian Blue Tits, *Cyanistes caeruleus*. Ibis 162:1146-1162.

Schmidt R, Kunc HP, Amrhein V, Naguib M, 2008. Aggressive responses to broadband trills are related to subsequent pairing success in nightingales. Behav Ecol 19:635-641.

Searcy WA, 1992a. Measuring responses of female birds to male song. Playback and studies of animal communication: Springer. p. 175-189.

Searcy WA, 1992b. Song repertoire and mate choice in birds. American Zoologist 32:71-80.

Searcy WA, Andersson M, 1986. Sexual selection and the evolution of song. Annual Review of Ecology and Systematics 17:507-533.

Searcy WA, Yasukawa K, 1990. Use of the song repertoire in intersexual and intrasexual contexts by male red-winged blackbirds. Behavioral Ecology and Sociobiology 27:123-128.

Shirihai H, Svensson L, 2018. Handbook of Western Palearctic Birds, Volume 1: Passerines: Larks to Warblers: Bloomsbury Publishing.

Sierro J, Sierro J, Slabbekoorn H, 2020. Experimental test of the communicative value of syllable diversity and syllable switching in the common chiffchaff. Animal Behaviour 165:11-21.

Slagsvold T, Dale S, Lampe HM, 1999. Does female aggression prevent polygyny? An experiment with pied flycatchers (*Ficedula hypoleuca*). Behavioral Ecology and Sociobiology 45:403-410.

Slagsvold T, Lifjeld JT, 1994. Polygyny in birds: the role of competition between females for male parental care. The American Naturalist 143:59-94.

Smith GT, 1996. Seasonal plasticity in the song nuclei of wild rufous-sided towhees. Brain research 734:79-85.

Smith GT, Brenowitz EA, Beecher MD, Wingfield JC, 1997a. Seasonal changes in testosterone, neural attributes of song control nuclei, and song structure in wild songbirds. Journal of Neuroscience 17:6001-6010.

Smith GT, Brenowitz EA, Wingfield JC, 1997b. Seasonal changes in the size of the avian song control nucleus HVC defined by multiple histological markers. Journal of Comparative Neurology 381:253-261.

Soma M, Garamszegi LZ, 2011. Rethinking birdsong evolution: meta-analysis of the relationship between song complexity and reproductive success. Behav Ecol 22:363–371.

Stadler H, 1951. Die stimmen der Blaumeise. Anzeiger der Ornithologische Gesellschaft in Bayern 4:155-170.

Stein RC, 1968. Modulation in bird sounds. The Auk 85:229-243.

Author. 2006. Seewave.

Suthers R, Goller F, Pytte C, 1999. The neuromuscular control of birdsong. Philosophical Transactions of the Royal Society B: Biological Sciences 354:927-939.

Suthers RA, 2004. How birds sing and why it matters. Nature’s music: the science of birdsong San Diego: Elsevier Academic Press. p. 272-295.

Suthers RA, Goller F, 1997. Motor correlates of vocal diversity in songbirds. Current ornithology: Springer. p. 235-288.

Suthers RA, Goller F, Hartley RS, 1996. Motor stereotypy and diversity in songs of mimic thrushes. Journal of neurobiology 30:231-245.

Suthers RA, Zollinger SA, 2008. From brain to song: the vocal organ and vocal tract. Neuroscience of birdsong:78-98.

Svensson L, 1992. Identification guide to European passerines: British Trust for Ornithology.

Symonds MR, Moussalli A, 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike’s information criterion. Behavioral Ecology and Sociobiology 65:13-21.

Theunissen FE, Doupe AJ, 1998. Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVC of male zebra finches. Journal of Neuroscience 18:3786-3802.

Thomas RJ, 2002. The costs of singing in nightingales. Animal Behaviour 63:959-966.

Tobias JA, Gamarra-Toledo V, Garcia-Olaechea D, Pulgarin PC, Seddon N, 2011. Year-round resource defence and the evolution of male and female song in suboscine birds: social armaments are mutual ornaments. Journal of Evolutionary Biology 24:2118-2138.

Tobias JA, Seddon N, 2009. Signal jamming mediates sexual conflict in a duetting bird. Current Biology 19:577-582.

Tomás G, 2015. Hatching date vs laying date: what should we look at to study avian optimal timing of reproduction? Journal of Avian Biology 46:107-112.

Vehrencamp SL, Yantachka J, Hall ML, de Kort SR, 2013. Trill performance components vary with age, season, and motivation in the banded wren. Behavioral Ecology and Sociobiology 67:409-419. doi: 10.1007/s00265-012-1461-x.

Wallschläger D, 1980. Correlation of song frequency and body weight in passerine birds. Cellular and Molecular Life Sciences 36:412-412.

Ward S, Speakman JR, Slater PJ, 2003. The energy cost of song in the canary, Serinus canaria. Animal Behaviour 66:893-902.

Wasserman F, Cigliano J, 1991. Song output and stimulation of the female in white-throated sparrows. Behavioral Ecology and Sociobiology 29:55-59.

Weary DM, Lemon RE, 1988. Evidence against the continuity versatility relationship in bird song. Animal Behaviour 36:1379-1383. doi: 10.1016/s0003-3472(88)80207-0.

Weatherhead PJ, Robertson RJ, 1979. Offspring quality and the polygyny threshold:" the sexy son hypothesis". The American Naturalist 113:201-208.

Webb WH, Brunton DH, Aguirre JD, Thomas DB, Valcu M, Dale J, 2016. Female song occurs in songbirds with more elaborate female coloration and reduced sexual dichromatism. Frontiers in Ecology and Evolution 4:22.

Welling P, Koivula K, Lahti K, 1995. The dawn chorus is linked with female fertility in the willow tit *Parus montanus*. Journal of Avian Biology 26:241-246.

Wesołowski T, 2004. The origin of parental care in birds: a reassessment. Behav Ecol 15:520-523.

Wilkins MR, Odom KJ, Benedict L, Safran RJ, 2020. Analysis of female song provides insight into the evolution of sex differences in a widely studied songbird. Animal Behaviour 168:69-82.

Williams GC, 1966. Natural Selection, the Costs of Reproduction, and a Refinement of Lack's Principle. The American Naturalist 100:687-690. doi: 10.1086/282461.

Author. 2017. Package ‘gamm4’.

Woolley SC, Doupe AJ, 2008. Social context–induced song variation affects female behavior and gene expression. PLoS Biol 6:e62.

Yaskin V, 2011. Seasonal changes in hippocampus size and spatial behavior in mammals and birds. Biology Bulletin Reviews 1:279-288.

Yaskin VA, 1984. Seasonal changes in brain morphology in small mammals. Winter ecology of small mammals 10:183-191.

Zahavi A, 1980. Ritualization and the evolution of movement signals. Behaviour 72:77-80.

Zollinger SA, Riede T, Suthers RA, 2008. Two-voice complexity from a single side of the syrinx in northern mockingbird Mimus polyglottos vocalizations. Journal of Experimental Biology 211:1978-1991.

Zollinger SA, Suthers RA, 2004. Motor mechanisms of a vocal mimic: implications for birdsong production. Philosophical Transactions of the Royal Society B: Biological Sciences 271:483-491.

Zuur A, Ieno E, Walker N, Saveliev A, Smith G, 2009. Mixed effects models and extensions in ecology with R. Gail M, Krickeberg K, Samet JM, Tsiatis A, Wong W, editors. New York: Spring Science and Business Media.