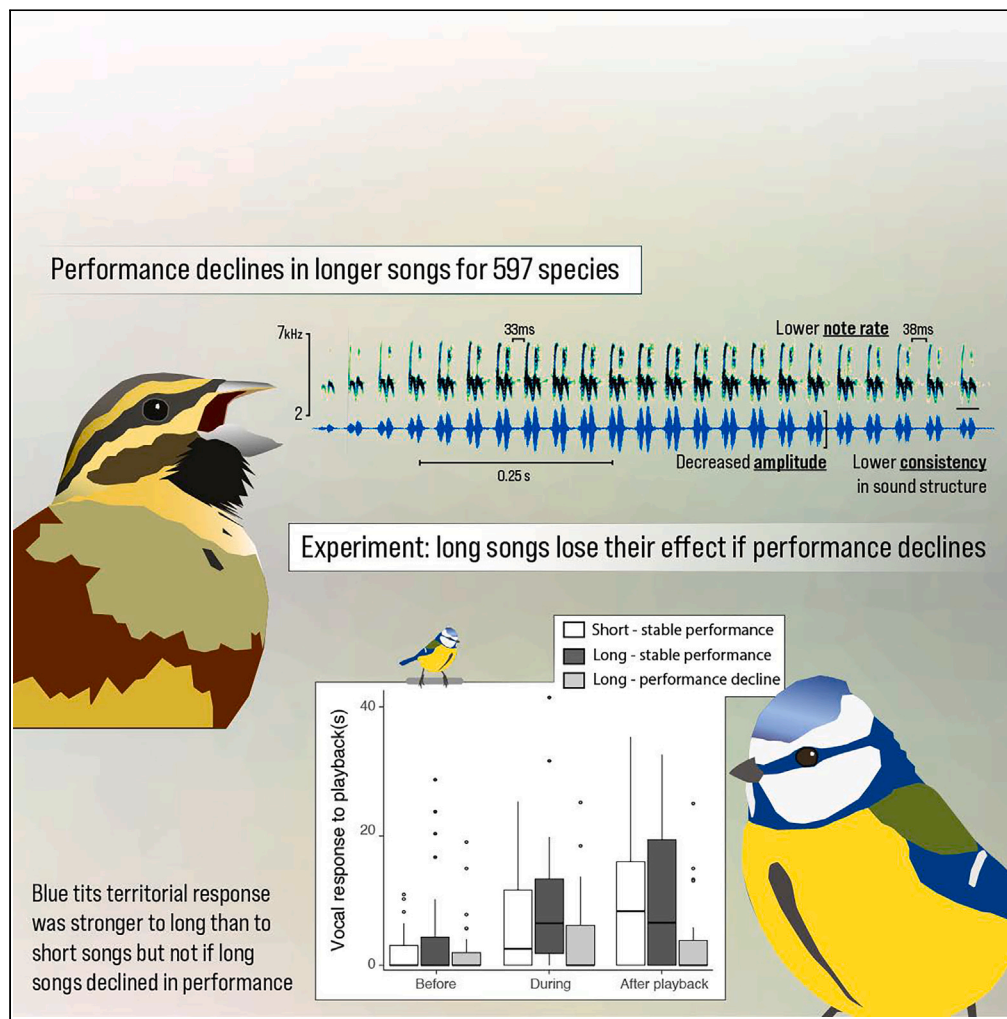


Article

# A limit to sustained performance constrains trill length in birdsong



Javier Sierra,  
Selvino R. de Kort,  
Ian R. Hartley

sierra.2.8@gmail.com

Highlights

We hypothesize there is a limit to sustained song performance in birds

We found that within-song performance declines at the end of songs in 597 species

Our experiment showed that long songs lose their effect if performance declines

Duration vs. quality of performance trade-off ensures honesty in communication



## Article

## A limit to sustained performance constrains trill length in birdsong

Javier Sierra,<sup>1,3,\*</sup> Selvino R. de Kort,<sup>2</sup> and Ian R. Hartley<sup>1</sup>

## SUMMARY

**In birds, song performance determines the outcome of contests over crucial resources. We hypothesized that 1) sustained performance is limited within song, resulting in a performance decline towards the end and 2) the impact of song length is compromised if performance declines. To test these hypotheses, we analyzed the songs of 597 bird species (26 families) and conducted a playback experiment on blue tits (*Cyanistes caeruleus*). Our multi-species analysis showed that song performance declines after sustained singing, supporting our hypothesis. If the performance decline is determined by individual attributes (i.e., physical condition), our results explain how trill length can honestly signal quality. Our experiment showed that longer trills of high performance elicited a stronger response during territorial interactions. However, long trills that declined in performance elicited a weaker response than short, high-performance trills. A trade-off between the duration and performance quality of a motor display can be an important aspect in communication across taxa.**

## INTRODUCTION

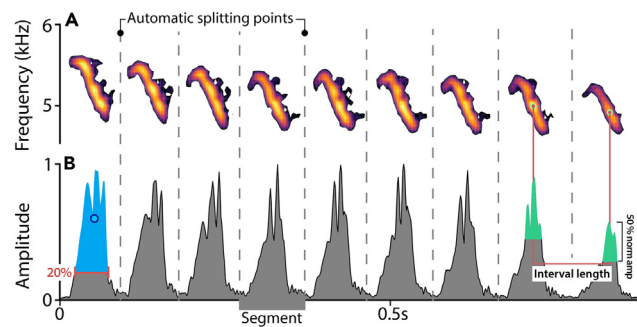
Many animals perform ritualized displays of stereotyped motor patterns to moderate social interactions<sup>1</sup> and motor performance of such displays is key for communication in many taxa.<sup>2,3</sup> The performance of motor displays has been proposed as an honest signal of whole-organism condition, not always because it is energetically costly, but because executing challenging motor patterns provides information on the functioning of multiple organ systems.<sup>4–6</sup>

Birdsong is famous for the diversity of sounds, but perhaps it is less well known that singing involves a skillful coordination of multiple muscle groups. Abdominal muscles regulate air pressure in the respiratory system while the phonatory muscles of the syrinx move at high speeds to modulate the frequency of sounds.<sup>7–9</sup> Ultimately, the muscles controlling the shape of the vocal tract and beak gape adjust precisely within milliseconds to potentiate or filter certain frequencies produced in the syrinx.<sup>10</sup> Most of this takes place hidden from view, but the song output is the manifestation of such motor patterns, and it can be used by receivers to assess the performance of the motor display. As with other ritualized displays,<sup>11</sup> motor performance of song may indicate the general condition of the neuro-motor system of the individual.<sup>6,12–14</sup> In birds, high song performance correlates with increased motivation,<sup>15,16</sup> higher social status,<sup>17,18</sup> increased reproductive output,<sup>19–21</sup> longevity<sup>22</sup> and sexual attractiveness.<sup>23–25</sup>

Motor performance of singing can be displayed in different song parameters (Figure 1). The sound amplitude of song is directly associated with the expiratory pressure controlled by abdominal muscles<sup>8</sup> and it increases during song development.<sup>26–28</sup> Dusky warblers (*Phylloscopus fuscatus*) that sing a larger proportion of their songs over a certain threshold of relative amplitude obtain more extra-pair copulations and have higher reproductive success.<sup>29</sup> Another aspect of song performance is signaling rates,<sup>30</sup> as females tend to prefer songs with faster note rates<sup>23,31</sup> and note rate is also relevant during territorial contests.<sup>32–34</sup> As in other displays by animals, repetitive songs are common in songbirds,<sup>6,35</sup> and many species produce trilled songs, where the same note type (i.e., identical sound elements) is repeated multiple times consecutively. Song performance can be assessed by the consistency in the repetition of the same note as the same motor pattern is repeated (vocal consistency).<sup>22,36,37</sup> Experimental evidence shows that birds can discriminate between songs of high and low vocal consistency<sup>18,22,38</sup> and birds with higher vocal consistency have higher social status and increased fitness.<sup>18,25,39</sup> Birds increase their vocal consistency during high-motivation contexts<sup>15,32,37</sup> and vocal consistency also increases during song maturation over the breeding season.<sup>40,41</sup>

The multidimensionality of birdsong means that performance parameters can have interacting effects, causing performance constraints.<sup>30</sup> A well-known trade-off occurs between the duration of inter-note intervals and the extent of mechanical reconfiguration of the vocal apparatus between notes.<sup>42,43</sup> Male songs performed close to this limit (defined as having low vocal deviation) elicit a more aggressive response from other males<sup>43</sup> and are preferred by females.<sup>23,31</sup> Furthermore, maximizing performance in terms of note rate and bandwidth can reduce singing precision as indicated by decreased vocal consistency,<sup>44</sup> although other studies showed that songs with high vocal deviation (lower performance) are more likely to have “mistakes,” defined as song structures that deviate from their standard form.<sup>39</sup>

<sup>1</sup>Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK<sup>2</sup>Ecology and Environment Research Centre, Department of Natural Sciences, Manchester Metropolitan University, Manchester M15 6BH, UK<sup>3</sup>Lead contact\*Correspondence: [sierra.2.8@gmail.com](mailto:sierra.2.8@gmail.com)  
<https://doi.org/10.1016/j.isci.2023.108206>



**Figure 1. Sound spectrogram of a trill part from a blue tit song**

(A), with time on the X axis and frequency on the Y axis. The associated amplitude envelope (B), with time on the X axis and the amplitude, normalized along the trill, on the Y axis. The splitting points (dashed vertical lines) between notes were detected automatically using a script in R. Vocal consistency was measured by comparing each note (segments between dashed lines) to all other notes using SPCC. In the portion of the segment above a 20% amplitude threshold, relative to the maximum amplitude in the trill (blue area), we measured mean amplitude (blue circle). The reference points for measuring interval length between notes were the middle time point within the portion of each note's waveform that was greater than 50% of the peak amplitude for that note (see green annotations at the right of the figure). Note rate was the inverse of interval length for each note, indicating the speed of note repetition in notes per second.

Another crucial parameter in birdsong is song length. Experimental evidence shows that longer songs elicit a stronger response during simulated territorial interactions.<sup>45,46</sup> In some species, song length varies seasonally, peaking during the fertile period,<sup>40,47</sup> is positively associated with an individual's physical condition<sup>48,49</sup> and with survival,<sup>50</sup> and is under positive sexual selection.<sup>20,51,52</sup> In other taxa such as amphibians and insects, longer calls are also perceived as more attractive.<sup>53,54</sup> However, many bird species that show a stronger response to longer songs also decrease song length during aggressive interactions.<sup>45,55,56</sup> Finally, it is unclear how variation in song length reflects individual quality as singing does not appear to be metabolically demanding.<sup>57,58</sup>

Here, we propose that sustained singing results in a decline in song performance, probably due to a decrease in air pressure within the respiratory system and maybe also to oxygen deprivation or muscle fatigue. To investigate this hypothesis, we focused on trilled structures within songs, which are common for a wide range of species of songbirds (see Results). As trills are made up of the same repeated note, it is possible to investigate how performance changes within the trill in relation with varying trill length, removing possible confounding factors associated with the use of different note types in the song. Trill length is strikingly variable within and between species and will likely depend on habitat structure,<sup>59</sup> motivational state<sup>16</sup> or species-specific singing skills.<sup>60</sup> For instance, domestic canaries (*Serinus canaria*) produce long trills of many notes, and to avoid air depletion during sustained singing, canaries take short mini-breaths in between notes within the trill.<sup>61</sup> Northern mockingbirds (*Mimus polyglottos*) trying to mimic the canary song do not take mini-breaths and seem unable to match the canaries' trills in length.<sup>60</sup> The resulting mimicry is different from the original canary song, as it is composed of short trills produced via pulsatile expiration (without mini-breaths) followed by long gaps of inspiration to restore air pressure.<sup>60</sup>

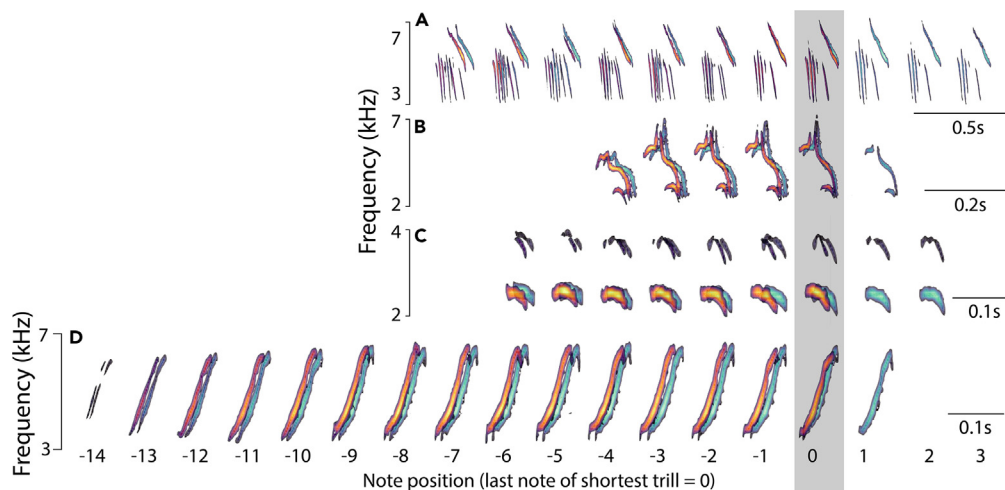
Despite variation in absolute trill length between and within species, we hypothesize that individual birds will be limited in their capacity to sing continuously with high performance, reaching a point in the song where continued singing leads to a decline in motor performance. This is similar to a trade-off between sound amplitude and note duration observed in white bellbirds (*Procnias albus*).<sup>62</sup> We also hypothesize that the impact of trill length in communication will be compromised for trills of declining performance. We expect this point of decline to vary between individuals, potentially reflecting inter-individual differences relevant during mate choice or agonistic interactions.

To support our hypotheses, we first look for a pattern of performance decline within the trill after sustained singing in a multi-species analyses which includes 597 bird species. We then test whether the information encoded in trill length is compromised by the performance level using a playback song experiment on wild blue tits. In blue tits, trill length varies seasonally peaking during the fertile period,<sup>40</sup> and has been associated with more extra-pair mating success<sup>20</sup> and increased survival in males.<sup>50</sup> Furthermore, vocal consistency, one measure of song performance,<sup>37</sup> also varies seasonally, peaking during the fertile period of the female, is associated with larger clutches, and is preferred by females.<sup>25</sup> During agonistic interactions, song consistency also plays a role in this species as male blue tits show a stronger response to high consistency trills (de Kort, In prep).

## RESULTS

### Multi-species song performance within trills

We checked 1994 species (all members of 26 families) in the Xeno-Canto repository and 492 species did not have any recordings under the "song" and "A quality" categories. From the remaining 1502 species ( $14.8 \pm 12.3$  recordings available per species) we analyzed the trills of 597 species ( $20.3 \pm 13.1$  recordings available per species, Figure 2), while we did not find suitable recordings of trilled songs in 925 species ( $11.4 \pm 10.5$  recordings available per species). This could be because we did not find suitable high-quality recordings for analysis (278 species were represented by only 3 recordings or fewer) or because those species did not produce trills in their songs. The model to study the change in performance at the beginning of a trill included 6317 trills ( $3.8 \pm 1.8$  trills per individual) in 1680 recordings (individuals) and 597 species. See Table S1 for a list of the species selected within each family and published data from this study at Figshare



**Figure 2. Spectrograms of trills from example species included in this study**

To understand the variation in performance at the end of trills of different lengths, we numbered all trills in relation to the shortest trill within our data for each individual and song type. The shortest trill is shown in orange, and a second, longer, trill is shown in green for each species. All “short trills” in orange end at position “zero,” which is the last note of the shortest trill for an individual and song type. Note positions above zero (+1 to +3 in this case) reflect the last notes of longer trills. The species shown are water pipit *Anthus spinoletta* (A), Oaxaca sparrow *Aimophila notostica* (B), aquatic warbler *Acrocephalus paludicola* (C) and pine warbler *Setophaga pinus* (D).

repository: <https://doi.org/10.6084/m9.figshare.24173367.v1> for a full table of the Xeno-Canto tracks analyzed and the specific recording ID code. To investigate change in performance at the end of trills, we selected all individuals with at least two trills analyzed for each song type. This subset included 5202 trills in 1511 different recordings (individuals) and  $3.8 \pm 1.7$  trills per song type per individual. The sample is reduced as we only selected those song types within individuals that had at least two trills sampled ( $3.5 \pm 1.6$  trills per individual and song type). Sometimes we selected multiple song types within one individual, but this was rare ( $1.1 \pm 0.31$  song types per individual). Mean trill length across all species was  $8.5 \pm 6.6$  notes and 95% of all trills were fewer than 21 notes in length, with a minimum of 3 notes and a maximum of 127 notes.

#### Change in song performance at the beginning of trills

All three performance variables changed significantly at the beginning of trills and these changes were not linear, as indicated by the effective degrees of freedom (EDF) that are significantly different from 1 (Tables 1 and 2).<sup>63</sup> The change in performance at the start of trills was not the same for all variables. The note rate was highest at the start, then declined steadily until an inflection point at note position 8, when the note rate became stable (Tables 1 and 2; Figures 3A and S2A). Both relative amplitude and vocal consistency increased significantly over the first three notes and became stable after the inflection point (Tables 1 and 2; Figures 3B, 3C, S2B and S2C). The estimated inflection point for the change in relative amplitude was at note position 4.3, (Tables 1 and 2; Figures 3B and S2B), while the inflection point for vocal consistency was estimated at note position 3.2 (Tables 1 and 2; Figures 3C and S2C).

#### Change in song performance after sustained singing

All three performance variables declined significantly, but non-linearly after sustained singing within trills (Tables 3 and 4; Figures 3D–3F and S2D–S2F). Note rate, relative amplitude and vocal consistency showed stable performance until approximately note position –2 (see Tables 3 and 4 for the estimated inflection points for each variable). This is two notes before the end in the case of the shortest trills within individual, but up to six notes before the end in some longer trills. After the inflection point, performance declined in note rate, relative amplitude and consistency (Tables 3 and 4; Figures 3D–3F and S2D–S2F). Our post-hoc analysis showed that all three performance variables (note rate, amplitude, and consistency) were significantly lower in the last four notes of long trills compared to the last four notes in the shortest trills, within individual and song type (Table S2 in supplemental information).

#### Playback experiment

We conducted 52 playback tests in 31 trials, testing 31 individual males. For practical reasons, not all trials could be completed with two tests. From the 31 trials, 21 trials included both the trill length and the consistency decline tests, 42 playback tests in total (Figure 4). In 12 trials, the trill length test was conducted first and in 9 trials the consistency decline test was first. Ten trials, on ten different males, only received one test, five received only the trill length test while five received only the consistency decline test.

Before playback, approach response (time spent within a 4 m range of to the loudspeaker) was not different between treatments. During the playback phase of the trill length test, blue tit males spent significantly more time within 4 m of the speaker in response to the long-stable treatment than to the short-stable treatment (Figure 5; Table 5). However, no difference was found in the approach response during playback

**Table 1. Results from the GAMM model fitted to investigate variation in vocal performance at the beginning of trills**

| Fixed effect                      | EDF   | F       | P      | Inflection point |
|-----------------------------------|-------|---------|--------|------------------|
| Note rate: Note position          | 3.208 | 131.842 | <0.001 | 8.5              |
| Relative amplitude: Note position | 4.656 | 92.546  | <0.001 | 4.3              |
| Vocal consistency: Note position  | 8.52  | 599.487 | <0.001 | 3.2              |

The table shows the Effective Degrees of Freedom (EDF) for the smooth terms and the test statistics derived from the frequentist properties of Bayesian confidence intervals for smooths (Marra & Wood 2012). The results show that performance varies significantly and not linearly at the start of trills. Analysis of the first derivative on the GAMM splines showed the inflection points after which the slope is not significantly different from zero.

between the short-stable and the long-decline treatment. After the playback, the approach response was lower to the long-decline than to the short-stable treatment, although this was a non-significant trend (Estimate =  $-8.4$ , 5% CI:  $-15.5$ , 95% CI:  $-1.3$ , Figure 5; Table 5). Overall, regardless of treatment, approach response was significantly higher during and after playback, compared to before playback. The order of stimuli within tests or the order of tests within trial did not affect the approach response significantly.

Before playback, vocal response was not different between treatments (Figure 5; Table 6). During the playback phase, the vocal response did not differ significantly between treatments (Figure 5; Table 6). After the playback, vocal response was significantly lower toward the long-decline than to the short-stable treatment, while no difference was found between the short-stable and the long-stable treatments in this phase (Figure 5; Table 6). Overall, regardless of treatment, vocal response was higher during and after playback, compared to before the playback. In this case, we found a significant effect of order of stimulus presentation, as males showed a greater response to the second presentation within a test, while no effect was found for the order of tests in a trial (Figure 5; Table 6).

## DISCUSSION

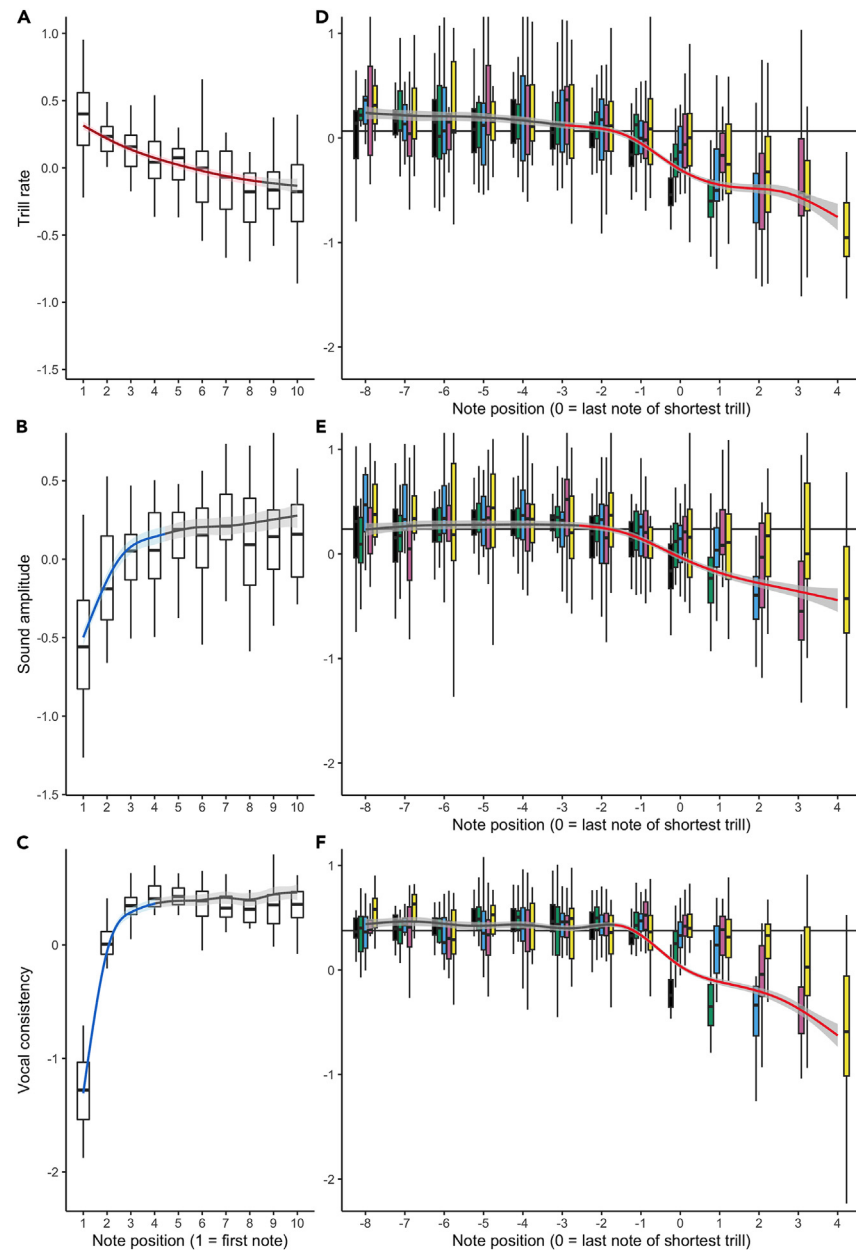
For 26 families of songbirds, we found that individual birds sang with high and stable performance up until the last two notes when producing their shortest trills. However, longer trills followed a progressive decline in performance at the end. The data support our hypothesis that sustained performance within song is limited and such a trade-off appears common in a large number of taxonomically diverse species of songbirds. Our playback experiment on territorial male blue tits showed that long trills elicit a stronger response than short trills for songs of high vocal performance. However, long trills that declined in performance in the middle of the trill elicit a weaker response than short trills of high and stable performance. These results show that trill length is meaningful in communication during territorial interactions, but that effect disappears if song performance declines within the trill. Identifying the limitations and constraints in motor displays is a crucial step in the search for meaningful measures of motor performance in songbirds and other taxa.<sup>12</sup>

Our multi-species analysis indicates that song performance is stable within individuals for most of the length of the shortest trills, while there is a decline in song performance at the end in longer trills. Our analysis further showed that the last notes of short trills show higher performance than the last notes of longer trills, supporting our prediction that extending the trill results in notes of lower performance. Male dark-eyed juncos (*Junco hyemalis*) increase both trill length and within-song performance during agonistic singing,<sup>16</sup> but these observations do not contradict our results, as a limit may still be present despite increased overall performance during high-motivation contexts. The playback experiment in blue tits shows that motor performance (i.e., vocal consistency) and trill length play a key role in communication during territorial interactions. We observed a stronger approach response during the playback phase to long than to short trills of high consistency. This agrees with previous studies that highlight trill length as a key feature of blue tit song.<sup>20,50</sup> In other songbird species, trill length has been positively associated with body condition,<sup>48</sup> social status,<sup>49</sup> and reproductive success.<sup>51,52</sup> However, when presented with long trills

**Table 2. Results from the GAMM model fitted to investigate variation in vocal performance at the beginning of trills**

| Fixed effect                      |          |         |          |         | Random effects       |               |
|-----------------------------------|----------|---------|----------|---------|----------------------|---------------|
| Variable                          | Estimate | 2.5% CI | 97.5% CI | T       | Variable             | Variance (SD) |
| Note rate: Intercept              | 0.099    | 0.083   | 0.116    | 11.958  | Species/Genus/Family | 0.059 (0.24)  |
| Note rate: Note position          | -0.419   | -0.469  | -0.369   | -16.508 | Residual             | 0.82 (0.68)   |
| Relative amplitude: Intercept     | 0.457    | 0.449   | 0.465    | 112.858 | Species/Genus/Family | 0.39 (0.15)   |
| Relative amplitude: Note position | 0.046    | 0.04    | 0.053    | 13.81   | Residual             | 0.85 (0.73)   |
| Vocal consistency: Intercept      | 0.945    | 0.941   | 0.95     | 436.718 | Species/Genus/Family | 0.34 (0.12)   |
| Vocal consistency: Note position  | -0.099   | -0.105  | -0.094   | -33.452 | Residual             | 0.79 (0.63)   |

The table shows the parametric estimates for the predictors, with the associated 95% confidence intervals (CI) and the T statistic. These coefficients show that all performance variables had a significant directional change along with note position as the 95% CI do not overlap with zero. However, note rate decreased significantly while relative amplitude and vocal consistency increased significantly. The estimated variance and the standard deviation (SD) explained by the random effects are shown in the last two columns.



**Figure 3. Variation of song performance within trills for 597 species of songbirds**

All variables were normalized within trill (see STAR Methods). (A–C) show the change relative to the beginning of the trill (first note = note 1), and (D–F) show the change relative to the end, taking the last note of the shortest trill as reference (note 0). Boxplots show median, higher and lower quartiles and 1.5 inter-quartile range for the three performance variables at each note position. The color of the boxes depicts different lengths of trill: black, for the shortest trill within individual and song type, green for 1-note longer trills, blue for 2-note longer trills, pink for 3-note longer trills and yellow for 4-note longer trills. Lines show the fitted values from the GAMMs with the confidence intervals as a shaded area around the lines. Blue lines indicate significantly positive change in performance (positive slope) whereas red lines indicate a significant performance decline (negative slope) and gray represents stable performance (slope is not significantly different from zero).

of declining consistency, the approach response of territorial males was equal to the response to short trills of stable consistency. Moreover, post-playback vocal response was significantly lower to long trills of declining consistency than to short trills of high, stable consistency. These experimental results show that the perceptual impact of long trills of declining performance is equal, or lower, to that of short trills of high performance. In distantly related taxa, such as gray frogs (*Dryophytes chrysoscelis*), a similar phenomenon has been described, as an attractive signal (high call rate) loses its advantage if produced inconsistently.<sup>64</sup> In birds, long songs do not appear to be metabolically costly,<sup>57,58</sup> raising

**Table 3. Results from the GAMM model fitted to investigate variation in vocal performance at the end of trills**

| Fixed effect                      | EDF   | F       | P      | Inflection point |
|-----------------------------------|-------|---------|--------|------------------|
| Note rate: Note position          | 7.854 | 195.87  | <0.001 | -2.1             |
| Relative amplitude: Note position | 5.107 | 78.887  | <0.001 | -2.0             |
| Vocal consistency: Note position  | 7.696 | 186.572 | <0.001 | -1.7             |

The table shows the Effective Degrees of Freedom (EDF) for the smooth terms and the test statistics derived from the frequentist properties of Bayesian confidence intervals for smooths (Marra & Wood 2012). The results show that performance varies significantly and not linearly at the start of trills. Analysis of the first derivative on the GAMM splines showed the inflection points after which the slope is not significantly different from zero.

the question of why trill length is relevant during interactions. The fact that the perception of trill length is affected by vocal performance within song may be a clue to the communicative value of trill length in birdsong.

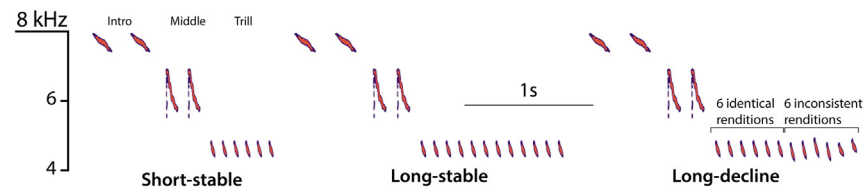
If the point where performance declines is determined by an individual's condition, physical attributes and/or motor skill, trill length would be an honest signal of individual quality since the communicative value of trill length is effectively constrained by the point of decline in performance. Although the physiological mechanisms underlying our observations are beyond the scope of this article, we propose that the decline in performance is likely associated with reduced sub-syringeal pressure and, potentially, with oxygen deprivation or muscle fatigue after sustained singing. To produce song, birds use air stored in their respiratory system to create sound via oscillations of the labia in the syrinx.<sup>65</sup> Depletion of air during sustained singing will reduce air pressure leading to decreased amplitude of vocalizations, as observed in our multi-species analysis. Reduced air pressure will also cause the syrinx to exhibit nonlinear oscillatory behaviors, decreasing control of the resulting vocalizations,<sup>66,67</sup> matching the observed decline in consistency. Taking mini-breaths in between notes will avoid air depletion, allowing birds to sing uninterrupted for a longer period of time. However, not all species seem to take mini-breaths during singing,<sup>60</sup> and some authors suggest that mini-breaths may not be sufficient to restore air pressure completely after each note in all cases.<sup>8,68</sup> It has been suggested that blood oxygenation during song is less efficient, as mini-breaths may not renew air in all areas of the respiratory system.<sup>8,68</sup> In fact, birds perform specific respiratory patterns immediately before and following song, indicating that special respiratory requirements may constrain song production in some species.<sup>69</sup> The decrease in note rate that we observed at the end of trills (longer intervals) could aid to increase air pressure by increasing volume of mini-breaths and compensate for lower oxygen levels by leaving more time for gas exchange to take place before air is used to produce song. Muscle fatigue could also explain the observed decline in performance. Studies conducted in mammals show that the sustained production of a motor task can cause muscle fatigue, leading to lower force,<sup>70</sup> decreased contraction speed,<sup>71</sup> slower muscle relaxation rates<sup>72</sup> and lower precision of task performance.<sup>73</sup> Although no similar data is available for vocal musculature in songbirds, these muscular changes match the observed decline in note rate, relative amplitude and vocal consistency within trills in birds.

Our data show that birds are able to start song with maximal performance for some parameters as they sing at maximum note rate from the start. This is not the case for vocal consistency and relative amplitude that increase over the first 3–4 notes of the trill before reaching a plateau. An increase in song performance within hours or days due to practice (i.e., warm-up effect) has been demonstrated in birds,<sup>22,74–76</sup> but these changes are difficult to compare to our observed pattern because of the difference in time scales. The performance decline that we observed seems dysfunctional, based on our experimental results, but this is not necessarily the case for the initial increase in performance. A “fade-in” of increasing performance could be due to a physiological constraint but also serve an introductory function, drawing the attention of the audience in preparation for the part of the trill that is produced with maximal performance. This idea is speculative, and the functional role of such fade-in pattern must be tested experimentally.

**Table 4. Results from the GAMM model fitted to investigate variation in vocal performance at the end of trills**

| Fixed effect                      |          |         |          |         | Random effects       |               |
|-----------------------------------|----------|---------|----------|---------|----------------------|---------------|
| Variable                          | Estimate | 2.5% CI | 97.5% CI | T       | Variable             | Variance (SD) |
| Note rate: Intercept              | -0.124   | -0.144  | -0.104   | -12.191 | Species/Genus/Family | 0.09 (0.30)   |
| Note rate: Note position          | -1.038   | -1.116  | -0.96    | -26.117 | Residual             | 0.65 (0.80)   |
| Relative amplitude: Intercept     | 0.461    | 0.456   | 0.466    | 189.362 | Species/Genus/Family | 0.13 (0.36)   |
| Relative amplitude: Note position | -0.058   | -0.066  | -0.049   | 13.428  | Residual             | 0.64 (0.80)   |
| Vocal consistency: Intercept      | 0.959    | 0.957   | 0.961    | 997.696 | Species/Genus/Family | 0.08 (0.28)   |
| Vocal consistency: Note position  | -0.075   | -0.081  | -0.069   | -25.821 | Residual             | 0.55 (0.74)   |

The table shows the parametric estimates for the predictors, with the associated 95% confidence intervals (CI) and the T statistic. These coefficients show that all performance variables declined significantly along with note position as the 95% CI do not overlap with zero. The estimated variance and the standard deviation (SD) explained by the random effects are shown in the last two columns.

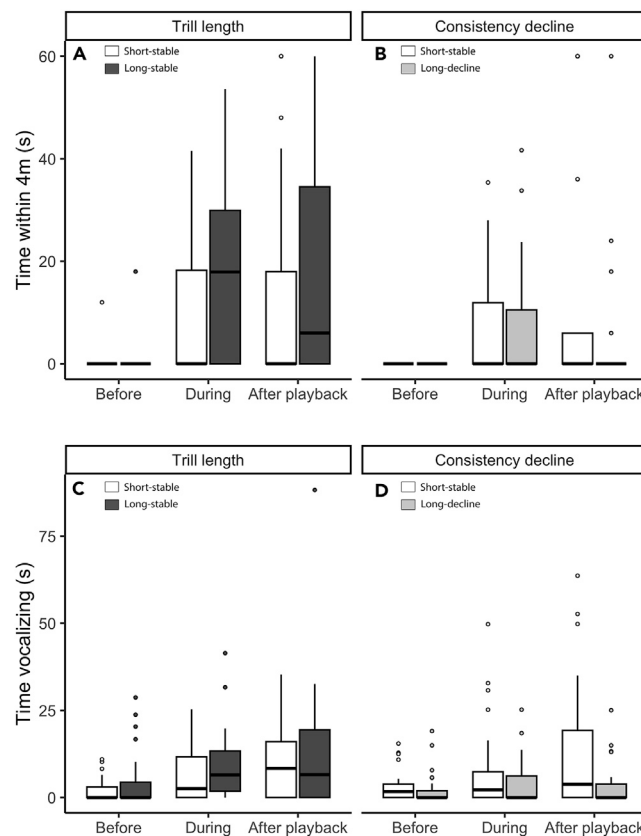


**Figure 4. Examples of the three playback song treatments used as stimuli in the experiment, all synthesized using R**

The short-stable and long-stable have identical renditions of the same note within the trill, only differing in the number of notes (trill length). The long-decline treatment has the same six identical notes in the first half, whereas the last six notes show lowered vocal consistency as we introduced variations in the note duration, note peak frequency and note bandwidth. The long-decline treatment represents a bird that would sing a long trill (12 notes) despite the decline in vocal consistency after the 6<sup>th</sup> note.

Overall, our findings highlight the importance of song performance in communication, as longer trills ending with low-performance notes are ineffective. Maximizing sound amplitude during singing will increase the active range, making signal transmission more effective<sup>59,77</sup> and has been shown to be more attractive to potential mates.<sup>29</sup> Singing a relatively fast song may also be more effective in the context of mate choice<sup>23,31,78</sup> and during territorial contests.<sup>32,33</sup> Producing precise motor patterns (vocal consistency) is also important for intra- and inter-sexual interactions in songbirds.<sup>37</sup> Our findings thus are consistent with recent studies on the importance of skilled moves (i.e., vocal consistency), rather than pure stamina (i.e., trill length), during fights.<sup>79</sup> Finally, if any of the proposed physiological mechanisms are responsible for the observed decline in performance, including the capacity to store air or muscle resistance to fatigue, trill length would convey information relevant to fitness.

Our hypothesis may be generally applicable to other signaling systems across taxa. Many animal displays involve a series of repetitive motor patterns<sup>1,35,64</sup> and, even if energetic costs are low, continued repetition can produce a decline in task performance. Physiological limitations on the sustained production of high-performance displays could be meaningful during many types of social interactions, potentially reflecting inter-individual differences relevant in mate choice and resource defense. Whether it is a dancing human,<sup>80,81</sup> an acrobatic



**Figure 5. Approach (A and B) and vocal (C and D) response of male blue tits before, during and after playback stimulation**

White represents the “short-stable” treatment, which is compared “long-stable” treatment (black) in one test while it is compared to “long-decline” treatment (gray) in the other test within trial. Boxplots show median, higher and lower quartiles and 1.5 inter-quartile range with dots indicating outliers.



**Table 5. Results from the model investigating approach response of male blue tits in the playback experiment**

| Fixed effect                                  | Estimate | 2.5% -  | 97.5% CI | T      | Random effect | Variance (SD) |
|-----------------------------------------------|----------|---------|----------|--------|---------------|---------------|
| Intercept (Short, Before, test 1, stimulus 1) | 1.229    | -5.009  | 7.46     | 0.381  | Individual    | 34.48 (5.87)  |
| Treatment (before playback)                   |          |         |          |        |               |               |
| Long vs. Short                                | -0.169   | -6.478  | 6.141    | -0.052 | Residual      | 166.4 (5.9)   |
| Long-decline vs. Short                        | 0.539    | -5.805  | 6.883    | 0.164  |               |               |
| Treatment (during playback)                   |          |         |          |        |               |               |
| Long vs. Short                                | 8.976    | 0.511   | 17.44    | 2.048  |               |               |
| Long-decline vs. Short                        | -0.55    | -9.015  | 7.915    | -0.126 |               |               |
| Treatment (after playback)                    |          |         |          |        |               |               |
| Long vs. Short                                | 6.115    | -2.349  | 14.58    | 1.396  |               |               |
| Long-decline vs. Short                        | -8.423   | -16.888 | 0.042    | -1.922 |               |               |
| Phase (During vs. Before)                     | 7.581    | 2.694   | 12.468   | 2.996  |               |               |
| Phase (After vs. Before)                      | 12.577   | 7.69    | 17.464   | 4.971  |               |               |
| Test order                                    | -0.605   | -3.465  | 2.255    | -0.409 |               |               |
| Stimulus order                                | -0.746   | -4.875  | 3.377    | -0.349 |               |               |

Provided are model estimates, the associated 95% confident intervals (CI) and the T statistic. The intercept shows the approach behavior before playback of the short treatment. The next rows after the intercept show the estimates for approach response for each treatment during each phase of the playback, compared to the response to the short-stable treatment of that phase.

bird,<sup>4</sup> a waving frog<sup>82</sup> a fighting crab<sup>83</sup> or a singing rodent<sup>84</sup> motor performance is key in communication. We suggest that a trade-off between the length of the display and the quality of motor performance is an important aspect in communication across taxa.

### Limitations to the study

This study overcame the difficulty of applying the same methodology to the analysis of songs from a wide range of species with highly variable song structure. However, in order to do this, we had to normalize the performance variables within songs, removing inter-species and inter-individual variation. Such variation can be valuable to investigate the relationship between absolute performance measures and body features and singing abilities. Although many species produced trills, we also found many other types of song structures that could not be included in this study. Further work is needed for a complete and comprehensive study of birdsong performance across species.

**Table 6. Results from the model investigating vocal response of male blue tits in the playback experiment**

| Fixed effect                | Estimate | 2.5% -  | 97.5% CI | T      | Random effect | Variance (SD) |
|-----------------------------|----------|---------|----------|--------|---------------|---------------|
| Intercept                   | -2.712   | -7.545  | 2.119    | -1.083 | Individual    | 19.84 (4.5)   |
| Treatment (before playback) |          |         |          |        |               |               |
| Long vs. Short              | 2.636    | -2.35   | 7.622    | 1.02   | Residual      | 96.4 (9.8)    |
| Long-decline vs. Short      | -1.517   | -6.44   | 3.405    | -0.595 |               |               |
| Treatment (during playback) |          |         |          |        |               |               |
| Long vs. Short              | -0.153   | -6.829  | 6.522    | -0.044 |               |               |
| Long-decline vs. Short      | -3.638   | -10.223 | 2.948    | -1.066 |               |               |
| Treatment (after playback)  |          |         |          |        |               |               |
| Long vs. Short              | -1.246   | -7.921  | 5.43     | -0.36  |               |               |
| Long-decline vs. Short      | -7.921   | -14.507 | -1.336   | -2.321 |               |               |
| Phase (During vs. Before)   | 4.973    | 1.145   | 8.801    | 2.507  |               |               |
| Phase (After vs. Before)    | 9.094    | 5.267   | 12.922   | 4.585  |               |               |
| Test order                  | 3.08     | 0.843   | 5.316    | 2.657  |               |               |
| Stimulus order              | 1.875    | -1.322  | 5.068    | 1.132  |               |               |

Provided are model estimates, the associated 95% confident intervals (CI) and the T statistic. The intercept shows the vocal behavior before playback of the short treatment. The next rows after the intercept show the estimates for time spent vocalizing for each treatment during each phase of the playback, compared to the response to the short-stable treatment of that phase.

**STAR★METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND PARTICIPANT DETAILS**
  - Multi-species analysis
  - Playback experiment in blue tits
- **METHOD DETAILS**
  - Multi-species analysis
  - Playback experiment in blue tits
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Multi-species analysis
  - Playback experiment in blue tits

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.108206>.

**ACKNOWLEDGMENTS**

We want to thank the referees for the comprehensive review of the study with many constructive comments. Also, we are grateful to the input provided by Dr. Anne Sue Zollinger on the first version of the article.

**AUTHOR CONTRIBUTIONS**

S.K. and J.S. provided the overall research concept and conceived the study; J.S. collected behavioral data, sound recordings and designed the experimental study with inputs from S.K. and I.R.H.; J.S. collected the multi-species dataset of birdsong recordings from xeno-canto; Breeding and ringing data were collected by I.R.H. and J.S.; J.S. developed the code for the acoustic and the statistical analysis; J.S. wrote the article with inputs from S.K. and I.R.H.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

Received: February 7, 2023

Revised: June 11, 2023

Accepted: October 11, 2023

Published: October 14, 2023

**REFERENCES**

1. Mowles, S.L., and Ord, T.J. (2012). Repetitive signals and mate choice: insights from contest theory. *Anim. Behav.* *84*, 295–304.
2. Lane, S.M., and Briffa, M. (2022). Skilful mating? Insights from animal contest research. *Anim. Behav.* *184*, 197–207.
3. Byers, J., Hebets, E., and Podos, J. (2010). Female mate choice based upon male motor performance. *Anim. Behav.* *79*, 771–778.
4. Barske, J., Schlinger, B.A., Wikelski, M., and Fusani, L. (2011). Female choice for male motor skills. *Proc. Biol. Sci.* *278*, 3523–3528. <https://doi.org/10.1098/rspb.2011.0382>.
5. Irschick, D.J., Herrel, A., Vanhooydonck, B., and Damme, R.V. (2007). A functional approach to sexual selection. *Funct. Ecol.* *21*, 621–626.
6. Botero, C.A., and de Kort, S.R. (2013). In Learned signals and consistency of delivery: a case against receiver manipulation in animal communication. *Animal communication theory: information and influence*, U.E. Stegmann, ed. (Cambridge University Press), pp. 281–296.
7. Suthers, R.A. (2004). In *How birds sing and why it matters. Nature's music: the science of birdsong*, R.P. Marler and H. Slabbekoorn, eds. (Elsevier Academic Press), pp. 272–295.
8. Suthers, R.A., and Goller, F. (1997). Motor correlates of vocal diversity in songbirds. In *Current ornithology* (Springer), pp. 235–288.
9. Zollinger, S.A., Riede, T., and Suthers, R.A. (2008). Two-voice complexity from a single side of the syrinx in northern mockingbird *Mimus polyglottos* vocalizations. *J. Exp. Biol.* *211*, 1978–1991. <https://doi.org/10.1242/jeb.014092>.
10. Riede, T., Suthers, R.A., Fletcher, N.H., and Blevins, W.E. (2006). Songbirds tune their vocal tract to the fundamental frequency of their song. *Proc. Natl. Acad. Sci. USA* *103*, 5543–5548.
11. Fusani, L., Barske, J., Day, L.D., Fuxjager, M.J., and Schlinger, B.A. (2014). Physiological control of elaborate male courtship: female choice for neuromuscular systems. *Neurosci. Biobehav. Rev.* *46*, 534–546. <https://doi.org/10.1016/j.neubiorev.2014.07.017>.
12. Cardoso, G.C. (2017). Advancing the inference of performance in birdsong. *Anim. Behav.* *125*, e29–e32.
13. Byers, B.E., and Kroodtsma, D.E. (2009). Female mate choice and songbird song repertoires. *Anim. Behav.* *77*, 13–22.

14. Dubbeldam, J. (1997). Intratelencephalic sensorimotor circuits in birds—What have feeding and vocalization in common? *Neth. J. Zool.* **48**, 199–212.
15. Kao, M.H., Doupe, A.J., and Brainard, M.S. (2005). Contributions of an avian basal ganglia–forebrain circuit to real-time modulation of song. *Nature* **433**, 638–643. <https://doi.org/10.1038/nature03127>.
16. Cardoso, G.C., Atwell, J.W., Ketterson, E.D., and Price, T.D. (2009). Song types, song performance, and the use of repertoires in dark-eyed juncos (*Junco hyemalis*). *Behav. Ecol.* **20**, 901–907.
17. Christie, P.J., Mennill, D.J., and Ratcliffe, L.M. (2004). Pitch shifts and song structure indicate male quality in the dawn chorus of black-capped chickadees. *Behav. Ecol. Sociobiol.* **55**, 341–348.
18. Botero, C.A., Rossman, R.J., Caro, L.M., Stenzler, L.M., Lovette, I.J., de Kort, S.R., and Vehrencamp, S.L. (2009). Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. *Anim. Behav.* **77**, 701–706.
19. Byers, B.E. (2007). Extra-pair paternity in chestnut-sided warblers is correlated with consistent vocal performance. *Behav. Ecol.* **18**, 130–136.
20. Kempnaers, B., Verheyen, G.R., and Dhondt, A.A. (1997). Extra-pair paternity in the blue tit (*Parus caeruleus*): female choice, male characteristics, and offspring quality. *Behav. Ecol.* **8**, 481–492.
21. Poesel, A., Foerster, K., and Kempnaers, B. (2001). The dawn song of the blue tit (*Parus caeruleus*) and its role in sexual selection. *Ethology* **107**, 521–531.
22. de Kort, S.R., Eldermire, E.R.B., Valderrama, S., Botero, C.A., and Vehrencamp, S.L. (2009). Trill consistency is an age-related assessment signal in banded wrens. *Proc. Biol. Sci.* **276**, 2315–2321.
23. Ballentine, B., Hyman, J., and Nowicki, S. (2004). Vocal performance influences female response to male bird song: an experimental test. *Behav. Ecol.* **15**, 163–168.
24. Wasserman, F.E., and Cigliano, J.A. (1991). Song output and stimulation of the female in white-throated sparrows. *Behav. Ecol. Sociobiol.* **29**, 55–59.
25. Siero, J., de Kort, S.R., and Hartley, I.R. (2023). Sexual selection for both diversity and repetition in birdsong. *Nat. Commun.* **14**, 3600. <https://doi.org/10.1038/s41467-023-39308-5>.
26. Nice, M.M. (1943). Studies in the life history of the song sparrow II. The behavior of the song sparrow and other passerines. *Trans. Linn. Soc. N. Y.* **6**, 1–328.
27. Thorpe, W., and Pilcher, P. (1958). The nature and characteristics of subsong. *Br. Birds* **51**, 509–514.
28. Brumm, H., and Hultsch, H. (2001). Pattern amplitude is related to pattern imitation during the song development of nightingales. *Anim. Behav.* **61**, 747–754. <https://doi.org/10.1006/anbe.2000.1664>.
29. Forstmeier, W., Kempnaers, B., Meyer, A., and Leisler, B. (2002). A novel song parameter correlates with extra-pair paternity and reflects male longevity. *Proc. Biol. Sci.* **269**, 1479–1485. <https://doi.org/10.1098/rspb.2002.2039>.
30. Podos, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: *Emberizidae*). *Evolution* **51**, 537–551. <https://doi.org/10.1111/j.1558-5646.1997.tb02441.x>.
31. Drăgănoiu, T.I., Nagle, L., and Kreutzer, M. (2002). Directional female preference for an exaggerated male trait in canary (*Serinus canaria*) song. *Proc. Biol. Sci.* **269**, 2525–2531. <https://doi.org/10.1098/rspb.2002.2192>.
32. Geberzahn, N., and Aubin, T. (2014). How a songbird with a continuous singing style modulates its song when territorially challenged. *Behav. Ecol. Sociobiol.* **68**, 1–12.
33. Phillips, J.N., and Derryberry, E.P. (2017). Vocal performance is a salient signal for male-male competition in white-crowned Sparrows. *Auk* **134**, 564–574. <https://doi.org/10.1642/auk-17-2.1>.
34. Linhart, P., Jaška, P., Petrusková, T., Petrušek, A., and Fuchs, R. (2013). Being angry, singing fast? Signalling of aggressive motivation by syllable rate in a songbird with slow song. *Behav. Processes* **100**, 139–145.
35. Price, J.J. (2013). Why is birdsong so repetitive? Signal detection and the evolution of avian singing modes. *Beyond Behav.* **150**, 995–1013.
36. Zahavi, A. (1980). Ritualization and the evolution of movement signals. *Beyond Behav.* **72**, 77–80.
37. Sakata, J.T., and Vehrencamp, S.L. (2012). Integrating perspectives on vocal performance and consistency. *J. Exp. Biol.* **215**, 201–209. <https://doi.org/10.1242/jeb.056911>.
38. Rivera-Gutierrez, H.F., Pinxten, R., and Eens, M. (2011). Songs differing in consistency elicit differential aggressive response in territorial birds. *Biol. Lett.* **7**, 339–342. <https://doi.org/10.1098/rsbl.2010.0962>.
39. Ferreira, A.C., Atwell, J.W., Whittaker, D.J., Ketterson, E.D., and Cardoso, G.C. (2016). Communication value of mistakes in dark-eyed junco song. *Am. Nat.* **188**, 289–305. <https://doi.org/10.1086/687520>.
40. Siero, J., de Kort, S.R., Riebel, K., and Hartley, I.R. (2022). Female blue tits sing frequently: a sex comparison of occurrence, context, and structure of song. *Behav. Ecol.* **33**, 912–925. <https://doi.org/10.1093/beheco/arc044>.
41. Smith, G.T., Brenowitz, E.A., Beecher, M.D., and Wingfield, J.C. (1997). Seasonal changes in testosterone, neural attributes of song control nuclei, and song structure in wild songbirds. *J. Neurosci.* **17**, 6001–6010. <https://doi.org/10.1523/JNEUROSCI.17-15-06001.1997>.
42. Podos, J., and Nowicki, S. (2004). In Performance limits on birdsong. *Nature's music: the science of birdsong*, R.P. Marler and H. Slabbekoorn, eds. (Elsevier Academic Press), pp. 318–342.
43. Geberzahn, N., and Aubin, T. (2014). Assessing vocal performance in complex birdsong: a novel approach. *BMC Biol.* **12**, 58. <https://doi.org/10.1186/s12915-014-0058-4>.
44. Vehrencamp, S.L., Yantachka, J., Hall, M.L., and de Kort, S.R. (2013). Trill performance components vary with age, season, and motivation in the banded wren. *Behav. Ecol. Sociobiol.* **67**, 409–419. <https://doi.org/10.1007/s00265-012-1461-x>.
45. Linhart, P., Slabbekoorn, H., and Fuchs, R. (2012). The communicative significance of song frequency and song length in territorial chiffchaffs. *Behav. Ecol.* **23**, 1338–1347.
46. Lattin, C., and Ritchison, G. (2009). Intra- and intersexual functions of singing by male blue grosbeaks: the role of within-song variation. *Wilson J. Ornithol.* **121**, 714–721.
47. Keating, H.R., and Reichard, D.G. (2021). Seasonal song variation in male Carolina Wrens (*Thryothorus ludovicianus*). *Wilson J. Ornithol.* **133**, 365–371.
48. Galeotti, P., Saino, N., Sacchi, R., and Møller, A.P. (1997). Song correlates with social context, testosterone and body condition in male barn swallows. *Anim. Behav.* **53**, 687–700. <https://doi.org/10.1006/anbe.1996.0304>.
49. Lambrechts, M., and Dhondt, A.A. (1986). Male quality, reproduction, and survival in the great tit (*Parus major*). *Behav. Ecol. Sociobiol.* **19**, 57–63.
50. Bijmens, L. (1988). Blue tit *Parus caeruleus* song in relation to survival, reproduction and biometry. *Hous. Theor. Soc.* **35**, 61–67.
51. Mennill, D.J., Badyaev, A.V., Jonart, L.M., and Hill, G.E. (2006). Male house finches with elaborate songs have higher reproductive performance. *Ethology* **112**, 174–180. <https://doi.org/10.1111/j.1439-0310.2006.01145.x>.
52. Nolan, P.M., and Hill, G.E. (2004). Female choice for song characteristics in the house finch. *Anim. Behav.* **67**, 403–410.
53. Welch, A.M., Semlitsch, R.D., and Gerhardt, H.C. (1998). Call duration as an indicator of genetic quality in male gray tree frogs. *Science* **280**, 1928–1930.
54. Shaw, K.L., and Herlihy, D.P. (2000). Acoustic preference functions and song variability in the Hawaiian cricket (*Laupala cerasina*). *Proc. Biol. Sci.* **267**, 577–584. <https://doi.org/10.1098/rspb.2000.1040>.
55. Nelson, D.A., and Poesel, A. (2012). Responses to variation in song length by male white-crowned sparrows. *Ethology* **118**, 24–32.
56. Siero, J., Siero, J., and Slabbekoorn, H. (2020). Experimental test of the communicative value of syllable diversity and syllable switching in the common chiffchaff. *Anim. Behav.* **165**, 11–21.
57. Ward, S., Speakman, J.R., and Slater, P.J. (2003). The energy cost of song in the canary, *Serinus canaria*. *Anim. Behav.* **66**, 893–902.
58. Oberweger, K., and Goller, F. (2001). The metabolic cost of birdsong production. *J. Exp. Biol.* **204**, 3379–3388. <https://doi.org/10.1242/jeb.204.19.3379>.
59. Slabbekoorn, H. (2004). In Singing in the wild: The ecology of birdsong. *Nature's music: the science of birdsong*, R.P. Marler and H. Slabbekoorn, eds. (Elsevier Academic Press), pp. 39–79.
60. Zollinger, S.A., and Suthers, R.A. (2004). Motor mechanisms of a vocal mimic: implications for birdsong production. *Proc. Biol. Sci.* **271**, 483–491. <https://doi.org/10.1098/rspb.2003.2598>.
61. Hartley, R.S., and Suthers, R.A. (1989). Airflow and pressure during canary song: direct evidence for mini-breaths. *J. Comp. Physiol.* **165**, 15–26.
62. Podos, J., and Cohn-Haft, M. (2019). Extremely loud mating songs at close range in white bellbirds. *Curr. Biol.* **29**, R1068–R1069.
63. Marra, G., and Wood, S.N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scand. J. Stat.* **39**, 53–74.

64. Tanner, J.C., and Bee, M.A. (2020). Inconsistent sexual signaling degrades optimal mating decisions in animals. *Sci. Adv.* 6, eaax3957. <https://doi.org/10.1126/sciadv.aax3957>.
65. Goller, F., and Larsen, O.N. (1997). A new mechanism of sound generation in songbirds. *Proc. Natl. Acad. Sci. USA* 94, 14787–14791. <https://doi.org/10.1073/pnas.94.26.14787>.
66. Jensen, K.K., Cooper, B.G., Larsen, O.N., and Goller, F. (2007). Songbirds use pulse tone register in two voices to generate low-frequency sound. *Proc. Biol. Sci.* 274, 2703–2710. <https://doi.org/10.1098/rspb.2007.0781>.
67. Fee, M.S., Shraiman, B., Pesaran, B., and Mitra, P.P. (1998). The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature* 395, 67–71. <https://doi.org/10.1038/25725>.
68. Suthers, R.A., and Zollinger, S.A. (2008). From brain to song: The vocal organ and vocal tract. In *Neuroscience of Birdsong*, H.P. Zeigler and P. Marler, eds. (Cambridge University Press), pp. 78–98.
69. Méndez, J.M., Duker, J., and Cooper, B.G. (2022). Preparing to sing: respiratory patterns underlying motor readiness for song. *J. Neurophysiol.* 128, 1646–1662. <https://doi.org/10.1152/jn.00551.2021>.
70. Bigland-Ritchie, B. (1981). EMG and fatigue of human voluntary and stimulated contractions. *Ciba Found. Symp.* 82, 130–156. <https://doi.org/10.1002/9780470715420.ch9>.
71. Lännergren, J., and Westerblad, H. (1989). Maximum tension and force velocity properties of fatigued, single Xenopus muscle fibres studied by caffeine and high K+. *J. Physiol.* 409, 473–490. <https://doi.org/10.1113/jphysiol.1989.sp017508>.
72. Westerblad, H., and Lännergren, J. (1991). Slowing of relaxation during fatigue in single mouse muscle fibres. *J. Physiol.* 434, 323–336. <https://doi.org/10.1113/jphysiol.1991.sp018472>.
73. Gandevia, S.C. (2001). Spinal and supraspinal factors in human muscle fatigue. *Physiol. Rev.* 81, 1725–1789. <https://doi.org/10.1152/physrev.2001.81.4.1725>.
74. Dinh, J.P., Peters, S., and Nowicki, S. (2020). Song performance improves with continued singing across the morning in a songbird. *Anim. Behav.* 167, 127–137.
75. Schraft, H.A., Medina, O.J., McClure, J., Pereira, D.A., and Logue, D.M. (2017). Within-day improvement in a behavioural display: wild birds ‘warm up’. *Anim. Behav.* 124, 167–174.
76. Adam, I., Riebel, K., Stal, P., Wood, N.B., Previs, M.J., and Elemans, C.P. (2023). Peak performance singing requires daily vocal exercise in songbirds. Preprint at bioRxiv. <https://doi.org/10.1101/2023.02.23.529633>.
77. Bee, M., and Gerhardt, H.C. (2007). Recognition and localization of acoustic signals. In *Hearing and sound communication in amphibians*, P.M. Narins, A.S. Feng, R.R. Fay, and A.N. Popper, eds. (Springer), pp. 113–146.
78. Holveck, M.-J., and Riebel, K. (2007). Preferred songs predict preferred males: consistency and repeatability of zebra finch females across three test contexts. *Anim. Behav.* 74, 297–309.
79. Lane, S.M., and Briffa, M. (2020). Perceived and actual fighting ability: determinants of success by decision, knockout or submission in human combat sports. *Biol. Lett.* 16, 20200443. <https://doi.org/10.1098/rsbl.2020.0443>.
80. Neave, N., McCarty, K., Freynik, J., Caplan, N., Hönekopp, J., and Fink, B. (2011). Male dance moves that catch a woman’s eye. *Biol. Lett.* 7, 221–224.
81. McCarty, K., Darwin, H., Cornelissen, P.L., Saxton, T.K., Tovée, M.J., Caplan, N., and Neave, N. (2017). Optimal asymmetry and other motion parameters that characterise high-quality female dance. *Sci. Rep.* 7, 42435–42439.
82. Hartmann, M.T., Giasson, L.O.M., Hartmann, P.A., and Haddad, C.F.B. (2005). Visual communication in Brazilian species of anurans from the Atlantic forest. *J. Nat. Hist.* 39, 1675–1685.
83. Lane, S.M., and Briffa, M. (2020). The role of spatial accuracy and precision in hermit crab contests. *Anim. Behav.* 167, 111–118. <https://doi.org/10.1016/j.anbehav.2020.07.013>.
84. Demartsev, V., Haddas-Sasson, M., Ilany, A., Koren, L., and Geffen, E. (2023). Male rock hyraxes that maintain an isochronous song rhythm achieve higher reproductive success. *J. Anim. Ecol.* 92, 1520–1531.
85. R Core Team (2022). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
86. Siero, J., de Kort, S.R., and Hartley, I.R. (2023). Sound properties affect measurement of vocal consistency in birdsong: Validation of the spectrogram cross correlation method (SPCC). *J. Acoust. Soc. Am.* 154, 699–708. <https://doi.org/10.1121/10.0020543>.
87. Ligges, U. (2013). *tuneR—Analysis of Music*.
88. Sueur, J., Aubin, T., and Simonis-Sueur, C. (2006). *Seewave*.
89. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*.
90. Wickham, H., and Wickham, M.H. (2019). Package “stringr”. R Package Version 1.4.0.
91. Wickham, H., François, R., Henry, L., and Müller, K. (2022). A grammar of data manipulation. R Package Version 1.0.10.
92. Wood, S., and Wood, M.S. (2015). Package ‘mgcv’. R Package Version 1, 729.
93. McGregor, P.K. (1988). Song length and ‘male quality’ in the chaffinch. *Anim. Behav.* 36, 606–608.
94. Araya-Salas, M., and Smith-Vidaurre, G. (2017). warbleR: An R package to streamline analysis of animal acoustic signals. *Methods Ecol. Evol.* 8, 184–191.
95. Mainwaring, M.C., and Hartley, I.R. (2009). Experimental evidence for state-dependent nest weight in the blue tit, *Cyanistes caeruleus*. *Behav. Processes* 81, 144–146. <https://doi.org/10.1016/j.beproc.2009.02.001>.
96. Redfern, C.P.F., and Clark, J.A. (2001). *Ringers’ Manual* (BTO).
97. Svensson, L. (1992). *Identification Guide to European Passerines* (British Trust for Ornithology).
98. Bijmens, L., and Dhondt, A. (1984). Vocalizations in a Belgian blue tit, *Parus c. caeruleus*, population. *Gerfaut* 74, 243–269.
99. Latimer, W. (1977). A comparative study of the songs and alarm calls of some *Parus* species. *Z. Tierpsychol.* 45, 414–433.
100. Poesel, A., Pedersen, S.B., and Dabelsteen, T. (2004). Dawn song of male blue tits as a predictor of competitiveness in midmorning singing interactions. *Acta Ethol.* 6, 65–71.
101. Cardoso, G.C., and Atwell, J.W. (2011). On the relation between loudness and the increased song frequency of urban birds. *Anim. Behav.* 82, 831–836.
102. Mazzoni, D., and Dannenberg, R. (2014). *Audacity*.
103. R Development Core Team (2016). R: A Language and Environment for Statistical Computing.
104. Clark, C.W., Marler, P., and Beeman, K. (2010). Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. *Ethology* 76, 101–115.
105. Briffa, M., and Lane, S.M. (2017). The role of skill in animal contests: a neglected component of fighting ability. *Proc. Biol. Sci.* 284, 20171596. <https://doi.org/10.1098/rspb.2017.1596>.
106. Coleman, S.W., Patricelli, G.L., Coyle, B., Siani, J., and Borgia, G. (2007). Female preferences drive the evolution of mimetic accuracy in male sexual displays. *Biol. Lett.* 3, 463–466. <https://doi.org/10.1098/rsbl.2007.0234>.
107. Cortopassi, K.A., and Bradbury, J.W. (2000). The comparison of harmonically rich sounds using spectrographic cross-correlation and principal coordinates analysis. *Bioacoustics* 11, 89–127.
108. Gerhardt, H.C. (1992). Conducting playback experiments and interpreting their results. In *Playback and studies of animal communication*, P.K. McGregor, ed. (Springer), pp. 59–77.
109. Fishbein, A.R., Löschner, J., Mallon, J.M., and Wilkinson, G.S. (2018). Dynamic sex-specific responses to synthetic songs in a duetting subsong passerine. *PLoS One* 13, e0202353. <https://doi.org/10.1371/journal.pone.0202353>.
110. Bremond, J.-C. (1976). Specific recognition in the song of Bonelli’s warbler (*Phylloscopus bonelli*). *Beyond Behav.* 58, 99–116.
111. Nowicki, S., Mitani, J.C., Nelson, D.A., and Marler, P. (1989). The communicative significance of tonality in birdsong: responses to songs produced in helium. *Bioacoustics* 2, 35–46.
112. Nowicki, S. (1987). Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere. *Nature* 325, 53–55. <https://doi.org/10.1038/325053a0>.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE                | SOURCE                      | IDENTIFIER                                                                                                                      |
|------------------------------------|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| <b>Deposited data</b>              |                             |                                                                                                                                 |
| List of tracks analysed            | This study                  | <a href="https://doi.org/10.6084/m9.figshare.24173367.v1">https://doi.org/10.6084/m9.figshare.24173367.v1</a>                   |
| Data set to fit statistical models | This study                  | <a href="https://doi.org/10.6084/m9.figshare.24173367.v1">https://doi.org/10.6084/m9.figshare.24173367.v1</a>                   |
| <b>Software and algorithms</b>     |                             |                                                                                                                                 |
| Birdsong repository                |                             | <a href="http://www.xeno-canto.org">www.xeno-canto.org</a>                                                                      |
| R software                         | R Core Team <sup>85</sup>   | <a href="https://cran.r-project.org/">https://cran.r-project.org/</a>                                                           |
| SPCC function                      | Sierro et al. <sup>86</sup> | <a href="https://github.com/javisierro/SPCC-function">https://github.com/javisierro/SPCC-function</a>                           |
| tuneR                              | Ligges <sup>87</sup>        | <a href="https://cran.r-project.org/web/packages/tuneR/index.html">https://cran.r-project.org/web/packages/tuneR/index.html</a> |
| seewave                            | Sueur et al. <sup>88</sup>  | <a href="https://rug.mnhn.fr/seewave/">https://rug.mnhn.fr/seewave/</a>                                                         |
| ggplot2                            | Wickham, <sup>89</sup>      | <a href="https://ggplot2.tidyverse.org/">https://ggplot2.tidyverse.org/</a>                                                     |
| stringr                            | Wickham, <sup>90</sup>      | <a href="https://stringr.tidyverse.org/">https://stringr.tidyverse.org/</a>                                                     |
| dplyr                              | Wickham, <sup>91</sup>      | <a href="https://dplyr.tidyverse.org/">https://dplyr.tidyverse.org/</a>                                                         |
| mgcv                               | Wood et al. <sup>92</sup>   | <a href="https://cran.r-project.org/web/packages/mgcv/index.html">https://cran.r-project.org/web/packages/mgcv/index.html</a>   |

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources should be directed to the lead contact, Dr. Javier Sierro ([sierro.2.8@gmail.com](mailto:sierro.2.8@gmail.com)).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

- Audio recordings from the multi-species analysis are taken from a public repository of bird sound recordings ([xeno-canto.org](http://xeno-canto.org)) and a list of the exact recordings analysed is deposited in a public data repository. The DOI is listed in the [key resources table](#).
- All original code has been deposited at Figshare and is publicly available as of the date of publication. DOIs are listed in the [key resources table](#).
- Any additional information required to re-analyse the data reported in this paper is available from the [lead contact](#) upon request.

### EXPERIMENTAL MODEL AND PARTICIPANT DETAILS

#### Multi-species analysis

##### Species selection

We selected 26 families of songbirds (i.e. Acrocephalidae, Aegithalidae, Cettiidae, Cinclidae, Emberizidae, Estrildidae, Fringillidae, Icteridae, Locustellidae, Maluridae, Mimidae, Motacillidae, Muscicapidae, Paridae, Parulidae, Passerellidae, Passeridae, Petroicidae, Phylloscopidae, Prunellidae, Regulidae, Remizidae, Sittidae, Troglodytidae, Turdidae and Vireonidae). We defined a trill as the consecutive repetition of the same note type at least three times and measured trill length as the number of notes per trill. This is a common practice in the birdsong research<sup>20,45,50–52,93</sup> and it is meaningful when comparing trills of the same individual and song type, because notes are highly stereotyped within an individual's repertoire. The *warbler* package in R<sup>94</sup> was used to download a maximum of 30 recordings per species from the Xeno-Canto repository ([www.xeno-canto.org](http://www.xeno-canto.org)), using the following criteria (recording category: 'song', recording quality = 'A'). The trills selected could be isolated and separated from other parts of the song by long silent gaps, for example in the ciril bunting (*Emberiza cirilus*, <https://xeno-canto.org/763783>) but could also be embedded within a continuous stream of song, as in the European goldfinch (*Carduelis carduelis*, <https://xeno-canto.org/787235>). Some species repeat structures made of two or three notes, as is the case in the great tit (*Parus major*, <https://xeno-canto.org/788731>). In these cases, only one note from the repeated structure was analysed in order to apply a comparable method for

analysing trills across species. Note that many bird species produce trills as well as other types of song structures that were not analysed in this study. Examples of trills that were analysed are shown in [Figure 2](#) and examples of songs that were not included in the analysis are shown in [Figure S1](#). Some songs with apparent trills were excluded from our analysis because the notes being repeated were similar but not of the same type, with varying note structure within the trill ([Figure S1](#)).

## Playback experiment in blue tits

### *Blue tit population and song recordings*

We conducted a field study on a population of wild blue tits breeding in nest boxes at Lancaster University campus, UK, as part of a long-term monitoring study.<sup>95</sup> Every year, breeding birds were caught and ringed with a unique combination of three coloured rings and one numbered metal ring for identification in the field.<sup>96</sup> During the breeding season, breeding pairs were sexed in the hand, based on the presence of a brood patch, present only in females, or cloacal protuberance, present only in males.<sup>97</sup> Blue tit song is usually composed of a few introductory notes followed by a trill ([Figure 4](#)).<sup>98,99</sup> Each individual has a repertoire of several stereotyped song types that have the same function<sup>100</sup> and these are usually shared between individuals within a population. From January to May in 2018 to 2020 (egg laying starts in April), song recordings were made daily using a Marantz PMD661 recorder (48kHz sampling rate and 24-bit depth) and a Sennheiser ME67 microphone.

To conduct our playback experiment, we tested the territorial response of 31 individual male blue tits from a wild population of birds at Lancaster University campus. All fieldwork involving the blue tits was approved by the Lancaster University animal welfare and ethical review board.

## METHOD DETAILS

### Multi-species analysis

#### *Numbering notes within trills*

To measure variation in vocal performance within song, we divided the trills into discrete units (notes) and designed a note numbering system to compare note positions in trills of different absolute length. To examine changes within the trills, we used two different note numbering systems, one relative to the beginning of the trill, and another relative to the end of a bird's shortest trill. Trills commonly show an increase in performance at the beginning, sometimes referred to as a fade-in effect.<sup>101</sup> Comparing the initial change in song performance in trills of distantly related species was done by aligning all trills from the start and numbering notes from the first repetition. However, investigating a decline after sustained singing by an individual is more difficult due to the great variability in absolute trill length between song types, individuals and species. Trills that are long for some species may be relatively short for others and, within species, trill length may vary considerably between individuals, across seasons<sup>40,47</sup> and with context.<sup>16</sup> Hence, we assessed the change in performance at the end of trills relative to the shortest trill length found for each individual ([Figure 2](#)). First, we randomly selected two or more trills of the same song type in one recording, that usually varied in length. The shortest trill was used as a reference and the position of its last note was assigned note position 'zero' ([Figure 2](#)). Preceding notes were given negative values in sequence from zero. Notes in all other trills were numbered in relation to the shortest trill, so that additional notes at the end of longer trills were given positive numbers ([Figure 2](#)). This allowed us to investigate changes in performance at the point where individual birds extend their trills relative to their individual minimum trill length for each song type. We included at least two trills of the same song type for each individual.

#### *Acoustic analysis*

We located trills manually in the recordings with the cursor in Audacity using both the spectrogram and the waveform<sup>102</sup> (window type: 'Hanning', window length 1024 samples, 90% overlap and -80 dB range). Time marks were loosely placed around the entire trill and exported as a text file that served to mark the location of trills for the acoustic analysis that was conducted in R software.<sup>87,88,103</sup> We ran a custom-made script in R to split automatically each trill into segments containing individual notes. Using an amplitude threshold, changes in the amplitude slope and minimum distance threshold between notes, the script finds the splitting points approximately in the middle of the silent gaps between notes. These splitting points divided the trill into segments that enclosed individual notes ([Figure 1](#)) but were not used to measure note intervals (see below). Each trill was checked visually by JS to delete miss-detections and add splitting points that were not identified automatically.

We assessed three measures of song performance: note rate, relative sound amplitude and vocal consistency. Note rate reflects the speed in muscle contraction rates. The relative sound amplitude reflects the changes in expiratory pressure within a trill, possibly associated with the volume of air stored, muscular characteristics of expiratory muscles or singing skills. Both were selected as indicators of muscle performance during phonation in birds.<sup>76</sup> We also measured vocal consistency as a measure of precision in the execution of motor patterns.<sup>6,22,104,105</sup> We quantified each variable for each repeated note of the trill.

Note rate at any point in the trill was the inverse of the interval length between the note and the preceding note. To measure the separation, in seconds, between two notes, we located the middle point of each note within the portion of each segment that was above a 50% relative amplitude, normalizing peak amplitude of each segment to a value of 1 (shown in green in the last two notes of the trill in [Figure 1](#)).

The amplitude of the sound was recorded in volts by the audio recorder and then translated by an analogue-to-digital converter into bits, depending on the bit depth of the recording. To measure changes in note amplitude along a trill, we first normalized the amplitude of the entire trill, where the highest amplitude of the trill had a value of 1 (*env* function in 'seewave',<sup>88</sup> 128 window length and 90% of overlap). We

calculated the mean amplitude of each note using only the portion of the note that was above 20% of the trill's peak amplitude within each segment (see the first note and blue designations in Figure 1).

Measurement of vocal consistency assesses the precision in execution of singing motor patterns. We compared each note within a trill to all other repetitions of the note, assuming that the note that is produced with highest precision will be, on average, more similar to all other notes. A clarifying analogy would be the task of assessing the precision of multiple shots on a white wall with the observer, but not the shooter, unaware of the exact target. In this case, it is difficult to assess which shots are closest to the target (higher precision) but we can have an approximate idea if we assume that the target is close to the centre marked by all the shots that were aimed at the same target. Thus, to measure vocal consistency, we calculated the acoustic similarity of each note, using the entire segment between splitting points, to all other notes in the trill, using a spectrogram cross-correlation algorithm (SPCC),<sup>18,86,104,106,107</sup> with temporal offset of 20 ms and a temporal resolution of 1 ms following previous authors.<sup>18,104,106</sup> 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. This was carried out using a custom-made function in R software (Github repository: <https://github.com/javisierro/SPCC-function>). Each note has a single, unique value of consistency that ranges between 0 and 1, calculated as the mean acoustic similarity of that note to all other notes in the trill (1 means that notes are exactly identical). A trill with overall high vocal consistency will be made up of notes whose individual consistency values are all high.

## Playback experiment in blue tits

### Experimental design

We carried out a song playback experiment in the field with three treatments to investigate: 1) the communicative value of trill length and 2) the impact of a performance decline in the perception of trill length, in this case a decline in consistency. In our experiment, the three treatments were: a "short-stable" treatment, which was a song with a trill of six notes of high, stable vocal consistency; a "long-stable" treatment, a song with a trill of twelve notes of high, stable vocal consistency and a "long-decline" treatment, a song with a 12-note trill where the first six notes are of high, stable vocal consistency and the last six notes are of lowered vocal consistency (Figure 4). The songs in all three treatments were identical in the first six notes of the trill. We designed our experiment to compare 'short-stable' versus 'long-stable' and 'short-stable' versus 'long-decline'. This would inform us about 1) the impact of trill length in communication, for songs of similar performance, and 2) if a decline in performance affects the perception of trill length. The third contrast (long-stable vs. long-decline) was not included as it is similar to previous experiments conducted in the same population, which showed that birds show a stronger response to high consistency than low consistency songs (S.R.K., unpublished data).

The playback stimuli were synthesized using natural blue tit song as a template. The use of synthetic stimuli allowed us to accurately manipulate several acoustic parameters and avoid confounding factors related to natural variation in sound signals and recording conditions.<sup>108</sup> Other studies have also used synthetic song to simulate territorial interactions in blue tits<sup>100</sup> and other species.<sup>109,110</sup> Studies conducted in this species consider song types to be functionally the same,<sup>21,98,100</sup> hence we chose a song type in our population that presented long trills, suitable to test our hypothesis<sup>111,112</sup> (Figure 4).

The songs from which stimuli were derived had three distinct parts: the introduction, the middle and the trill (Figure 4). To create the synthetic stimuli, we first analysed 708 songs of this song type from 35 individuals ( $20.2 \pm 14.9$  trills per individual) that had a trill of  $9.1 \pm 4.0$  notes in length. We analysed these songs by manually marking the start and end times of individual notes using the cursor in Audacity, with the multi-view mode that shows both the spectrogram and the waveform of the audio file. From this we derived the length of each note and the length of silent intervals between notes within the trill. These time marks served to locate and cut each individual note in R software as a normalized WAV file to conduct a detailed acoustic analysis. For each note, we computed the normalized power spectrum (window size: 1024 samples, max. dB = 0) and measured peak frequency, as the frequency with the highest amplitude, and sound bandwidth, as the distance between the maximum and the minimum frequency of the note. The maximum and minimum frequencies were the highest and lowest frequency above a -15 dB amplitude threshold in the power spectrum, following Podos (1997).<sup>30</sup> Finally, we measured the percentage difference in note length, peak frequency and note bandwidth of each note, relative to the mean note length, mean peak frequency and mean bandwidth of the trill. The 75%-95% quantile of this variation was 8.3 – 17.5% for note length, 1.7 – 3.2% for peak frequency and 25.3 – 60.5% for note bandwidth. These were used as the range of inconsistencies introduced to recreate low-consistency notes in the songs for the long-decline playback treatment.

For the synthesis of artificial song, we selected 31 high-quality recordings from 31 different males (one for each trial) to use as template for the stimuli. For each note, we tracked the fundamental frequency of one introduction, one middle and one trill note to recreate the same shape synthetically. While preserving the spectrographic shape of each note, we synthesized each note to match the note's peak frequency with the mean peak frequency of the population for each note type. Furthermore, the duration of synthetic notes and note rate matched the mean of the population. In this way, all playback stimuli recreated the natural song of 31 different individual males, avoiding pseudo-replication, but they were all adjusted to the same note duration, note peak frequency and note rate, to remove these potential confounding factors from the experiment.<sup>12,108</sup> All songs, regardless of treatment, had two introductory and two middle notes before the trill. All stimuli presented within one trial were paired based on the same original recording.

A complete stimulus presentation of 60 seconds consisted of ten songs of one of the treatments, delivered at a song rate of one song every six seconds. In a stimulus from a short-stable or a long-stable treatment, all 10 songs were equal within stimuli, whereas in a stimulus from the long-decline treatment, all ten songs were individually distinct since the variation introduced aimed to represent imprecisions and therefore the inconsistencies of the last six notes within a trill were fully randomized in each song.

### Song playback trials

The experiment was conducted between February and April 2020. Each trial consisted of two separate tests, carried out on two different days, leaving one week in between tests for each subject. Each test consisted of the presentation of two stimuli, either a short-stable versus a long-stable stimulus, hereafter the “trill length” test and a short-stable versus long-decline stimulus, hereafter the “consistency decline” test. The short-stable treatment was present in both tests and only the other treatment changed, choosing either the long-stable or the long-decline treatment. Both the order of stimulus presentation within a test, as well as the test order within a trial, were alternated to control for a possible order effect. Playback stimuli were broadcast 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 as WAV files in the internal memory with an alias filename hiding the treatment, making the field observer blind to the treatment.

Each test began by placing the speaker within 10 m of the nest box occupied by a focal bird. A perimeter of 4 m radius 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, and therefore we waited for several minutes until such mobbing behaviour stopped. The test began by locating the breeding male, identifiable by its coloured leg rings, and recording the vocal behaviour of the subject using a Marantz PMD661 recorder (48kHz sampling rate and 24-bit depth) and a Sennheiser ME67 microphone. Simultaneously, the approach behaviour was registered by dictating onto a separate voice recorder each time the bird entered and left the 4 m perimeter around the speaker. We recorded one minute of natural baseline behaviour (before playback), followed by one minute of the first treatment stimulus (during playback), and continued recording for another minute after the playback ended (after playback). Once the “after” phase ended, there was a two-minute interval before presentation of the complementary stimulus within the test, using the same procedure.

All recording files were named using an alias that concealed the treatments used during each test. Using Audacity software,<sup>102</sup> we linked the audio recordings of the birds’ vocalizations with the voice notes of their behaviour. Then, we marked the start and end times of all vocalizations as well as when the birds went in and out of the 4 m radius of the speaker using the labelling tool. After exporting the time marks as .txt files we measured the total time of vocalizations and the total time spent within 4 m of the speaker in R software.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Multi-species analysis

#### Analysis of performance change within trills

All measures are presented as mean  $\pm$  one SD, unless otherwise indicated. Statistical analyses were carried out in R software.<sup>103</sup> Packages used in the statistical analysis included *mgcv*<sup>92</sup> and for data management and visualization we used *stringr*,<sup>90</sup> *dplyr*<sup>91</sup> and *ggplot2*.<sup>89</sup> To measure the relative change in performance within the trill, each variable was centred (subtracting the within-trill mean) and rescaled (dividing by the within-trill standard deviation). This resulted in normalized note rate, amplitude and consistency within each trill, allowing us to compare the relative change in performance within trills with different acoustic structures.<sup>86</sup>

**Analysis of changes at the beginning of trills.** Using our multi-species database, we investigated the naturally occurring variation in song performance within trills. For the analysis of change in performance relative to the beginning of the trill, we numbered all notes in the trill starting with the first (first note = 1, as described in “Numbering notes within trills”) and restricted the sample to the first ten notes in the trill, as 75% of all trills were equal to or shorter than ten notes in length. Trills longer than ten notes were also included in this analysis but notes above note position ten were removed. For trills that were shorter or equal to 10 notes, the last two notes were excluded to remove the potential change at the end of trills from this analysis. We fitted three Generalized Additive Mixed Models (GAMM) as these are suitable for measuring change within the sequence of notes within a trill. We calculated the mean note rate, mean relative amplitude and mean vocal consistency at each note position along trills of the same individual. Therefore, the data used to fit the model had only one observation within individual and note position. We controlled for pseudo-replication by defining a hierarchical structure of the data through the random effects, nesting individuals within species, species within genus and genera within family. After fitting the model, we calculated the inflection points in the GAMM splines, as the point where the direction of change varied within trill. The inflection points were defined as the note position where the slope of the change (i.e. first derivative of the spline function) was significantly different from zero, after estimating the 95% confidence intervals of the slope at each point in the splines ([https://rpubs.com/hrlai/gam\\_inflection](https://rpubs.com/hrlai/gam_inflection)). We fitted three models, one model for each of the three performance variables (rate, amplitude and consistency). From these models we determined: 1) whether there was significant change in song performance along the trill, 2) whether this change was a directional increase or decrease in performance and 3) the position within the trill (note position) of any inflection point(s).

**Analysis of changes after sustained singing.** To investigate a change in performance at the end of the trill after sustained singing, we fitted another set of three GAMMs using the same method. In this case, notes in the trill were numbered with respect to the position of the last note in the shortest trill (see “Numbering notes within trills” above) and we excluded the first two notes of all trills to remove the effect of initial changes in performance from this analysis. For the analysis of performance decline, we restricted the sample to those song types within an individual that had at least two trills sampled, as the shortest trill is used as a reference. By comparing the songs of a specific song type from the same individual we were able to search for patterns of variation in performance in relation to differences in trill length, removing inter-individual variation and other confounding factors (i.e. different performance constraints associated with different song types).



We calculated the mean note rate, mean relative amplitude, and mean vocal consistency at each note position, relative to the end of an individual's shortest trill (see [Figure 3](#)). These models allowed us to test our prediction that performance declines after sustained singing, particularly if trills become longer. In our sample, 75% of trills used as input for the models had nine notes or fewer, therefore, to avoid note positions with relatively small sample sizes, we restricted the sample to note positions between -8 and +4, where zero is the final note of the shortest trill found for each specific song type within individual. This excludes some notes at the beginning of trills in those cases where the shortest trill is longer than nine notes, but the variation in performance at the start of trills was considered separately in our analysis. Within individuals and song types, 95% of trills were fewer than four notes longer than the shortest trill. Hence there were few trills that were five or more notes longer than the shortest, and those songs were excluded from the analysis.

We had previously noticed a large drop in performance for the very last note in most trills. This fade-out effect<sup>101</sup> could be mistaken for a pattern of progressive decline as trills become longer when in fact it is only the very last note that decreases in performance, regardless of trill length. Such a pattern would not be indicative of a limitation to sustained performance within song. Post-hoc, we fitted three Linear Mixed-Effects Models (LMM) to test our prediction that the last notes of long trills (+4-note trills) were lower in performance than the last notes of the shortest trills. For this we compared note positions from -3 to 0 in short trills (black in [Figures 3D–3F](#)) versus note positions from 1 to +4 in the long trills (in yellow in [Figures 3D–3F](#)). These three LMM models each included one of the performance variables (rate, amplitude or consistency) as a function of trill length (short vs. long trills), nesting individuals within species and family using random effects to control pseudo-replication.

### Playback experiment in blue tits

#### *Statistical analysis of playback experiment*

To investigate the changes in behaviour of territorial male blue tits in response to our playback stimulation, we fitted two LMM models, one to describe the approach behaviour and one to describe the vocal behaviour. The approach behaviour model was fitted on the time spent within 4 m of the speaker as a function of treatment, phase, order of test within trial and order of treatment presentation within test. A similar model was fitted with the total duration (s) of all vocalizations as the response variable. We included the full interaction between treatment and phase to compare the short-stable treatment in each test with the long-stable and long-decline treatments respectively for each test. Test order was a two-level categorical variable indicating which test (short-stable vs. long-stable or short-stable vs. long-decline) was conducted first within a trial. Similarly, treatment order was a two-level categorical variable indicating which treatment was presented first within a test.

The identity of the subject male was included as a random effect to group together observations of the same subject and avoid pseudo-replication. We considered there was a significant effect on the response variable if the 95% confidence intervals (CI) did not overlap with zero. We inferred a non-significant trend if the 95% CI overlapped with zero but not the 90% CI. A non-significant trend is not shown in the tables, but it is mentioned in the text where appropriate. Since the model structure was based on the *a priori* experimental design, we did not carry out a model selection process.