Please cite the Published Version

Sierro, Javier, De Kort, Selvino R and Hartley, Ian R (2023) Sound properties affect measurement of vocal consistency in birdsong: validation of the spectrogram cross correlation method (SPCC). The Journal of the Acoustical Society of America, 154 (2). pp. 699-708. ISSN 0001-4966

DOI: https://doi.org/10.1121/10.0020543

Publisher: Acoustical Society of America

Version: Accepted Version

Sound properties affect measurement of vocal consistency in birdsong: validation of the Spectrogram Cross-Correlation method (SPCC)

Javier Sierro *[1], Selvino R. de Kort [2] & Ian R. Hartley [1]

1 Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK

2 Ecology and Environment Research Centre, Department of Natural Sciences, Manchester Metropolitan University, Manchester, UK

Keywords: Acoustic similarity; Vocal performance; acoustic methods; measuring birdsong; animal bioacoustics; motor performance; motor precision

Abstract

In songbirds, singing with precision (vocal consistency) has been proposed to reflect whole-organism performance. Vocal consistency is measured using Spectrogram Cross-Correlation (SPCC), assessing the acoustic similarity between subsequent renditions of the same note. To test whether the SPCC is sensitive to the acoustic discrepancies found in birdsong, we created a set of 40,000 synthetic sounds, that were designed based on the song of 345 species. This set included 10,000 reference sounds and 30,000 inexact variants with known differences in frequency, bandwidth or duration with respect the reference sounds. We found that SPCC is sensitive to acoustic discrepancies within the natural range of vocal consistency, supporting this method as a tool to assess vocal consistency in songbirds. Importantly, the sensitivity of SPCC was significantly affected by the bandwidth of sounds. The predictions derived from the analysis of synthetic sounds were then validated using 954 song recordings from 345 species (20 families). Based on psychoacoustic studies from birds and humans, we propose that the sensitivity of SPCC to acoustic discrepancies mirrors a perceptual bias in sound discrimination. Nevertheless, we suggest the tool be used with care, since sound bandwidth varies largely between singing styles and therefore SPCC scores may not be comparable.

I. INTRODUCTION

Birdsong is arguably one of the most complex acoustic signals in animal communication. Although it is known for the diversity of songs and motifs, singing involves the execution of complex motor patterns through the coordination of various muscle systems (Suthers, 2004). As in other animal displays, motor performance of song relays important information about a bird's quality that is relevant during social interactions (Byers *et al.*, 2010; Sakata and Vehrencamp, 2012; Botero and de Kort, 2013). One important aspect of motor performance is *precision*, the ability to produce the same act with minimal variation (Lane and Briffa, 2021). In birdsong, precision can be measured as vocal consistency, which refers to the ability to produce the same note without variation (de Kort *et al.*, 2009; Sakata and Vehrencamp, 2012).

A note is a short acoustic structure with a stereotypic shape within an individual's repertoire, generally defined as a continuous trace in the spectrogram (Knudsen and Gentner, 2010). When a bird produces subsequent renditions of the same note, it is executing the same motor pattern multiple times (Allan and Suthers, 1994; Suthers *et al.*, 1996). Hence, small discrepancies in the acoustic structure among renditions of the same note within song must be due to variation in the brain and muscle activation patterns during the execution. Most movements performed during singing occur inside the body, hidden from view, but the song output is the manifestation of these motor patterns. By measuring the acoustic similarity between two renditions of the same note type, we can assess the precision in which the same motor pattern has been executed, referred to as vocal consistency (Cardoso, 2017). Other types of variation in vocal output, such as learning accuracy or syntactical arrangement are not included here as vocal consistency. In some species it has been shown that vocal consistency is a signal of fitness related to reproductive success, perhaps associated with the neuro-motor skills of the individual (Sakata and Vehrencamp, 2012; Botero and de Kort, 2013; Sierro, 2022). Furthermore, vocal consistency varies seasonally in relation to the breeding season, similar to seasonal changes in hormone levels and brain structures, and with age (Smith *et al.*, 1997; Ballentine *et al.*, 2004; Botero *et al.*, 2009; de Kort *et al.*, 2009; Cramer, 2013; Vehrencamp *et al.*, 2013; Sierro *et al.*, 2022) which further supports the importance of vocal consistency in communication of birds.

Playback studies have shown that songbirds react differently to high and low consistency songs (de Kort *et al.*, 2009; Rivera-Gutierrez *et al.*, 2011). In fact, songbirds are highly sensitive to minute variations in the acoustic structure of sounds (Margoliash, 1983; Theunissen and Doupe, 1998; Lawson *et al.*, 2018; Fishbein *et al.*, 2019). Birds can identify frequency discrepancies between sounds as small as 1% and they are most sensitive to sounds within the range of 2-5 kHz, with decreasing sensitivity towards lower and higher frequencies, resembling in general terms the audiogram curve of humans (Dooling *et al.*, 2000; Knudsen and Gentner, 2010). Field studies show that spectral characteristics of song seem crucial in species recognition (Falls, 1963; Bremond, 1976; Fletcher and Smith, 1978; Nelson, 1989). In the temporal dimension, songbirds are able to discriminate differences in duration when sounds are at least 14-23% different in duration, with shorter sounds being generally more difficult to discriminate (Maier and Klump, 1990). These results are similar to those found in humans (Maier and Klump, 1990), although birds seem to be more sensitive to temporal discrepancies in complex sounds (Dooling *et al.*, 2002).

Since birds are highly sensitive to minute acoustic discrepancies, the method to measure vocal consistency must be equally sensitive. A commonly used method is the

Spectrogram Cross-Correlation (SPCC) algorithm that measures the acoustic similarity between two sounds represented by two spectrograms (Clark *et al.*, 1987). A spectrogram is essentially a double matrix with frequency in the Y-axis, time in the X-axis and the sound amplitude in each time-frequency bin. Two spectrogram matrices can be overlaid to estimate a correlation coefficient, as a measure of similarity between the two sounds, but there are many options for how these two spectrograms are aligned, a common problem when comparing time series. In the SPCC, this problem is solved by the second step in the method, the cross-correlation algorithm, which computes multiple correlations of both spectrograms at different temporal alignments. By definition, such optimization process will result in a lower sensitivity of the method to detect temporal discrepancies. After this process, the peak correlation coefficient from all correlations computed is selected as the acoustic similarity score between the two sounds (Clark *et al.*, 1987). The cross-correlation algorithm is essentially an optimizer (in the temporal dimension) that provides the similarity score between two sounds, rendering an acoustic similarity score from 0 (no similarity) to 1 (identical).

The SPCC has been shown to be a suitable tool to measure vocal consistency (Khanna *et al.*, 1997), reflecting biologically meaningful variation in birdsong such as individual differences or age variation (de Kort *et al.*, 2009; Rivera-Gutierrez *et al.*, 2012; Cramer, 2013). However, it is unclear how sensitive the method is to acoustic discrepancies found within the range of vocal consistency in birds. There are also reservations as to whether it provides an objective, universal tool to measure vocal consistency regardless of the singing style or song attributes (Cardoso, 2017). This a common problem in the study of vocal performance, since different singing styles might impose different physiological challenges and therefore the assessment of vocal performance is difficult to generalize (Cardoso, 2017). The bounded, standardized and unit-less nature of the SPCC similarity score has been an argument for the universality of the index, but it is still possible that the temporal or spectral properties of the sounds influence the SPCC response to acoustic discrepancies.

Here, we investigate the response of SPCC to acoustic discrepancies in a controlled set of synthetic sounds that can be defined and manipulated. These synthetic sounds emulate whistle-like vocalizations of songbirds when upper harmonics are filtered out by the vocal tract (Nowicki, 1987; Nowicki *et al.*, 1989; McGregor and Dabelsteen, 1996; Fletcher and Tarnopolsky, 1999). We used this set of synthetic sounds to test: 1) if the SPCC method is sensitive to acoustic discrepancies within the range of natural variation found in birdsong and 2) whether the SPCC response is influenced by the spectral or temporal properties of sound. Because the cross-correlation algorithm of SPCC acts as an optimizer in the temporal dimension, we expect that the SPCC sensitivity to temporal discrepancies will be lower than to spectral discrepancies. We then tested the findings and predictions derived from the analysis of synthetic sounds in a data base of natural song recordings from 345 different species of songbirds (20 different families) from around the world. Finally, we compare the quantitative properties of SPCC with published data on the perception of acoustic discrepancies by birds, evaluating the validity of this method to provide a biologically meaningful measure of vocal consistency.

II. METHODS

A. Natural variation in birdsong

To create the synthetic sounds that simulated bird notes, we used data derived from the analysis of 954 different recordings from 345 species that belong to 20 different families

(Acrocephalidae, Cettiidae, Cinclidae, Emberizidae, Estrildidae, Fringillidae, Icteridae, Mimidae, Motacillidae, Muscicapidae, Paridae, Passerellidae, Passeridae, Petroicidae, Phylloscopidae, Remizidae, Sittidae, Troglodytidae, Turdidae and Vireonidae). For all 20 families, we reviewed the song of all species (1,815 species in total) by listening to at least two recordings from the Xeno-Canto repository (www.xeno-canto.org). Then, we selected all those species that produced trills, defined as the consecutive repetition of the same note type at least 5 times. A note was defined as a continuous trace in the spectrogram, and the sample includes a large diversity of note's shapes (Figure 1). From each species, we selected a maximum of 5 different recordings, with high signal-to-noise ratio and selected a maximum of 5 different trills.

In each trill, we measured the duration of individual notes manually and tracked the fundamental frequency (window size: 512 samples; 90% overlap, amplitude threshold; 15%). The fundamental frequency (F0) is a series of values measuring the peak frequency of a note at each time point (window) (Figure 2B). The F0 range was defined as the distance in kHz between the highest and the lowest values of the F0, hereafter referred to as bandwidth (Figure 2B). The central frequency was defined as the equidistant point in the F0 range, hereafter refer to only as frequency (Figure 2B). To measure the within-trill variation, we calculated the percentage difference between each note and the mean duration, mean bandwidth and mean frequency of all notes in the trill. Estimating percentages with zero in the denominator can be problematic, but we did not encounter any case where the mean bandwidth of all notes within a trill was zero, see the next section.

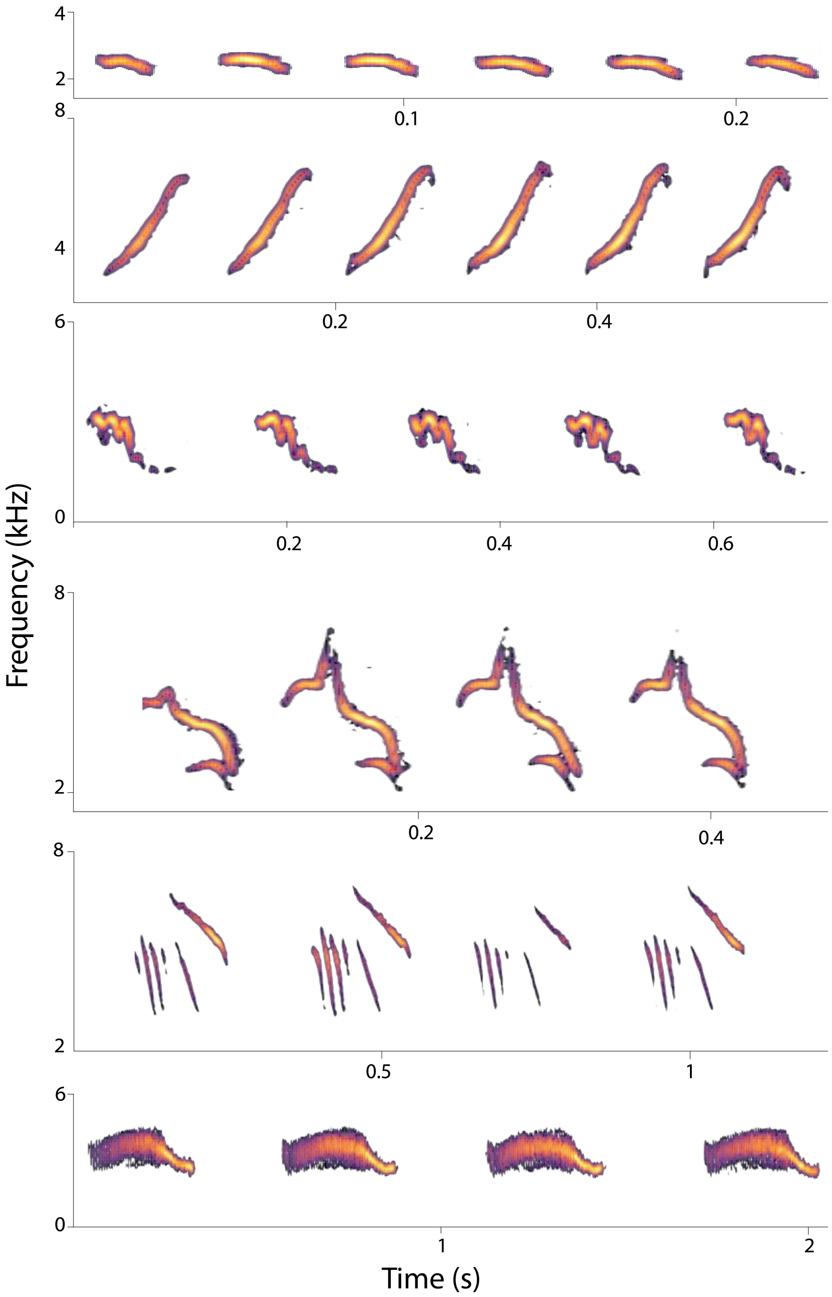

FIG 1 – Spectrograms showing different types of bird sounds included in our multi-species analysis. From top to bottom, *Acrocephalus paludicola, Setophaga pinus, Acrocephalus atyphus*, *Aimophila notosticta*, *Anthus spinoletta*, *Locustella montis*.

A. Study design and sound synthesis

To investigate the response of the SPCC score to acoustic discrepancies in frequency, bandwidth and duration, we created a set of 10,000 reference sounds that were tonal sounds with a gradient of possible frequency modulations, including pure tones, and no harmonics. The frequency modulation followed a shape based on a sine function (see Figure S1). These synthetic sounds had a central frequency of 4.1 kHz, matching the mean frequency measured in natural birdsong. They ranged from 0 kHz bandwidth (pure tone) to 1.64 kHz bandwidth, matching the mean bandwidth measured in birdsong, and a note length between 28 and 172 ms, matching the natural range in note length measured in birdsong as mean ± one standard deviation (SD). For each reference sound we synthesized 3 inexact copies, one for each treatment group, hereafter frequency, bandwidth and duration treatments (Figure 2). Each variant differed from the reference sound in just one parameter. For the frequency treatment, we created inexact variants that had the same spectrographic shape, bandwidth and duration but with a higher or lower frequency (Figure 2A). For the bandwidth treatment, we created inexact copies that differed in bandwidth from the reference sound, by stretching or shrinking the reference sound in the frequency spectrum while keeping the duration and frequency unchanged (Figure 2B). Finally, in the duration treatment, we stretched or contracted the reference sound in the temporal dimension to create an inexact variant that differed only in duration, but with the same bandwidth and frequency (Figure 2C). The full synthesis process as well as the following acoustic analyses were conducted in R software (Sueur *et al.*, 2006; Ligges, 2013; R Core Team, 2022).

The range of the variation introduced between a reference and a variant sound was derived from the naturally occurring variation between notes of the same trill measured in our birdsong database. In real birdsong, we measured the absolute difference in frequency, bandwidth and duration between notes of the same trill, relative to the mean frequency, bandwidth and duration of all notes within that trill. The absolute difference was transformed to a percentage relative to the mean frequency, bandwidth or duration found in that trill. Then, we calculated the mean of the differences per species and took the 75% quartile of the variation in frequency (6.0%), bandwidth (43.3%) and note duration (15.4%) as the maximum variation introduced between reference and variant sounds in each treatment of the set of synthetic sounds. Then, for each variant sound we calculated the frequency and the duration as a percentage of the model frequency and duration. In the case of bandwidth, we calculated a range of possible bandwidths for variants, ranging from 0 to 0.71 Hz, which is 43.3% of the maximum bandwidth (i.e. 1.64 kHz). A random value within this range was then added or subtracted to the bandwidth defined for the model sound. We did this because estimating a percentage of 0 kHz, or very low bandwidth sounds like pure tones, would lead to very small variations in bandwidth and therefore a bias throughout the range of bandwidth discrepancies.

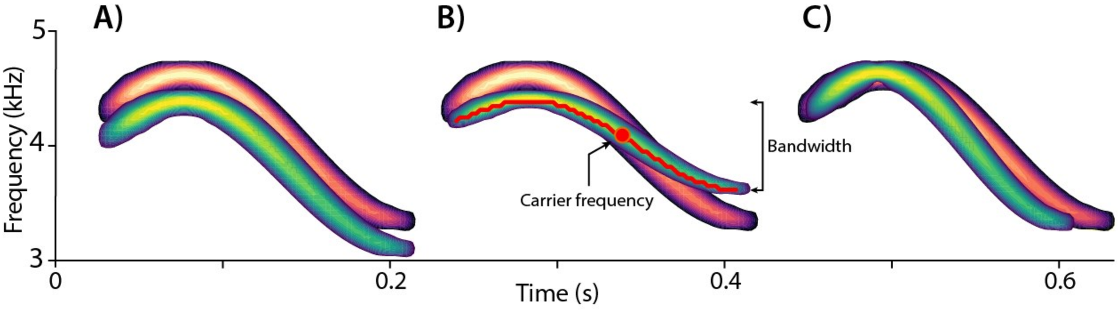

FIG 1 – Spectrograms of a synthetic sound built as a reference (red) and three inexact variants (green), one for the frequency treatment (A), one for the bandwidth treatment (B), and one for the duration treatment (C). Maximum, minimum and frequency indicated in 1B, as measured in the fundamental frequency (shown in a red line)

C. Measuring sound similarity with the Spectrogram Cross-Correlation algorithm

We measured the acoustic similarity between each synthetic sound (reference) and each variant using the SPCC algorithm (Clark *et al.*, 1987; Cortopassi and Bradbury, 2000). First, we computed the spectrogram matrices using an FFT algorithm with a window size of 512 samples, 80 % overlap between successive windows and 'Hanning' window type (Figure 3A). The algorithm overlays two spectrogram matrices at multiple (consecutive) time offsets, calculating a correlation coefficient at each point (Figure 3B). Plotting each correlation coefficient per time offset will produce a curve (Figure 3C), with the peak correlation in the curve taken as the acoustic similarity between those two sounds.

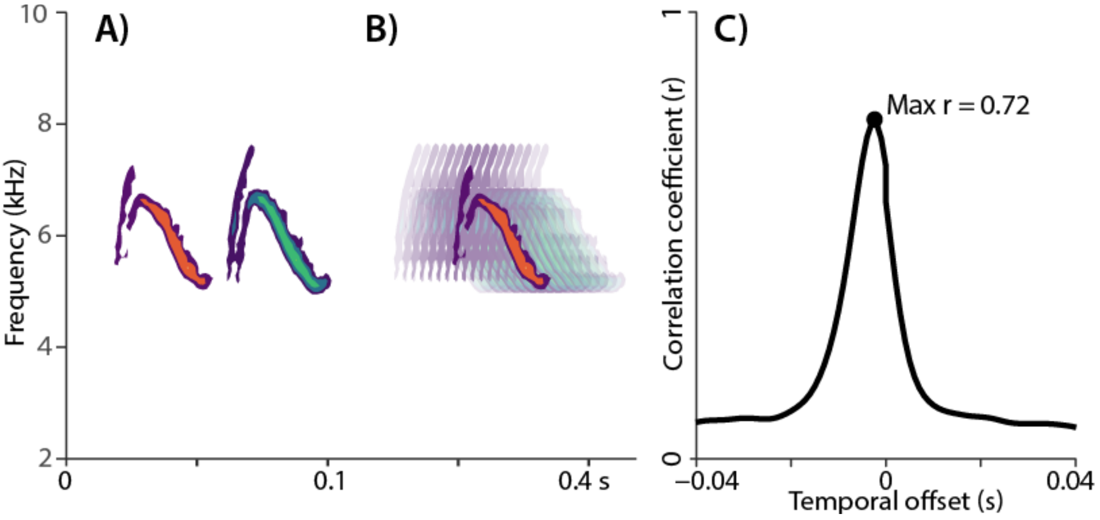

FIG 2 – Example of SPCC algorithm used to compare two notes of the trill of a blue tit (*Cyanistes caeruleus*). The two notes to be compared (A) are overlaid at different time offsets during the SPCC (B), producing multiple correlations coefficients, one at each of these alignments (C). The maximum correlation is taken as the SPCC score (C).

D. Statistical modelling of SPCC response

All measures are presented as mean ± one SD, unless otherwise indicated. Statistical analysis was carried out in R software (Bates *et al.*, 2015; R Core Team, 2022).

We fitted Linear Models (LMs) to the SPCC score, the response variable, as a function of the difference between variant-reference sound pairs, taking the variation in frequency, bandwidth and duration as a percentage. In the case of variation in frequency and duration, the percentage was taken using the model as a reference (denominator). In the case of bandwidth, the model could be a pure tone (i.e. 0 Hz of bandwidth) and, to avoid having 0 as a denominator, we selected the highest value of bandwidth (between the model and the variant) as a reference (denominator) to estimate the percentage difference in bandwidth between model and variant. This solved the problem, as by definition there was no case where both variant and model were pure tones.

Three models were fitted, one for each treatment. The estimated parameter for the variable “variant-reference difference” would indicate the SPCC sensitivity to acoustic discrepancies. In the models, we also included the absolute bandwidth and note duration of the reference sound and the full interaction with the variant-reference differences, to explore how the acoustic structure of the note influenced the SPCC sensitivity. These variables, bandwidth and duration of model sound, were scaled and center to allow us to compare the impact regardless of different units (Gelman, 2008).

Based on preliminary analysis and given the bounded distribution of SPCC score between 0 and 1, we transformed the response variable using an arcsine and a logit function. Both transformations seemed appropriate in some part of the distribution range but neither led to a reasonably good fit throughout the entire range. We observed that there was a change in the slope or curve (SPCC sensitivity) towards larger values of variant-reference difference, particularly in the frequency and the bandwidth treatments. Thus, we decided to fit two models in each case, splitting the range of acoustic discrepancies into two parts after calculating the break point by fitting a segmented model (Muggeo, 2008). Data were then split into two groups: one with small acoustic differences, those variant-reference pairs with a difference below the estimated break point, and another with large acoustic differences for those variant-reference pairs with acoustic differences larger than the break point (Figure 4). In the frequency and the bandwidth treatments, we fitted a LM with an arcsine transformation of the SPCC score for the small differences group, while for large differences group we fitted an LM with a logistic transformation of the SPCC score. For the duration treatment, a single model with an arcsine transformation fitted well for the entire range of acoustic differences. We considered a variable to have a significant impact on the SPCC score if the 95% confidence intervals (CI) did not overlap with zero.

D. Testing the results with real birdsong

We investigated whether the conclusions derived from the analysis of synthetic sounds were reflected in real data using the multi-species song data. To this end, we first classified all notes as being narrowband sounds, with a bandwidth lower than 100 Hz, and broadband sounds, those with a bandwidth higher than 100 Hz. Then, all notes were classified as similar in frequency if the difference in their frequency was less than 63 Hz with respect to the mean trill frequency, or different in frequency If the difference between the note and the trill central was larger than 63 Hz. The 63 Hz frequency threshold was the median variation in frequency in all notes from the birdsong data set, with respect to mean frequency within trill, dividing the whole sample approximately in half. Similarly, all notes were classified as different duration if the difference between note duration and mean trill note duration was larger than 4%. Again, this threshold was the median difference in note duration in our birdsong data. This analysis allowed us to explore the impact of bandwidth in measuring vocal consistency when two notes were different in frequency or in duration. We used a Mann-Whitney U test to compare the SPCC scores of broad and narrowband trills with the same and with different frequency. Similarly, we compared the SPCC scores of notes of narrow and broad bandwidth that were different in duration, but not in frequency.

III. RESULTS

A. SPCC response to discrepancies in synthetic sounds

We found that the relationship between SPCC score and acoustic discrepancies fitted an arcsine curve in the duration treatment and for small acoustic differences of the bandwidth and frequency treatments. In the case of large acoustic differences in the frequency and the bandwidth treatment, the observed pattern best fitted a logistic curve. The breakpoints detected by the segmented models were 3.4% ± 0.12% SE in the case of frequency discrepancies and 21.3% ± 0.36% SE in the case of bandwidth discrepancies. In general, qualitative results from the arcsine and logistic models in the frequency and bandwidth treatments were very similar, henceforth we will refer to the arcsine curves (Table 1), although for completeness the logit models are presented in Table S1.

In all cases, the SPCC method was sensitive to acoustic discrepancies between reference-variant pairs, as the SPCC score showed a significant negative correlation with the acoustic discrepancies in frequency, bandwidth and duration generated between the reference-variant pairs (Figure 5, Table 1). The SPCC method was most sensitive to differences in frequency, with a mean decrease of 22% in SPCC score with an increment of 1% in frequency difference (Figure 5, Table 1). SPCC was less sensitive to differences in bandwidth as SPCC score decreased by a mean of 4.7% with a 1% increment in bandwidth differences, and finally, SPCC was least sensitive to differences in duration, as SPCC decreased by a mean of 1.8% with a 1% increment in the difference in duration (Figure 5, Table 1). Note that these estimates considered the average change in SPCC throughout the range of variations. We also found that the SPCC score was influenced by the bandwidth of the sounds being compared in all treatments, but the direction and size of the effect of bandwidth varied across treatments (Figure 5, Table 1). In the frequency treatment, where sounds were only different in frequency, the SPCC score was generally higher if the reference sound had a broad bandwidth than if it was narrowband sound (Figure 5). This is shown in the model by the positive, significant impact of model bandwidth and its interaction with variant-reference difference (Table 1). The steeper down slope in the SPCC response for narrowband notes in the frequency treatment is shown in Figure 5A, with the bandwidth shown by a gray gradient. See also a visual explanation in Figure 6A-B. For the bandwidth treatment, the impact of bandwidth was similar to the frequency treatment but smaller (Table 1). In the duration treatment, the impact of bandwidth was opposite, as the same difference in duration rendered a higher SPCC score in narrowband sounds (Figure 5B & Figure 5C-D). In general, shorter sounds rendered higher SPCC scores in all treatments, as shown by the negative effect of note duration and its interactions with the reference-variant difference (Table 1). This means that SPCC was less sensitive to acoustic discrepancies of shorter sounds, although this effect was relatively small. Finally, we see a significant interaction in all models of both bandwidth and duration with the model-variant difference (Table 1). This indicates that the impact of bandwidth and duration explained before is not homogeneous thorughout the range of acoustic discrepancies but increases with increasing acoustic discrepancies. Such effect is represented in Figure 3A-C as all three lines showing sensitivity for sounds of different bandwidth converge in the upper left corner.

Our detailed quantitative analysis allows to derive the exact sensitivity of SPCC throughout the range of acoustic discrepancies, while considering the effect of bandwidth and duration. To derive the exact values one can apply the estimated coefficients using a linear model: $SPCC = \alpha p + \beta w - \gamma d + \delta(pw) + \psi(pd)$. P is the percentage difference between sounds while the bandwidth and duration of the model sound are represented by w and d respectively. Then, α is the model-variant coefficient, β is the bandwidth coefficient, γ is the duration coefficient, δ is the coefficient for the model-variant interaction with bandwidth and ψ is the coefficient for the model-variant interaction with duration. In the models shown in Table 1 and S1, the explanatory variables are scaled and center so we can compare the impact of each predictor. In order to get the real values for sensitivity we provide the estimates derived from a model with the original, non-scaled variables (Tables S2 & S3).

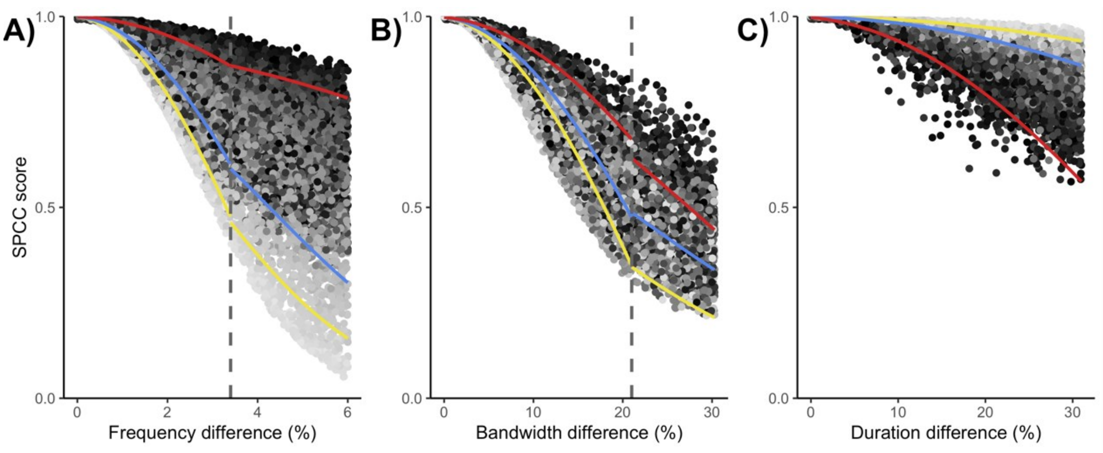

FIG 3 – Response of the SPCC score to acoustic discrepancies in frequency, defined as the equidistant point between maximum and minimum points of the F0 (A), bandwidth, defined as the distance in kHz between the maximum and minimum frequencies of the F0 (B), and sound duration in milliseconds (C). The gray gradient of the points shows the bandwidth of the reference sound from 0 kHz, i.e. a pure tone (light gray) to 1.6 kHz (black). For each treatment, lines represent the predicted values from the model adjusted to different bandwidth (0 kHz in yellow, 0.5 kHz in blue and 1.7 kHz in red). SPCC algorithm is most sensitive to frequency discrepancies in sounds, as shown by the steeper down slope in (A) regarding the frequency treatment. The duration treatment in (C) shows the shallowest slope, indicating that SPCC is least sensitive to temporal discrepancies. Figure (A) shows the impact of bandwidth in the SPCC response when

dealing with frequency discrepancies. Here, SPCC score of narrowband notes (light gray) decrease in a steeper slope than broadband sounds (black). This effect is opposite in the case of SPCC response to discrepancies in duration, where narrowband sounds (light gray) have a very shallow slope compared to broadband sounds (black).

B. SPCC and bandwidth in real birdsong

The birdsong database included 28,266 notes of 3,100 trills in 954 different recordings from 345 species in 20 families (17.3 ± 13.5 species per family). As predicted by our analysis of synthetic sounds, we found that SPCC scores were significantly higher in broadband notes than in narrowband notes if they differed in frequency (Broadband: 0.80 ± 0.11, Narrowband: 0.68 ± 0.20 SPCC score, W = 13819, $P < 0.001$, 5% CI = -0.13, 95% CI = -0.05; Figure 5A and 5A-B) but not if they were similar in frequency (Broadband: 0.85 ± 0.09, Narrowband: 0.87 ± 0.08 SPCC score, W = 39891, P = 0.004, 5% CI = 0.009, 95% =0.045, Figure 5B and 5A-B). Similarly, analysis of real birdsong confirmed our findings on the impact of bandwidth on SPCC between sounds of different duration. In this case, broadband sounds showed significantly lower SPCC scores than narrowband sounds, for the same difference in duration (Broadband: 0.84 ± 0.10, Narrowband: 0.87 ± 0.09 SPCC score, W = 31169, $P < 0.001$, 5% CI = 0.017, 95% =0.052, Figure 5C and 5C-D).

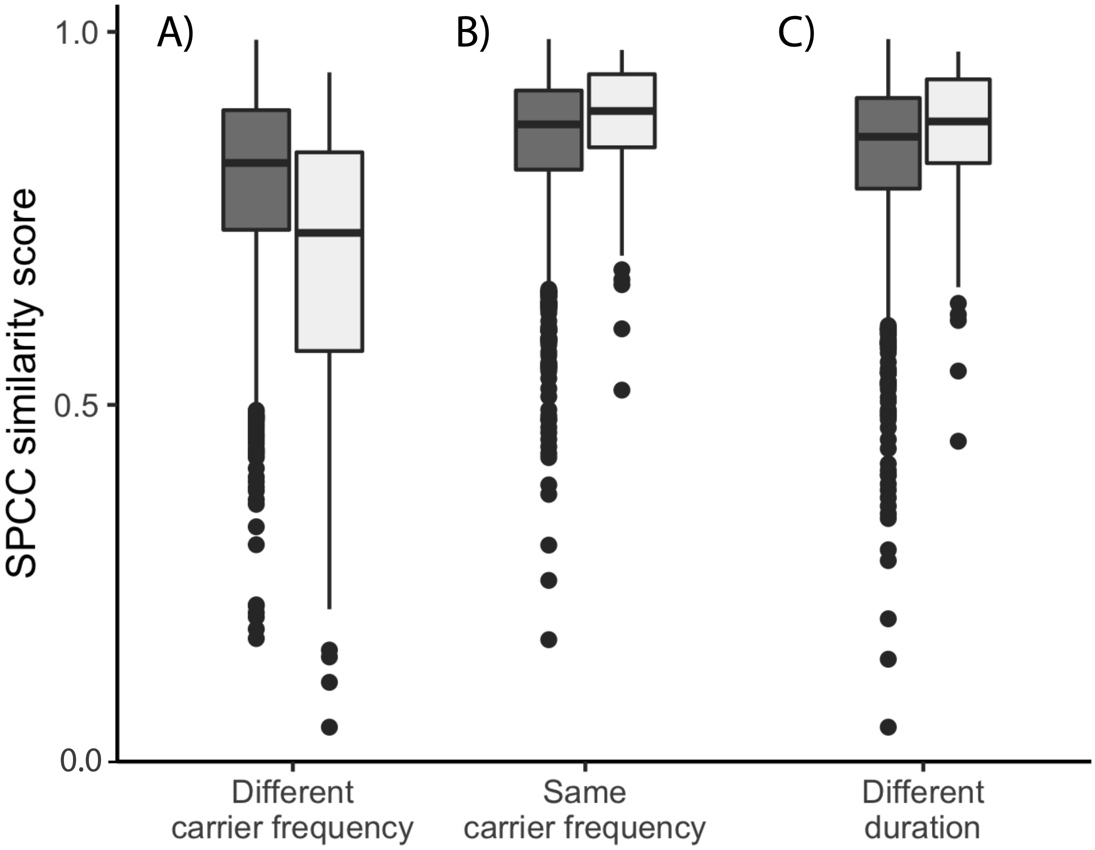

FIG 4 – Differences in SPCC score between broad band sounds (dark grey) and narrow band sounds (light grey), measured in natural song of 345 different species. As predicted by our analysis of synthetic sounds, SPCC scores of narrow band sounds with different frequency, i.e. different F0 but same shape and duration, is lower than in broadband sounds (A). However, if frequency is the same, narrow band sounds have higher SPCC scores (B). When two narrowband sounds differ in duration (but with the same frequency) they show higher SPCC scores than two broadband sounds of different duration (C).

TABLE 1 – Output of the model investigating the SPCC response to acoustic differences in frequency, duration and bandwidth, with an arcsine transformation of the SPCC score. For each fixed effect, the model estimate, the lower and higher CI and the T statistic are shown. The estimate of the parameter of reference-variant difference indicates the slope in the correlation between the SPCC score and the programmed difference between synthetic sounds, i.e., the sensitivity of the SPCC. The bandwidth of the sounds being compared has a significant impact on the SPCC score, especially in the frequency and duration treatment but with opposite effects. The duration of the sound shows a significant impact on the SPCC score as shorter sounds tend to have higher SPCC values, but the effect size is small.

Treatment	Parameters	Estimate	*T*	CI 5%	CI 95%	*P*
Frequency	Intercept	0.911	874.2	0.91	0.913	< 0.0001
	Model-variant difference	-0.22	-168.9	-0.222	-0.217	< 0.0001
	Bandwidth	0.08	61.1	0.078	0.083	0.26
	Duration	-0.015	-11.7	-0.018	-0.013	0.92
	Model-variant difference : Bandwidth	0.044	33.7	0.042	0.047	< 0.0001
	Model-variant difference : Duration	-0.009	-6.8	-0.012	-0.006	< 0.0001
Bandwidth	Intercept	0.998	742.3	0.998	0.998	< 0.0001
	Model-variant difference	-0.047	-269.6	-0.047	-0.046	< 0.0001
	Bandwidth	0.000	0.2	-0.003	0.004	0.83431
	Duration	0.000	0.2	-0.003	0.004	0.81524
	Model-variant difference : Bandwidth	0.002	14.4	0.002	0.003	< 0.0001
	Model-variant difference : Duration	-0.002	-14.1	-0.003	-0.002	< 0.0001
Duration	Intercept	0.999	1307.1	0.999	0.999	< 0.0001
	Model-variant difference	-0.018	-282.5	-0.018	-0.018	< 0.0001
	Bandwidth	-0.012	-10.6	-0.015	-0.01	< 0.0001
	Duration	-0.004	-3.6	-0.006	-0.002	< 0.001
	Model-variant difference : Bandwidth	-0.004	-64.9	-0.004	-0.004	< 0.0001
	Model-variant difference : Duration	-0.001	-22.7	-0.002	-0.001	< 0.0001

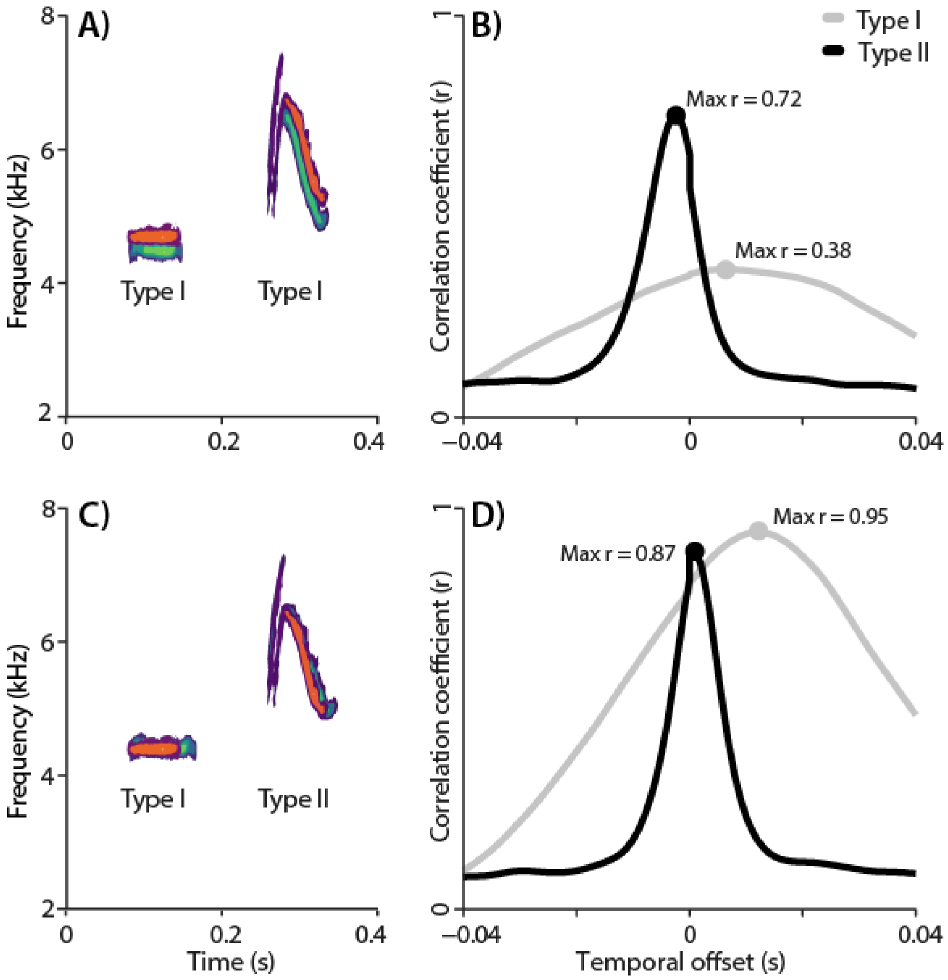

FIG 5 – Visual representation of the impact of bandwidth on SPCC sensitivity to differences in frequency, using natural notes recorded from blue tit song. In green, two notes types arbitrarily used as a reference. Another rendition of each note type is overlaid using red colors. (A) depicts the two note types (green) and variants (red) that differed mainly in frequency, with the associated cross-correlation curve in (B). The broadband note (type II) produces a high SPCC score by shifting the red note earlier in time. This is shown by the peak in correlation before zero in the X-axis in the cross-correlation curve for note type II in (B). Hence, for the same difference in frequency, the SPCC score is lower in narrowband notes in grey (type I), compared to broadband notes in black (type II). (C) depicts two pairs of notes that differ in duration, but not in frequency, with the respective SPCC curves on (D). In this case, the red note in the narrowband note (type I) shows a high overlap regardless of the difference in duration, whereas lengthening a broadband note (type II) will change the shape of the note and therefore reduce the SPCC score. In this case, (D) shows that for the same difference in

duration, narrowband notes in grey (type I) render a slightly higher SPCC score than broadband notes in black (type II).

IV. DISCUSSION

Our results support the use of SPCC to measure vocal consistency in birds, since the acoustic similarity score derived from SPCC correlated significantly with the known acoustic discrepancies between synthetic sounds as found in natural birdsong. As expected from the optimizing algorithm in SPCC, the SPCC sensitivity to spectral differences was higher than to temporal differences, when both parameters were within the range of natural variation in vocal consistency found in birds. The relationship between SPCC and acoustic discrepancies (sensitivity) was not linear and best fitted an arcsine curve or a logistic curve. We also found that, in the case of spectral discrepancies (frequency and bandwidth), the sensitivity of SPCC decreased as the note bandwidth increased. This means that spectral discrepancies between narrowband sounds were easier to detect than those in broadband sounds. The opposite pattern was found when measuring differences in duration. Differences in note duration between broadband sounds were easier to detect than those in narrowband sounds. In general, shorter sounds produced higher SPCC scores, suggesting that SPCC is less sensitive when dealing with shorter sounds. The findings derived from the analysis of synthetic sounds were confirmed in our analysis of birdsong in 345 different species as: 1) broadband sounds had lower SPCC than narrowband sounds when notes differed in frequency and 2) narrowband sounds of different duration had higher SPCC scores than broadband sounds with the same difference in duration. Quantifying the SPCC response along the range of acoustic discrepancies found in birdsong allows us to compare the properties of SPCC with the perceptual abilities found in birds, based on literature. Furthermore, such a quantitative analysis permits researchers to determine the suitability of the method for their study model and scientific question.

We found that the response of SPCC along the range of acoustic discrepancies was not linear, which is likely due to the frequency resolution of the spectrograms that limits detectability of small acoustic differences. As differences between two sounds approach the frequency resolution, such differences are more difficult to detect and therefore the sensitivity of SPCC is reduced. The frequency resolution is determined by the chosen window length of the FFT algorithm. Increasing the window length would increase frequency resolution and thus SPCC sensitivity to small spectral discrepancies but, in turn, temporal resolution would be lower, compromising sensitivity of SPCC to temporal differences. Choosing the appropriate window length is an important step depending on the target of the study (Khanna *et al.*, 1997; De Kort *et al.*, 2002).

In birds, frequency discrimination threshold is estimated in 1% (Dooling, 1982). In our simulated data, the SPCC score of acoustic similarity decreased by 4.4 % when two sounds of intermediate bandwidth differ by 1% in frequency, strongly supporting this method to measure the smallest frequency discrepancies perceived by birds. In contrast with a 1% discrimination threshold for frequency differences, birds are only able to detect discrepancies in duration when two sounds are at least 14% different in duration, going up to 23% for short sounds of 100 ms (Maier and Klump, 1990). For a 14% difference in duration between two sounds, the SPCC similarity score decrease by 3.1%, again supporting the use of SPCC to assess the smallest temporal differences as perceived by birds. Hence, the sensitivity of SPCC to detect temporal discrepancies is in practice relatively similar to the frequency sensitivity if we consider the hearing capacities of birds (Knudsen and Gentner, 2010). Technically, a lower sensitivity of SPCC to temporal discrepancies is inherent to the method due to the cross-correlation algorithm. By computing multiple comparisons at different time offsets, the SPCC maximizes the chances of finding a match (i.e. optimization), while reducing the sensitivity to temporal discrepancies. However, this step is important to solve the problem of aligning two time series during their comparison. There are alternative methods to solve the alignment problem (i.e., Dynamic Time Warping; DTW) but, unless the optimization acts in the three dimensions (i.e. conducting a second cross-correlation in the frequency axis), this step will always cause differential sensitivity in the acoustic similarity score between spectral, amplitude or temporal discrepancies.

Another consequence of the cross-correlation algorithm, computing multiple comparisons in time (X-axis), implies that the frequency bandwidth (Y-axis) influences the SPCC score. Considering two sounds that differ in frequency, two pure tones of zero bandwidth will be represented by two parallel lines in the spectrogram. These two lines will never overlap regardless of the cross-correlation process sliding two notes along the temporal dimension, rendering low SPCC scores (Figure 6). On the other hand, broadband sounds of different frequency can be partly matched during SPCC if the difference in frequency is smaller than the bandwidth (Figure 6). The better fit of a logistic curve to large acoustic differences indicates that there is a threshold over which SPCC is relatively insensitive to increasing differences, as the logistic curve will approach zero asymptotically. Nevertheless, this is close to the upper range of the natural variation in vocal consistency, which implies a minor issue in the use of the method.

When considering differences in note duration between sounds, two pure tones of different duration are essentially two overlapping lines, meaning that the shape of the note does not vary by changing the duration and thus, SPCC renders high scores.

However, the spectrographic shape of a sound with modulating frequency will change substantially by changing the note's duration, meaning that the SPCC score will decrease considerably in response to differences in duration. These examples show the impact of bandwidth on the SPCC response, indicating that the same difference in frequency or duration is not reflected with a similar decrease in SPCC if measured in two pairs of sounds with different bandwidths.

At first, the impact of bandwidth on SPCC sensitivity may appear a flaw, implying that measurement of vocal consistency is biased, but this bias may not be a drawback if birds show similar perception of acoustic differences. In fact, it is expected that sensitivity to detect acoustic discrepancies by birds or other animals will not follow a linear response and will likely be affected by sound structure, as found in the SPCC response. Common starlings (*Sturnus vulgaris*), show lower discrimination thresholds when presented with two pure tones than when presented with a frequency modulated tone (Langemann and Klump, 1992). In humans, the threshold of frequency discrimination increases significantly with increasing frequency modulation (Dooley and Moore, 1988). Similarly, when two pure tones of different frequency are presented in sequence, the threshold of frequency discrimination is lower than when those two tones are presented by modulating the first frequency into the second frequency (Fastl, 1978). These studies strongly suggest that assessing acoustic differences is more difficult when the sounds to be compared have frequency modulations. In this sense, the impact of bandwidth in the SPCC score could mirror the perception of acoustic discrepancies in frequency, if birds follow similar perceptual patterns (Knudsen and Gentner, 2010). Other psychoacoustic studies in common starlings also show that sensitivity to frequency differences is higher for longer sounds (Maier and Klump, 1990), again similar to our findings that SPCC sensitivity is higher for longer sounds.

If ability to detect vocal inconsistencies is higher in narrowband sounds, receivers could show a preference for narrowband trills to assess motor performance skills faster and more accurately. From the sender's perspective, less skilled birds could in turn use broadband trills to 'hide' their mistakes, as inconsistencies are difficult to perceive. In line with this idea, common nightingales (*Luscinia megarhynchos*) produce narrowband trills (whistle songs) that are important in mate attraction, and vocal consistency within those trills indicates male quality (Bartsch *et al.*, 2016). It has been shown that individuals with higher vocal consistency produced more narrowband trills (Bartsch *et al.*, 2016), which suggests that less skilled individuals could hide their mistakes by avoiding narrowband trills. Common nightingales also produce fast trills of broadband tones during simulated intrasexual conflicts (Schmidt *et al.*, 2008), a type of song that is challenging and indicates muscle speed (Podos, 1997; Podos *et al.*, 2016). Hence, it seems possible that individual song repertoire (i.e. diversity of song types within individuals) may serve to demonstrate neuro-motor skills in relation to different performance constraints (Cardoso, 2017). In this case, narrowband trills may display *precision* (Cardoso, 2017; Lane and Briffa, 2021) while fast broadband trills may display *speed* (Podos and Nowicki, 2004; Lane and Briffa, 2021). This could help explain the lack of ecological correlates of some performance parameters in studies that use multiple song types (Cardoso, 2012).

In conclusion, our results support the use of the SPCC method to measure vocal consistency in birdsong elements and possibly in other taxa. Our findings further support multiple field studies that found meaningful correlations between vocal consistency measured by SPCC and individual features or ecological factors. Despite these results in support of SPCC as a biologically meaningful measure of vocal consistency, there are some concerns. We found that the sensitivity of SPCC was not linear along the range of naturally occurring vocal (in)consistency and that sensitivity to detect acoustic discrepancies is significantly affected by frequency bandwidth. We suggest that these patterns found in SPCC sensitivity may reflect a similar perceptual pattern in acoustic discrimination in bird hearing. Further empirical studies are needed to explore bird perception of vocal consistency and how it is affected by acoustic structure of sound. Despite this, we recommend caution when comparing absolute values of SPCC scores if the songs analyzed have different spectral structure (e.g. emitted by different species). If appropriate, a possible solution would be to normalize or standardize SPCC scores using statistical techniques to compare vocal consistency. Finally, we highlight the importance of understanding and validating the methods of measuring song performance to provide meaningful measures that can be generalized (Cardoso, 2017).

V REFERENCES

Allan, S. E., and Suthers, R. A. (**1994**). "Lateralization and motor stereotype of song production in the brown-headed cowbird," J. Neurobiol. **25**, 1154-1166.

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.

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

Bates, D. M., Maechler, M., Bolker, B., and Walker, S. (**2015**). "Fitting Linear Mixed-Effects models using 'lme4'," Journal of Statistical Software **67**, 1-48.

Botero, C. A., and de Kort, S. R. (**2013**). "Learned signals and consistency of delivery: a case against receiver manipulation in animal communication," in *Animal communication theory: information and influence*, pp. 281-296.

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.

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

Byers, J., Hebets, E., and Podos, J. (**2010**). "Female mate choice based upon male motor performance," Anim. Behav. **79**, 771-778.

Cardoso, G. C. (**2012**). "Paradoxical calls: the opposite signaling role of sound frequency across bird species," Behav Ecol **23**, 237-241.

Cardoso, G. C. (**2017**). "Advancing the inference of performance in birdsong," Anim. Behav. **125**, e29-e32.

Clark, C. W., Marler, P., and Beeman, K. (**1987**). "Quantitative analysis of animal vocal phonology: an application to swamp sparrow song," Ethology **76**, 101-115.

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.

Cramer, E. R. A. (**2013**). "Measuring consistency: spectrogram cross-correlation versus targeted acoustic parameters," Bioacoustics-the International Journal of Animal Sound and Its Recording **22**, 247-257.

De Kort, S. R., Den Hartog, P. M., and Ten Cate, C. (**2002**). "Diverge or merge? The effect of sympatric occurrence on the territorial vocalizations of the vinaceous dove Streptopelia vinacea and the ring-necked dove S. capicola," J. Avian Biol. **33**, 150-158.

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," Philos. Trans. R. Soc. Lond., B, Biol. Sci. **276**, 2315-2321.

Dooley, G. J., and Moore, B. C. (**1988**). "Detection of linear frequency glides as a function of frequency and duration," The Journal of the Acoustical Society of America **84**, 2045-2057.

Dooling, R. J. (**1982**). "Auditory perception in birds," in *Acoustic communication in birds*, pp. 95-130.

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

Dooling, R. J., Lohr, B., and Dent, M. L. (**2000**). "Hearing in birds and reptiles," in *Comparative hearing: Birds and reptiles* (Springer), pp. 308-359.

Falls, J. B. (**1963**). "Properties of bird song eliciting responses from territorial males," in *Proceedings of the International Ornithological Congress*, pp. 259-273.

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

Fishbein, A. R., Idsardi, W. J., Ball, G. F., and Dooling, R. J. (**2019**). "Sound sequences in birdsong: how much do birds really care?," Philos. Trans. R. Soc. Lond., B, Biol. Sci. **375**, 20190044.

Fletcher, L. E., and Smith, D. G. (**1978**). "Some parameters of song important in conspecific recognition by gray catbirds," The Auk **95**, 338-347.

Fletcher, N. H., and Tarnopolsky, A. (**1999**). "Acoustics of the avian vocal tract," The Journal of the Acoustical Society of America **105**, 35-49.

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

Khanna, H., Gaunt, S., and McCallum, D. (**1997**). "Digital spectrographic cross-correlation: tests of sensitivity," Bioacoustics **7**, 209-234.

Knudsen, D. P., and Gentner, T. Q. (**2010**). "Mechanisms of song perception in oscine birds," Brain and language **115**, 59-68.

Lane, S. M., and Briffa, M. (**2021**). "Skilful mating? Insights from animal contest research," Anim. Behav. **184**, 197-207.

Langemann, U., and Klump, G. M. (**1992**). "Frequency discrimination in the European starling (*Sturnus vulgaris*): A comparison of different measures," Hear. Res. **63**, 43-51.

Lawson, S. L., Fishbein, A. R., Prior, N. H., Ball, G. F., and Dooling, R. J. (**2018**). "Relative salience of syllable structure and syllable order in zebra finch song," Animal cognition **21**, 467-480.

Ligges, U. (**2013**). "tuneR–analysis of music."

Maier, E. H., and Klump, G. M. (**1990**). "Auditory duration discrimination in the European starling (*Sturnus vulgaris*)," The Journal of the Acoustical Society of America **88**, 616-621.

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

McGregor, P. K., and Dabelsteen, T. (**1996**). "Communication networks," in *Ecology and Evolution of Acoustic Communication in Birds* (Cornell University Press, Ithaca, NY).

Muggeo, V. (**2008**). "segmented: an R Package to Fit Regression Models with Broken-Line Relationships," R news **1**, 20-25.

Nelson, D. A. (**1989**). "Song frequency as a cue for recognition of species and individuals in the field sparrow (*Spizella pusilla*)," Journal of Comparative Psychology **103**, 171.

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

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.

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

Podos, J., Moseley, D. L., Goodwin, S. E., McClure, J., Taft, B. N., Strauss, A. V., Rega-Brodsky, C., and Lahti, D. C. (**2016**). "A fine-scale, broadly applicable index of vocal performance: frequency excursion," Anim. Behav. **116**, 203-212.

Podos, J., and Nowicki, S. (**2004**). "Performance limits on birdsong," in *Nature's music: the science of birdsong* (Elsevier Academic Press, San Diego), pp. 318-342.

R Core Team (**2022**). "R: A language and environment for statistical computing," (R Foundation for Statistical Computing, Vienna, Austria).

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.

Rivera-Gutierrez, H. F., Pinxten, R., and Eens, M. (**2012**). "Tuning and fading voices in songbirds: age-dependent changes in two acoustic traits across the life span," Anim. Behav. **83**, 1279-1283.

Sakata, J. T., and Vehrencamp, S. L. (**2012**). "Integrating perspectives on vocal performance and consistency," J. Exp. Biol. **215**, 201-209.

Schmidt, R., Kunc, H. P., Amrhein, V., and Naguib, M. (**2008**). "Aggressive responses to broadband trills are related to subsequent pairing success in nightingales," Behav Ecol **19**, 635-641.

Sierro, J. (**2022**). "The role of vocal consistency in bird communication, a case study in the blue tit (*Cyanistes caeruleus*)," (Lancaster University, Lancaster).

Sierro, 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, arac044.

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," Journal of Neuroscience **17**, 6001-6010.

Sueur, J., Aubin, T., and Simonis-Sueur, C. (**2006**). "Seewave," in *Université Paris XI-MNHN, Paris*.

Suthers, R. A. (**2004**). "How birds sing and why it matters," in *Nature's music: the science of birdsong* (Elsevier Academic Press, San Diego), pp. 272-295.

Suthers, R. A., Goller, F., and Hartley, R. S. (**1996**). "Motor stereotypy and diversity in songs of mimic thrushes," J. Neurobiol. **30**, 231-245.

Theunissen, F. E., and Doupe, A. J. (**1998**). "Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVC of male zebra finches," Journal of Neuroscience **18**, 3786-3802.

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.

Supporting information

TABLE S1 – Output of the model investigating the SPCC response to acoustic differences in frequency, duration and bandwidth, with a logistic transformation of the SPCC score. For each fixed effect, the model estimate, the lower and higher CI of the estimate and the T statistic are shown. The estimate of the parameter of reference-variant difference indicates the slope in the correlation between the SPCC score and the programmed difference between synthetic sounds, i.e., the sensitivity of the SPCC to detect acoustic differences in each treatment. The slope is significantly lower than zero, indicating that the acoustic similarity from SPCC is sensitive to acoustic variation in the three treatments. The SPCC is most sensitive to spectral differences (frequency and bandwidth) than in temporal differences (duration). The bandwidth of the sounds being compared has a significant impact on the SPCC score, especially in the frequency and duration treatment but with opposite effects. The duration of the sound shows a significant impact on the SPCC score as shorter sounds tend to have higher SPCC values, but the effect size is relatively small.

Treatment	Parameters	Estimate	*T*	CI 5%	CI 95%	*P*
Frequency	Intercept	0.74	39.8	0.811	0.659	< 0.0001
	Model-variant difference	-0.384	-32.3	-0.406	-0.361	< 0.0001
	Bandwidth	0.193	3.3	0.08	0.304	< 0.001
	Duration	-0.056	-1	-0.17	0.058	0.33347
	Model-variant difference : Bandwidth	0.102	8.3	0.078	0.125	< 0.0001
	Model-variant difference : Duration	-0.007	-0.5	-0.031	0.018	0.60
Bandwidth	Intercept	0.991	23.7	1	0.962	< 0.0001
	Model-variant difference	-0.069	-24.8	-0.075	-0.064	< 0.0001
	Bandwidth	0.305	4.4	0.17	0.435	< 0.0001
	Duration	-0.045	-0.6	-0.184	0.095	0.53
	Model-variant difference : Bandwidth	-0.007	-2.7	-0.013	-0.002	0.008
	Model-variant difference : Duration	-0.004	-1.5	-0.01	0.001	0.13

TABLE S2 – Output of the model investigating the SPCC response to acoustic differences in frequency, duration and bandwidth, with an arcsine transformation of the SPCC score. In this case, the variables "bandwidth" and "duration" were not scaled. The estimates from these models can be used to derive the sensitivity of SPCC at the chosen range of acoustic discrepancies and for a given bandwidth and sound duration

Treatment	Parameters	Estimate	*T*	CI 5%	CI 95%	*P*
Frequency	Intercept	0.999	183.3	0.998	1	< 0.0001
	Model-variant difference	-0.278	-66.7	-0.286	-0.27	< 0.0001
	Bandwidth	0	1.1	0	0	0.26
	Duration	0.006	0.1	-0.117	0.129	0.92
	Model-variant difference : Bandwidth	0	33.7	0	0	< 0.0001
	Model-variant difference : Duration	-0.216	-6.8	-0.277	-0.154	< 0.0001
Bandwidth	Intercept	0.998	247.9	0.997	0.999	< 0.0001
	Model-variant difference	-0.045	-83.8	-0.046	-0.044	< 0.0001
	Bandwidth	0	0.2	0	0	0.83
	Duration	0.011	0.2	-0.084	0.106	0.82
	Model-variant difference : Bandwidth	0	14.4	0	0	< 0.0001
	Model-variant difference : Duration	-0.059	-14.1	-0.067	-0.051	< 0.0001
Duration	Intercept	1	427.1	1	1	< 0.0001
	Model-variant difference	-0.007	-36.6	-0.008	-0.007	< 0.0001
	Bandwidth	0	-10.6	0	0	< 0.0001
	Duration	-0.1	-3.6	-0.155	-0.045	< 0.001
	Model-variant difference : Bandwidth	0	-64.9	0	0	< 0.0001
	Model-variant difference : Duration	-0.035	-22.7	-0.038	-0.032	< 0.0001

TABLE S3 – Output of the model investigating the SPCC response to acoustic differences in frequency, duration and bandwidth, with a logistic transformation of the SPCC score. In this case, the variables "bandwidth" and "duration" were not scaled. The estimates from these models can be used to derive the sensitivity of SPCC at the chosen range of acoustic discrepancies and for a given bandwidth and sound duration

Treatment	Parameters	Estimate	*T*	CI 5%	CI 95%	*P*
Frequency	Intercept	-0.983	239.8	-0.975	-0.989	< 0.0001
	Model-variant difference	-0.787	-166.1	-0.794	-0.78	< 0.0001
	Bandwidth	0.257	13.7	0.221	0.293	< 0.0001
	Duration	0.001	0	-0.036	0.038	0.96
	Model-variant difference : Bandwidth	0.09	16.5	0.079	0.101	< 0.0001
	Model-variant difference : Duration	-0.025	-4.6	-0.036	-0.014	< 0.0001
Bandwidth	Intercept	0.923	4.9	0.648	0.997	< 0.0001
	Model-variant difference	-0.044	-4.7	-0.062	-0.026	< 0.0001
	Bandwidth	0.001	4.4	0	0.001	< 0.0001
	Duration	-0.882	-0.6	0.962	0.763	0.53
	Model-variant difference : Bandwidth	0	-2.7	0	0	0.008
	Model-variant difference : Duration	-0.102	-1.5	-0.23	0.029	0.13

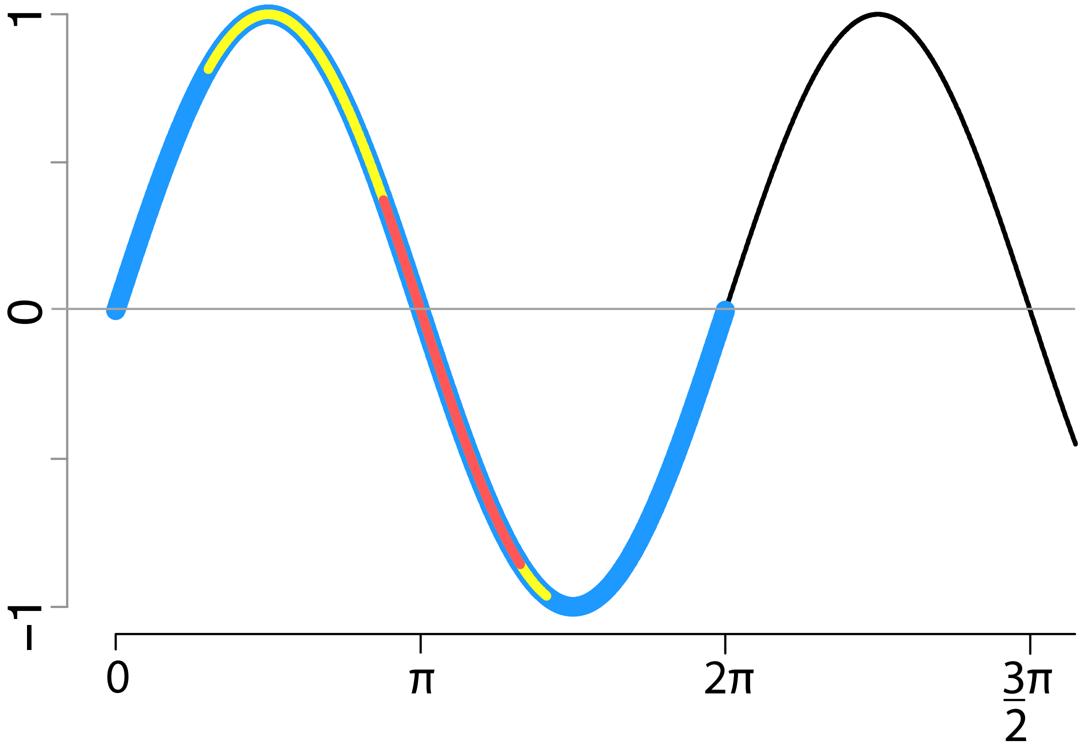

FIG S1 – The shape of the synthetic sounds was based on a sine shape with a starting point in $\sin(x\pi)$ and ending point in $\sin(y\pi)$, where x could be a value between 0 to 1 and y could be a value from 1 to 2. The entire range is depicted in blue in the figure, while the yellow and red traces show two possible shapes that would derive from the process. The selected shape would then be transported so a central frequency of 4.1 kHz and then adjusted to match a randomly selected bandwidth between 0 and 1.64 kHz. In the case of a 0 kHz bandwidth, the resulting sound would be a pure tone, regardless of the original shape.