



**Manchester  
Metropolitan  
University**

---

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

---

**Downloaded from:** <https://e-space.mmu.ac.uk/632488/>

**Version:** Accepted Version

**Publisher:** Acoustical Society of America

**DOI:** <https://doi.org/10.1121/10.0020543>

**Usage rights:** In Copyright

Please cite the published version

<https://e-space.mmu.ac.uk>

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

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

5

6 1 Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK

7 2 Ecology and Environment Research Centre, Department of Natural Sciences,  
8 Manchester Metropolitan University, Manchester, UK

9

10 **Keywords:** Acoustic similarity; Vocal performance; acoustic methods; measuring

11 birdsong; animal bioacoustics; motor performance; motor precision

12 **Abstract**

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

29

## 30 I. INTRODUCTION

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

41 A note is a short acoustic structure with a stereotypic shape within an individual's  
42 repertoire, generally defined as a continuous trace in the spectrogram (Knudsen and  
43 Gentner, 2010). When a bird produces subsequent renditions of the same note, it is  
44 executing the same motor pattern multiple times (Allan and Suthers, 1994; Suthers *et*  
45 *al.*, 1996). Hence, small discrepancies in the acoustic structure among renditions of the  
46 same note within song must be due to variation in the brain and muscle activation  
47 patterns during the execution. Most movements performed during singing occur inside  
48 the body, hidden from view, but the song output is the manifestation of these motor  
49 patterns. By measuring the acoustic similarity between two renditions of the same note  
50 type, we can assess the precision in which the same motor pattern has been executed,  
51 referred to as vocal consistency (Cardoso, 2017). Other types of variation in vocal  
52 output, such as learning accuracy or syntactical arrangement are not included here as  
53 vocal consistency. In some species it has been shown that vocal consistency is a signal

54 of fitness related to reproductive success, perhaps associated with the neuro-motor  
55 skills of the individual (Sakata and Vehrencamp, 2012; Botero and de Kort, 2013; Sierro,  
56 2022). Furthermore, vocal consistency varies seasonally in relation to the breeding  
57 season, similar to seasonal changes in hormone levels and brain structures, and with age  
58 (Smith *et al.*, 1997; Ballentine *et al.*, 2004; Botero *et al.*, 2009; de Kort *et al.*, 2009;  
59 Cramer, 2013; Vehrencamp *et al.*, 2013; Sierro *et al.*, 2022) which further supports the  
60 importance of vocal consistency in communication of birds.

61 Playback studies have shown that songbirds react differently to high and low consistency  
62 songs (de Kort *et al.*, 2009; Rivera-Gutierrez *et al.*, 2011). In fact, songbirds are highly  
63 sensitive to minute variations in the acoustic structure of sounds (Margoliash, 1983;  
64 Theunissen and Doupe, 1998; Lawson *et al.*, 2018; Fishbein *et al.*, 2019). Birds can  
65 identify frequency discrepancies between sounds as small as 1% and they are most  
66 sensitive to sounds within the range of 2-5 kHz, with decreasing sensitivity towards  
67 lower and higher frequencies, resembling in general terms the audiogram curve of  
68 humans (Dooling *et al.*, 2000; Knudsen and Gentner, 2010). Field studies show that  
69 spectral characteristics of song seem crucial in species recognition (Falls, 1963;  
70 Bremond, 1976; Fletcher and Smith, 1978; Nelson, 1989). In the temporal dimension,  
71 songbirds are able to discriminate differences in duration when sounds are at least 14-  
72 23% different in duration, with shorter sounds being generally more difficult to  
73 discriminate (Maier and Klump, 1990). These results are similar to those found in  
74 humans (Maier and Klump, 1990), although birds seem to be more sensitive to temporal  
75 discrepancies in complex sounds (Dooling *et al.*, 2002).

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

78 Spectrogram Cross-Correlation (SPCC) algorithm that measures the acoustic similarity  
79 between two sounds represented by two spectrograms (Clark *et al.*, 1987). A  
80 spectrogram is essentially a double matrix with frequency in the Y-axis, time in the X-  
81 axis and the sound amplitude in each time-frequency bin. Two spectrogram matrices  
82 can be overlaid to estimate a correlation coefficient, as a measure of similarity between  
83 the two sounds, but there are many options for how these two spectrograms are  
84 aligned, a common problem when comparing time series. In the SPCC, this problem is  
85 solved by the second step in the method, the cross-correlation algorithm, which  
86 computes multiple correlations of both spectrograms at different temporal alignments.  
87 By definition, such optimization process will result in a lower sensitivity of the method  
88 to detect temporal discrepancies. After this process, the peak correlation coefficient  
89 from all correlations computed is selected as the acoustic similarity score between the  
90 two sounds (Clark *et al.*, 1987). The cross-correlation algorithm is essentially an  
91 optimizer (in the temporal dimension) that provides the similarity score between two  
92 sounds, rendering an acoustic similarity score from 0 (no similarity) to 1 (identical).  
93 The SPCC has been shown to be a suitable tool to measure vocal consistency (Khanna *et*  
94 *al.*, 1997), reflecting biologically meaningful variation in birdsong such as individual  
95 differences or age variation (de Kort *et al.*, 2009; Rivera-Gutierrez *et al.*, 2012; Cramer,  
96 2013). However, it is unclear how sensitive the method is to acoustic discrepancies  
97 found within the range of vocal consistency in birds. There are also reservations as to  
98 whether it provides an objective, universal tool to measure vocal consistency regardless  
99 of the singing style or song attributes (Cardoso, 2017). This a common problem in the  
100 study of vocal performance, since different singing styles might impose different  
101 physiological challenges and therefore the assessment of vocal performance is difficult

102 to generalize (Cardoso, 2017). The bounded, standardized and unit-less nature of the  
103 SPCC similarity score has been an argument for the universality of the index, but it is still  
104 possible that the temporal or spectral properties of the sounds influence the SPCC  
105 response to acoustic discrepancies.

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

## 122 **II. METHODS**

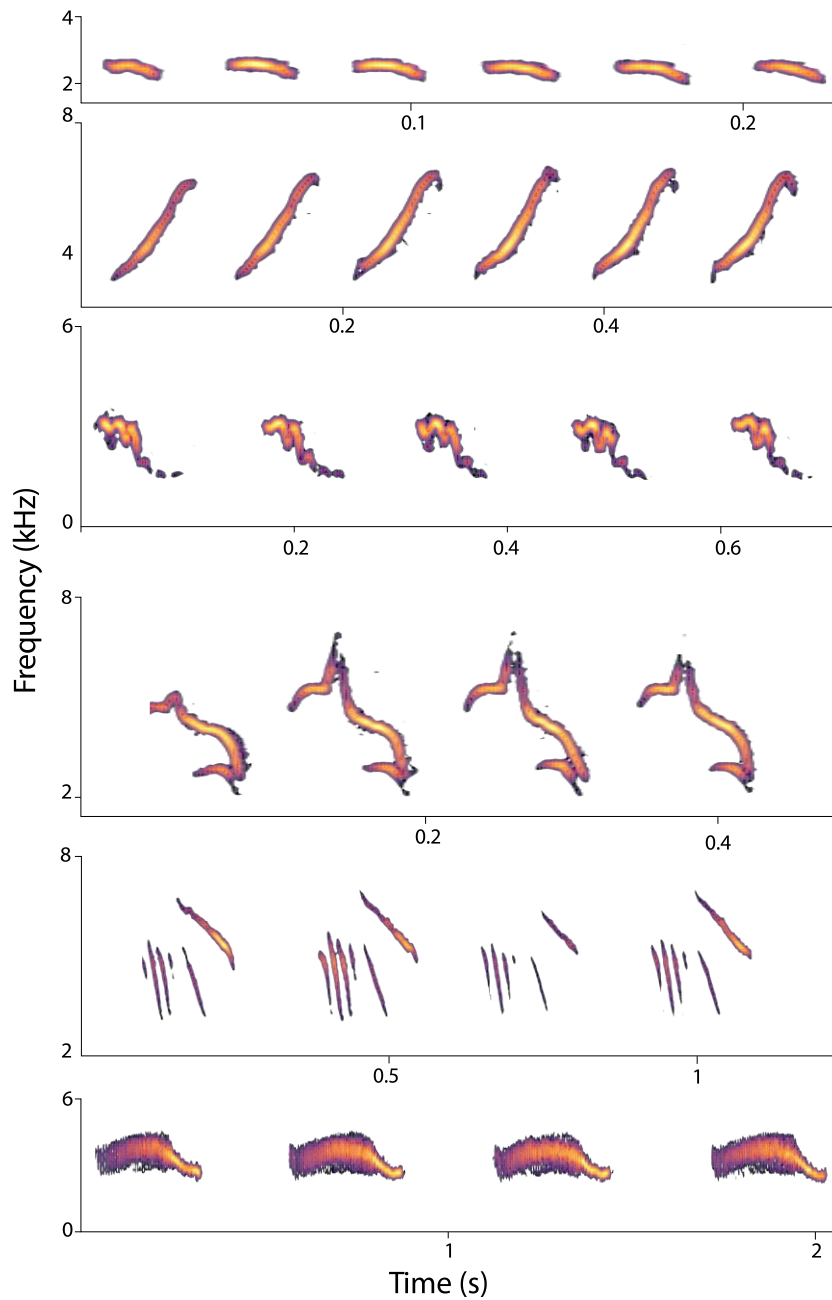
### 123 **A. Natural variation in birdsong**

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

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

136 In each trill, we measured the duration of individual notes manually and tracked the  
137 fundamental frequency (window size: 512 samples; 90% overlap, amplitude threshold;  
138 15%). The fundamental frequency (F0) is a series of values measuring the peak  
139 frequency of a note at each time point (window) (Figure 2B). The F0 range was defined  
140 as the distance in kHz between the highest and the lowest values of the F0, hereafter  
141 referred to as bandwidth (Figure 2B). The central frequency was defined as the  
142 equidistant point in the F0 range, hereafter refer to only as frequency (Figure 2B). To  
143 measure the within-trill variation, we calculated the percentage difference between  
144 each note and the mean duration, mean bandwidth and mean frequency of all notes in  
145 the trill. Estimating percentages with zero in the denominator can be problematic, but  
146 we did not encounter any case where the mean bandwidth of all notes within a trill was  
147 zero, see the next section.





148

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

152 **A. Study design and sound synthesis**

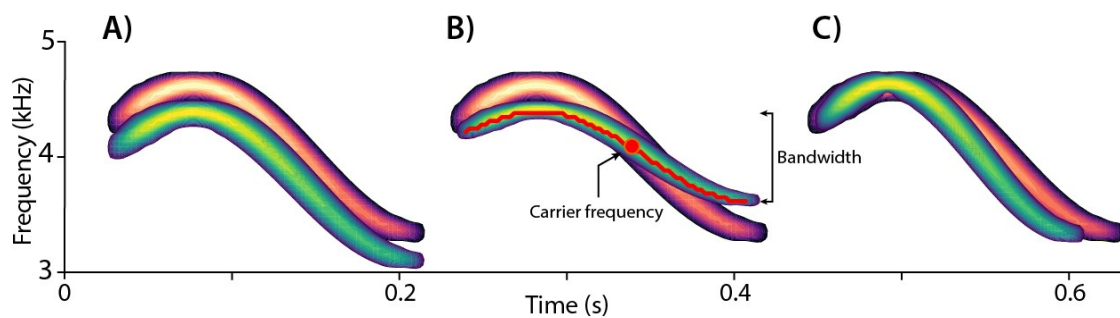
153 To investigate the response of the SPCC score to acoustic discrepancies in frequency,  
 154 bandwidth and duration, we created a set of 10,000 reference sounds that were tonal  
 155 sounds with a gradient of possible frequency modulations, including pure tones, and no

156 harmonics. The frequency modulation followed a shape based on a sine function (see  
157 Figure S1). These synthetic sounds had a central frequency of 4.1 kHz, matching the  
158 mean frequency measured in natural birdsong. They ranged from 0 kHz bandwidth (pure  
159 tone) to 1.64 kHz bandwidth, matching the mean bandwidth measured in birdsong, and  
160 a note length between 28 and 172 ms, matching the natural range in note length  
161 measured in birdsong as mean  $\pm$  one standard deviation (SD). For each reference sound  
162 we synthesized 3 inexact copies, one for each treatment group, hereafter frequency,  
163 bandwidth and duration treatments (Figure 2). Each variant differed from the reference  
164 sound in just one parameter. For the frequency treatment, we created inexact variants  
165 that had the same spectrographic shape, bandwidth and duration but with a higher or  
166 lower frequency (Figure 2A). For the bandwidth treatment, we created inexact copies  
167 that differed in bandwidth from the reference sound, by stretching or shrinking the  
168 reference sound in the frequency spectrum while keeping the duration and frequency  
169 unchanged (Figure 2B). Finally, in the duration treatment, we stretched or contracted  
170 the reference sound in the temporal dimension to create an inexact variant that differed  
171 only in duration, but with the same bandwidth and frequency (Figure 2C). The full  
172 synthesis process as well as the following acoustic analyses were conducted in R  
173 software (Sueur *et al.*, 2006; Ligges, 2013; R Core Team, 2022).

174 The range of the variation introduced between a reference and a variant sound was  
175 derived from the naturally occurring variation between notes of the same trill measured  
176 in our birdsong database. In real birdsong, we measured the absolute difference in  
177 frequency, bandwidth and duration between notes of the same trill, relative to the mean  
178 frequency, bandwidth and duration of all notes within that trill. The absolute difference  
179 was transformed to a percentage relative to the mean frequency, bandwidth or duration

180 found in that trill. Then, we calculated the mean of the differences per species and took  
 181 the 75% quartile of the variation in frequency (6.0%), bandwidth (43.3%) and note  
 182 duration (15.4%) as the maximum variation introduced between reference and variant  
 183 sounds in each treatment of the set of synthetic sounds. Then, for each variant sound  
 184 we calculated the frequency and the duration as a percentage of the model frequency  
 185 and duration. In the case of bandwidth, we calculated a range of possible bandwidths  
 186 for variants, ranging from 0 to 0.71 Hz, which is 43.3% of the maximum bandwidth (i.e.  
 187 1.64 kHz). A random value within this range was then added or subtracted to the  
 188 bandwidth defined for the model sound. We did this because estimating a percentage  
 189 of 0 kHz, or very low bandwidth sounds like pure tones, would lead to very small  
 190 variations in bandwidth and therefore a bias throughout the range of bandwidth  
 191 discrepancies.

192



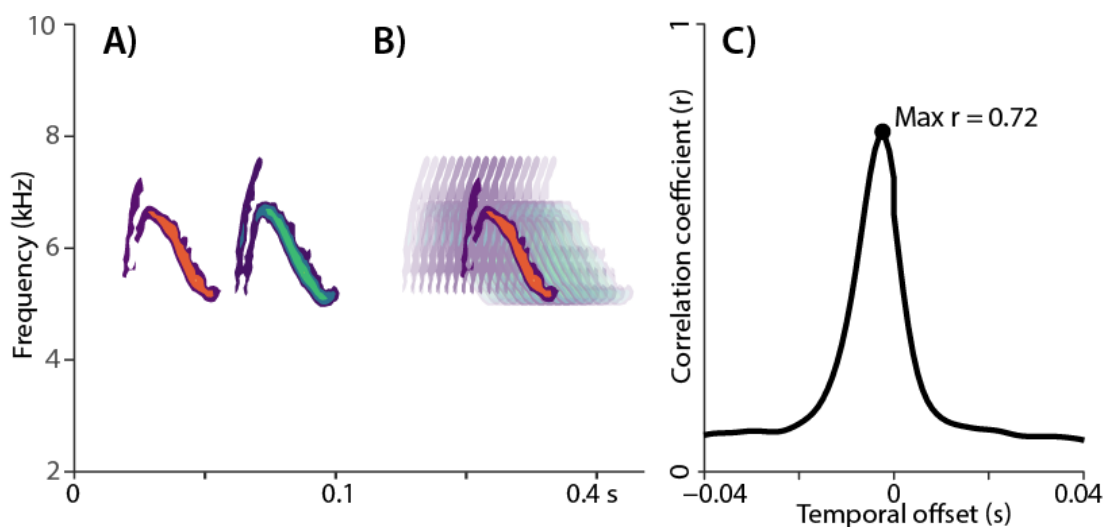
193

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

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

199 We measured the acoustic similarity between each synthetic sound (reference) and  
 200 each variant using the SPCC algorithm (Clark *et al.*, 1987; Cortopassi and Bradbury,

201 2000). First, we computed the spectrogram matrices using an FFT algorithm with a  
202 window size of 512 samples, 80 % overlap between successive windows and 'Hanning'  
203 window type (Figure 3A). The algorithm overlays two spectrogram matrices at multiple  
204 (consecutive) time offsets, calculating a correlation coefficient at each point (Figure 3B).  
205 Plotting each correlation coefficient per time offset will produce a curve (Figure 3C),  
206 with the peak correlation in the curve taken as the acoustic similarity between those  
207 two sounds.



208

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

#### 213 **D. Statistical modelling of SPCC response**

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

216 We fitted Linear Models (LMs) to the SPCC score, the response variable, as a function of  
217 the difference between variant-reference sound pairs, taking the variation in frequency,  
218 bandwidth and duration as a percentage. In the case of variation in frequency and  
219 duration, the percentage was taken using the model as a reference (denominator). In

220 the case of bandwidth, the model could be a pure tone (i.e. 0 Hz of bandwidth) and, to  
221 avoid having 0 as a denominator, we selected the highest value of bandwidth (between  
222 the model and the variant) as a reference (denominator) to estimate the percentage  
223 difference in bandwidth between model and variant. This solved the problem, as by  
224 definition there was no case where both variant and model were pure tones.

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

232 Based on preliminary analysis and given the bounded distribution of SPCC score  
233 between 0 and 1, we transformed the response variable using an arcsine and a logit  
234 function. Both transformations seemed appropriate in some part of the distribution  
235 range but neither led to a reasonably good fit throughout the entire range. We observed  
236 that there was a change in the slope or curve (SPCC sensitivity) towards larger values of  
237 variant-reference difference, particularly in the frequency and the bandwidth  
238 treatments. Thus, we decided to fit two models in each case, splitting the range of  
239 acoustic discrepancies into two parts after calculating the break point by fitting a  
240 segmented model (Muggeo, 2008). Data were then split into two groups: one with small  
241 acoustic differences, those variant-reference pairs with a difference below the  
242 estimated break point, and another with large acoustic differences for those variant-  
243 reference pairs with acoustic differences larger than the break point (Figure 4). In the

244 frequency and the bandwidth treatments, we fitted a LM with an arcsine transformation  
245 of the SPCC score for the small differences group, while for large differences group we  
246 fitted an LM with a logistic transformation of the SPCC score. For the duration  
247 treatment, a single model with an arcsine transformation fitted well for the entire range  
248 of acoustic differences. We considered a variable to have a significant impact on the  
249 SPCC score if the 95% confidence intervals (CI) did not overlap with zero.

#### 250 **D. Testing the results with real birdsong**

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

### 268 III. RESULTS

#### 269 A. SPCC response to discrepancies in synthetic sounds

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

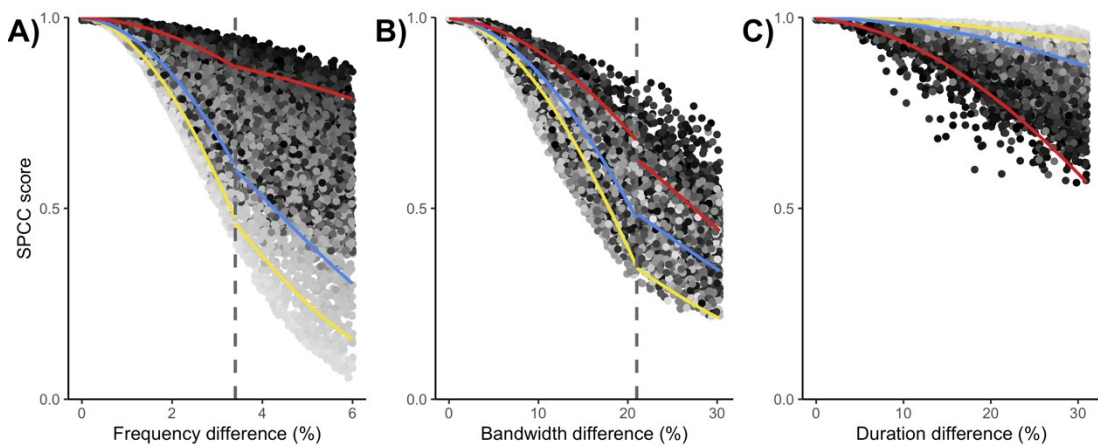
280 In all cases, the SPCC method was sensitive to acoustic discrepancies between  
281 reference-variant pairs, as the SPCC score showed a significant negative correlation with  
282 the acoustic discrepancies in frequency, bandwidth and duration generated between  
283 the reference-variant pairs (Figure 5, Table 1). The SPCC method was most sensitive to  
284 differences in frequency, with a mean decrease of 22% in SPCC score with an increment  
285 of 1% in frequency difference (Figure 5, Table 1). SPCC was less sensitive to differences  
286 in bandwidth as SPCC score decreased by a mean of 4.7% with a 1% increment in  
287 bandwidth differences, and finally, SPCC was least sensitive to differences in duration,  
288 as SPCC decreased by a mean of 1.8% with a 1% increment in the difference in duration  
289 (Figure 5, Table 1). Note that these estimates considered the average change in SPCC  
290 throughout the range of variations. We also found that the SPCC score was influenced  
291 by the bandwidth of the sounds being compared in all treatments, but the direction and

292 size of the effect of bandwidth varied across treatments (Figure 5, Table 1). In the  
293 frequency treatment, where sounds were only different in frequency, the SPCC score  
294 was generally higher if the reference sound had a broad bandwidth than if it was  
295 narrowband sound (Figure 5). This is shown in the model by the positive, significant  
296 impact of model bandwidth and its interaction with variant-reference difference (Table  
297 1). The steeper down slope in the SPCC response for narrowband notes in the frequency  
298 treatment is shown in Figure 5A, with the bandwidth shown by a gray gradient. See also  
299 a visual explanation in Figure 6A-B. For the bandwidth treatment, the impact of  
300 bandwidth was similar to the frequency treatment but smaller (Table 1). In the duration  
301 treatment, the impact of bandwidth was opposite, as the same difference in duration  
302 rendered a higher SPCC score in narrowband sounds (Figure 5B & Figure 5C-D). In  
303 general, shorter sounds rendered higher SPCC scores in all treatments, as shown by the  
304 negative effect of note duration and its interactions with the reference-variant  
305 difference (Table 1). This means that SPCC was less sensitive to acoustic discrepancies  
306 of shorter sounds, although this effect was relatively small. Finally, we see a significant  
307 interaction in all models of both bandwidth and duration with the model-variant  
308 difference (Table 1). This indicates that the impact of bandwidth and duration explained  
309 before is not homogeneous throughout the range of acoustic discrepancies but  
310 increases with increasing acoustic discrepancies. Such effect is represented in Figure 3A-  
311 C as all three lines showing sensitivity for sounds of different bandwidth converge in the  
312 upper left corner.

313 Our detailed quantitative analysis allows to derive the exact sensitivity of SPCC  
314 throughout the range of acoustic discrepancies, while considering the effect of  
315 bandwidth and duration. To derive the exact values one can apply the estimated



316 coefficients using a linear model:  $SPCC = \alpha p + \beta w - \gamma d + \delta(pw) + \psi(pd)$ . P is the  
 317 percentage difference between sounds while the bandwidth and duration of the model  
 318 sound are represented by w and d respectively. Then,  $\alpha$  is the model-variant coefficient,  
 319  $\beta$  is the bandwidth coefficient,  $\gamma$  is the duration coefficient,  $\delta$  is the coefficient for the  
 320 model-variant interaction with bandwidth and  $\psi$  is the coefficient for the model-variant  
 321 interaction with duration. In the models shown in Table 1 and S1, the explanatory  
 322 variables are scaled and center so we can compare the impact of each predictor. In order  
 323 to get the real values for sensitivity we provide the estimates derived from a model with  
 324 the original, non-scaled variables (Tables S2 & S3).



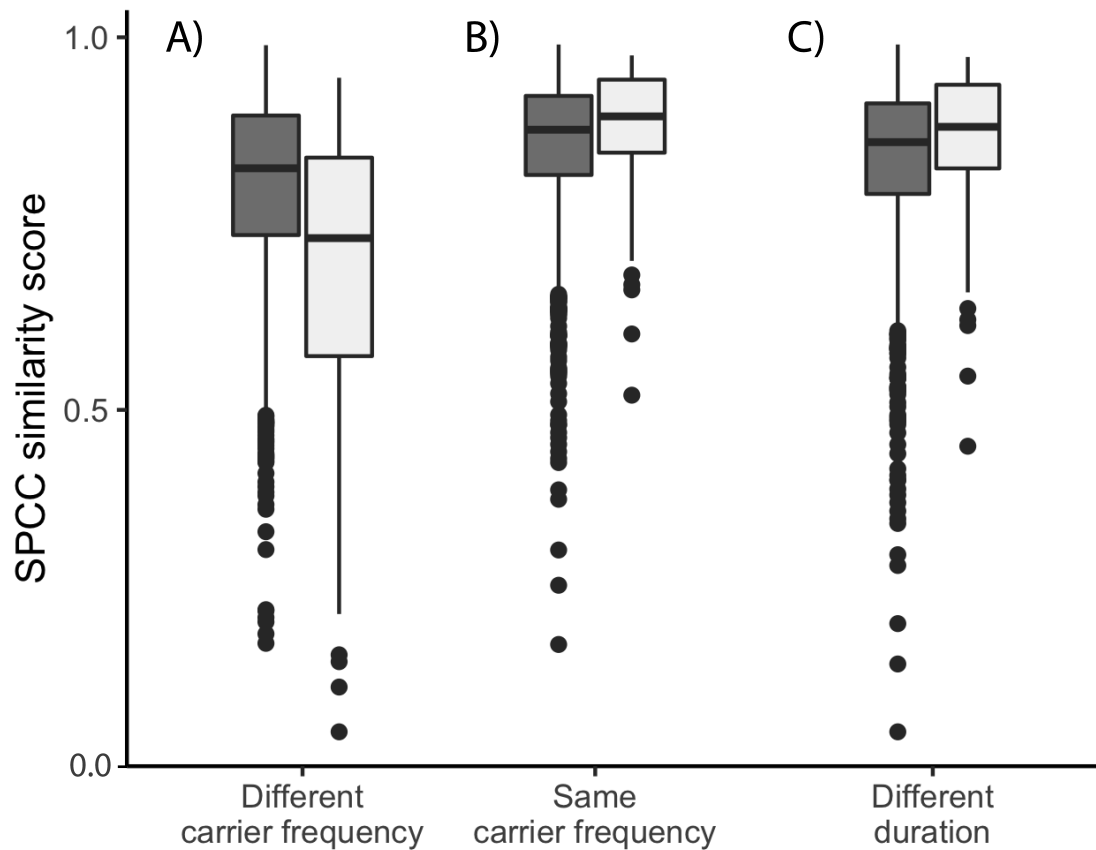
325

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

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

#### 341 **B. SPCC and bandwidth in real birdsong**

342 The birdsong database included 28,266 notes of 3,100 trills in 954 different recordings  
343 from 345 species in 20 families ( $17.3 \pm 13.5$  species per family). As predicted by our  
344 analysis of synthetic sounds, we found that SPCC scores were significantly higher in  
345 broadband notes than in narrowband notes if they differed in frequency (Broadband:  
346  $0.80 \pm 0.11$ , Narrowband:  $0.68 \pm 0.20$  SPCC score,  $W = 13819$ ,  $P < 0.001$ , 5% CI = -0.13,  
347 95% CI = -0.05; Figure 5A and 5A-B) but not if they were similar in frequency (Broadband:  
348  $0.85 \pm 0.09$ , Narrowband:  $0.87 \pm 0.08$  SPCC score,  $W = 39891$ ,  $P = 0.004$ , 5% CI = 0.009,  
349 95% =0.045, Figure 5B and 5A-B). Similarly, analysis of real birdsong confirmed our  
350 findings on the impact of bandwidth on SPCC between sounds of different duration. In  
351 this case, broadband sounds showed significantly lower SPCC scores than narrowband  
352 sounds, for the same difference in duration (Broadband:  $0.84 \pm 0.10$ , Narrowband:  $0.87$   
353  $\pm 0.09$  SPCC score,  $W = 31169$ ,  $P < 0.001$ , 5% CI = 0.017, 95% =0.052, Figure 5C and 5C-  
354 D).



355

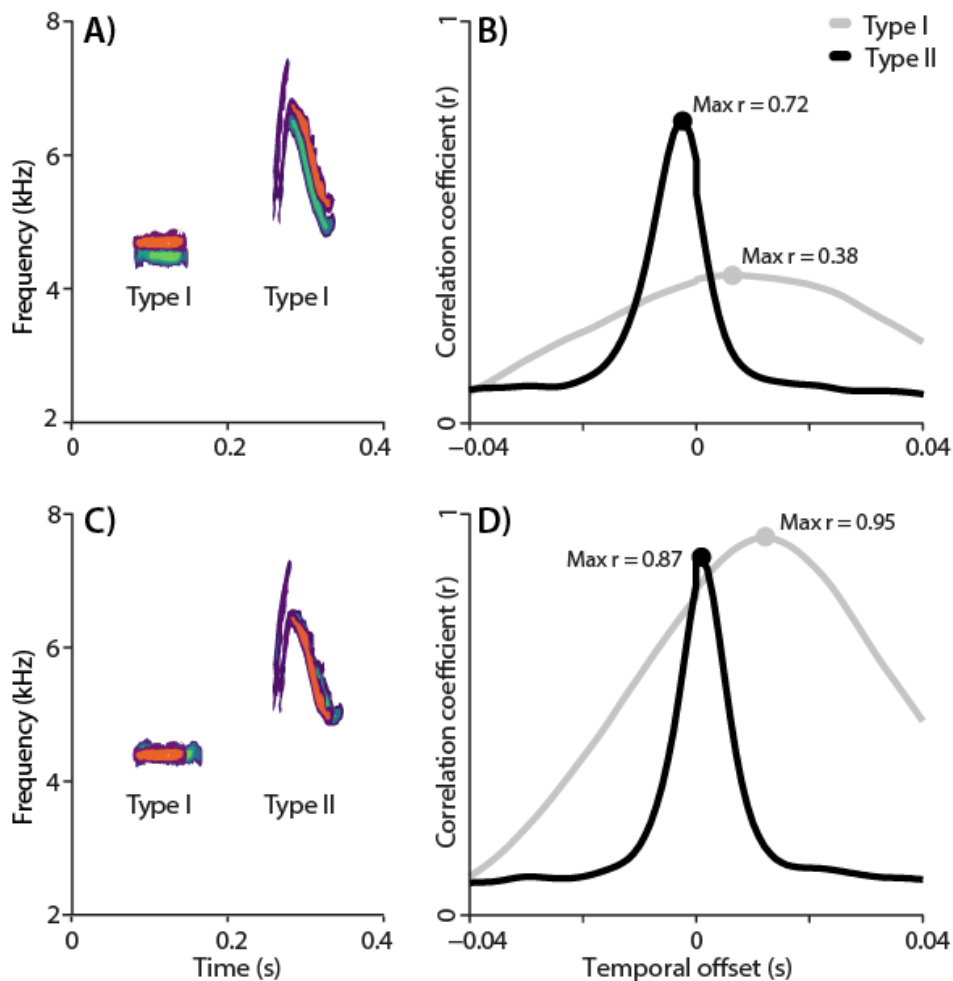
356 **FIG 4** – Differences in SPCC score between broad band sounds (dark grey) and narrow  
 357 band sounds (light grey), measured in natural song of 345 different species. As predicted  
 358 by our analysis of synthetic sounds, SPCC scores of narrow band sounds with different  
 359 frequency, i.e. different F0 but same shape and duration, is lower than in broadband  
 360 sounds (A). However, if frequency is the same, narrow band sounds have higher SPCC  
 361 scores (B). When two narrowband sounds differ in duration (but with the same  
 362 frequency) they show higher SPCC scores than two broadband sounds of different  
 363 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 :<br>Bandwidth | 0.044    | 33.7   | 0.042  | 0.047  | < 0.0001 |
|           | Model-variant difference :<br>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 :<br>Bandwidth | 0.002    | 14.4   | 0.002  | 0.003  | < 0.0001 |
|           | Model-variant difference :<br>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 :<br>Bandwidth | -0.004   | -64.9  | -0.004 | -0.004 | < 0.0001 |
|           | Model-variant difference :<br>Duration  | -0.001   | -22.7  | -0.002 | -0.001 | < 0.0001 |

364

365



366

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

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

#### 383 **IV. DISCUSSION**

384 Our results support the use of SPCC to measure vocal consistency in birds, since the  
385 acoustic similarity score derived from SPCC correlated significantly with the known  
386 acoustic discrepancies between synthetic sounds as found in natural birdsong. As  
387 expected from the optimizing algorithm in SPCC, the SPCC sensitivity to spectral  
388 differences was higher than to temporal differences, when both parameters were within  
389 the range of natural variation in vocal consistency found in birds. The relationship  
390 between SPCC and acoustic discrepancies (sensitivity) was not linear and best fitted an  
391 arcsine curve or a logistic curve. We also found that, in the case of spectral discrepancies  
392 (frequency and bandwidth), the sensitivity of SPCC decreased as the note bandwidth  
393 increased. This means that spectral discrepancies between narrowband sounds were  
394 easier to detect than those in broadband sounds. The opposite pattern was found when  
395 measuring differences in duration. Differences in note duration between broadband  
396 sounds were easier to detect than those in narrowband sounds. In general, shorter  
397 sounds produced higher SPCC scores, suggesting that SPCC is less sensitive when dealing  
398 with shorter sounds. The findings derived from the analysis of synthetic sounds were  
399 confirmed in our analysis of birdsong in 345 different species as: 1) broadband sounds  
400 had lower SPCC than narrowband sounds when notes differed in frequency and 2)  
401 narrowband sounds of different duration had higher SPCC scores than broadband  
402 sounds with the same difference in duration. Quantifying the SPCC response along the  
403 range of acoustic discrepancies found in birdsong allows us to compare the properties  
404 of SPCC with the perceptual abilities found in birds, based on literature. Furthermore,

405 such a quantitative analysis permits researchers to determine the suitability of the  
406 method for their study model and scientific question.

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

417 In birds, frequency discrimination threshold is estimated in 1% (Dooling, 1982). In our  
418 simulated data, the SPCC score of acoustic similarity decreased by 4.4 % when two  
419 sounds of intermediate bandwidth differ by 1% in frequency, strongly supporting this  
420 method to measure the smallest frequency discrepancies perceived by birds. In contrast  
421 with a 1% discrimination threshold for frequency differences, birds are only able to  
422 detect discrepancies in duration when two sounds are at least 14% different in duration,  
423 going up to 23% for short sounds of 100 ms (Maier and Klump, 1990). For a 14%  
424 difference in duration between two sounds, the SPCC similarity score decrease by 3.1%,  
425 again supporting the use of SPCC to assess the smallest temporal differences as  
426 perceived by birds. Hence, the sensitivity of SPCC to detect temporal discrepancies is in  
427 practice relatively similar to the frequency sensitivity if we consider the hearing  
428 capacities of birds (Knudsen and Gentner, 2010). Technically, a lower sensitivity of SPCC

429 to temporal discrepancies is inherent to the method due to the cross-correlation  
430 algorithm. By computing multiple comparisons at different time offsets, the SPCC  
431 maximizes the chances of finding a match (i.e. optimization), while reducing the  
432 sensitivity to temporal discrepancies. However, this step is important to solve the  
433 problem of aligning two time series during their comparison. There are alternative  
434 methods to solve the alignment problem (i.e., Dynamic Time Warping; DTW) but, unless  
435 the optimization acts in the three dimensions (i.e. conducting a second cross-correlation  
436 in the frequency axis), this step will always cause differential sensitivity in the acoustic  
437 similarity score between spectral, amplitude or temporal discrepancies.

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

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



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

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

476 for longer sounds (Maier and Klump, 1990), again similar to our findings that SPCC  
477 sensitivity is higher for longer sounds.

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

497 In conclusion, our results support the use of the SPCC method to measure vocal  
498 consistency in birdsong elements and possibly in other taxa. Our findings further support  
499 multiple field studies that found meaningful correlations between vocal consistency

500 measured by SPCC and individual features or ecological factors. Despite these results in  
501 support of SPCC as a biologically meaningful measure of vocal consistency, there are  
502 some concerns. We found that the sensitivity of SPCC was not linear along the range of  
503 naturally occurring vocal (in)consistency and that sensitivity to detect acoustic  
504 discrepancies is significantly affected by frequency bandwidth. We suggest that these  
505 patterns found in SPCC sensitivity may reflect a similar perceptual pattern in acoustic  
506 discrimination in bird hearing. Further empirical studies are needed to explore bird  
507 perception of vocal consistency and how it is affected by acoustic structure of sound.  
508 Despite this, we recommend caution when comparing absolute values of SPCC scores if  
509 the songs analyzed have different spectral structure (e.g. emitted by different species).  
510 If appropriate, a possible solution would be to normalize or standardize SPCC scores  
511 using statistical techniques to compare vocal consistency. Finally, we highlight the  
512 importance of understanding and validating the methods of measuring song  
513 performance to provide meaningful measures that can be generalized (Cardoso, 2017).

#### 514 **V REFERENCES**

- 515 Allan, S. E., and Suthers, R. A. (1994). "Lateralization and motor stereotype of song  
516 production in the brown-headed cowbird," *J. Neurobiol.* **25**, 1154-1166.
- 517 Ballentine, B., Hyman, J., and Nowicki, S. (2004). "Vocal performance influences female  
518 response to male bird song: an experimental test," *Behav Ecol* **15**, 163-168.
- 519 Bartsch, C., Hultsch, H., Scharff, C., and Kipper, S. (2016). "What is the whistle all about?  
520 A study on whistle songs, related male characteristics, and female song  
521 preferences in common nightingales," *Journal of Ornithology* **157**, 49-60.
- 522 Bates, D. M., Maechler, M., Bolker, B., and Walker, S. (2015). "Fitting Linear Mixed-  
523 Effects models using 'lme4'," *Journal of Statistical Software* **67**, 1-48.

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

527 Botero, C. A., Rossman, R. J., Caro, L. M., Stenzler, L. M., Lovette, I. J., de Kort, S. R., and  
528 Vehrencamp, S. L. (2009). "Syllable type consistency is related to age, social  
529 status and reproductive success in the tropical mockingbird," *Anim. Behav.* **77**,  
530 701-706.

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

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

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

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

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

541 Cortopassi, K. A., and Bradbury, J. W. (2000). "The comparison of harmonically rich  
542 sounds using spectrographic cross-correlation and principal coordinates  
543 analysis," *Bioacoustics* **11**, 89-127.

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

547 De Kort, S. R., Den Hartog, P. M., and Ten Cate, C. (2002). "Diverge or merge? The effect  
548 of sympatric occurrence on the territorial vocalizations of the vinaceous dove  
549 Streptopelia vinacea and the ring-necked dove *S. capicola*," *J. Avian Biol.* **33**, 150-  
550 158.

551 de Kort, S. R., Eldermire, E. R. B., Valderrama, S., Botero, C. A., and Vehrencamp, S. L.  
552 (2009). "Trill consistency is an age-related assessment signal in banded wrens,"  
553 *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **276**, 2315-2321.

554 Dooley, G. J., and Moore, B. C. (1988). "Detection of linear frequency glides as a function  
555 of frequency and duration," *The Journal of the Acoustical Society of America* **84**,  
556 2045-2057.

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

559 Dooling, R. J., Leek, M. R., Gleich, O., and Dent, M. L. (2002). "Auditory temporal  
560 resolution in birds: Discrimination of harmonic complexes," *J. Acoust. Soc. Am.*  
561 **112**, 748-759.

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

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

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

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

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

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

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

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

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

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

583 Langemann, U., and Klump, G. M. (1992). "Frequency discrimination in the European  
584 starling (*Sturnus vulgaris*): A comparison of different measures," *Hear. Res.* **63**,  
585 43-51.

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

589 Ligges, U. (2013). "tuneR—analysis of music."

590 Maier, E. H., and Klump, G. M. (1990). "Auditory duration discrimination in the European  
591 starling (*Sturnus vulgaris*)," *The Journal of the Acoustical Society of America* **88**,  
592 616-621.

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

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

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

600 Nelson, D. A. (1989). "Song frequency as a cue for recognition of species and individuals  
601 in the field sparrow (*Spizella pusilla*)," Journal of Comparative Psychology **103**,  
602 171.

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

605 Nowicki, S., Mitani, J. C., Nelson, D. A., and Marler, P. (1989). "The communicative  
606 significance of tonality in birdsong: responses to songs produced in helium,"  
607 Bioacoustics **2**, 35-46.

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

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

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

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

617 Rivera-Gutierrez, H. F., Pinxten, R., and Eens, M. (2011). "Songs differing in consistency  
618 elicit differential aggressive response in territorial birds," Biol. Lett. **7**, 339-342.

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

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

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

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

629 Sierro, J., de Kort, S. R., Riebel, K., and Hartley, I. R. (2022). "Female blue tits sing  
630 frequently: a sex comparison of occurrence, context, and structure of song,"  
631 Behav Ecol, arXiv:2204.044.

632 Smith, G. T., Brenowitz, E. A., Beecher, M. D., and Wingfield, J. C. (1997). "Seasonal  
633 changes in testosterone, neural attributes of song control nuclei, and song  
634 structure in wild songbirds," Journal of Neuroscience **17**, 6001-6010.

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

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

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



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

644 Vehrencamp, S. L., Yantachka, J., Hall, M. L., and de Kort, S. R. (2013). "Trill performance  
645 components vary with age, season, and motivation in the banded wren," *Behav.*  
646 *Ecol. Sociobiol* **67**, 409-419.

647

648

649 **Supporting information**

650

**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 :<br>Bandwidth | 0.102    | 8.3   | 0.078  | 0.125  | < 0.0001 |
|           | Model-variant difference :<br>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 :<br>Bandwidth | -0.007   | -2.7  | -0.013 | -0.002 | 0.008    |
|           | Model-variant difference :<br>Duration  | -0.004   | -1.5  | -0.01  | 0.001  | 0.13     |

651

652

**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

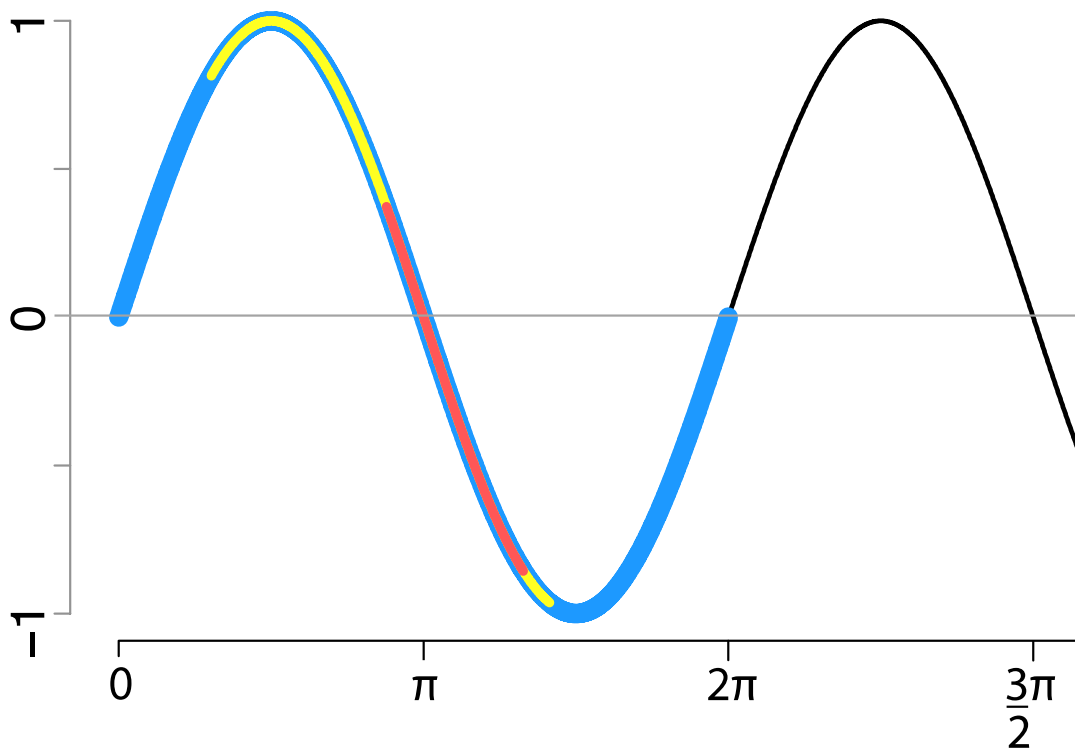
| Treatment | Parameters                              | Estimate | <i>T</i> | CI 5%  | CI 95% | <i>P</i> |
|-----------|---|----------|----------|--------|--------|----------|
| Frequency | Intercept                               | 0.999    | 183.3    | 0.998  | 1      | < 0.0001 |
|           | Model-variant 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 :<br>Bandwidth | 0        | 33.7     | 0      | 0      | < 0.0001 |
|           | Model-variant difference :<br>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 :<br>Bandwidth | 0        | 14.4     | 0      | 0      | < 0.0001 |
|           | Model-variant difference :<br>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 :<br>Bandwidth | 0        | -64.9    | 0      | 0      | < 0.0001 |
|           | Model-variant difference :<br>Duration  | -0.035   | -22.7    | -0.038 | -0.032 | < 0.0001 |

654  
655  
656  
657

**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 | <i>T</i> | CI 5%  | CI 95% | <i>P</i> |
|-----------|---|----------|----------|--------|--------|----------|
| 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 :<br>Bandwidth | 0.09     | 16.5     | 0.079  | 0.101  | < 0.0001 |
|           | Model-variant difference :<br>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 :<br>Bandwidth | 0        | -2.7     | 0      | 0      | 0.008    |
|           | Model-variant difference :<br>Duration  | -0.102   | -1.5     | -0.23  | 0.029  | 0.13     |

660  
661  
662



663  
664 **FIG S1** – The shape of the synthetic sounds was based on a sine shape with a starting  
665 point in  $\sin(x\pi)$  and ending point in  $\sin(y\pi)$ , where  $x$  could be a value between 0 to 1  
666 and  $y$  could be a value from 1 to 2. The entire range is depicted in blue in the figure,  
667 while the yellow and red traces show two possible shapes that would derive from the  
668 process. The selected shape would then be transported so a central frequency of 4.1  
669 kHz and then adjusted to match a randomly selected bandwidth between 0 and 1.64  
670 kHz. In the case of a 0 kHz bandwidth, the resulting sound would be a pure tone,  
671 regardless of the original shape.