Animal Behaviour 125 (2017) e29–e32

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Forum

Advancing the inference of performance in birdsong

Gonçalo C. Cardoso [a, b, *]

[a] CIBIO – Research Centre in Biodiversity and Genetic Resources, University of Porto, Campus Agrário de Vairão, Vairão, Portugal
[b] Behavioural Ecology Group, Department of Biology, University of Copenhagen, Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 18 July 2016
Initial acceptance 21 September 2016
Final acceptance 4 October 2016
Available online 17 January 2017
MS. number: AF-16-00638R

Keywords:
bioacoustics
birdsong
communication
methods
performance

It is appealing to integrate different acoustic traits to infer differences in performance demands among birdsongs, and to use this as a tool for investigating which roles song performance plays in communication. But inferring performance from acoustic measurements introduces a degree of interpretation that can cause disagreement. Here I give an overview of approaches to assess song performance, associated methodological issues, and ways of addressing them. I note advantages and limitations of performance metrics derived from physiological principles or from acoustic trade-offs, discuss issues with the scaling of performance metrics, and with choosing and adapting metrics to different study species and research goals. Throughout I emphasize that these metrics provide tentative assessments of performance, and that empirical results should be interpreted by comparison to alternative hypotheses.

Birdsong is one of the most diverse sexual signals in nature. Many song traits have been correlated with aspects of individual quality, mating success or motivation, and several of those song traits have associated costs or pose performance challenges (Gil & Gahr, 2002). The diversity of song traits, and of ways of combining them, can make it difficult to decide how to quantify birdsong in a way meaningful for communication. In this regard, assessing song performance is appealing because it offers a way of integrating information from many acoustic traits to derive hypotheses on how birds may best evaluate the quality of songs, and doing so in a way that can be customized to species differing in song and in the singing constraints they are subject to (Podos, Huber, & Taft, 2004).

However, assessing song performance often introduces a layer of interpretation and data transformation in between the acoustic measurements used in research and the conclusions taken, which can give rise to disagreement. In this issue, Kroodsma (2017) criticizes the state of the art, raising issues with the empirical evidence and with the methodological rationale of work using the metric ‘vocal deviation’. Here I will focus on methodological issues, both those noted by Kroodsma and additional ones, and give an overview of problems and solutions when assessing performance either using ‘vocal deviation’ (a metric based on the compromise between frequency modulation and rate of syllable repetitions; Podos, 1997, 2001) or other metrics.

SONG PERFORMANCE IS INFERRED AND ASSESSED, NOT MEASURED

Song performance refers to the degree of challenge to the motor system, the respiratory system or other physiological processes involved in singing. Measuring performance would require very precise knowledge on the physiology of singing, such as to translate acoustic differences into quantitative physiological demands. Current knowledge on song production mechanisms is insufficient to do this for complex birdsongs, but suffices to make informed inferences on the directions in which changing song traits may be more or less demanding. For example, longer continuous singing may pose ventilation challenges (Suthers & Zollinger, 2008), louder songs require building larger airsac pressures (Goller & Cooper, 2008), wider frequency modulation should require more movement of the vocal tract (Goller & Cooper, 2008; Suthers & Zollinger, 2008), and two-voiced sounds or precise repetition of syllables are hurdles of neuromotor coordination (Sakata & Vehrencamp, 2012; Suthers & Zollinger, 2008).

But quantifying such acoustic traits, or using a metric that combines more than one trait, does not measure performance. It tentatively assesses (i.e. places a quantitative value on)

* Correspondence: G. C. Cardoso, CIBIO – Research Centre in Biodiversity and Genetic Resources, University of Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal.
E-mail address: gcardoso@cibio.up.pt.

http://dx.doi.org/10.1016/j.anbehav.2016.11.034
0003-3472/

performance, based on the inference that acoustic changes in a certain direction should be more physiologically demanding, at least along part of the quantitative scale. Assessing differs from measuring, in that it places quantitative values indirectly, based on indexes, classifications, judgement, etc. Awareness of the tentative nature of performance metrics is important because it prompts researchers to use them critically, for example by testing the adequacy of a metric to their species, adapting and refining them if possible, and considering alternative interpretations to empirical findings. Most of the proposed solutions to the problems below follow from viewing performance metrics as tentative tools that must be validated by whether they improve empirical insight on communication.

PROBLEMS AND RECOMMENDATIONS

Single Acoustic Traits versus Composite Metrics

Aspects of song performance can be reflected by simple acoustic traits (e.g. song length, sound amplitude, etc.), but many such traits trade-off with each other within songs (e.g. Cardoso & Mota, 2009; Nemeth et al., 2013; Podos, 1997). This means that a high-performance song can either exaggerate a single acoustic trait or use a demanding combination of traits. We therefore face the choice of assessing performance with measurements of a single acoustic trait versus metrics that combine several acoustic traits. The former has the advantages of simplicity and ease of interpretation, but the potential disadvantage of capturing only a small part of performance differences.

Composite metrics have the advantage of improved comprehensiveness, but may have a less straightforward interpretation. This problem is substantially ameliorated using post hoc tests asking whether the results obtained for composite metrics of performance are indeed best interpreted as due to a synergistic effect of different acoustic traits (i.e. results are clearer compared to analyses of individual acoustic traits alone), or can be equally understood with one of those acoustic traits or with simpler metrics. Such tests can ground the interpretation of results, but have rarely been used (Cardoso & Atwell, 2016; Cardoso, Atwell, Ketterson, & Price, 2009; Geberzahn & Aubin, 2014a,b; Podos et al., 2016).

For example, skylarks, *Alauda arvensis*, reacting to a song playback use higher-performance singing, as assessed by 'sound density' (a metric of respiratory performance based on singing long syllables with short intervals; Geberzahn & Aubin, 2014a) and by 'vocal gap deviation' (a metric of motor performance based on the speed of frequency changes during intervals within song; Geberzahn & Aubin, 2014b). These composite metrics appear better tuned to capture performance differences relevant for communication, compared to the individual acoustic traits used by them, because song differences when reacting to playbacks are not detected when analysing separately those individual acoustic traits (Geberzahn & Aubin, 2014a,b).

Finding the Direction of Performance Metrics: Physiological Principles

Metrics of performance derived a priori from physiological principles have the advantage of generality, because many sound production mechanisms are shared among bird species. Examples of this approach are 'song consistency', which quantifies the ability to render accurate repetitions of a song or syllable (Sakata & Vehrencamp, 2012), 'sound density' and similar metrics (e.g. 'percentage peak performance'; Forstmeier, Kempanaers, Meyer, & Leisler, 2002) that assess the relation between the length of sound and intervals within song, or 'frequency excursion', which assesses the rate of frequency modulation (Podos et al., 2016). Many measurements of single acoustic traits with implications for performance also fall in this category.

But, even if some physiological principles are universal, bird species differ so much in song that the challenges they experience should also differ (Podos et al., 2004). Depending on which song traits are typical of a species, some physiological constraints will be limiting, while others are less relevant because they are not approached. For example, high sound amplitude requires building higher airsac pressures (Goller & Cooper, 2008) that are likely more challenging to achieve instantaneously than gradually, and should thus be limited by the brevity of sounds. Accordingly, short syllables or syllables split into multiple elements reach lower amplitudes in the songs of several *Serinus* finches and related species (Cardoso & Mota, 2009). But the trade-off between sound amplitude and number of within-syllable elements weakens and disappears for species with simpler syllables (Cardoso & Mota, 2009), meaning that comparing song performance based on this compromise between sound amplitude and syllable complexity may only be relevant for some species.

Therefore, using metrics of performance based on general physiological principles is no guarantee that those metrics are relevant for a particular study species. This problem is reduced when choosing metrics guided by evidence that a particular aspect of vocal performance should be limiting for a focal species. For example, high sound frequency and fast note rate in great tits, *Parus major*, are associated with occasional disruptions in singing, suggesting that those song traits place performance challenges (Lambrechts, 1997), or swamp sparrows, *Melospiza georgiana*, tutored with trills with increased syllable repetition rates produce variants suggestive that, all else being equal, trill rate limits their performance (Podos, 1996). Less direct guidance for likely relevant aspects of performance may be that a species has an exaggerated song trait. For example, we have looked for communication functions of sound frequency and syllable rate in serins (*Serinus serinus*; Cardoso, Mota, & Depraz, 2007; Funghi, Cardoso, & Mota, 2015) guided by this species having the highest sound frequency and shortest syllable intervals among related finches (Cardoso, Hu, & Mota, 2012; Cardoso & Mota, 2007). Using a metric without evidence that it addresses a limiting aspect of performance for the study species can be attempted, and may be validated by finding that animals use or respond to variation in that aspect of song performance. But negative results are ambiguous to interpret in this circumstance, as they can be due to inadequacy of the metric chosen.

Finding the Direction of Performance Metrics: Acoustic Trade-offs

This type of metric is ambitious in that acoustic trade-offs are used not only to select likely limiting aspects of performance, but also to compute the performance metric itself. Examples of these metrics are 'vocal deviation' (Podos, 2001), 'vocal gap deviation' (Geberzahn & Aubin, 2014b), and metrics based on multidimensional trade-offs of acoustic traits with an aspect of song output (e.g. with the brevity of intervals or sound amplitude; Cardoso, Atwell, Ketterson, & Price, 2007; Cardoso et al., 2009). The rationale is to infer performance trade-offs from the co-distribution of acoustic traits across a large sample of songs, and then assess how individual songs are positioned in relation to these trade-offs. This type of metric can be made comprehensive by integrating effects of many acoustic traits in a multidimensional analysis. Comprehensiveness, however, can be advantageous or not depending on the comparisons intended (see below).

Out of many potential performance limits, this approach focuses on the trade-offs that actually affect singing in a focal