On amplitude and frequency in birdsong: a reply to Zollinger et al

Gonçalo C. Cardoso[a,*], Jonathan W. Atwell[b]

[a]CIBIO—Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Vairão, Portugal

[b]Department of Biology, Indiana University, Bloomington, IN, U.S.A.

Keywords

Zollinger et al. (2012) provided an evaluation of methods to measure sound frequency and amplitude of vocalizations, and a criticism of methods and conclusions in a previous article of ours (Cardoso & Atwell 2011a). Our main finding was that higher amplitude was associated with lower minimum frequency across repetitions of the same syllables or songs in dark-eyed juncos, *Junco hyemalis*. We agree that this discussion of methods is beneficial, and here we further discuss these methods, noting when and why they should be used.

We also found flaws that render the criticisms unwarranted. Thus, here we also explain the adequacy of the methods we used and the correctness of conclusions, both regarding our study system, and the more general conclusion regarding the hypothesis that higher-frequency song in urban birds might be a side-effect of singing with higher amplitude (Nemeth & Brumm 2010). Our response follows the outline of material as presented in Zollinger et al. (2012), using the same main headings.

FREQUENCY MEASUREMENTS

Zollinger et al. (2012) explain three problems when measuring minimum or maximum frequencies from spectrograms. The first is that spectrograms increase the apparent bandwidth of sound traces. Sounds with higher amplitude may be represented with darker and wider traces on spectrograms, which could compromise comparisons of frequency measurements, even within recordings. But, with the spectrogram settings we used, this effect is negligible, at most a change in frequency of 1 pixel (Fig. 1a), and thus, is not a concern for our results (cf. magnitude of differences in Figure 2 of Cardoso & Atwell 2011a). In some circumstances (e.g. high-amplitude sound with low dynamic range in spectrograms) extended shading may appear, which is clearly identified as arte-factual and does not affect comparisons (Fig. 1b).

---

[*]Correspondence: G. C. Cardoso, CIBIO—Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal. gcardoso@cibio.up.pt (G. C. Cardoso).

The second problem is that relative sound amplitude is represented on spectrograms with a colour scale, such that lower-amplitude sounds may not be visible. The extreme frequencies of syllables can have lower amplitude than the middle frequencies and become invisible on spectrograms, especially for the lower-amplitude syllables in a recording, as Zollinger et al. (2012) illustrate in their Figure 2 (reproduced here as Fig. 2a). In most sound analysis software, the upper limit for the colour scale of relative amplitude is adjustable. Researchers usually set this limit not too low, in order to maximize contrast in the amplitude range of interest (usually the higher-amplitude parts of vocalizations), or to produce 'clean' spectrograms with reduced contrast in the amplitudes comprising background noise. This also reduces contrast in the lower-amplitude parts of vocalizations and can create the problem in Fig. 2a. To make valid minimum or maximum frequency measurements from spectrograms, one has to set this limit low enough (i.e. reduce the dynamic range of the spectrogram) to render those parts of the vocalization visible, so that the extreme frequencies reach a plateau, and decreasing the dynamic range further will not reveal more sound in the vocalizations. As noted in our Methods (Cardoso & Atwell 2011a, page 833), we measured frequency after 'having adjusted the dynamic range (greyscale) of each recording's spectrogram (up to 70 dB) to visualize the entire frequency range of syllables'. Although the example in Fig. 2 is the worst scenario for measuring minimum frequency (amplitude and frequency decrease together and gradually), by adjusting the dynamic range of the spectrogram, the visible frequencies converge and are then comparable across renditions of different amplitudes (Fig. 2b, c). Adjusting the dynamic range of spectrograms to prevent the artefact in Fig. 2a is thus an integral part of measuring frequency from spectrograms.

Figure 2 shows another important point. The extreme frequencies measured from spectrograms (in this case 3.1 kHz for minimum frequency) are not necessarily the absolute extreme frequencies of the vocalization. We agree with Zollinger et al. (2012) that the absolute extreme frequency can be an elusive concept. For example, in Fig. 2, the synthesized frequency sweep goes down to 3 kHz with amplitude diminishing gradually until zero. Thus, 3 kHz (the 'true' minimum frequency) would not be possible to measure or perceive even at a short distance. Instead, the minimum and maximum frequencies we measured on spectrograms reflect the frequency information that is available at biologically meaningful distances for communication.

The third issue with respect to frequency measurements relates to comparisons across different recordings: the colour scale of relative amplitude in spectrograms will often not be comparable across recordings when it is set based on peak amplitude, since peak amplitude is affected by noise. We did not compare across recordings (Cardoso & Atwell 2011a), but this could be addressed to a large extent by adjusting the dynamic range of each recording as described above.

Thus, we maintain that our measurements of frequency allow correct comparison across syllables or songs within recordings. Some examples may elucidate the nature of frequency differences in juncos. Figure 1 of our original article (Cardoso & Atwell 2011a) contains three song spectrograms, and previous articles have over 15 additional spectrograms from the same populations (Cardoso et al. 2007, 2008, 2009; Cardoso & Atwell 2011b). Inspection of these spectrograms shows that the frequency bandwidth of syllables within a

song often does not match its relative amplitude (darkness), and that syllable amplitude usually does not decrease gradually at the extreme frequencies, causing the artefact in Fig. 2a. For syllable types that 'fade out' at the extreme frequencies, differences in the visible frequencies are usually associated with subtle changes in syllable shape, indicating that they are not simple omissions of the softer endings. The adjustment of dynamic range described above would disambiguate difficult cases.

Advantages and Limitations of Different Methods

In field recordings, namely from urban habitats, vocalizations can be embedded in background noise of overlapping frequencies, especially in the lower frequencies. In these circumstances, minimum frequency can often be measured from spectrograms, which contain information on the shape of the signal, but not with methods based on amplitude thresholds that only use information from the power spectrum of the vocalization. This is because the cumulative amplitude of noise across the syllable or vocalization more easily surpasses the amplitude of the signal (e.g. Figure 1c, din Cardoso & Atwell 2011a). The signal-to-noise ratio necessary to discriminate the signal is much smaller when measuring frequency from spectrograms, because the signal is discriminated against instantaneous noise (i.e. in the time frames where the signal reaches minimum frequency), not against cumulative noise across the vocalization. This is an important advantage for recordings from urban or other noisy habitats, and explains why the method is commonly used in those cases. Measuring frequency from spectrograms has the disadvantages of being manual and having a degree of subjectivity. These are shared with most techniques of ethological observation, and can be addressed via standard methodological precautions (Martin & Bateson 1986); for example, averaging multiple measurements (in our case, each individual syllable) and observer blindness relative to the hypotheses, to name two that we used.

Measuring minimum and maximum frequency based on amplitude thresholds relative to the peak amplitude of the syllable, as Zollinger et al. (2012) advise, has the advantages of being automatic, objective and fast. But it also has limitations and should not be viewed as the preferred method in all cases. In an early application of this method to birdsong using commercial and archival recordings, Podos (1997) chose a threshold of −24 dB relative to peak amplitude of the syllable. This is an excellent threshold because it captures practically all sound energy in vocalizations, but it requires very high signal-to-noise ratios. We have used this method for other material: when using commercial recordings, we used the same −24 dB threshold and excluded only very few recordings (<1% in Hu & Cardoso 2009; none in Cardoso 2010), and when we selected archival recordings based on sound quality (still very high signal-to-noise ratios, although not as consistently high as in commercial recordings), we settled for a threshold of −15 dB as a compromise with recording quality (Cardoso & Hu 2011), which is still very inclusive. In field studies, however, compromises would generally have to be more severe, and in the literature, thresholds as low as −3 dB have been used (Sockman 2009). This still provides indexes of frequency bandwidth for simple signals but, when the focus of the work is specifically on minimum frequency, one may not want to compromise.

An alternative suggestion was to use algorithms that extract frequency contours of the syllables in spectrograms and obtain the extreme frequencies of those contours (Zollinger et al. 2012). This has the same strength as that of measurements from spectrograms (using information on the shape of the signal) and would offer additional advantages of objectivity and automation. We agree. But in practice this requires an equal amount of human intervention to check that the limits of the extracted contour are consistent, rather than some contours being incomplete due to sound amplitude dropping below noise. In that case, an artefact similar to the one illustrated in Fig. 2awould apply, and quality checks (identical to the adjustment of dynamic range described above) would be necessary.

AMPLITUDE MEASUREMENTS

At several points in their article, Zollinger et al. (2012) allude to the need to measure absolute amplitude of sound, as opposed to relative amplitude, to draw valid conclusions. This is not so. In our case, amplitude differences were sufficient to conclude that minimum frequency and amplitude of syllables within a recording covaried negatively. Absolute amplitude would be useful to investigate additional aspects of this finding (e.g. whether covariation is mostly due to a certain region of the amplitude range; see Cardoso & Atwell 2011a, page 835, second paragraph), but it is not necessary to determine the direction of covariation. Although phrased in connection with the absolute versus relative amplitude dichotomy, the single objection to amplitude measurements put forward by Zollinger et al. (2012) is unrelated to this dichotomy. Their objection is inaccuracy of measurements due to changes in recording conditions (e.g. head movements). Since we compared amplitude in a song or bout of the same perched bird, a single calibration per comparison would convert all relative measurements into absolute measurements, leaving this type of within-recording inaccuracy unaffected. Measuring amplitude can be affected by wind and air turbulence that introduce fluctuations in sound propagation (especially at long distances: Richards & Wiley 1980; Brenowitz 1982), and by changes in head orientation that affect recorded amplitude of directional sounds (especially at short distances: Brenowitz 1982; Patricelli et al. 2008).We anticipated, assessed and dealt with this accuracy issue in three ways.

First, inaccuracies in measurements of amplitude are random relative to the hypotheses tested, and only make tests conservative. Regarding this, Zollinger et al. (2012) noted that it is incorrect to use a conservative analysis in which lack of support for a hypothesis is used to reject that hypothesis. We did reject the hypothesis that increased amplitude explains the higher minimum frequency of urban songs in juncos, but based on support for the opposite of its prediction (negative rather than positive covariation in amplitude and minimum frequency), not based on lack of support for the hypothesis (in which case the conclusion would depend on demonstrating that we had the accuracy and statistical power to detect its prediction). Second, the opportunity for changes in recording conditions is much smaller within songs (duration of 1–2 s) than among songs (we measured sequences of five songs, which take, on average, half a minute to sing; Titus 1998). We used this difference to assess whether measurement error affected our analyses meaningfully, in which case comparisons among songs would be more conservative and show weaker covariation of amplitude and frequency. We found no evidence for such an effect of measurement error (Cardoso & Atwell 2011a). Third, our amplitude differences fit the expectations for birdsong (e.g.

Brumm 2004; Anderson et al. 2008), and for ease of interpretation we excluded the very few outliers above 14 dB (a large but normal amplitude difference). Zollinger et al. (2012) expressed scepticism about our amplitude measurements because of those outliers. Outliers were not unexpected. Absolute amplitude in juncos ranges from around 43 dB at 1 m for 'short-range song' to 80–85 dB for 'long-range song' (Nolan et al. 2002), a 40 dB interval. Similarly, in another emberizid, the song sparrow, *Melospiza melodia*, amplitudes were measured from about 50 to 85 dB at 1 m, a 35 dB difference (Anderson et al. 2008), including 'soft' and 'broadcast' song (usually named 'short-range' and 'long-range' song in juncos; Titus 1998; Nolan et al. 2002; Reichard et al. 2011). Despite these categorical names, amplitude varies continuously through the range (Anderson et al. 2008), low-amplitude songs can either be structurally distinct or be the long-range songs sung softly (Nolan et al. 2002; Anderson et al. 2008; Reichard et al. 2011), and occasionally birds change between high- and low-amplitude songs in a song bout (Nolan et al. 2002; Anderson et al. 2008). The most extreme outlier in our study (a 48 dB difference) was larger than previously reported differences in amplitude, which was expected since we had a much larger sample (over 1000 recordings) and selected the recording with the largest amplitude difference for each bird.

Overall, we find Zollinger et al.'s (2012) criticisms of amplitude measurements misplaced. Zollinger et al. (2012, page e6) stated that they 'do not mean to dismiss entirely the validity of measures of relative amplitude' but attempt to draw a line between comparing adjacent syllables, which they deem valid, versus comparing other syllables in a song or songs within a bout, as we did. It is at most a fragile position. Sources of error such as those described cannot be eliminated fully (with calibrated or uncalibrated recordings) but are unrelated to the song traits themselves, and are easily assessed and accommodated statistically.

ISSUES WITH TERMINOLOGY

We agree that terminology should have been more precise. We explained that we measured differences in relative amplitude, but then we often used 'loudness' or 'louder' (in the colloquial sense) interchangeably, including in the figure axes, where this should not be done.

ISSUES WITH DATA INTERPRETATION

Zollinger et al. (2012) argued four ideas about data interpretation. The first is that voluntary control of minimum frequency and involuntary positive covariation with amplitude might coexist, with the latter only detected when pushing vocal production to extremes. We agree with this possibility (see Cardoso & Atwell 2011a, page 835, second paragraph), but our data comprised both small and large amplitude differences and, opposite to Zollinger et al.'s suggestion, we found that minimum frequency decreased the most for larger increases in amplitude.

The second idea is similar: oscines have physiological mechanisms that control frequency independently of amplitude, but this need not imply absence of a more basic physiological link between amplitude and frequency (as more easily observed in nonoscines). We agree

and did not suggest otherwise, but rather provided a similar explanation (see Cardoso & Atwell 2011a, page 831, second paragraph, and page 835, fourth paragraph). The question, then, is whether the suggested positive covariation with amplitude is a relevant explanation for the observed changes in oscine song minimum frequency. This is an empirical question that we tested as such.

Third, Zollinger et al. contested our statement that there is 'no evidence that higher-frequency song types are louder' (Cardoso & Atwell 2011a, page 835). We agree with part of this criticism and amend our statement here (see Appendix).

Fourth, Zollinger et al. argued that we cannot conclude from our results whether the increased song frequency of urban birds is adaptive. This is true. But, from our results with juncos (negative covariation of amplitude and minimum frequency), we only concluded that 'louder singing does not explain the higher minimum frequency of urban dark-eyed juncos' (Cardoso & Atwell 2011a, page 831, abstract). Then, as stated in our Introduction, we moved on to discuss 'oscine vocal physiology and the vocal behaviour of urban birds, to evaluate whether the observed differences in urban birdsong frequency are more likely to be explained as by-products of increased amplitude, or as functional adjustments in response to noise' (Cardoso & Atwell 2011a, page 832). There, we did conclude that observed differences in urban birdsong are best explained as functional adjustments to noise. This was based on five arguments: (1) we found weak covariation between frequency and amplitude, which leaves much room for independent adjustments; (2) the mechanism of frequency control in oscines suggests the same; (3) the vast majority of urban bird species increase minimum frequency, not the frequency of whole songs, as predicted by the amplitude side-effect hypothesis (or other alternative hypotheses that Zollinger et al. 2012 mention); (4) differences in frequency can be long term; (5) species with higher-frequency vocalizations are, on average, less impacted ecologically by anthropogenic noise. All arguments except the first are based on the literature (see references in Cardoso & Atwell 2011a). We summarized this correctly at the end of the abstract (Cardoso & Atwell 2011a, page 831): 'We discuss oscine vocal physiology and details of the behaviour of urban birds, both of which we argue are consistent with the increased frequency of urban birdsong generally being a functional adjustment to noise, rather than a consequence of singing louder'.

CONCLUSIONS

We agree that it is useful to discuss methods for measuring frequency and amplitude of animal vocalizations, but we disagree with the prescriptive evaluations of Zollinger et al. (2012). Comparing minimum or maximum frequencies from measurements on spectrograms is correct if the necessary precautions are taken, and it has advantages for recordings that contain noise, as typical of studies in urban environments. The need for absolute over relative measurements of amplitude depends on the analysis intended. It would be detrimental to the field if comparing relative amplitude were perceived as undesirable, as it can be used to test a variety of hypotheses, and it is readily obtainable from many animal species in the wild.

In the preceding section we also reiterate the bases for the conclusions in our original article. Note that we did not take issue with the main finding of Nemeth & Brumm (2010), that increasing amplitude overcomes noise more efficiently than increasing frequency. This is not at odds with increased minimum frequency being a direct response to noise because, while it may afford lower benefits than increasing amplitude, its cost is likely to be smaller (Cardoso & Atwell 2011a; Slabbekoorn et al. 2012). We only disagreed with Nemeth & Brumm's (2010) suggestion that the increased frequency of urban songbirds could be reinterpreted as a side-effect of increased amplitude. For reasons stated above, this suggestion might only apply to a minority of species.

Acknowledgments

References

Amador A, Goller F, Mindlin GB. Frequency modulation during song in a suboscine does not require vocal muscles. Journal of Neurophysiology. 2009; 99:2383–2389.

Anderson RC, Searcy WA, Peters S, Nowicki S. Soft song in song sparrows: acoustic structure and implications for signal function. Ethology. 2008; 114:662–676.

Beckers GJL, Suthers RA, ten Cate C. Mechanisms of frequency and amplitude modulation in ring dove song. Journal of Experimental Biology. 2003; 206:1833–1843. [PubMed: 12728005]

Brenowitz EA. The active space of red-winged blackbird song. Journal of Comparative Physiology. 1982; 147:511–522.

Brumm H. The impact of environmental noise on song amplitude in a territorial bird. Journal of Animal Ecology. 2004; 73:434–440.

Cardoso GC. Loudness of birdsong is related to the body size, syntax and phonology of passerine species. Journal of Evolutionary Biology. 2010; 23:212–219. [PubMed: 19888936]

Cardoso GC, Atwell JW. On the relation between loudness and the increased song frequency of urban birds. Animal Behaviour. 2011a; 82:831–836.

Cardoso GC, Atwell JW. Directional cultural change by modification and replacement of memes. Evolution. 2011b; 65:295–300. [PubMed: 20722726]

Cardoso GC, Hu Y. Birdsong performance and the evolution of simple (rather than elaborate) sexual signals. American Naturalist. 2011; 178:679–686.

Cardoso GC, Mota PG. Song diversification and complexity in canaries and seedeaters (*Serinus* spp.). Biological Journal of the Linnean Society. 2007; 92:183–194.

Cardoso GC, Mota PG. Loudness of syllables is related to syntax and phonology in the songs of canaries and seedeaters. Behaviour. 2009; 146:1649–1663.

Cardoso GC, Atwell JW, Ketterson ED, Price TD. Inferring performance in the songs of dark-eyed juncos (*Junco hyemalis*). Behavioral Ecology. 2007; 18:1051–1057.

Cardoso GC, Mamede AT, Atwell JW, Mota PG, Ketterson ED, Price TD. Song frequency does not reflect differences in body size among males in two oscine species. Ethology. 2008; 114:1084–1093.

Cardoso GC, Atwell JW, Ketterson ED, Price TD. Song types, song performance, and the use of repertoires in dark-eyed juncos (*Junco hyemalis*). Behavioral Ecology. 2009; 20:901–907.

Cardoso GC, Hu Y, Mota PG. Birdsong, sexual selection, and the flawed taxonomy of canaries, goldfinches and allies. Animal Behaviour. 2012; 84:111–119.

Clement, P, Harris, A, Davies, J. Finches and Sparrows. Princeton, New Jersey: Princeton University Press; 1993.

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

Dunning, JB. CRC Handbook of Avian Body Masses. Boca Raton, Florida: Taylor & Francis; 2008.

Goller, F, Cooper, BG. Peripheral mechanisms of sensorimotor integration during singing. In: Zeigler, HP, Marler, P, editors. Neuroscience of Birdsong. Cambridge: Cambridge University Press; 2008. 99–114.

Halfwerk W, Slabbekoorn H. A behavioural mechanism explaining noise-dependent frequency use in urban birdsong. Animal Behaviour. 2009; 78:1301–1307.

Hu Y, Cardoso GC. Are bird species that vocalize at higher frequencies preadapted to inhabit noisy urban areas? Behavioral Ecology. 2009; 20:1268–1273.

Lambrechts, MM. Organization of birdsong and constraints on performance. In: Kroodsma, DE, Miller, EH, editors. Ecology and Evolution of Acoustic Communication in Birds. Ithaca, New York: Cornell University Press; 1996. 305–320.

Martin, P, Bateson, P. Measuring Behaviour: an Introductory Guide. Cambridge: Cambridge University Press; 1986.

Nelson BS. Avian dependence on sound pressure level as an auditory distance cue. Animal Behaviour. 2000; 59:57–67. [PubMed: 10640367]

Nemeth E, Brumm H. Birds and anthropogenic noise: are urban songs adaptive? American Naturalist. 2010; 176:465–475.

Nemeth E, Zollinger SA, Brumm H. Effect sizes and the integrative understanding of urban bird song: a reply to Slabbekoorn et al. American Naturalist. 2012; 180:146–152.

Nolan, V, Ketterson, ED, Cristol, DA, Rogers, CM, Clotfelter, ED, Titus, RC, Schoech, SJ, Snajdr, E. Dark-eyed junco (Junco hyemalis). In: Poole, A, Gill, F, editors. The Birds of North America. No. 716. Philadelphia/Washington, DC: Academy of Natural Sciences/American Ornithologists' Union; 2002.

Patricelli GL, Dantzker MS, Bradbury JW. Acoustic directionality of red-winged blackbird (Agelaius phoeniceus) song relates to amplitude and singing behaviours. Animal Behaviour. 2008; 76:1389–1401.

Podos J. A performance constraint on the evolution of trilled vocalization in a songbird family (Passeriformes: Emberizidae). Evolution. 1997; 51:537–551. [PubMed: 28565357]

Podos J, Huber SK, Taft B. Bird song: the interface of evolution and mechanism. Annual Review of Ecology, Evolution and Systematics. 2004; 35:55–87.

Reichard DG, Rice RJ, Vanderbilt CC, Ketterson ED. Deciphering information encoded in birdsong: male songbirds with fertile mates respond most strongly to complex, low-amplitude songs used in courtship. American Naturalist. 2011; 178:478–487.

Richards DG, Wiley RH. Reverberations and amplitude fluctuations in the propagation of sound in a forest: implications for animal communication. American Naturalist. 1980; 115:381–399.

Ritschard M, Brumm H. Effects of vocal learning, phonetics and inheritance on song amplitude in zebra finches. Animal Behaviour. 2011; 82:1415–1422.

Slabbekoorn H, Yang X, Halfwerk W. Birds and anthropogenic noise: singing higher may matter. American Naturalist. 2012; 180:142–145.

Sockman KW. Annual variation in vocal performance and its relationship with bill morphology in Lincoln's sparrows, Melospiza lincolnii. Animal Behaviour. 2009; 77:663–671. [PubMed: 20160859]

Titus RC. Short-range and long-range songs: use of two acoustically distinct song classes by dark-eyed juncos. Auk. 1998; 115:386–393.

Wallschläger D. Correlation of song frequency and body weight in passerine birds. Experientia. 1980; 36:412.

Zollinger SA, Podos J, Nemeth E, Goller F, Brumm H. On the relationship between, and measurement of, amplitude and frequency in birdsong. Animal Behaviour. 2012; 84:e1–e9.

Zuccon D, Prŷs-Jones R, Rasmussen PC, Ericson PGp. The phylogenetic relationships and generic limits of finches (Fringillidae). Molecular Phylogenetics and Evolution. 2012; 62:581–596. [PubMed: 22023825]

Appendix

Covariation of Frequency and Amplitude in Repertoires

We stated that there is 'no evidence that higher-frequency song types are louder' (Cardoso & Atwell 2011a, page 835), to which Zollinger et al. (2012) gave valid counterexamples. Some examples are inconclusive (increasing frequency and amplitude in noise can be either physiologically linked or independent responses, and the flexibility of frequency in the Paridae indicates independent control rather than linkage with amplitude), but five of the cited studies did find positive covariation of dominant frequency and amplitude within repertoires (Nelson 2000; Beckers et al. 2003; Goller & Cooper 2008; Ritschard & Brumm 2011; Nemeth et al. 2012). We discussed the first two studies by Nelson (2000) and Beckers et al. (2003), and an additional one by Amador et al. (2009) in our original paper, on oscine calls and nonoscine song, to introduce this possibility for oscine song too. The last two studies by Ritschard & Brumm (2011), on zebra finches, *Taeniopygia guttata*, and by Nemeth et al. (2012), on great tits, *Parus major*, were published after our article and are instructive. Both studies showed positive covariation between amplitude and frequency of notes specifically in the lower frequency range of each species. This evokes the idea that signal performance, in this case amplitude, decreases near physiological limits (Lambrechts 1996). The zebra finch in particular has much lower-frequency songs than expected for its small body size (Wallschläger 1980), and the physiological limit suggested to explain the result was the inadequate resonance frequency of the zebra finch's small vocal tract to match the lower-frequency sounds (Ritschard & Brumm 2011). Different species will sing nearer their lower or upper physiological limits and should show lower performance at the corresponding end of the frequency range (Podos et al. 2004).

Consistent with this, in several finches of the genus *Serinus* (lato sensu, Zuccon et al. 2012), syllables with wide frequency range have higher amplitude (Cardoso & Mota 2009): syllables with wide frequency range must include the middle frequencies where amplitude is putatively higher, while syllables with narrow frequency range will less often include the middle frequencies. Dominant frequency of syllables is also often significantly related to amplitude in *Serinus* finches, but the direction of the linear effects changes from species to species (Cardoso & Mota 2009), perhaps because some species sing closer to their upper or lower physiological limits. It is difficult to evaluate where each species sings relative to its potential frequency range, particularly in this group where song frequency is not related to body size across species and, on the contrary, evolves rapidly with negligible phylogenetic signal (Cardoso & Mota 2007; Cardoso et al. 2012). But we can use the allometry between dominant frequency and body size for passerines at large (Wallschläger 1980) as an approximation to evaluate whether each species uses frequencies above or below its predicted optimum (Fig. A1a). We find a suggestive trend, albeit not significant, in the expected direction: amplitude tends to increase with the frequency of syllables for species with lower song frequency relative to their size and to decrease for species with higher song frequency relative to their size (Pearson correlation: $r_{22} = -0.31$, $P = 0.15$; Fig. A1b).

Thus, contrary to our statement of 'no evidence that higher-frequency song types are louder' (Cardoso & Atwell 2011a, page 835), when looking at different syllables or notes within

oscine song, we have the full range of situations: in some species, syllable amplitude does not covary with frequency (e.g. Cardoso et al. 2007; Patricelli et al. 2008), and in other species, amplitude and frequency show positive or negative covariation (Christie et al. 2004; Goller & Cooper 2008; Cardoso & Mota 2009; Ritschard & Brumm 2011; Nemeth et al. 2012). It is premature to interpret these differences, but the results with zebra finches and *Serinus* spp. are consistent with the idea that signal performance lowers near the physiological limit of each species (Lambrechts 1996; Podos et al. 2004).

The context of our statement was the discussion of how great tits respond to noise: switching to higher-frequency song types, which seemed unrelated to amplitude. The demonstration that higher-frequency notes in great tit song have higher amplitude (Nemeth et al. 2012) does raise the possibility that changing song types is driven by amplitude differences, but this still seems unlikely because great tits also switch to lower-frequency song types when exposed to high-frequency noise (Halfwerk & Slabbekoorn 2009).
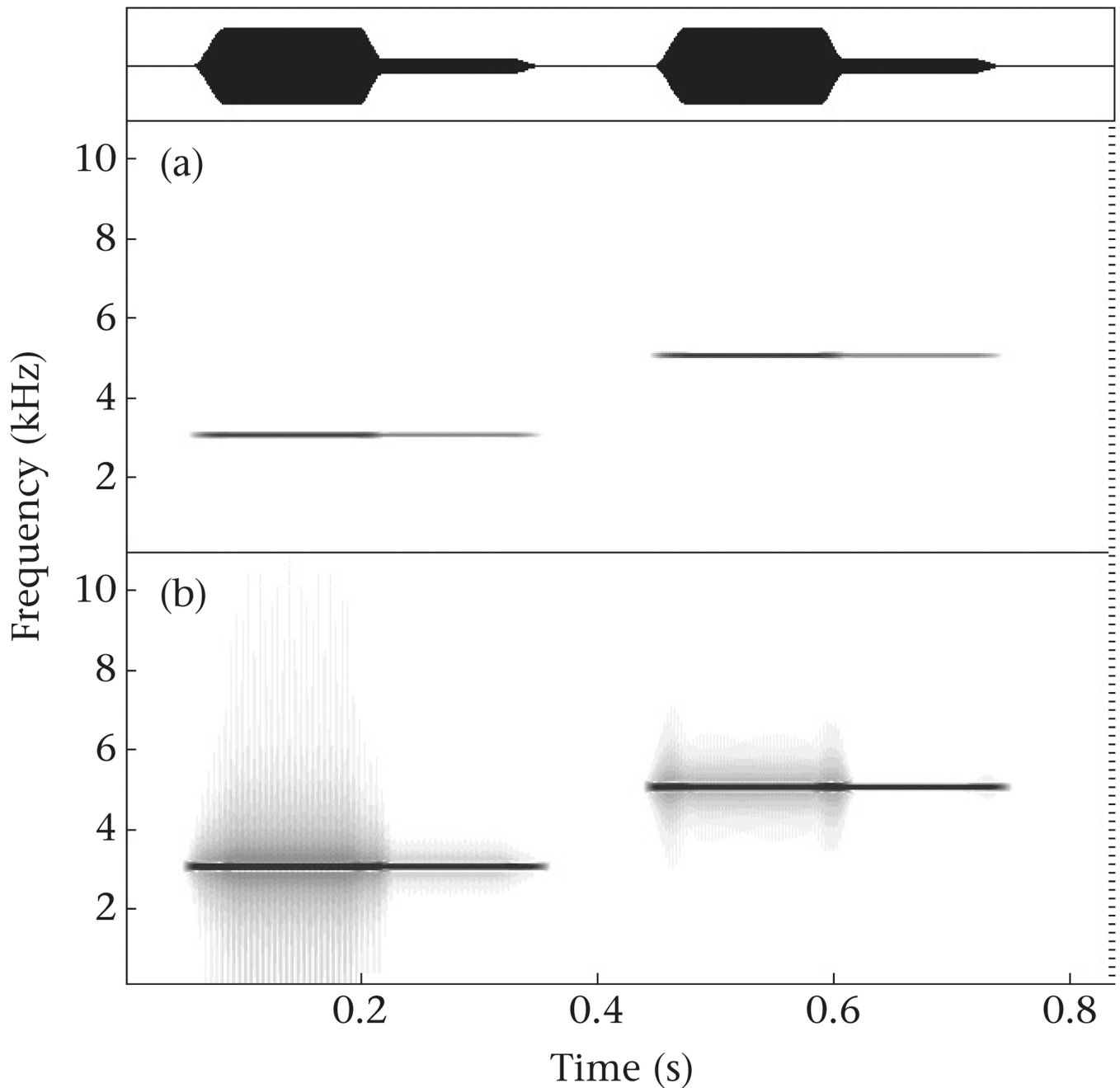
Figure 1.
(a) Spectrogram of pure tones increasing midway by 14 dB, the maximum amplitude differences in our analyses, showing only a negligible change in extreme frequencies as measured from spectrograms. (b) The same pure tone sounds after decreasing the dynamic range of the spectrogram to produce shading around the higher-amplitude sounds. Waveform in the top panel. Pure tones and spectrograms generated with Avisoft SASLab-Pro v.5.2 (Avisoft Bioacoustics, Berlin). Spectrogram settings are the same as those used for measurements and illustrations in our original article: FFT length of 512, hamming window, 100% window duration, 93.75% overlap and asymptotic greyscale of relative amplitude (i.e.

contrast = 'char1.grd' in Avisoft SASLab-Pro) on sound files with 22.050 Hz sampling rate. Black lines at the right side of spectrograms mark 5-pixel intervals.
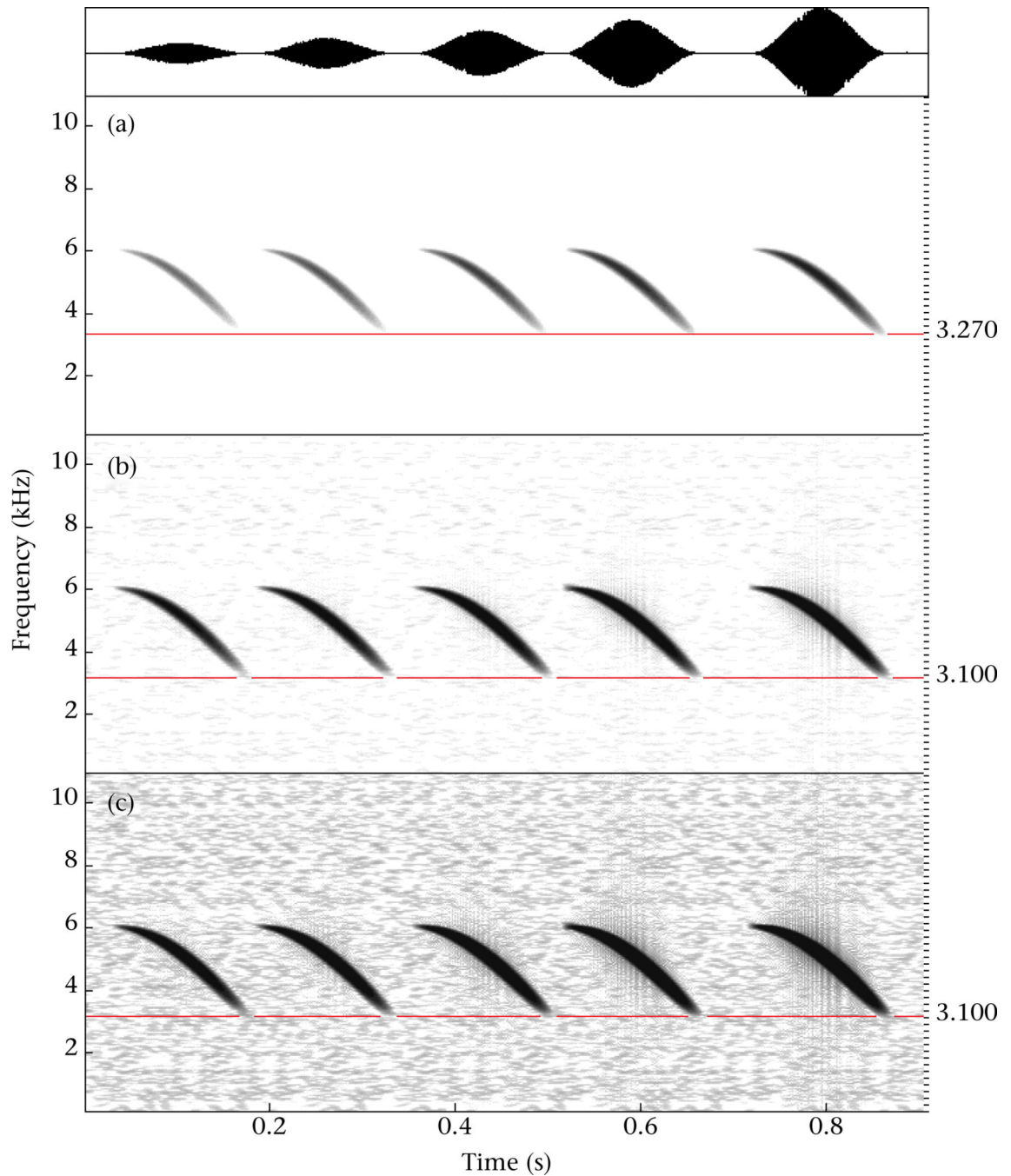
Figure 2.
(a) Spectrogram of five frequency-modulated sweeps with a tapering amplitude envelope and a 3.5 dB difference between sweeps, for a total of 14 dB between the first and last sweep. The sound file is the same as in Figure 2 of Zollinger et al. (2012), but down-sampled to 22.050 Hz sampling rate. Spectrogram settings are the same as those used in our analyses (see Fig. 1 legend), and the spectrogram dynamic range is set to obtain a decreasing visible minimum frequency across sweeps similar to that in Figure 2 of Zollinger et al. (2012). (b) The same spectrogram as in (a) after decreasing its dynamic range; note that the

measurable minimum frequencies of all sweeps converged. (c) The same spectrogram after further decreasing its dynamic range; note that no further sound below the measurable minimum frequency appears. A horizontal reference line is drawn in each spectrogram by the measured minimum frequency of the last sweep. Waveform in the top panel.
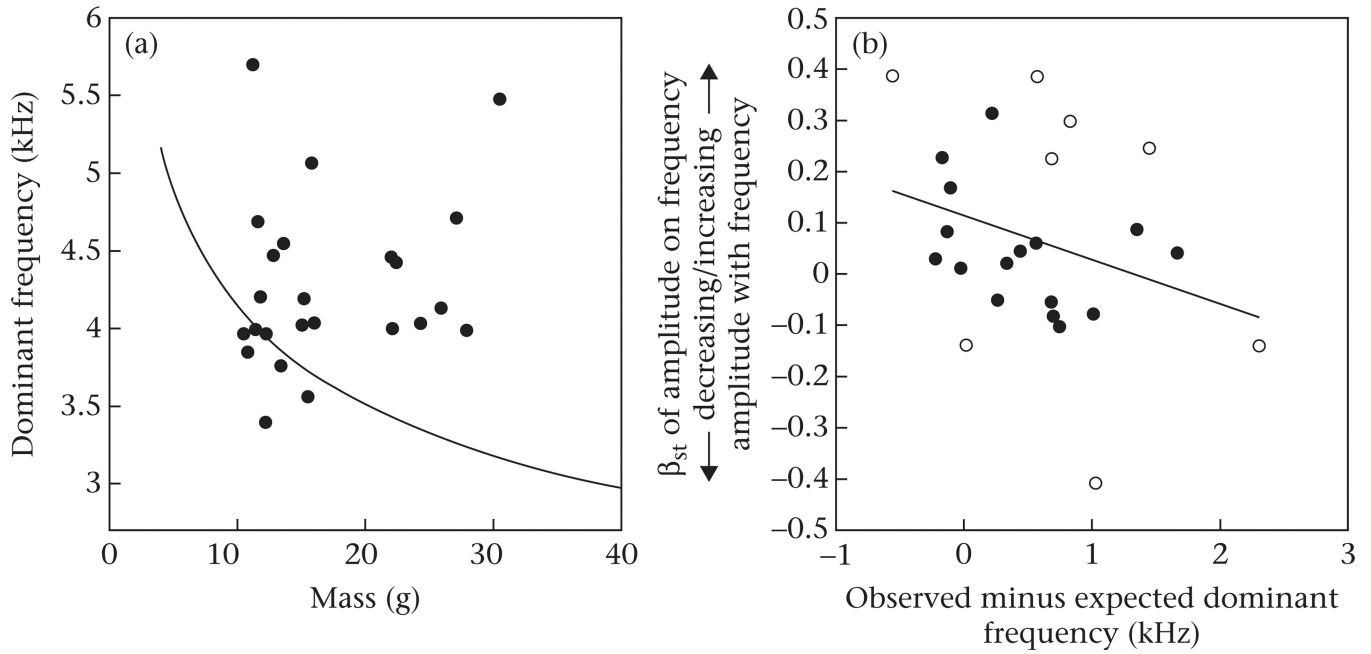
Figure A1.

(a) Dominant song frequency and body mass of *Serinus* spp. (dots) compared with the allometry between those traits for passerines (curved line). (b) Standardized regression coefficients (β_{st}) quantifying the direction and strength of covariation between amplitude and dominant frequency of syllables, plotted against the difference between dominant song frequency and predicted frequency for each species' body mass. The regression line across species is indicated; we report a simple linear regression because phylogenetic signal for song frequency is negligible (Cardoso & Mota 2007; Cardoso et al. 2012). White dots denote species with significant covariation of frequency and amplitude after controlling for multiple comparisons (Cardoso & Mota 2009). Dominant song frequency and standardized regression coefficients from Cardoso & Mota (2009), allometry between body mass and dominant frequency from Wallschläger (1980), and body masses from Dunning (2008), with one missing value substituted by predicted mass using the linear regression equation of mass on body length for the remaining 23 species (lengths for all 24 species from Clement et al. 1993).