Animal Behaviour 84 (2012) e1–e9

Contents lists available at SciVerse ScienceDirect

Animal Behaviour

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

Forum

On the relationship between, and measurement of, amplitude and frequency in birdsong

Sue Anne Zollinger[a,*], Jeffrey Podos[b], Erwin Nemeth[a], Franz Goller[c], Henrik Brumm[a]

[a] *Communication and Social Behaviour Group, Max Planck Institute for Ornithology, Seewiesen, Germany*
[b] *Department of Biology, Graduate Program in Organismic and Evolutionary Biology, University of Massachusetts, Amherst, MA, U.S.A.*
[c] *Department of Biology, University of Utah, Salt Lake City, UT, U.S.A.*

ARTICLE INFO

Article history:
Received 17 November 2011
Initial acceptance 2 January 2012
Final acceptance 17 April 2012
Available online 3 September 2012
MS. number: AF-11-00939R

Keywords:
amplitude
bioacoustics
birdsong
FFT artefacts
frequency
methodology
noise
signal-to-noise ratio
spectrogram

A growing number of studies ask whether and how bird songs vary between areas with low versus high levels of anthropogenic noise. Across numerous species, birds are seen to sing at higher frequencies in urban versus rural populations, presumably because of selection for higher-pitched songs in the face of low-frequency urban noise, or because birds can avoid masking directly by shifting to higher-frequency sounds (Fernández-Juricic et al. 2005; Slabbekoorn & den Boer-Visser 2006; Nemeth & Brumm 2009; Gross et al. 2010; Potvin et al. 2010). In addition to changing song frequency, birds are also reported to respond to increased background noise by singing at higher amplitudes (Brumm & Zollinger 2011). Nightingales, *Luscinia megarhynchos*, for example, sing with a higher sound pressure level in areas with intense traffic noise as compared to quieter locations (Brumm 2004). While frequency- and amplitude-based responses to ambient noise are often considered independently, the two might also vary in tandem because of shared production mechanics. For instance, in phonating birds, increased vocal amplitudes and frequencies might both be achieved by elevated air pressure at the sound source, the syrinx (Beckers et al. 2003; Goller & Cooper 2008; Suthers & Zollinger 2008; Riede et al. 2010; Zollinger et al. 2011). Nemeth & Brumm (2010) recently presented data suggesting that, in the face of low-frequency noise, birds could recover favourable signal-to-noise ratios (SNRs) more efficiently by elevating amplitude rather than frequency. Could it be that birds circumvent environmental noise mainly by elevating amplitude, with frequency occurring as a nonfunctional by-product?

In a recent publication in *Animal Behaviour*, Cardoso & Atwell (2011a) addressed this question by testing the relationship between frequency and amplitude across a broad sample of field-recorded songs. They conducted their analyses on songs of 151 dark-eyed juncos, *Junco hyemalis thurberi*, a sample that has been featured in several prior manuscripts (Cardoso et al. 2007, 2009; Cardoso & Atwell 2011b). In the *Animal Behaviour* study, Cardoso & Atwell (2011a) extracted amplitude and frequency measures from uncalibrated field recordings, and tested for relationships between these measures at two levels: across songs produced within bouts

---

[*] Correspondence: S. A. Zollinger, Communication and Social Behaviour Group, Max Planck Institute for Ornithology, 82319 Seewiesen, Germany.
E-mail address: zollinger@orn.mpg.de (S. A. Zollinger).

http://dx.doi.org/10.1016/j.anbehav.2012.04.026

and within individual songs. For the former, they compared the frequencies of the highest- and lowest-amplitude syllables in each recorded bout, and for the latter, they compared the frequencies of high-amplitude versus low-amplitude syllables within songs. For both analyses, the authors report positive relationships between amplitude and maximum frequency, consistent with the hypothesis that frequency and amplitude should covary. However, the authors also report some negative relationships between amplitude and minimum frequency in both analyses, which runs contrary to the covariance hypothesis. This led the authors to conclude that frequency shifts in populations exposed to anthropogenic noise are best explained as ‘a functional adjustment to noise, rather than a consequence of singing louder’ (Cardoso & Atwell 2011a, page 831). Unfortunately, in our view, the Cardoso & Atwell (2011a) manuscript has several errors in methodology and measurement that make it difficult to eliminate alternative hypotheses for the patterns reported. We believe that these errors undermine the validity of the results reported and the conclusions reached. One particular problem is that one of the variables they tested, amplitude, was not actually measured and thus conclusions about the covariance between amplitude and frequency are probably not meaningful. We present this comment not in confrontation, but rather as a chance to constructively discuss and reflect on how acoustic data are garnered and interpreted, especially with respect to studies on how bird songs may evolve in response to ecological factors.

FREQUENCY MEASUREMENTS

Frequency is perhaps the most common feature used in analyses of acoustic variation. There are several tools and methods for measuring sound frequencies. Two of the more commonly used tools are zero-crossing counts from waveforms (possible for sounds with consistent frequencies), and extraction of peak frequencies or frequency bandwidths from power or amplitude spectra. A third method, used by some researchers including Cardoso & Atwell (2011a), involves extracting frequency measures via cursors or by hand from spectrograms. Spectrograms are excellent and valuable tools for visualizing and describing complex acoustic signals, comparing frequency contours, or scanning through recordings to identify sounds of interest. Spectrograms plot successive slices of a sound as a series of amplitude spectra strung together in sequence. They show how, over time (*X* axis), sounds change in both frequency (*Y* axis) and amplitude (*Z* axis), with the *Z* axis represented either by a colour or greyscale spectrum. In spite of their visual appeal, however, spectrograms express an inherent trade-off between temporal and frequency resolution and also express amplitude variation on a relative scale, in the absence of calibration to known values of sound pressure level (SPL). For both reasons, spectrograms can be challenging tools for extracting accurate, repeatable and unbiased measures of duration, or maximum and minimum frequency. This can undermine their use in comparative analyses of frequency and amplitude, particularly for uncalibrated field recordings. In the case of Cardoso & Atwell’s study, we argue that their main finding reported, that songs with the highest amplitude have the highest maximum and lowest minimum frequencies, might instead be explained as an artefact of their reliance on visual inspection of spectrograms for frequency analyses, and on their use of recordings that lack amplitude calibration.

Even in the early analogue years of spectrogram analysis, problems with accurate measurement of frequency and time parameters were well recognized. Greenewalt (1968) devoted a chapter in the beginning of his classic book on birdsong to spectrograms and on how best to utilize the then-novel tool to visualize and reliably measure animal sounds. Notably, Greenewalt included several cautionary figures illustrating how increases in signal amplitude can exaggerate the frequency range of the spectrographic display, and thus lead to measurement errors such as a false conclusion that frequency bandwidth increases in step with amplitude (Greenewalt 1968). Twenty years later, Beecher (1988) published a thorough description of common spectrogram filter settings and described sources of error that impede accurate measurement of time and frequency from spectrograms. His very clear guide outlines how to select spectrographic settings based on the types of sounds being analysed so as to minimize the inevitable errors introduced by the method. More contemporary textbooks of bioacoustics (e.g. Bradbury & Vehrencamp 1998; Gerhardt & Huber 2002) offer excellent reviews on spectrogram construction, and of how different window and filter sizes and shapes may affect the resulting output. Many modern acoustics and digital signal-processing texts also offer detailed equations, filter functions and solutions to problems encountered in processing and analysing digital sound files, and we encourage curious readers and those new to the bioacoustics field to refer to these comprehensive and clearly written guides (e.g. Greenewalt 1968; Beecher 1988; Smith 1997; Bradbury & Vehrencamp 1998; Gerhardt & Huber 2002).

Returning to the issue at hand, we see at least three problems with the practise of extracting frequency measurements visually from spectrograms. First, depending on the way spectrograms are rendered, sounds can often appear to have greater ranges of frequencies than actually exist in their analogue waveforms. To produce a spectrogram, a waveform is analysed in a series of short time ‘windows’. Each of these windows is analysed using a fast Fourier transform (FFT) algorithm that calculates the frequencies present within that window. Consider a synthesized sine wave with an oscillation frequency of 1000 Hz. While the signal contains only a single frequency, the spectrogram may give the appearance of additional frequencies above and below the 1000 Hz wave, rendered in greyscale spectrograms by black and grey pixels in a horizontal band. The thickness of this band will depend on the amplitude of the sound as well as user-manipulated settings including window size and shape sample rate and FFT resolution. The presence of the band suggests distinct ‘minimum’ and ‘maximum’ frequencies with substantial bandwidth in between (Fig. 1a). However, as we know, the source signal contains only a pure sine tone of exactly 1000 Hz. This singular frequency is, by contrast, revealed with precision in the other acoustic analysis domains: high-resolution power spectra (Fig. 1b) and the waveform itself (Fig. 1d), in which frequency can be calculated through zero-crossings. This is to say, apparent sound energy at frequencies other than 1000 Hz is an artefact of how the spectrogram was constructed.

A second problem is that, within any given spectrogram, it is not possible to compare accurately the frequency content of sounds of different amplitudes. With an increase in amplitude, spectral information becomes visible on a spectrogram that is not visible in a softer rendition of the very same sound (assuming the spectrogram settings remain constant). This can lead to fundamental measurement errors if one attempts to determine frequency bandwidth solely by visual inspection of uncalibrated spectrograms. Figure 2 illustrates a series of the same frequency-modulated (FM) sweeps modelled on a typical dark-eyed junco song syllable. The sweeps have identical maximum frequencies, minimum frequencies and durations, but are repeated at different amplitudes within the file. Using constant spectrogram settings, these sweeps appear to increase in both minimum and maximum frequency as they increase in amplitude (Fig. 2a). By contrast, when minimum and maximum frequency are measured on amplitude spectra, by taking the values at a fixed number of decibels below

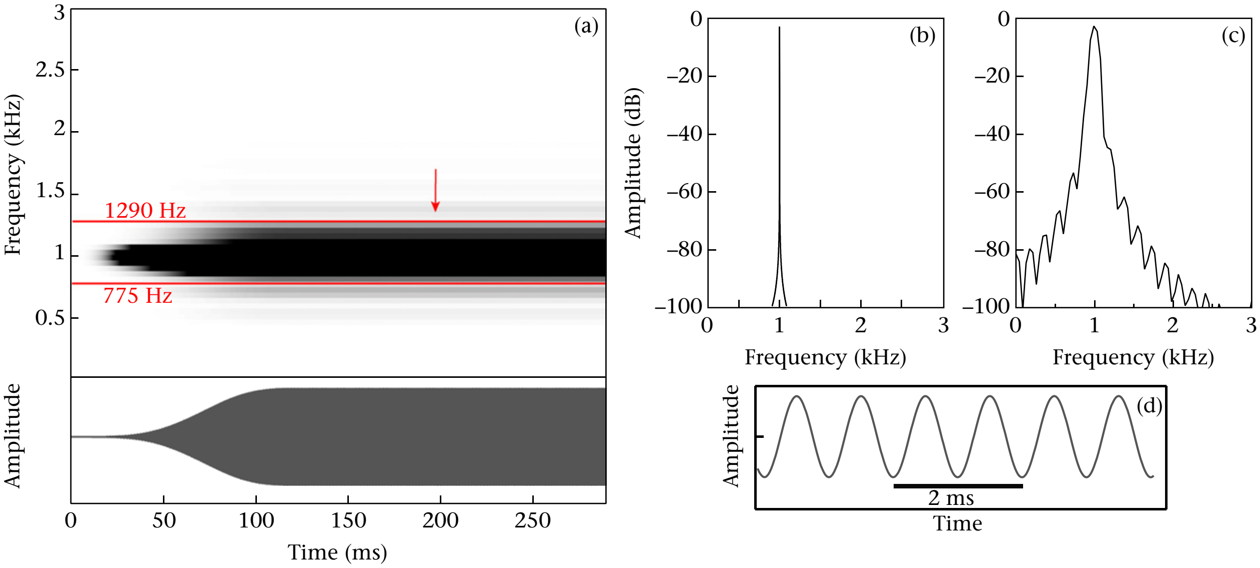

Figure 1. As a by-product of the way they are calculated and drawn, spectrograms can display misleading information about what frequency content is actually present in the analogue signal. A synthesized 1000 Hz sine wave is represented here by a spectrogram (a), power spectra (b, c) and waveform (d). The spectrogram displays the sine wave as a number of dark pixels, the width of which depends on the amplitude of the signal, analysis window size and shape, and contrast display settings chosen by the user. The sound file contains only the one frequency, as is clearly seen under closer inspection of the waveform (d), which consists of one cycle per ms (1000 cycles per second), and by the high-frequency resolution power spectra taken across the entire waveform (0.973 Hz frequency resolution, 1.29 Hz bandwidth). However, at more typical spectrogram settings, a series of individual spectra are calculated across shorter time windows with lower-frequency resolution. In (c), the spectrum from one analysis window used to calculate the spectrogram in (a), taken at the position of the red arrow in the spectrogram. Both the spectrum (c) and the spectrogram were calculated with a 1024-point FFT length, using a Hamming window and 100% window duration, resulting in a 56 Hz bandwidth and 43 Hz frequency resolution. (This procedure gives the equivalent frequency and temporal resolution used by Cardoso & Atwell (2011a), who down-sampled their files to 22.05 kHz and then plotted them as a spectrogram using a 512-point FFT length.) As a by-product of calculating the FFT across the data points contained in each short analysis window, the spectra in each spectrogram bin contains additional frequencies that are not actually present in the original signal. Thus, if using the method employed by Cardoso & Atwell (2011a) to measure a maximum and minimum frequency by visually placing cursors on the apparent upper and lower limits of the sound (red horizontal cursors on spectrogram in (a)), one might falsely conclude that this pure sine tone wave had a bandwidth of more than 500 Hz. The 1 kHz tone was generated digitally using SASLab Pro v.5.1 (Avisoft Bioacoustics, Berlin, Germany) using a sampling rate of 44.1 kHz, with 16-bit depth.

the peak amplitude, the results reflect the fact that the frequency content is identical for all five sweeps, irrespective of SPL (Fig. 2b–d). In the example illustrated (Fig. 2), we chose a measurement threshold for the power spectra based on the syllable in the recording with the lowest SNR (sweep b in the spectrogram, and the spectra in panel (b)), and determined that −20 dB from the peak was as far down as we could measure the bandwidth of the syllable and still be above the noise floor. We then used this threshold to compare bandwidths of all the sweeps in the recording. It is useful to note that both the frequency contour (FM pattern) of the syllable and the shape of the amplitude envelope can affect the magnitude of potential measurement errors. For example, in the cosine-shaped downsweep that we analysed in Fig. 2, the relatively linear FM shape at the syllable offset leads to substantial variation in measured minimum frequency with changing amplitude, whereas the relatively flat FM shape at the syllable onset leads to less severe amplitude-associated inaccuracies in measures of maximum frequency. Overall, we conclude that differences in minimum and maximum frequency observed in spectrograms might simply occur as an artefact of how spectrograms display sounds of varying amplitude, and this provides an alternative, nonbiological explanation for Cardoso & Atwell's (2011a) main findings.

A third problem concerns comparing sounds that are identical in both frequency and amplitude. At constant spectrogram display settings, noises other than the target signals in a sound file, such as constant background noises or transient high-intensity sounds from other animals or abiotic sources, can confound reliable measurements from spectrograms. In typical sound analysis programs, spectrograms encode amplitude values by greyscale or colour scale and display the sounds with the highest amplitude with the darkest colour (in positive greyscale display settings). For this reason, two sounds that are identical in both frequency and amplitude at the source can appear to differ markedly if there is some other, higher-amplitude sound in the background of one of the recordings that is not present in the other. We illustrate this phenomenon in Fig. 3, in which a syllable appears to be much quieter, and thus seems to have a narrower bandwidth, in a spectrographic representation that includes another background sound of higher amplitude than the focal syllable. This problem can be minimized by recording reference sounds of known SPL (measured, for example, in the field with a calibrated SPL meter), and using these reference signals to calibrate the colour scale in the spectrogram, allowing sounds of the same amplitude in different recordings to be rendered as the same colour (or grey tone).

Critically, all three of these errors in frequency measurement can be circumvented by using power spectra rather than spectrograms. From power spectra, minimum and maximum frequency can be measured reliably from the bandwidth at some set number of decibels below the peak amplitude (e.g. Figs 2, 3). This method is well established and is relatively standard practise among bioacousticians studying animal sounds across taxa (e.g. Podos 1997; Fischer et al. 1998; Gerhardt & Huber 2002; Templeton et al. 2005; Siemers & Kerth 2006; DuBois et al. 2009). We see two indications that Cardoso & Atwell (2011a) attempted to validate the accuracy of their spectrogram frequency measures, although we do not find that either offered a compelling solution. First, although not used in their analyses, Cardoso & Atwell illustrate power spectra alongside spectrograms (their Figure 1) to support their contention that 'minimum frequency can be measured visually on spectrograms' when the signal-to-noise ratio of the recording makes it difficult to identify the signal in a power spectrum (their Figure 1 legend). The

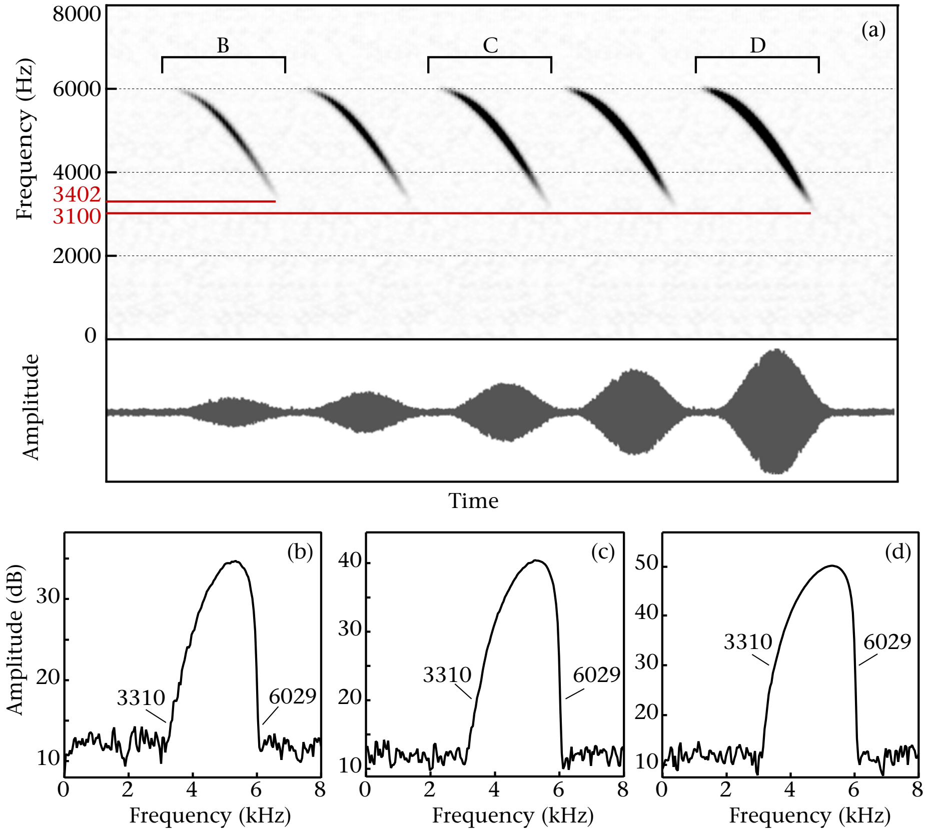

Figure 2. Five frequency-modulated sweeps, modelled after a typical dark-eyed junco song syllable in frequency contour, bandwidth and duration. All of these sweeps are identical in the frequency and time domain (3–6 kHz cosine modulated downsweeps), but vary in amplitude by 3.5 dB increments. Visual inspection of the spectrogram suggests that minimum or maximum frequency of these five sweeps differ dramatically between the lowest- and highest-amplitude renditions. However, all syllables pictured have the exact same minimum and maximum frequency. To further illustrate the magnitude of this problem, we measured by manual placement of cursors on the spectrogram the 'minimum frequency' of the first and last syllables (which differ in amplitude by 14 dB). Using this method, the minimum frequencies appear to vary by about 300 Hz. In contrast, the minimum and maximum frequency values, measured at a standard decibel threshold (here −20 dB) below the peak in power spectra are identical, despite an overall difference in peak amplitude (b–d), calculated across the first, third and fifth syllables in the spectrogram (brackets B, C and D). The magnitude of this problem varies with the pattern of frequency modulation (FM sweeps are more prone to measurement error than constant frequency tones) and amplitude envelope shape (more tapered envelopes will lead to greater errors than ones with more abrupt onsets). Syllables were generated using SASLab Pro v.5.1, following the parameters indicated in the legend of Fig. 1.

horizontal lines they include to demarcate minimum frequency are still placed subjectively, however, given that there is no reference to the maximum amplitude. Recordings may simply be unsuitable for making accurate measures of frequency if the signal is not identifiable from background noise (even after high-pass filtering). Second, Cardoso & Atwell argue in the text that by adjusting spectrogram greyscale settings, one can overcome varying background noise levels in order to 'visualize the entire frequency range of syllables' (Cardoso & Atwell 2011a, page 833). As our examples in Figs 1and 2 illustrate, however, it is simply not possible to determine frequency bandwidth with accuracy by visual inspection of uncalibrated spectrograms, particularly when the sound varies in amplitude. Unless the frequency range or amplitude at the source is known, adjusting the greyscale settings by eye can allow you to 'visualize' either a greater frequency range (e.g. Fig. 1), or a smaller frequency range (e.g. Figs 2, 3) than actually exists in the sound. In addition, we are unable to see how this method could be applied uniformly and objectively across different songs.

We do not mean to suggest that spectrograms are not useful for frequency analysis; on the contrary, spectrograms have their specific utility. For example, spectrograms provide information about how a signal changes over time that is not evident in a power spectrum. As noted by Cardoso & Atwell (2011a), spectrograms are also particularly useful for describing signal frequency contours and timing patterns when signal-to-noise ratios are poor. In such cases, power spectra are often not as useful in detecting signals because their frequency peaks tend to get lost against the background noise, particularly for spectra calculated across a broad temporal window. However, this is not to suggest that field recordings with low SNRs are unusable, but one must understand the limitations of the data that can be reliably measured in such recordings when developing initial hypotheses or interpreting the measurements made from such recordings. Numerous good examples of using spectrograms in analysing frequency contours and timing patterns can be found in the marine mammal vocalization literature. Several research groups have published algorithms for extracting and comparing frequency contours of whale calls in noisy underwater recordings (Buck & Tyack 1993; Deecke et al. 1999; Roch et al. 2011). In these algorithms, signal contours are derived by first finding frequency peaks that exceed a set amplitude threshold for each overlapping analysis window along the duration of the vocalization. By contrast, segments of the signal with particularly low amplitude, for example, call onset and offset, which do not surpass the set detection threshold (i.e. cannot be reliably discriminated from background noise) are discarded. Thus, rather than trying to extrapolate data from the extremes of their signals (e.g. maximum and minimum frequencies or onset and offset times), which

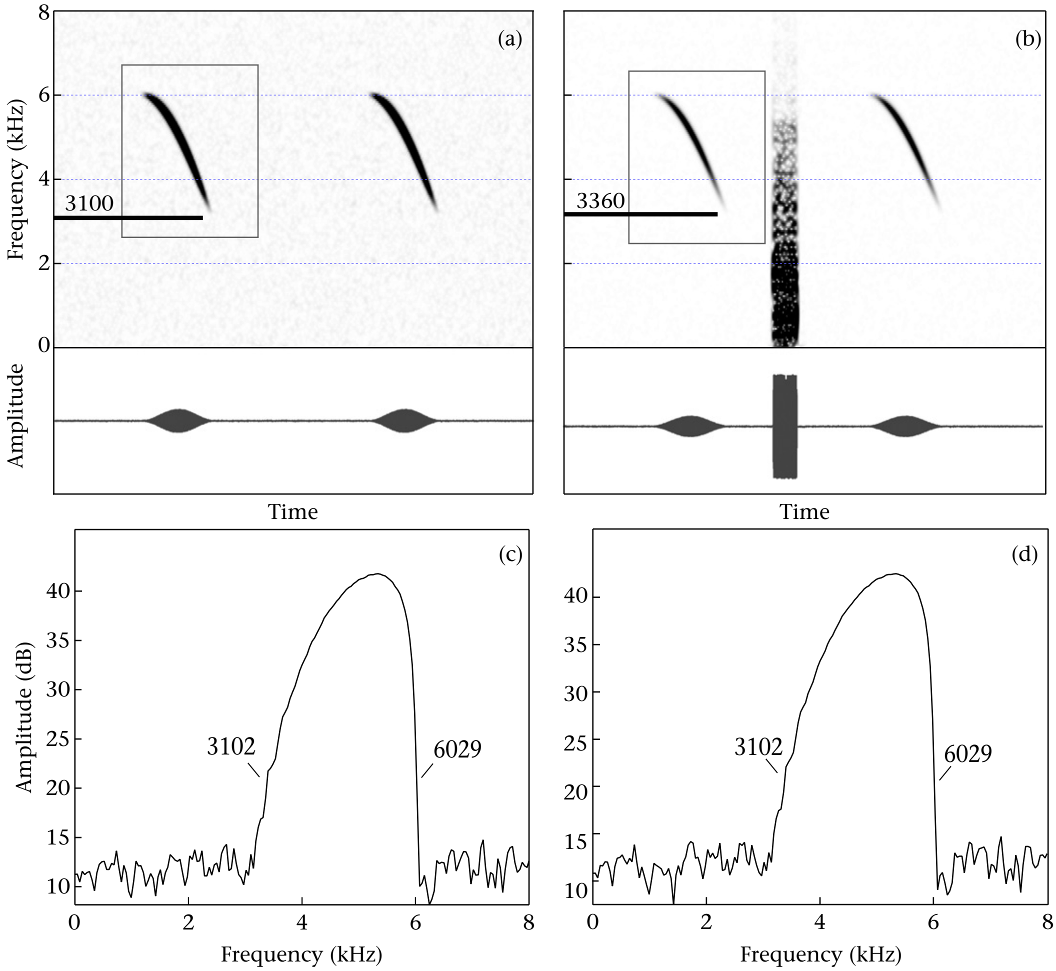

Figure 3. Four frequency-modulated sweeps identical in both frequency and amplitude, illustrated by spectrograms (a, b) and power spectra (c, d) drawn from the first syllables in (a) and (b) (boxed areas). However, in (b), a brief loud sound (a short blast from a car horn) occurred between the two FM sweeps. Calibration of the spectrogram greyscale to the high-amplitude background sound gives the false impression that the sweeps are quieter and have a higher minimum frequency in (b) than in (a), despite both being plotted with identical spectrogram display settings. Thus, background noises with high SPL values, including continuous high-intensity noises such as from traffic or wind, or transient noises such as other animal sounds or abiotic sources, will affect the relative amplitude scaling and can lead to profound measurement errors. For details of syllable generation and spectrogram settings see Fig. 1 legend.

cannot be accurately measured, these studies focus on data that can be extracted (i.e. the timing and shape of frequency contours).

A similar approach could have been used to compare frequency content between different renditions of junco song syllables by Cardoso & Atwell (2011a), if rather than focusing on minimum and maximum frequency (measurements that are often difficult to extract from noisy recordings), they looked for frequency shifts across the syllable contours that fell above a detectable threshold. That said, since they did not actually measure amplitude, it would still not be possible to draw conclusions about a covariance between frequency and amplitude.

AMPLITUDE MEASUREMENTS

Measures of amplitude are ideally taken using calibrated equipment such as sound pressure level meters or measuring microphones. Calibration can also be achieved, although without as much precision, through regular recordings of reference sounds of known SPL. Yet in the absence of calibrated recordings, the amplitude (voltage) of recordings has little meaning, because amplitude tends to vary in accordance with a broad range of factors besides the actual amplitude of the source. Such factors include the distance between the animal and the microphone, the orientation of the animal relative to the microphone, environmental conditions such as wind, reflections from the ground or other surfaces, the type and sensitivity of the microphone, and the gain settings and properties of the audio recorder.

Cardoso & Atwell's (2011a) sample made use of noncalibrated recordings, which disallows direct measures of amplitude. As a work-around, the authors decided to focus on relative measures of amplitude, either across songs (presumably separated in time by seconds or minutes), or across syllables within song (temporal separation of $<\sim 2$ s). These comparisons presume that external factors (listed above) that might otherwise alter amplitude values either (1) do not vary enough to obscure actual amplitude modulations by the vocalizing animals, or (2) do not vary in a way that would bias the test of the hypothesis at hand.

We find it hard to accept this first presumption at face value. To support this presumption, the authors state that their recordings were made with 'the microphone in a static position, so that the distance to the bird remained the same during the bout and the relative sound amplitude of songs in the recording could be compared' (Cardoso & Atwell's 2011a, page 832). However, even if the bird remained at a fixed distance to the microphone during the entire recording, and even if record gain levels were held constant, the simple act of the bird turning its head or body during the song bout could modulate the intensity of sound reaching the microphone as much as 18 dB, depending on frequency and body size (Larsen & Dabelsteen 1990; Brumm 2002) (note that this is a greater difference than the 14 dB maximum amplitude difference Cardoso & Atwell used for within-song comparisons). Songbirds (including dark-eyed juncos; S. A. Zollinger, personal observation) indeed often rapidly change their head orientation during a single song (Clark 1976; Brumm & Todt 2003). Differences in relative SPL resulting from orientation of the bird make it impossible to determine whether the bird's vocal amplitude is actually changing, or whether he is singing the same amplitude but simply moving his head or body relative to the microphone. Of course, with greater separation in time between compared notes, the likelihood that external factors affecting amplitude measures will vary becomes greater (again independently of actual amplitude changes by the bird).

We do not mean to dismiss entirely the validity of measures of relative amplitude. For example, the relative amplitude of different frequency components in a note (e.g. fundamental frequencies versus harmonic overtones) may be compared fairly using amplitude spectra when recorded in acoustically controlled environments (e.g. Nowicki 1987). Furthermore, depending on the type of question asked, one might reasonably assess the relative amplitude of notes or syllables that always occur in immediate succession. In this sort of comparison, temporal changes in recording conditions or bird orientation would only weaken the effect size if the relative amplitudes of the two are actually different at the source, therefore one would be unlikely to measure a positive correlation when there is none. However, as the problems we describe above are, in practise, highly frequency dependent and prone to measurement biases, we still recommend caution in comparisons of adjacent notes. Prior studies on the amplitude of successive notes (e.g. Beebee 2004; Christie et al. 2004) focused on relative amplitude and the question of whether amplitude increased or decreased across successive notes. It seems a very different thing, however, to use relative amplitude differences as a basis for asking how other features (such as frequency) covary in their production together with amplitude. It also seems a very different thing to state that amplitude measurement errors are 'conservative' in an analysis in which a lack of support for a hypothesis is interpreted as support for an untested alternative hypothesis.

We also here convey scepticism regarding the variation between within-song amplitude reported by Cardoso & Atwell (2011a) in their data set. The values reported seem excessive from a biological viewpoint for within-song repetitions of the same syllable type. Certainly the example of junco song presented in the study (Figure 1b in Cardoso & Atwell 2011a) does not show such large variation between successive renditions of the same syllable. A difference of 14 dB would mean that some sounds were perceptually three times louder than others (to an average human observer) and 25 times higher in sound intensity. At the extreme end of their variation, the authors report that they measured (but excluded as outliers) a difference within one bird of 48 dB. This seems rather unbelievable to us from a biological point of view. If the mean SPL of junco song at 1 m from the bird were 80 dB (re. 20 μPa) which seems a reasonable estimate based on published calibrated measurements of birdsong amplitudes (Dabelsteen 1981; Nelson 2000; Brumm 2009; Brumm & Ritschard 2011), a difference of 48 dB would mean that some syllables would have been 24 dB greater in amplitude, and others 24 dB less than the mean level. This variation in sound pressure level equals more than a 250-fold difference in sound pressure, or converted into sound intensity, one syllable that is roughly 63 000 times more intense than the other. While it is possible that juncos can produce vocalizations that differ that much in SPL, it seems more likely to us that these differences result from changing orientation of the bird relative to the microphone, or from fluctuating recording conditions. In support of differences of this magnitude, the authors cite a study by Anderson et al. (2008), which found large differences in amplitude (up to 36 dB) between songs of song sparrows, *Melospiza melodia*. However, the Anderson et al. study compared structurally different song types, 'broadcast songs' and 'soft songs', rather than renditions of the same syllable types within broadcast songs, as in Cardoso & Atwell (2011a).

ISSUES WITH TERMINOLOGY

Some terminology in Cardoso & Atwell (2011a) seems to be insufficiently precise. For example, 'loudness' is often used as if it were interchangeable with amplitude or intensity. Loudness is a troublesome term, because it is commonly used colloquially as a synonym for intensity, sound pressure level or amplitude. We do not take issue with the casual use of 'loudness' in the introduction or discussion as it makes the manuscript more accessible to nonspecialist readers. However, it is easy to let it slip into technical writing even when it may not be appropriate. Loudness is a psychophysical term that refers to the perception of the sound by the auditory system of a listener, and is therefore subjective and differs between individuals and between species (Moore 2001). As the human (or bird) ear is not equally sensitive to all frequencies, it is problematic to use loudness as an experimental variable. However, the use of loudness in the Results or Figures, when loudness was not what was measured, can be misleading. A specific example is found in the horizontal axes of Cardoso & Atwell's (2011a) Figure 2, which is labelled as 'loudness difference between songs (dB)'. Loudness, when that is what is actually measured (e.g. in a psychoacoustic study), is expressed in the units phons or sones and not in decibels. In this case the authors should indicate what was measured, such as the difference in voltage or relative amplitude in decibels.

ISSUES WITH DATA INTERPRETATION

Lombard Effect versus Voluntary Amplitude Control

Setting aside methodological questions for the moment, even if it was shown that acoustic frequency does not covary positively with amplitude, the conclusions as drawn by Cardoso & Atwell (2011a) seem to us incorrect in their logic. In particular, we do not think that such a relationship, if demonstrated, would be sufficient to eliminate the hypothesis that frequency shifts observed in noisy environments can occur as a by-product of the Lombard effect.

Control of and variation in vocal amplitude is a normal and typical component of vocal production in most animals. Just imagine what opera would be like if humans were unable to stay on pitch as their vocal amplitude changed, or conversely if each note could only be produced at a certain amplitude. As Cardoso & Atwell (2011) note, birds sing and call at a wide range of amplitudes in their normal lives (Brumm & Todt 2004; Anderson et al. 2008; Reichard et al. 2011). This dynamic range of amplitude is an integral part of a bird's communication system, and a vocalizing individual can voluntarily adjust both frequency and amplitude in normal communicative interactions (Brumm & Todt 2004; Christie et al. 2004; Brumm & Slater 2006; Geberzahn et al. 2009).

In contrast, the Lombard effect is an involuntary, reflex-like response to an increase in background noise (Zollinger & Brumm 2011). Because the presence of background noise interferes with the vocalizers' own sensory feedback, Lombard speech differs from the 'normal' variation in vocal amplitude that might occur during a conversation in a quiet room. Lombard speech differs from 'loud speech' when the speaker intentionally elevates the level of his or her voice to greater extremes, even when the loud speech reaches the same intensity level as noise-induced Lombard speech (reviewed by Jenssen et al. 2005). When these two kinds of 'louder-than-normal' speech are compared within individuals, the fundamental frequency (F0) of their voice typically increases more dramatically during Lombard speech than during voluntary 'loud speech' even when the amplitude of their voice is elevated by the same degree (Gramming et al. 1988; Stanton et al. 1988; Åkerlund et al. 1992). In addition, the degree of spectral changes during Lombard speech varies considerably between individuals and in different social and communicative contexts. For example, when two individuals are engaged in a dialogue, the Lombard effect is stronger (vocal intensity and fundamental frequency are higher) than during undirected speech (Garnier et al. 2010). So while frequency and amplitude are related mechanistically, humans are able to decouple this relationship to some degree and in certain contexts. Although not explicitly stated in their manuscript, we assume that the recordings used by Cardoso & Atwell (2011a) were of typical broadcast songs recorded in typical natural habitat and not from noisy urban areas or during periods of extreme background noise, and therefore, are more representative of normal variation in dynamic range for this species, rather than of juncos pushing their vocal production to extremes, as might occur in the presence of loud noise or unusual social contexts.

Relationship between Amplitude and Frequency

There are several different aspects of this relationship that need to be considered. For one, Cardoso & Atwell's (2011a) statement that one should not expect a physiological link between frequency and amplitude modulation is overly simplistic. As the authors correctly state, different muscle groups control airflow, syringeal aperture and tension of the labia (the vibratory tissues in the syrinx that create the sound pressure waves by interrupting the flow of air from the lungs) (Goller & Cooper 2004; Suthers & Zollinger 2008). However, the increase of subsyringeal airsac pressure required to phonate at a higher amplitude will also result in a higher frequency of vibration of the labia in the absence of voluntary counter-adjustments because higher pressure exerts more tensing force on the vibrating tissues than does lower pressure. In ring doves, *Streptopelia risoria*, direct measurements of subsyringeal airsac pressure found a very strong relationship between airsac pressure and vocalization frequency (Beckers et al. 2003). A strong correlation between subsyringeal pressure and vocalization frequency was also found in a suboscine bird, the great kiskadee, *Pitangus sulphuratus* (Amador et al. 2008), providing further evidence that driving pressure and frequency are biomechanically linked. In songbirds, the inter-relationship between amplitude and frequency control is complex. First, it is likely that the production mechanism for different sounds also influences to what degree fundamental frequency is influenced by the driving airsac pressure (e.g. Jensen et al. 2007; Sitt et al. 2008). Second, independence of frequency and amplitude therefore postulates that songbirds perfectly compensate for the effect of varying pressure by adjusting the tension-regulating muscle force. It is not clear, at present, whether songbirds compensate through adjustments in neural control or allow frequency to vary somewhat with sound amplitude. We cannot rule out the latter possibility, and there is no reason to assume that all species use similar compensatory mechanisms if they are used at all.

Cardoso & Atwell (2011a, page 835) state that there is 'no evidence that higher-frequency song types are louder' and cite a handful of studies that did not actually measure amplitude, but rather infer it using similar methods as employed in their study. While studies that directly test covariance between amplitude and frequency are few, there are a growing number that report positive correlations between the two variables in a variety of bird species, both in response to noise exposure (i.e. the Lombard effect: Leonard & Horn 2005; Osmanski & Dooling 2009) and within the vocal repertoire (Nelson 2000; Beckers et al. 2003; Goller & Cooper 2008; Ritschard & Brumm 2011; Nemeth et al. 2012), although a negative correlation has been reported for certain song types in black-capped chickadees, *Parus atricapillus* (Christie et al. 2004). Even in studies in which rises in amplitude and frequency do not appear to be correlated, it is not clear whether small changes in frequency caused by relatively small pressure changes would have been detected (see Riede et al. 2010). Furthermore, there are published accounts of significant variation in frequency of song syllables in naturally singing birds (e.g. chickadees: Horn et al. 1992; Ratcliffe & Weisman 1985; coal tits, *Periparus ater*: Goller 1987), suggesting that frequency might be under less precise control than often presumed, and that Lombard-related frequency shifts may be difficult to detect within the normal range of frequency variation.

False Dichotomy

One of the primary conclusions made by Cardoso & Atwell (2011a) is that the lack of a causal relationship between frequency and amplitude means, therefore, that the observed frequency shifts in urban species must be adaptive. This seems a false dichotomy to us, because lack of support for one hypothesis does not necessarily imply support for an untested alternative hypothesis.

More specifically, song amplitude is only one of several, mutually nonexclusive candidates for a possible selected trait with which song frequency may vary. Others include differences in hormonal state, breeding stage or territory density (Brumm & Naguib 2009; Nemeth & Brumm 2009). Indeed, Hamao et al. (2011) found that minimum song frequencies varied with territory density in urban great tits, *Parus major* (but see Ripmeester et al. 2009 for a dissimilar result in common blackbirds, *Turdus merula*). Moreover, Cardoso & Atwell's conclusion that higher urban frequency is best explained as an adaptation to noise is also not justified because there is no evidence to date suggesting that the increased frequency of city songs is maintained by selection in connection with acoustic masking by noise. On the contrary, as alluded to above, a recent study showed that the observed small increases in the frequency of songs of city birds have only a weak releasing effect in traffic noise (Nemeth & Brumm 2010). At least in the two investigated species (great tits and common blackbirds) an increase in amplitude is a more effective way to enlarge communication distances than increases in frequency. To demonstrate that elevated minimum frequencies are an adaptation to noise, one needs to show that there is (1) direct selection on song frequency in noisy areas (i.e. males with higher-frequency songs have higher fitness) and (2) that this selective adaptation is causally related to a release from masking by noise.

CONCLUSIONS

The relationship between amplitude and frequency in animal vocalizations is complex and is in need of further empirical inquiry. The study by Cardoso & Atwell (2011a) addresses an interesting and important gap in current knowledge. Unfortunately, we feel that it lacks the experimental rigour required to draw any new conclusions about how birds regulate amplitude and frequency during song. We emphasize that we still do not know whether the higher minimum frequencies of urban species (or higher peak frequencies as observed in common blackbird songs; Nemeth & Brumm 2009) are the result of the Lombard effect, we only know that the study by Cardoso & Atwell (2011a) does not satisfactorily allow the rejection of this hypothesis.

On a broader note, we are troubled by the recent increase in published bioacoustics studies that rely on ‘eye-balling’ frequency and amplitude measures in uncalibrated spectrograms. We fear that continuing to publish studies that either inadvertently or wilfully ignore established methodologies will have a profound adverse effect on the way the research field is viewed by the rest of the scientific community. The use of poor practises should not be justified by reference to existing published work that relied on the same flawed methodology and interpretations. To understand the relation between amplitude and frequency, amplitude must be measured. Acoustic recording and analysis protocols should be carefully designed according to the question being asked, and the auditory sensitivities of the animal being studied. It can certainly save time to re-use recordings that were made for previous studies, but in some cases one simply cannot extract the parameters needed to test the hypothesis in question from certain recordings. As we have pointed out in this comment, this is particularly true for amplitude measurements, which cannot be resurrected from uncalibrated song recordings.

Moreover, we hope that we have clearly illustrated the reasons why uncalibrated spectrograms are generally unsuitable for making accurate and repeatable measurements of minimum and maximum frequency and bandwidth. By making these frequency measurements from power spectra rather than from spectrograms, many of these methodological problems can be avoided. In addition, the use of carefully calibrated SPL measurements or reference signals to calibrate spectrogram displays would allow measurements of frequency or duration to be made with a minimum of error. We again encourage those investigating acoustic signals to consider the basic principles of acoustics and the proper tools for analysis when recording and measuring frequency and amplitude.

We thank Gabriël J. L. Beckers, Gonçalo C. Cardoso, Brian S. Nelson and an anonymous referee for constructive comments on the manuscript. H.B.’s research programme is supported by the German Research Foundation, DFG (BR 2309/6-1). F.G.’s research programme is supported by the National Institutes of Health, NIH (DC 06876). J.P.’s research programme is supported by the National Science Foundation, NSF (IOS 1028964).

References

Åkerlund, L., Gramming, P. & Sundberg, J. 1992. Phonetogram and averages of sound pressure levels and fundamental frequencies of speech: comparison between female singers and nonsingers. *Journal of Voice*, **6**, 55–63.

Amador, A., Goller, F. & Mindlin, G. B. 2008. Frequency modulation during song in a suboscine does not require vocal muscles. *Journal of Neurophysiology*, **99**, 2383–2389.

Anderson, R. C., Searcy, W. A., Peters, S. & Nowicki, S. 2008. Soft song in song sparrows: acoustic structure and implications for signal function. *Ethology*, **114**, 662–676.

Beckers, G. J. L., Suthers, R. A. & ten Cate, C. 2003. Mechanisms of frequency and amplitude modulation in ring dove song. *Journal of Experimental Biology*, **206**, 1833–1843.

Beebee, M. D. 2004. Variation in vocal performance in the songs of a wood-warbler: evidence for the function of distinct singing modes. *Ethology*, **110**, 531–542.

Beecher, M. D. 1988. Spectrographic analysis of animal vocalizations: implications of the ‘uncertainty principle’. *Bioacoustics*, **1**, 187–208.

Bradbury, J. W. & Vehrencamp, S. L. 1998. *Principles of Animal Communication*. Sunderland, Massachusetts: Sinauer.

Brumm, H. 2002. Sound radiation patterns in nightingale (*Luscinia megarhynchos*) songs. *Journal für Ornithologie*, **143**, 468–471.

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

Brumm, H. 2009. Song amplitude and body size in birds. *Behavioral Ecology and Sociobiology*, **63**, 1157–1165.

Brumm, H. & Naguib, M. 2009. Environmental acoustics and the evolution of bird song. In: *Advances in the Study of Behavior. Vol. 40* (Ed. by M. Naguib, K. Zuberbühler, N. S. Clayton & V. M. Janik), pp. 1–33. Amsterdam: Elsevier.

Brumm, H. & Ritschard, M. 2011. Song amplitude affects territorial aggression of male receivers in chaffinches. *Behavioral Ecology*, **22**, 310–316.

Brumm, H. & Slater, P. J. B. 2006. Animals can vary signal amplitude with receiver distance: evidence from zebra finch song. *Animal Behaviour*, **72**, 699–705.

Brumm, H. & Todt, D. 2003. Facing the rival: directional singing behaviour in nightingales. *Behaviour*, **140**, 43–53.

Brumm, H. & Todt, D. 2004. Male–male vocal interactions and the adjustment of song amplitude in a territorial bird. *Animal Behaviour*, **67**, 281–286.

Brumm, H. & Zollinger, S. A. 2011. The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour*, **148**, 1173–1198.

Buck, J. R. & Tyack, P. L. 1993. A quantitative measure of similarity for *Tursiops truncatus* signature whistles. *Journal of the Acoustical Society of America*, **94**, 2497–2506.

Cardoso, G. C. & Atwell, J. W. 2011a. On the relation between loudness and the increased song frequency of urban birds. *Animal Behaviour*, **82**, 831–836.

Cardoso, G. C. & Atwell, J. W. 2011b. Directional cultural change by modification and replacement of memes. *Evolution*, **65**, 295–300.

Cardoso, G. C., Atwell, J. W., Ketterson, E. D. & Price, T. D. 2007. Inferring performance in the songs of dark-eyed juncos (*Junco hyemalis*). *Behavioral Ecology*, **18**, 1051–1057.

Cardoso, G. C., Atwell, J. W., Ketterson, E. D. & Price, T. D. 2009. Song types, song performance, and the use of repertoires in dark-eyed juncos (*Junco hyemalis*). *Behavioral Ecology*, **20**, 901–907.

Christie, P. J., Mennill, D. J. & Ratcliffe, L. M. 2004. Pitch shifts and song structure indicate male quality in the dawn chorus of black-capped chickadees. *Behavioral Ecology and Sociobiology*, **55**, 341–348.

Clark, G. A. 1976. Taxonomic differences in head movements of vocalizing passerines. *Bird-Banding*, **47**, 123–128.

Dabelsteen, T. 1981. The sound pressure level in the dawn song of the blackbird (*Turdus merula*) and a method for adjusting the level in experimental song to the level in natural song. *Zeitschrift für Tierpsychologie*, **56**, 137–149.

Deecke, V. B., Ford, J. K. B. & Spong, P. 1999. Quantifying complex patterns of bioacoustic variation: use of a neural network to compare killer whale (*Orcinus orca*) dialects. *Journal of the Acoustical Society of America*, **105**, 2499–2507.

DuBois, A. L., Nowicki, S. & Searcy, W. A. 2009. Swamp sparrows modulate vocal performance in an aggressive context. *Biology Letters*, **5**, 163–165.

Fernández-Juricic, E., Poston, R., De Collibus, K., Morgan, T., Bastain, B., Martin, C., Jones, K. & Treminio, R. 2005. Microhabitat selection and singing behavior patterns of male house finches (*Carpodacus mexicanus*) in urban parks in a heavily urbanized landscape in the western U.S. *Urban Habitats*, **3**, 49–69.

Fischer, J., Hammerschmidt, K. & Todt, D. 1998. Local variation in Barbary macaque shrill barks. *Animal Behaviour*, **56**, 623–629.

Garnier, M., Henrich, N. & Dubois, D. 2010. Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language and Hearing Research*, **53**, 588–608.

Geberzahn, N., Goymann, W., Muck, C. & ten Cate, C. 2009. Females alter their song when challenged in a sex-role reversed bird species. *Behavioral Ecology and Sociobiology*, **64**, 193–204.

Gerhardt, H. C. & Huber, F. 2002. *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions*. Chicago: University of Chicago Press.

Goller, F. 1987. Der Gesang der Tannenmeise (*Parus ater*): Beschreibung und kommunikative Funktion. *Journal of Ornithology*, **128**, 291–310.

Goller, F. & Cooper, B. G. 2004. Peripheral motor dynamics of song production in the zebra finch. *Annals of the New York Academy of Sciences: Behavioral Neurobiology of Birdsong*, **1016**, 130–152.

Goller, F. & Cooper, B. G. 2008. Peripheral mechanisms of sensorimotor integration during singing. In: *Neuroscience of Birdsong* (Ed. by H. P. Zeigler & P. Marler), pp. 99–114. Cambridge: Cambridge University Press.

Gramming, P., Sundberg, J., Ternströmm, S., Leandersson, R. & Perkins, W. 1988. Relationship between changes in voice pitch and loudness. *Journal of Voice*, **2**, 118–126.

Greenewalt, C. H. 1968. *Bird Song: Acoustics and Physiology*. Washington, D.C.: Smithsonian Institution Press

Gross, K., Pasinelli, G. & Kunc, H. P. 2010. Behavioral plasticity allows short-term adjustment to a novel environment. *American Naturalist*, **176**, 456–464.

Hamao, S., Watanabe, M. & Mori, Y. 2011. Urban noise and male density affect songs in the great tit, *Parus major. Ethology Ecology & Evolution*, **23**, 111–119.

Horn, A. G., Leonard, M. L., Ratcliffe, L., Shackleton, S. A. & Weisman, R. G. 1992. Frequency variation in songs of black-capped chickadees (*Parus atricapillus*). *Auk*, **109**, 847–852.

Jensen, K. K., Cooper, B. G., Larsen, O. N. & Goller, F. 2007. Songbirds use pulse tone register in two voices to generate low-frequency sound. *Proceedings of the Royal Society B*, **274**, 2703–2710.

Jenssen, M., Köster, O. & Gfroerer, S. 2005. Influence of vocal effort on average and variability of fundamental frequency. *Speech, Language and the Law*, **12**, 174–212.

Larsen, O. N. & Dabelsteen, T. 1990. Directionality of blackbird vocalization: implications for vocal communication and its further study. *Ornis Scandinavica*, **21**, 37–45.

Leonard, M. L. & Horn, A. G. 2005. Ambient noise and the design of begging signals. *Proceedings of the Royal Society B*, **272**, 651–656.

Moore, B. C. J. 2001. Loudness, pitch and timbre. In: *Blackwell Handbook of Perception* (Ed. by E. B. Goldstein), pp. 408–436. Oxford: Blackwell.

Nelson, B. S. 2000. Avian dependence on sound pressure level as an auditory distance cue. *Animal Behaviour*, **59**, 57–67.

Nemeth, E. & Brumm, H. 2009. Blackbirds sing higher-pitched songs in cities: adaptation to habitat acoustics or side-effect of urbanization? *Animal Behaviour*, **78**, 637–641.

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

Nemeth, E., Zollinger, S. & Brumm, H. 2012. Effect sizes and the integrative understanding of urban bird song (A reply to Slabbekoorn et al). *American Naturalist*, **180**, 146–152, http://dx.doi.org/10.1086/665994.

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

Osmanski, M. S. & Dooling, R. J. 2009. The effect of altered auditory feedback on control of vocal production in budgerigars (*Melopsittacus undulatus*). *Journal of the Acoustical Society of America*, **126**, 911–919.

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

Potvin, D. A., Parris, K. M. & Mulder, R. A. 2010. Geographically pervasive effects of urban noise on frequency and syllable rate of songs and calls in silvereyes (*Zosterops lateralis*). *Proceedings of the Royal Society B*, **278**, 2464–2469.

Ratcliffe, L. & Weisman, R. G. 1985. Frequency shift in the fee bee song of the black-capped chickadee. *Condor*, **87**, 555–556.

Reichard, D. G., Rice, R. J., Vanderbilt, C. C. & Ketterson, E. D. 2011. Deciphering information encoded in birdsong: male songbirds with fertile mates respond most strongly to complex, low-amplitude songs used in courtship. *American Naturalist*, **178**, 478–487.

Riede, T., Fisher, J. H. & Goller, F. 2010. Sexual dimorphism of the zebra finch syrinx Indicates adaptation for high fundamental frequencies in males. *PLoS One*, **5**, e11368.

Ripmeester, E. A. P., Mulder, M. & Slabbekoorn, H. 2009. Habitat-dependent acoustic divergence affects playback response in urban and forest populations of the European blackbird. *Behavioral Ecology*, **21**, 876–883.

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

Roch, M. A., Brandes, S. T., Patel, B., Barkley, Y., Baumann-Pickering, S. & Soldevilla, M. S. 2011. Automated extraction of odontocete whistle contours. *Journal of the Acoustical Society of America*, **130**, 2212–2223.

Siemers, B. M. & Kerth, G. 2006. Do echolocation calls of wild colony-living Bechstein's bats (*Myotis bechsteinii*) provide individual-specific signatures? *Behavioral Ecology and Sociobiology*, **59**, 443–454.

Sitt, J. D., Amador, A., Goller, F. & Mindlin, G. B. 2008. Dynamical origin of spectrally rich vocalizations in birdsong. *Physical Review E*, **78**, 011905.

Slabbekoorn, H. & den Boer-Visser, A. 2006. Cities change the songs of birds. *Current Biology*, **16**, 2326–2331.

Smith, S. W. 1997. *The Scientist and Engineer's Guide to Digital Signal Processing*. San Diego: California Technical.

Stanton, B. J., Jamieson, L. H. & Allen, G. D. 1988. Acoustic–phonetic analysis of loud and Lombard speech in simulated cockpit conditions. *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*, **2**, 675–678.

Suthers, R. A. & Zollinger, S. A. 2008. From brain to song: the vocal organ and vocal tract. In: *Neuroscience of Birdsong* (Ed. by H. P. Zeigler & P. Marler), pp. 78–98. Cambridge: Cambridge University Press.

Templeton, C. N., Greene, E. & Davis, K. 2005. Allometry of alarm calls: black-capped chickadees encode information about predator size. *Science*, **308**, 1934–1937.

Zollinger, S. A. & Brumm, H. 2011. The Lombard effect. *Current Biology*, **21**, R614–R615.

Zollinger, S. A., Goller, F. & Brumm, H. 2011. Metabolic and respiratory costs of increasing song amplitude in zebra finches. *PLoS One*, **6**, e23198.