Applied Acoustics 73 (2012) 1138–1145

Contents lists available at SciVerse ScienceDirect

Applied Acoustics

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

Obtaining calibrated sound pressure levels from consumer digital audio recorders

D.J. Mennitt *, K.M. Fristrup

Natural Sounds and Night Skies Division, National Park Service, 1201 Oakridge Drive, Suite 100, Fort Collins, CO 80525, United States

ARTICLE INFO

Article history:
Received 5 January 2011
Received in revised form 9 May 2012
Accepted 14 May 2012
Available online 7 June 2012

Keywords:
Digital audio recorder
Calibration
Sound level meter
Acoustic metrics
Soundscape

ABSTRACT

Audio recording of environmental sound is an increasingly efficient method for autonomously sensing many ecological and anthropogenic processes. The increasing capabilities of consumer digital audio recorders (DARs), especially increases in storage capacity and reductions in power consumption, enable continuous audio recordings exceeding 1 month in duration with packages that are relatively small and inexpensive. To augment the ability of these systems to document the range of sounds present at a location, this paper examines two methods for calibrating recorders to measure sound levels. Compressed audio recorded by a DAR can be processed to yield relatively consistent measures of one-third octave band L_{eq} values within a limited frequency and dynamic range. This was evaluated by synchronizing data with a Type-1 sound level meter. The calibration is stable over a 23 day deployment outdoors with wide variation in ambient temperature and humidity. When considering aggregate acoustic metrics over time or a wide bandwidth such as an hourly A-weighted L_{50}, the results can be quite accurate (within 1 dBA).

1. Introduction

Noise pollution is a pervasive issue in the United States, extending far beyond airport and freeway boundaries into many National Parks and wilderness areas. The instruments and methods developed to address urban noise problems present some significant limitations for noise monitoring over large areas and in remote settings. The cost, power consumption, and limited data storage of precision sound level meters severely constrain the scope of their application, though they serve as valuable reference instruments. In contrast, advances in consumer digital audio products offer opportunities to obtain weeks of high quality audio recordings in small, inexpensive packages. These recordings are more valuable for wilderness noise studies when they are calibrated to yield sound level data.

Calibrated recordings are also valuable for biological research and conservation. The time–frequency structure of animal vocalizations has been studied for decades using spectrograms, and there is an extensive literature addressing the biological significance of the patterns that this analysis reveals [1]. However, very few studies have addressed intentional variation in the intensity of animal vocalizations, or the behavioral significance of highly directional signals [2]. Calibrated recordings can reveal additional dimensions of animal communication, and broader understanding of variation in vocal levels would enable scientists to apply distance sampling methods [3] when individual sounds can be distinguished, or translate the sound level of a chorus into the density of animals. Unobtrusive audio recorders can document changes in behavior, shifts in spatial distributions, and track population dynamics. Acoustical monitoring is especially suitable for monitoring phenological changes in ecosystem processes due to climate change.

In community and industrial settings, noise control problems are characterized by measuring sound pressure levels [4], such as those measured by a sound level meter (SLM). However, the extent and complexity of national park acoustical environments requires more data than is collected by an SLM [5,6]. Most sources are not adequately represented by one-third octave band spectral measurements. Audio recordings allow human listeners to identify sounds, and waveform data enable higher resolution time–frequency methods of automated detection, identification, and localization of sound sources.

Digital audio recorders have played central roles in extensive spatial surveys for rare or endangered species [7–13]. These studies required compact packages that could be readily transported to remote locations, long recording durations to capture adequate samples of rare acoustic events, and relative low cost to enable coverage of numerous sites. An appealing feature of these surveys is that they capture information about all audible phenomena in the environment, along with the sounds from species of special interest. Retrospective calibration of these recording systems would enhance the value of these archived data.

Consumer DARs have several characteristics that must be considered during the design of acoustical sampling plans and the selection of equipment. Intrinsic noise levels, dynamic range, bandwidth limits, sensitivity to environmental conditions, and build quality are important features, along with cost. The suitability of any DAR for

* Corresponding author. Tel.: +1 970 267 2114; fax: +1 970 267 2109.
E-mail address: daniel_mennitt@partner.nps.gov (D.J. Mennitt).

0003-682X/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apacoust.2012.05.006

sound level measurements will depend upon the application and the environment; in some situations an extremely inexpensive unit may be the best choice, despite the limitations this imposes on the accuracy of the measurements. The capabilities of DARs to record accurate time histories of acoustic signals has been studied [14] and there is precedent for calculating sound level metrics from audio data [15]. This paper evaluates methods for obtaining calibrated, 1-s, one-third octave sound level measurements from compressed audio recordings. Limitations and advantages of using MP3 compression, a lossy audio format, were investigated.

Two methods of calibrating a DAR using measurements made with a Type 1 SLM [16] were evaluated: a substitution method in controlled laboratory conditions and coincident field calibration. The laboratory calibration simplified the data analysis, because identical, stable sound levels could be presented to both instruments. However, field calibration tests were performed for two reasons. First, facilities like these may not be available to many user groups. Second, the coincident field calibration tests were able to span weeks of time and a wide range of atmospheric conditions, to investigate the stability of calibration.

However, the coincident field calibration posed an analytical challenge. Ambient sound levels fluctuate rapidly, and the MP3 compression could plausibly introduce artifacts in this setting that would not be observed under more static laboratory conditions. In order to observe these potential effects, methods were devised to enable precise alignment of 1 s segments from the digital audio recordings to the corresponding 1-s, one-third octave sound level measurements from the SLM. The time alignment method presented below has a precision of a single sample. With this time alignment, the accuracy of calibrated DAR measurements could be evaluated for 1-s, one-third octave spectrum measurements, broadband summaries (A-weighted metrics), and long term averages. As predicted by the Central Limit Theorem, broadband measurements and long-term averages show much closer agreement than short term spectrum measurements.

2. The digital audio recorder

2.1. Zoom H2

The calibration methods developed in this paper can be applied to any digital audio recorder. Many DARs are commercially available; this study utilized the Zoom H2 which is inexpensive, low power (<1 W) and simple to use. The H2 records data to an externally removable flash memory card in a variety of file formats. For this work, highly compressed audio in MP3 format at 64 kilobits per second (kbps) was used. The dynamic range of the analog–digital conversion was 16 bits.

The unit is equipped with four internal cardioid pattern microphones; the left and right front and left and right rear microphones each form a stereo pair in an XY configuration. For the results detailed herein, all four microphone outputs are summed to obtain a more omnidirectional response in the transverse plane. The coherent summation of four channels also had the benefit of reducing electronic noise and increasing the dynamic range of the device, nominally by 6 dB.

2.2. MP3 processing

Audio compression algorithms, like MP3, offer two important benefits for long term acoustical monitoring: efficient use of storage capacity and reduction in energy consumption (writing data to storage is energetically expensive). Lossy compression algorithms such as MP3 achieve high compression ratios: a 64 kbps MP3 file requires approximately 20 GB of storage per month, 22 times less than a WAV file over the equivalent duration. However, low bitrate MP3 compression results in a reduced bandwidth. Because this work utilized MP3s encoded at 64 kpbs, all analyses were limited to the one-third octave bands between 25 Hz and 6.3 kHz.

The reference data from the Type 1 SLM were 1-s one-third octave band L_{eq} values. L_{eq} is the equivalent continuous sound level defined as

$$L_{eq} = 10\log_{10}\left\{1/T\int_{t_1}^{t_2} p_B^2(t)/p_0^2 dt\right\} dB \quad (1)$$

where T is the averaging time interval, t_1 and t_2 are the start and end times of the interval, $p_B(t)$ is the sound pressure level in a given bandwidth, and p_0 is the reference sound pressure level.

To efficiently calculate 1-s one-third octave band L_{eq} values from the digital audio waveform data, a fast algorithm utilizing the Fast Fourier Transform was employed as opposed to time domain decimation and infinite impulse response filtering as specified by the ANSI standard [17]. The frequency response of the Butterworth filters specified by the ANSI standard were precalculated and stored. One second of data was transformed into the frequency domain and each one-third octave band magnitude was obtained by multiplication with the individual filter response followed by a summation across all frequency bins. Results of this algorithm showed excellent agreement with the standard while reducing processing time by a factor of four. Overall, data was processed 90 times faster than real-time. The penalty for speed was poor resolution in the low frequency bands because the data were not decimated. As a result, the fast method overestimated the energy in each one-third octave band relative to the standard. This effect is quantified in Fig. 1 which shows the average error, defined as the difference between the two methods, considering a white Gaussian noise input. This bias was corrected by including compensation in the calibration.

Although information is discarded during MP3 compression and the narrowband spectrum of a complex signal will not match its compressed counterpart, the relative power in a given one-third octave band is similar. The overall wideband power has an even better agreement. During the identification of tonal and noise maskers, the perceptual codec combines into a single noise masker all of the energy from individual spectral components that have not contributed to a tonal masker within the same band [18]. To illustrate this point, a 64 kbps MP3 file was created from a 16 bit

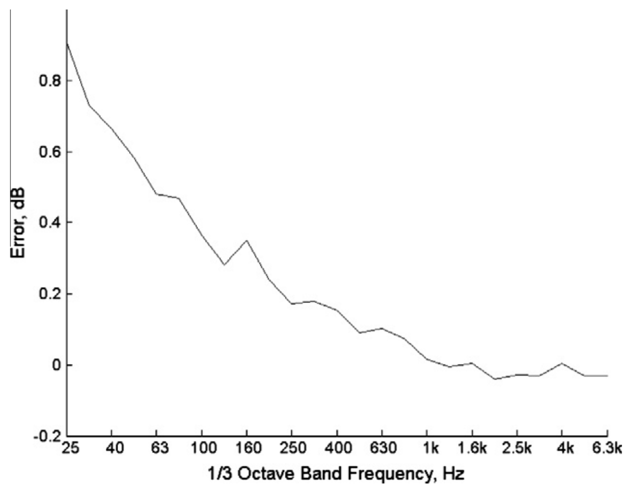

Fig. 1. Average difference between the fast method and the ANSI standard of calculating one-third octave band power levels considering a white Gaussian noise input.

WAV file sampled at 44.1 kHz and the original and compressed spectrums were compared. The original WAV file was a recording of the soundscape near a heavily used trail in wilderness and contains many sources such as wind, flowing water, birdsong, aircraft, human voices, and footsteps. Therefore, it is a diverse signal which spans a wide bandwidth and includes transient sources. Narrowband, one-third octave band, and wideband 1-s L_{eq} values were calculated for each signal. 25 one-third octave bands between 25 Hz and 6.3 kHz and 7058 narrowband frequencies between 22 and 7079 Hz were considered. The difference between the wideband L_{eq} values and the maximum difference amongst all of the one-third octave bands and individual narrowband bins, the worst case scenario, were stored for each second. These values over a 600 s example are plotted in Fig. 2. Because the MP3 algorithm preserves power in a band-wise sense but not in a tonal sense, the average difference in the wideband L_{eq} is zero whereas the average maximum difference in the narrowband L_{eq} is 35 dB. The average maximum difference between the original and compressed one-third octave spectrums is 0.7 dB.

A simple test was performed to verify the linearity of 1-s one-third octave band L_{eq} values calculated from an H2 recording. A pink noise signal with step changes in amplitude over time (input signal) was presented to the line-in input of the DAR. Fig. 3a shows the 1 kHz one-third octave band L_{eq} values of this signal which remain constant for 6 s and then decrease by 6 dB. The recorded MP3 was decompressed into WAV format (output signal). L_{eq} values of the input and output signal were calculated and the two sets are plotted against one another in Fig. 3b. The values plotted are referenced to 0 dB full scale (0 dBFS) of each individual device [19]. Zero dBFS is the maximum possible level and is achieved by a square wave with a full-scale peak value whereas all other signals result in negative numbers (e.g. a sine wave with full scale peak amplitude is −3 dB ref 0 dBFS by this convention). Because the levels are independently referenced to the full scale of each individual device, there exists an arbitrary mismatch between the input and output (in this case roughly 20 dB).

Fig. 3b shows a clear linear relationship over a dynamic range of approximately 80 dB. Signals exceeding the bit depth of the analog-to-digital converter or interacting with the device noise floor introduced non-linearity outside of this range. Also notice the scatter in the linear range; this was due to a time misalignment between the two signals (i.e. t_1 of Eq. (1) is different for each signal). For a given step level of power input, five points are clustered together at the corresponding output level. A regression line affirms this linear relationship. However, the sixth point represents the last second of integration which spans two steps of power level input. If the transitions between signal levels had matched up with boundaries of the 1-s data segments used to measure signal levels, then all six 1-s measurements would have fallen on the expected line. This illustrates the importance of addressing the time alignment issue when comparing sound level measurements from a sound level meter to those obtained from a DAR, which is addressed in Section 3.2."

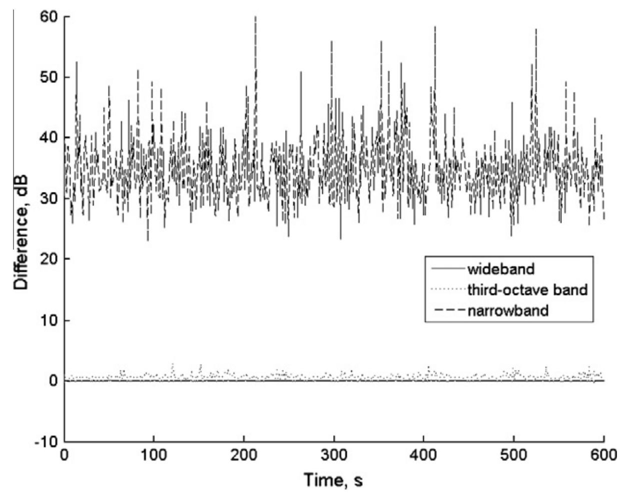

Fig. 2. Maximum difference in the 1-s L_{eq} between values calculated from a WAV file and 64 kbps MP3.

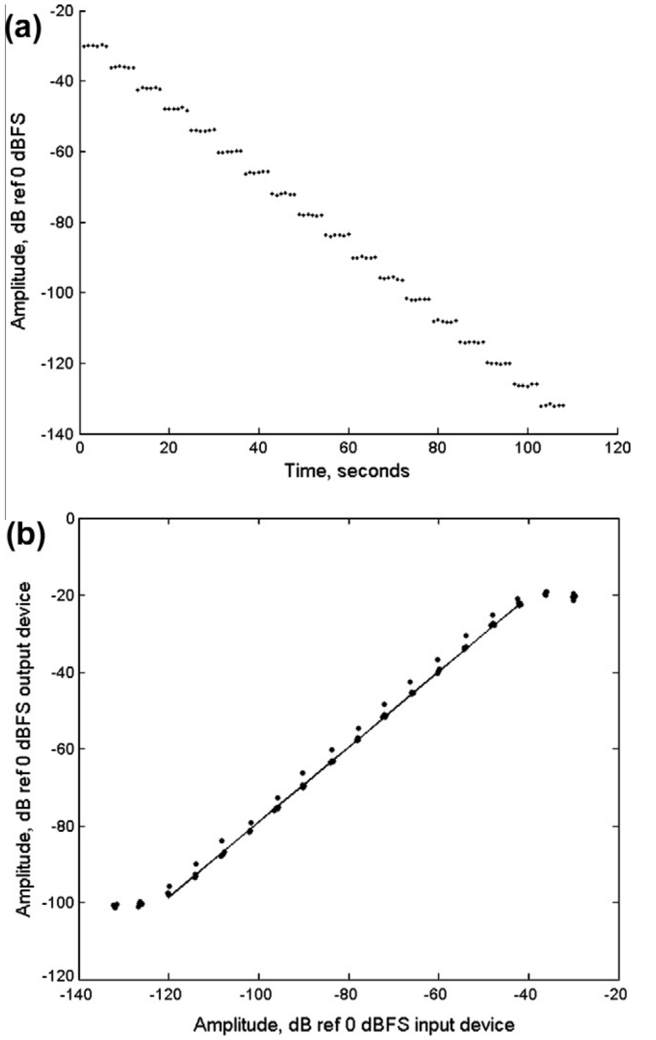

Fig. 3. 1-s L_{eq} values for the test signal input to the DAR (a) and the same values plotted against the corresponding output L_{eq} (b) for the 1 kHz one-third octave band. Other one-third octave bands similar.

3. Calibration methods

3.1. Substitution method

The substitution method involves placing the system of interest in a known acoustic field as measured by a reference system. The measured deviation from the reference can then be used to compensate subsequent measurements. In this case, the substitution method as specified by IEC 60268-4 [20], was used to measure the directional characteristics, linearity, frequency response and sensitivity, of the H2 DAR.

Measurements took place in a semi-anechoic chamber at the Engineering Research Center at Colorado State University. The reference system consisted of a PCB 377B02 microphone and RME Fireface UC digital audio interface. Golay codes [21] were used to efficiently measure system transfer functions; these were emitted by a pair of mid and low range woofers to adequately cover the frequency range of interest. Both the reference system and test system were measured. The transfer function of the reference system was then used to deconvolve the test system output to yield the frequency response of the DAR. Testing occurred during neutral conditions (average temperature of 25 °C and 43% relative humidity).

A sample of 20 Zoom H2 DARs was tested. These were measured along with an environmental shroud – a combination rain hat and wind fairing – which had the additional effect of attenuating the response in the 1–3 kHz range. As explained in Section 2, the four microphone channels were summed which yielded a fairly omnidirectional response as shown by the azimuthal directivity in Fig. 4. Interference effects were not present in the frequency range of interest due to the close proximity of microphones. Assuming a diffuse sound field, the measurements at 24 angles of incidence are averaged to give the substitution method calibration coefficients; a correction for the fast one-third octave band algorithm was also included. The one-third octave band frontal-incidence calibration coefficients for 20 DARs are shown in Fig. 5; there was some scatter due to the method and DAR build quality. Although the coefficients of the majority of devices were consistently within 3 dB of each other, calibration of individual devices provides greater accuracy. Nonetheless, a generic calibration could be applied to any given device for a modest increase in uncertainty.

The noise floor of the DAR was evaluated by isolating the device in a sealed case in a vehicle in a quiet wilderness area at night. Confirmation that the recorded levels were due to the self-noise of the DAR and not influenced by the sound pressure level present was achieved by measuring the test environment with a low noise Type-1 SLM. The self-noise measurements, calibrated by the substitution method coefficients, appear in Fig. 6. Based on these measurements, the lower limit of the Zoom H2 dynamic range is 45 dB (28 dBA). The input gain of the Zoom H2 has a negligible effect on the noise floor. The maximum SPL before clipping occurs is dependent on the crest factor of the input signal; considering a 1 kHz sine wave the upper limit of the dynamic range is 112 dB (A-weighted value is equivalent).

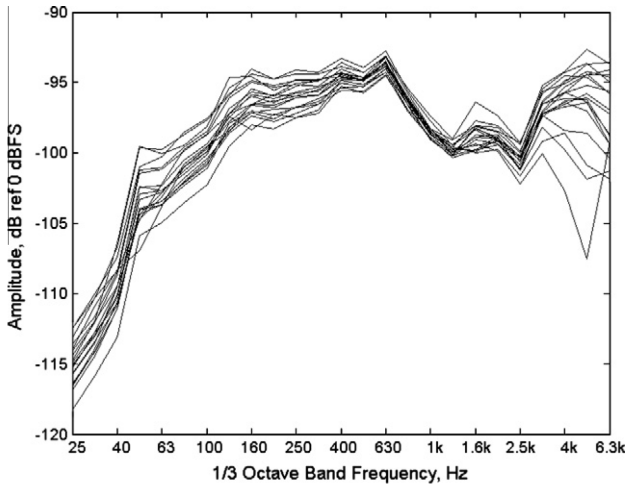

Fig. 5. One-third octave band frontal-incidence calibration coefficients of 20 H2 Zoom DARs using the substitution method.

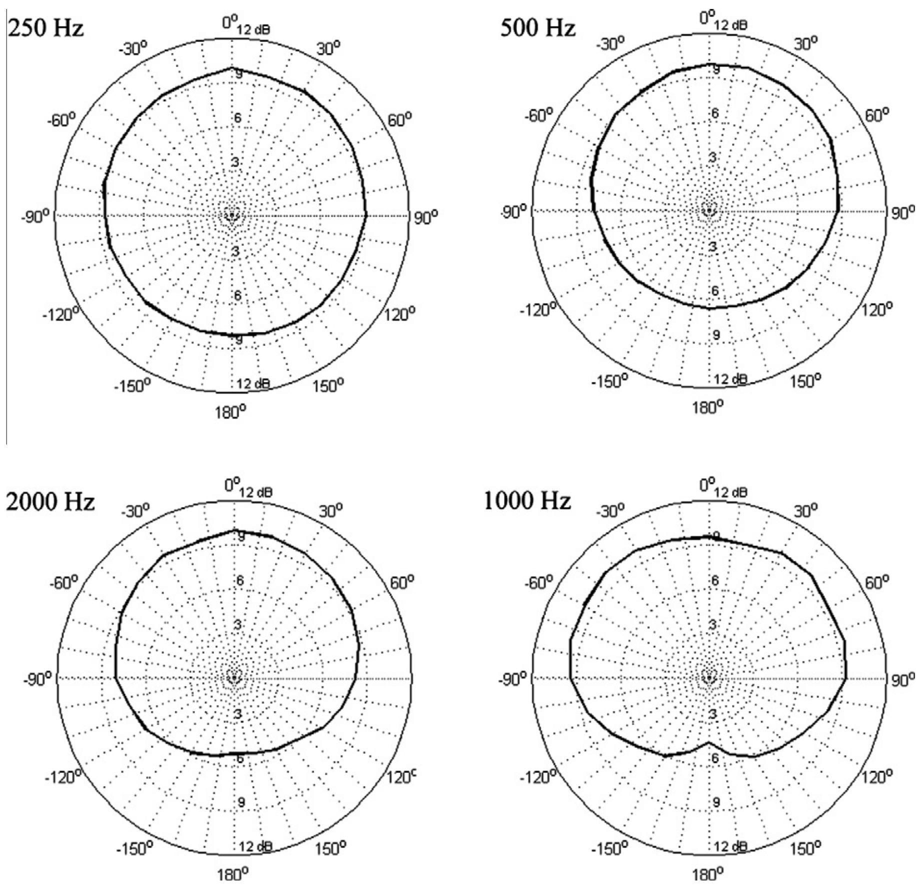

Fig. 4. Directivity of the DAR in the 250 Hz, 500 Hz, 1000 Hz, and 2000 Hz one-third octave band (clockwise from top left).

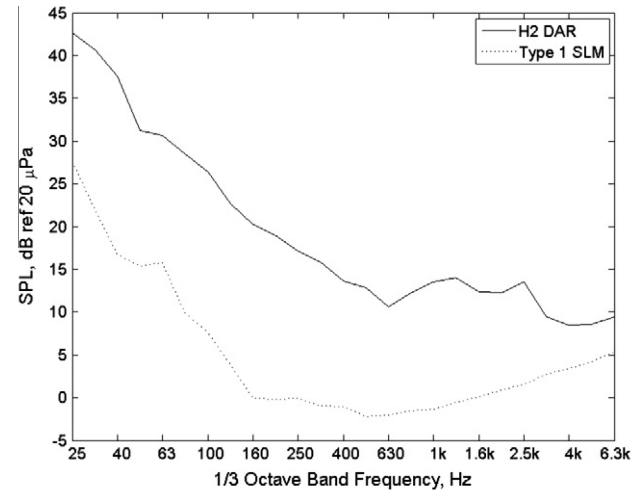

Fig. 6. Self-noise of the average H2 Zoom DAR. The Type-1 SLM measurement confirms that the influence of environmental noise is negligible.

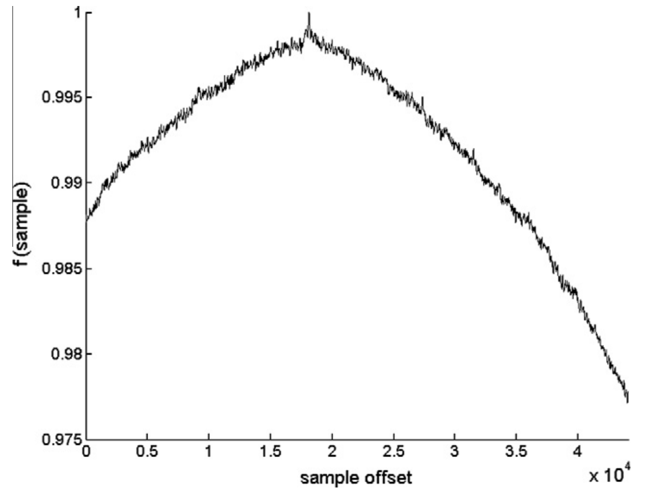

Fig. 7. Magnitude of the cross-correlation coefficient for an exhaustive search of the sample shift necessary to align two 1-s L_{eq} records.

3.2. Coincident field calibration

For coincident field calibration, SLM measurements were used as training data to calibrate the DAR output. The DAR was positioned near a Type-1 SLM such that the microphones were both 1.5 m above ground and 1 m apart transversely. Assuming distant sound sources, uniform ground surface, and no nearby scattering objects, the two closely spaced devices were exposed to identical sound pressure levels.

In order to minimize the factors that could degrade the cross-calibration of the DAR data, a method was devised to estimate the 1-s data blocks in the DAR that corresponded to the SLM sound level measurements. The SLM calculated energy averages over 1 s time intervals. Sound levels fluctuated in the natural environment, especially when there were prominent transient sounds, so the magnitude of SLM measurements depended on precisely which second the integration spanned. The SLM and DAR samples could not be synchronized without substantial hardware modifications.

In order to retrospectively identify the 1 s sample frames in the DAR that corresponded to the SLM measurements, a cross-correlation procedure was utilized to obtain the best match between the SLM one-third octave band levels and the levels computed from the DAR data across a range of time offsets. The cross correlation of individual one-third octave band 1-s L_{eq} histories proved sensitive to a time shift on the order of one sample. The following example illustrates the method employed to handle the registration problem during coincident field calibration.

Reference data was a WAV format recording from a site in Haleakala National Park. This data was recorded in MP3 format on a Zoom H2, and 1-s one-third octave band L_{eq} values were calculated from both digital audio files. As the reference data in this test case was originally in continuous audio format, the exact sample shift between 1-s L_{eq} frames could be found to confirm the registration. During a typical application however, only the 1-s L_{eq} values are available to the reference device.

The time synchronization occurs on two scales. First, the records were coarsely aligned to the nearest second by correlating the unaligned 1/3rd octave spectra. For the fine scale alignment, the reference spectral sequence was searched for periods of high variance to ensure that high signal to noise ratio events were available to compare between records. A 150 s record length provided good accuracy with reasonable processing time. To test a potential offset, in samples, the 150 s long one-third octave band histories from the test and reference data were cross-correlated. The maxima of the individual one-third octave band cross correlation functions were evaluated to determine a common lag; the values of all bands at this lag were then weighted and averaged to yield a composite metric of correlation. The weights favored high signal to noise ratios. The shape of the objective function in Fig. 7 demonstrates that the magnitude of the composite metric increases in proximity to the exact value. The function is continuous and the global maximum corresponds to the true sample shift, however, there were multiple local maxima. At a sample rate of 44100 samples per second it was computationally prohibitive to always employ an exhaustive search. Instead, an algorithm based on golden section and parabolic interpolation was used to efficiently find the function maximum over a fixed interval [22] (this function is implemented in Matlab as FMINBND).

Because the DAR output had the correct proportional relationship, a zeroth-order ordinary least squares fit solution was the calibration coefficient, c_b, for each one-third octave band.

$$c_b = \frac{1}{N}\sum_{i=1}^{N}(L_{y,i} - L_{r,i}) \tag{2}$$

where $L_{y,i}$ and $L_{r,i}$ are the ith test and reference 1-s L_{eq} values respectively. N is the total number of 1-s L_{eq} values available for a given one-third octave band.

4. Results and discussion

The results detailed in this section focus on data collected in Dinosaur National Monument near Rippling Brook, a side canyon extending off the Green River. Sound monitoring equipment consisting of a Larson Davis 831 Type-1 SLM and Zoom H2 DAR was deployed 30 May–21 June 2010. During this time, the nearby air temperature at the measurement site varied from 8.7 °C at night to 37.8 °C during the day. Relative humidity ranged from 10.1% to 100%. This was a fairly diverse soundscape with hourly median dBA values spanning 36–61 dBA over the 514 h monitored. The majority of this energy was concentrated in bands below 1.6 kHz and common sources include water, wind, birds, insects, and aircraft.

4.1. Comprehensive coincident field calibration

For this study, a set of calibration coefficients were solved for each hour due to clock drift and changing environmental conditions. The registration method presented in Section 3.2 aligned the DAR data to the SLM measurements to the nearest sample. The regularity of the estimated clock drift between devices (on the order of 200 ms per hour) and agreement between reference and test levels suggested that the DAR data were adequately aligned to the SLM measurements.

After transforming the raw compressed audio data to 1-s one-third octave band L_{eq} values and synchronizing the time histories, calibration coefficient training data were limited to the linear range of the DAR. For devices that respond linearly to changes in amplitude, calibration need occur at only a single amplitude, commonly 94 dB SPL. However, the coincident field calibration method exerts little control over the acoustic energy present. Because of the inherent measurement uncertainty, it was preferred to use all available data in the linear range as discussed in Section 2.2. The received signal, $p_y(t)$, is the summation of the desired signal and noise. The minimum level to consider for calibration can be determined given that the noise floor of the system is known. Assuming the signal and noise are uncorrelated

$$\langle p_y^2 \rangle = \langle p_s^2 \rangle + \langle p_n^2 \rangle \tag{3}$$

where the brackets denote the time average power of the signal. The sound level of the received signal is defined as $L_y = 10\log_{10}\frac{\langle p_y^2\rangle}{p_0^2}$. Although L_y is generally dominated by L_s, at low sound pressure levels the effect of L_n can be significant; this is dependent on the relative magnitudes of L_n and L_s. The magnitude of the effect of L_n on L_y is

$$\delta = L_y - L_s = 10\log_{10}\frac{\langle p_y^2\rangle}{\langle p_s^2\rangle} = 10\log_{10}\left(1 + \frac{\langle p_n^2\rangle}{\langle p_s^2\rangle}\right) \tag{4}$$

Rearranging Eq. (4) yields the minimum L_s for a given L_n to have an effect of magnitude δ:

$$L_{s_{MIN}} = L_n - 10\log_{10}\left(10^{\frac{\delta}{10}} - 1\right) \tag{5}$$

Therefore δ is the allowed influence of the noise floor on the received level and controls the tradeoff between the quantity and quality of data available for calibration. For this work, δ = 1 dB. When limiting training data, it was crucial to avoid data below the minimum level as measured by both the reference meter, L_r, and the DAR, L_y. An example of 1 h of data from the 6.3 kHz one-third octave band is shown in Fig. 8. The minimum level was used as the threshold of a simple binary test.

The mean and variance of calculated calibration coefficients are shown in Fig. 9. These statistics represent a sample size of 149–367, depending on the one-third octave band. Of the 514 h available, coefficients were not calculated during periods of rain or when insufficient energy was present. As evident by the variance, the DAR calibration coefficients were remarkably robust to time and varying environmental conditions over the 23-day period, especially in the mid frequency range. In general, the variance was due to calibration coefficients calculated during hours with low energy present and therefore with few seconds sufficiently above the noise floor. Some of the variance can also be explained by artifacts in the recording, local sounds (e.g. physical contact with one system), and perhaps how wind affected the two systems. Variance at the low frequency is partly attributable to the relatively high self-noise and low quality of the microphones. Agreement with the substitution method is fair, less than 3 dB difference for almost all bands, suggesting that field coincident calibration is a viable option.

Using each hour's calibration coefficients, the hourly L_{eq}, L_{10}, L_{50}, and L_{90} SPL metrics were calculated for each one-third octave band and for the wideband A-weighted level. A red energy distribution assumption was used to extrapolate the missing values from the one-third octave bands outside of the 25 Hz–6.3 kHz range. The

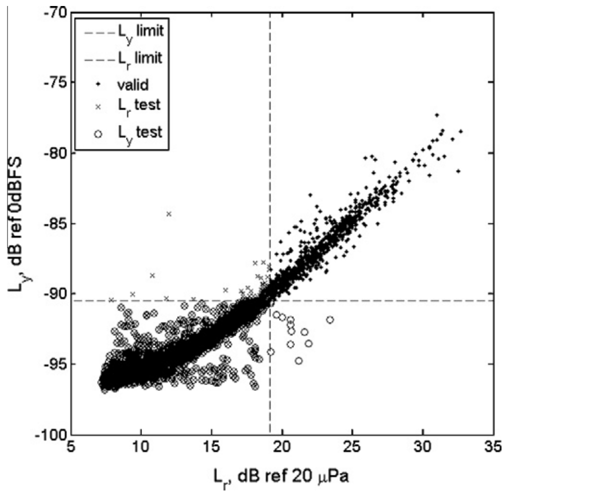

Fig. 8. Corresponding test, P_y, and reference, P_r, 6.3 kHz one-third octave band 1-s L_{eq} values from 1 h along with amplitude thresholds.

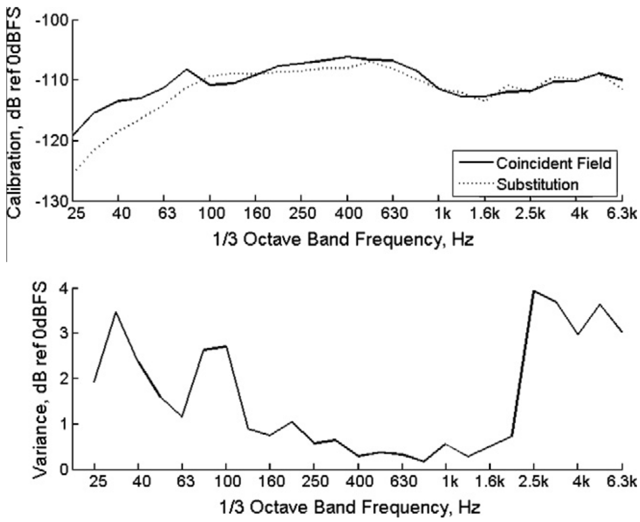

Fig. 9. Mean (top panel) and variance (bottom panel) of the exhaustive coincident field calibration study. The calibration results from the substitution method from Section 3.1 are included.

Table 1
Mean RMS error (dBA) of hourly wideband A-weighted statistics.

	L_{10}	L_{50}	L_{90}	L_{eq}
Comprehensive coincident field calibration	0.94	0.85	0.86	0.82
One hour (worst) coincident field calibration	0.76	0.91	1.12	0.89
Substitution method	0.70	0.73	0.86	0.70

spectral density of red noise is inversely proportional to frequency squared such that power decreases by 6 dB per octave. The error in DAR hourly metrics approached zero when the measurements were above the recorder's noise floor (Table 1). Fig. 10 illustrates the noise floor effects for the 50 Hz, 630 Hz, and 4 kHz one-third octave bands. Although the error of the wideband A-weighted hourly metrics is generally mean zero, the accuracy of these metrics also suffered at low input levels.

4.2. Practical calibration methods

In real applications, an exhaustive coincident field calibration is redundant. The low variance in the exhaustive coincident field

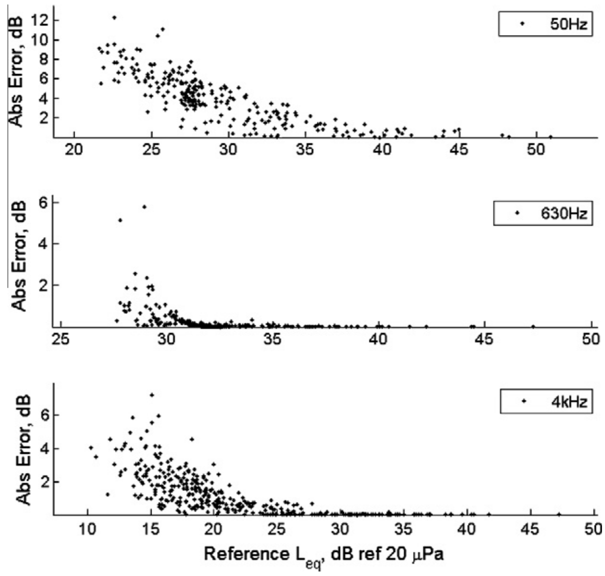

Fig. 10. Error in calculating the hourly one-third octave band L_{eq} plotted versus the reference L_{eq} level. Only reference levels encountered during the 23-day test period are considered.

calibration coefficients, especially in mid-frequency bands, suggests that any daytime hour will give a fairly accurate calibration, especially for calculating aggregate wideband A-weighted sound pressure level metrics. The results of using the worst hour, defined as the one most distant from the mean coefficients, for the entire 23 day period is included in Table 1. Using any given hour to perform the calibration does not result in a significant reduction in accuracy relative to the optimal dynamic calibration from the entire data set. In these data, the worst hour coincident field calibration method even had a lower error in estimating the L_{10}. This is because the comprehensive average set includes nighttime hours with very low levels whereas the "worst hour" was drawn from a restricted set of daytime hours during which there were sufficient SPL to calculate all 25 one-third octave band coefficients. To obtain an accurate coincident field calibration, use up to 1 h of data during an interval when sound levels fall largely above the noise floor and below the clipping limit of the recorder.

The substitution method is another alternative and requires no field measurements. This calibration can bias the measurements when each recorder is not individually calibrated, and because laboratory conditions likely differ from the field site. The hourly A-weighted sound pressure level metrics were calculated for the entire 23 day period using the substitution method calibration coefficients for the recorder used in the coincident field calibration trial (Table 1). Although the generic calibration coefficients determined by the substitution method introduced a small bias or offset, the average error was lower than the comprehensive coincident calibration. The substitution method can be arranged to avoid problems induced by the recorder noise floor and clipping level.

5. Conclusions

These results indicate that one-third octave band level measurements can be obtained from MP3 compressed audio recording made with consumer digital recorders. The lossy MP3 compression format largely preserves the one-third octave spectrum power of the original signal. While there was appreciable variation between the reference and calculated L_{eq} values on the 1 s scale (typically less than 5 dB), hourly summary statistics were very accurate.

Calibration coefficients given by the substitution method or a short coincident field calibration were fairly consistent. A single hour of coincident field calibration can provide an accurate basis for measurements throughout a 23 day period in terms of the metrics studied. However, that hour must be characterized by sound levels high enough to avoid DAR noise floor effects and low enough to avoid clipping. In this study, a method was devised to optimally align 1 s sound samples in the DAR recording with the corresponding SLM measurements. This data alignment provided substantial improvement in calibration accuracy. However, the close agreement reported in Table 1 suggests that hourly order statistics like L_{50} may be sufficient to map DAR measurements into equivalent SLM values without precise time alignment. The substitution method may be attractive for some applications because the range of sound levels used for calibration can be controlled. However, this method requires some care to ensure that sound levels presented during sequential tests truly are equivalent.

With appropriate calibration, consumer audio recorders can provide sound level measurements. Their principal limitation is intrinsic noise levels – likely determined by the microphone/preamplifier quality – especially for sites that have very low ambient sound levels. At high compression ratios (low bitrates), bandwidth is also compromised. However, the digital recorder tested yielded consistent performance over a long time period across a wide range of environmental conditions.

Acknowledgements

We would especially like to thank Damon Joyce, Katy Warner, and the entire Natural Sounds Program team for their assistance in this work. We would also like to acknowledge with sincere appreciation the reviewers for their time, constructive criticism and insight.

References

[1] Celis-Murillo A, Deppe JL, Allen MF. Using soundscape recordings to estimate bird species abundance, richness, and composition. J Field Ornithol 2009;80:64–78.

[2] Patricelli GL, Dantzker MS, Bradbury JW. Differences in acoustic directionality among vocalizations of the male red-winged blackbird (Agelaius pheoniceus) are related to function in communication. Behav Ecol Sociobiol 2007;61:1099–110.

[3] Buckland ST, Anderson DR, Burnham KP, Laake JL. Distance sampling. Encyclopedia of biostatistics. John Wiley & Sons Ltd.; 2005.

[4] Bies DA, Hansen CH. Engineering noise control. 2nd ed. London: Spon Press; 2002.

[5] Miller NP. US national parks and management of park soundscapes: a review. Appl Acoust 2008;69:77–92.

[6] Maher RC. Acoustics of national parks and historic sites: the 8760 h MP3 file. 127th Conv Aud Eng Soc 2009:7893.

[7] Clark CW, Gillespie D, Nowacek DP, Parks SE. Listening to their world; acoustics for monitoring and protecting right whales in an urbanized ocean. In: Kraus SD, Rolland RM, editors. The urban whale: North Atlantic right whales at the crossroads. Cambridge, MA: Harvard University Press; 2006. p. 559.

[8] Thompson ME, Schwager SJ, Payne KB. Heard but not seen: an acoustic survey of the African forest elephant population at Kakum conservation area, Ghana. African J Ecol 2010;48:224–31.

[9] Hill GE, Mennill DJ, Rolek BW, Hicks TL, Swiston KA. Evidence suggesting that ivory-billed woodpeckers (Campephilus principalis) exist in Florida. Avian Conserv Ecol – Écol et Conserv des Oiseaux, 5 January 2011; 2006. p. 3. <http://www.ace-eco.org/vol1/iss3/art2/>

[10] Fristrup KM, Clark CW. Acoustic monitoring of threatened and endangered species in inaccessible areas. Final report for SERDP project SI-1185, 5 January 2011; 2009. <http://www.serdp.org/content/download/6754/88257/file/SI-1185-FR.pdf>.

[11] Clark CW, Fristrup KM. Advanced technologies for acoustic monitoring of bird populations. Final report for SERDP project RC-1461, 5 January 2011; 2009. <http://www.serdp.org/content/download/8819/106844/file/RC-146one-TR.pdf>.

[12] Reed SE. Non-invasive methods to assess co-occurrence of mammalian carnivores. Southwest Nat 2011;56:231–40.

[13] Charif RA, Cortopassi KA, Figueroa HK, Fitzpatrick JW, Fristrup KM, Lammertink M, et al. Notes and double knocks from Arkansas. Science 2005;309:1489.

[14] Miyara F, Accolti E, Pasch V, Cabanellas S, Yanitelli M, Miechi P, et al. Suitability of a consumer digital recorder for use in acoustical measurements. Inter-Noise 2010, Lisbon Portugal; 13–16 June, 2010.
[15] Bean LW. Use of a digital voltmeter to measure the output of a sound level meter. Appl Acoust 1980;13(2):151–7.
[16] ANSI S1.4-1983. Specification for sound level meters; 2006.
[17] Specifications for octave-band and fractional-octave-band analog and digital filters. ANSI S1. One-1986; 1993.
[18] Painter T, Spanias A. Perceptual coding of digital audio. Proc IEEE 2000;88:451–513.
[19] AES standard method for digital audio engineering – measurement of digital audio equipment, AES17-1998 (r2009); 2010.
[20] International electrotechnical commission. sound system equipment – Part 4: microphones, IEC standard 60268-4; 2004.
[21] Zhou B, Green DM. Characterization of external ear impulse responses using Golay codes. J Acoust Soc Am 1992;92:1169–71.
[22] Brent RP. Algorithms for minimization without derivatives. Englewood Cliffs, NJ: Prentice-Hall; 1972.