Audio data compression affects acoustic indices and reduces detections of birds by human listening and automated recognisers

Alexander G. MacPhail, Daniel A. Yip, Elly C. Knight, Richard Hedley, Michelle Knaggs, Julia Shonfield, Emily Upham-Mills & Erin M. Bayne

Published online: 12 Dec 2023.

Submit your article to this journal ↗

Article views: 1105

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

Audio data compression affects acoustic indices and reduces detections of birds by human listening and automated recognisers

Alexander G. MacPhail [a*], Daniel A. Yip [b*], Elly C. Knight[a*], Richard Hedley[c], Michelle Knaggs[d], Julia Shonfield[e], Emily Upham-Mills[f] and Erin M. Bayne[a]

[a]Alberta Biodiversity Monitoring Institute, University of Alberta, Edmonton, AB, Canada; [b]Canadian Wildlife Service, Environment and Climate Change Canada, Delta, BC, Canada; [c]Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada; [d]Department of Renewable Resources, University of Alberta, Edmonton, AB, Canada; [e]LGL Limited Environmental Research Associates, Cambridge, ON, Canada; [f]Ministry of Water, Land and Resource Stewardship, Government of British Columbia, Nanaimo, BC, Canada

ABSTRACT

Increasing popularity in passive acoustic monitoring and the ease with which researchers can accumulate large quantities of acoustic data has resulted in challenges for audio recording storage, archiving, and management. Reductions in file size can be achieved by lowering sample rate and compressing to different formats; however, how these processes affect audio data quality, and the resulting interpretation of wildlife data is not well understood. We investigated the effect of sampling rate and lossy compression of audio recordings to MP3 from their native WAV format on the performance of four commonly applied avian bioacoustic applications: community listening, distance estimation, automated recognition, and acoustic indices. Compression to MP3 decreased the number of detections, including a reduction in total abundance of individuals when transcribing audio files for community listening and lower precision and recall for automated recognisers. Sampling rate reduction introduced systematic bias to acoustic indices and had an influence on precision and recall for recognisers as well. We recommend against the use of MP3 compression to reduce file volume and suggest other lossless forms of audio compression where an exact copy of the original recording can be recovered.

Introduction

Ecologists are increasingly using passive acoustic monitoring as an important tool for assessing acoustically active wildlife (Shonfield and Bayne 2017; Gibb et al. 2019; Sugai et al. 2019b). Technological advances have reduced the cost of acquiring passive acoustic sensors while increasing the ability to record sounds over extended durations (Gibb et al. 2019). The amount of data we can collect now greatly outpaces the rate at which human observers can transcribe audio recordings. To combat this challenge, developments in

CONTACT Daniel A. Yip ✉ Daniel.Yip@ec.gc.ca

*These authors equal contributors to this work.

automated processing methods have increased our ability to process large volumes of data. Two major developments include 1) computer-based, automated recognition algorithms (Knight et al. 2017; Kahl et al. 2021), where scores are assigned to acoustic signals based on their similarity to the target or focal species, and 2) the calculation of acoustic indices (Sueur et al. 2008), where measures of different patterns in energy distribution within an audio recording help with identifying specific sounds, species, or events from audio recordings.

One of the biggest remaining challenges in acoustic monitoring is the storage of these large amounts of audio recordings (Shonfield and Bayne 2017; Gibb et al. 2019; Sugai et al. 2019b). For example, monitoring programs can rapidly accumulate terabytes or petabytes of audio data (Truskinger et al. 2014), which should be archived in long-term data storage to 1) be in an organised, consistent location while awaiting manual processing or transcription by human observers, and 2) use for other objectives as needed, particularly as future advances in automated recognition and processing technology and methods allow for the extraction of additional information. Current standards in bioacoustic research recommend the storage of audio recordings in the highest quality format, typically Waveform Audio File Format (WAV) or using other lossless compression algorithms (Villanueva-Rivera et al. 2011; Browning et al. 2017). While this standard retains the most information, costs associated with easily accessible, reliable, and long-term storage of large-volume WAV files can quickly become prohibitive. Furthermore, processing efficiency can be reduced due to the transfer of data to and from storage locations and the computational demands of analysing large files (Truskinger et al. 2014).

Reduction in data volume to reduce data storage requirements can be achieved by reducing the intensity of recording schedules, or adjusting recording settings such as sampling rate, depending on research objectives (Sugai et al. 2019a). Sampling rate is defined as the number of amplitude measurements recorded per unit time, measured in Hz, and must be at least double the maximum frequency of the focal signal according to the Nyquist-Shannon sampling theorem (Sugai et al. 2019a). Sampling rate corresponds to the maximum frequency that is captured in an audio recording, and file volume also increases linearly with sampling rate, so selecting an appropriate frequency depending on the species to be sampled can reduce file size substantially. Data can also be deleted after processing, but this is often undesirable as future research questions could be addressed using the same audio dataset.

Alternatively, data can be compressed to substantially reduce data storage requirements. Audio data compression falls into two categories: lossless and lossy. Lossless compression algorithms, such as FLAC, retain all information from the original audio recording but typically offers less reduction in data volume relative to lossy compression. Lossy compression algorithms offer significantly greater storage savings but can degrade audio quality at higher compression levels. Furthermore, lossy compression algorithms cannot recover information that is removed during the compression process. Current bioacoustic protocols recommend against storage of audio data in lossy formats (Villanueva-Rivera et al. 2011; Browning et al. 2017); however, the influence on acoustic identification of wildlife is unclear. Trade-offs between data volume and data compression are likely required to make long-term storage of audio recordings sustainable.

In this study, we explore the effect of changes to sampling rate and the use of lossy MPEG-1 Audio Layer III (hereafter 'MP3') encoded audio to reduce the volume of

acoustic data and investigate the effect on acoustic wildlife identification. A typical 10-min acoustic wildlife survey recorded at a standard sample rate of 44,100 Hz using 16 bits per sample in WAV format has a file size of approximately 100 megabytes (Villanueva-Rivera et al. 2011). Sampling rate is directly proportional to file volume (i.e. a recording in WAV format at half the sampling rate is half the volume), and MP3 encoding, one of the most widely used forms of lossy audio compression, can reduce file size by 60–90%. This compression method relies on removing extraneous or redundant data within an audio recording based on the typical auditory resolution of the human ear through pattern recognition and signal prediction (Jayant et al. 1993). Data that are not perceivable to listeners are eliminated and signals that fall within the optimum range for human hearing are prioritised relative to signals outside the range. File size is also reduced by reducing or removing lower quality signals when they overlap with higher quality signals, both spatially and temporally. These factors become more prevalent as signal complexity increases. By assigning more emphasis to higher amplitude or intensity signals and reducing emphasis on signals that are masked, MP3 compression will keep components of an audio recording perceived to be the most representative of the original recording while reducing the amount of information and minimising perceptual distortion to the human auditory system.

The principal concern of audio data compression methods for identifying wildlife is that data compression may alter or remove important information for successfully detecting or identifying species or ecological events. Audio compression has been shown to have little effect on bird identification by human observers when compression quality is high (Rempel et al. 2005) and decrease precision when measuring fine-scale signal structure elements (Araya-Salas et al. 2019), but to our knowledge, there are no empirical studies that attempt to disentangle the effects of compression rate or investigate the influence on common automated methods for processing audio wildlife data. Additionally, since MP3 encoding reduces bitrate by removing information based on human perception, computer-based processing methods such as automated recognition or acoustic indices may be more sensitive to information loss that is not perceived by human listeners, and therefore more strongly influenced by data compression methods. Likewise, recommendations for sample rate are well documented, but based on the perception of human listeners. Impacts on automated processes, which can draw on elements of the spectrogram outside of the range of human hearing and the target signals (i.e. harmonics and whole soundscapes), are not well understood. Therefore, our objectives were to determine how the use of a lossy compression algorithm (WAV to MP3) and variation in sample rate may affect the detection and identification of wildlife from audio recordings. We investigated the effects of audio compression and sampling rate reduction for four different common ecological applications of soundscape audio recordings, 1) human listening for community composition ('community listening'), 2) detection distance of animals by human listeners ('perceptibility'), 3) automated recognition of focal species ('recognizers'), and 4) calculation of acoustic indices ('acoustic indices'). We hypothesised that audio compression and sampling rate reduction will 1) reduce the species richness and abundance of birds detected in community listening, 2) reduce the perceptibility of human listeners, 3) reduce the precision and recall of automated recognisers, and 4) alter the calculation of acoustic indices.

Materials and methods

We obtained audio recordings collected with Wildlife Acoustics Song Meter (Wildlife Acoustics, Maynard, Massachusetts, USA) autonomous recording units (ARUs) in northern Alberta, from the WildTrax data repository (www.wildtrax. ca). All recordings were collected between 1 May and 20 July (2013–2017) when breeding birds are actively singing. Recordings were collected with SM2+ ARUs for the perceptibility application, SM2+ and SM4 ARUs for the community listening and acoustic indices applications, and SM4 ARUs for the recogniser application. All original recordings were collected in stereo WAV format at a sample rate of 44,100 Hz and standard bitrate of 1411 kbps. We used the LAME MPEG Audio Layer III (MP3) encoder licenced under the LGPL to compress those original files using constant bitrate to MP3 with a 320 bitrate for all four applications as well as MP3 with a 96 bitrate for community listening, recognisers, and acoustic indices (Table 1, Figure 1). We then downsampled those recordings to 22,050 and 32,000 Hz for the recogniser and acoustic index applications. We investigated the effects of sample rate on our two automated applications only because sample rate recommendations for human listening are well documented and effort for human processing of audio data is high. We recorded the file size for a 1-min recording for each combination of compression type and sample rate to provide context on data storage savings (Table 2).

Community listening

We selected 56 3-min audio recordings for community listening and compressed the original files to MP3 at 320 kbps and 96 kbps. The recordings were randomly selected

Table 1. Sample size of compression type and sample rate treatments on audio recordings collected with autonomous recordings in northern Alberta and processed with four different common ecological applications of soundscape audio recordings.

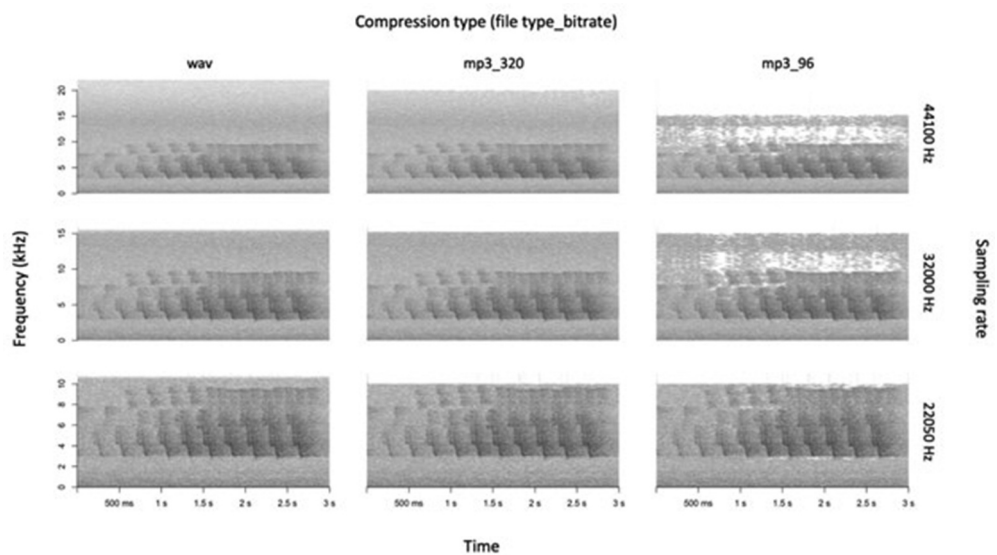| Compression rate (Hz)	Compression type (file type_bitrate)	No. of recordings processed per application			
		1. Community listening	2. Perceptibility	3. Recognizers	4. Acoustic indices
44,100	WAV	56	3150	100 per species	142
44,100	MP3_320	56	3150	100 per species	142
44,100	MP3_96	56	0	100 per species	142
32,000	WAV	0	0	100 per species	142
32,000	MP3_320	0	0	100 per species	142
32,000	MP3_96	0	0	100 per species	142
22,050	WAV	0	0	100 per species	142
22,050	MP3_320	0	0	100 per species	142
22,050	MP3_96	0	0	100 per species	142

Figure 1. Spectrograms of an audio recording of an Ovenbird (*Seiurus aurocapilla*) song collected in northern Alberta on J3 June 2015 under various compression type and sample rate treatments.

Table 2. File size and percentage of 44100 Hz WAV file size for 1-min audio files under various combinations of sample rate (Hz) and compression type (combination of file type and MP3 bitrate).

		Sample rate (Hz)		
		22050	32000	44100
Compression type (file type_bitrate)	WAV	5.3 MB (50.0%)	7.7 MB (72.6%)	10.6 MB (100.0%)
	MP3_320	1.6 MB (15.1%)	2.4 MB (22.6%)	3.3 MB (31.1%)
	MP3_96	0.7 MB (6.6%)	0.7 MB (6.6%)	0.7 MB (6.6%)

from the Bioacoustic Unit's multi-disciplinary data sets to reflect habitat and species communities across northern Alberta. These recordings were randomly assigned to three observers who annotated species occurrence to the level of individual animals for each minute of each recording, following a standardised acoustic recording analysis protocol (The Bioacoustic Unit 2019). Each observer was highly experienced, with a history of conducting point count surveys and having processed over 100 h of audio data from Western Canada. Observers had no knowledge of the file format, file size, location, time of survey, or species present. We calculated species richness as the total number of bird species detected to the species level in each recording and abundance as the total number of individual birds detected to the species level in each recording.

We modelled the effect of audio compression on community listening using mixed-effects linear regression in a Bayesian framework with vague priors (normal(0,10)) in the package *brms* (Bürkner 2017). For each response variable (richness, abundance), we built a model with compression treatment as a categorical variable and recording name and observer as random effects on the intercept. All models were run for three chains with a burn-in of 1,000 and a subsequent 20,000 iterations. We inspected all fixed and random effects for chain convergence, and ensured all rhat values were less than 1.10 and all

effective sample size ratios were greater than 0.1. We used 95% credible intervals to determine the presence of between-group differences with 95% confidence on average (Hespanhol et al. 2019).

Perceptibility

We used a sound broadcast experiment (see Yip et al. 2017) to test whether file compression influenced an observer's ability to detect different species at varying distances. We broadcast 23 different avian species of various frequencies with a speaker, re-recorded them using an ARU, and compressed the original files to MP3 at 320 kbps. These species included: Clay-coloured Sparrow (*Spizella pallida*), Black-and-white Warbler (*Mniotilta varia)*, Lincoln's Sparrow (*Melospiza lincolnii*), Brown-headed Cowbird (*Molothrus ater*), Red-breasted Nuthatch (*Sitta canadensis*), Dark-eyed Junco (*Junco hyemalis*), White-throated Sparrow (*Zonotrichia albicollis*), Cape May Warbler (*Setophaga tigrina*), Common Raven (*Corvus corax*), Belted Kingfisher (*Megaceryle alcyon*), Olive-sided Flycatcher (*Contopus cooperi*), Pine Siskin (*Spinus pinus*), Tennessee Warbler (*Oreothlypis peregrina*), Warbling Vireo (*Vireo gilvus*), Rose-breasted Grosbeak (*Pheucticus ludovicianus*), Ovenbird (*Seiurus aurocapilla*), Yellow Rail (*Coturnicops noveboracensis*), Northern Saw-whet Owl (*Aegolius acadicus*), Boreal Owl (*Aegolius funereus*), Long-eared Owl (*Asio otus*), Great Gray Owl (*Strix nebulosa*), and Barred Owl (*Strix varia*). We then randomised WAV and MP3 sounds and presented them to two observers who recorded whether a sound was detected and the species identification of that sound by sight, using visual scanning of spectrograms, and by sound, using standardised headphones and volume. Observers were aware of the sounds that could be detected but not the order or distance.

We estimated detection distance for each species or tone with generalised linear mixed-effects models in the package lme4 (Bates et al. 2015). We used a binomial distribution and logit link with detection of a sound as the response variable, various combinations of compression method and distance of sound as predictor variables, and species and observer as random effects on intercept. We ranked models by AICc and selected the model with the lowest AICc, or the simplest model when multiple candidate models had a deltaAICc < 2 (Arnold 2010, Table 3). We then calculated the effective detection radius (Buckland et al. 2001) for each compression treatment (WAV, MP3–320 kbps) using binomial generalised linear models with a 'cloglog' link function and Monte Carlo simulation (see Yip et al. 2017 for details). We interpreted statistical differences in detection distance between WAV and MP3 using 83% confidence intervals (Krzywinski and Altman 2013).

Table 3. Model selection for mixed effects models of detectability of audio recordings of 23 different avian species and tones recorded at various distances ('distance') for uncompressed and compressed acoustic recordings ('compression').

Model	K	AICc	DAICc	Wt
p = distance + compression	5	4534.4	1.64	0.30
p = distance	4	4532.8	0.00	0.70
p = compression	4	6169.8	1637.02	0.00
p = 1	3	6168.0	1635.26	0.00

Recognizers

To investigate the effects of file compression on the performance of automated recognisers, we used single-species convolutional neural network recognisers (CNNs, previously described in Yip et al. 2020) to detect vocalising Ovenbirds (*Seiurus aurocapilla*) and Common Nighthawks (*Chordeiles minor*). We selected these two bird species because they differ in acoustic signal complexity that could interact with the effects of compression and sample rate. Ovenbirds have a complex multi-phrased song that is 2.5–4.0 s in length (Porneluzi et al. 2011), whereas Common Nighthawks produces a simple single-note vocalisation that is ~0.3 s in length. We obtained 100 10-min ARU recordings with confirmed detections for each focal species (from Knight et al. 2020). We compressed each recording for all combinations of recording quality and sampling rate (Table 1). We processed each recording through the matching single-species CNN and validated all detections above a classification probability of 0.1 manually. We calculated precision as the proportion of detections that were true positives in each recording at each sample rate and recording quality combination. We built a 'gold standard' set of detections by pooling the detections across all versions of recording by detection time stamp and calculated recall as the proportion of those gold standard detections that were true positives in each recording at each sample rate and recording quality combination. We calculated score as the mean confidence score value per recording at each sample rate and recording quality combination.

We modelled the effect of audio compression on recogniser performance using mixed effects beta regression in a Bayesian framework with vague priors (normal(0,10)) in the package *brms* (Bürkner 2017). For each combination of species (Common Nighthawk, Ovenbird) and metric (precision, recall, score), we built a model with sample rate and compression treatment as categorical variables, as well as an interaction between the two and recording name as a random effect on intercept. All models were run for three chains with a burn-in of 1,000 and a subsequent 10,000 iterations. We inspected all fixed and random effects for chain convergence, and ensured all rhat values were less than 1.10 and all effective sample size ratios were greater than 0.1. We used 95% credible intervals to determine the presence of between-group differences with 95% confidence on average (Hespanhol et al. 2019).

Acoustic indices

We selected 142 10-min recordings and compressed the original files to MP3 at 320 kbps and 96 kbps and reduced the sample rate to 32,000 Hz and 22,050 Hz. We calculated the acoustic complexity index (ACI; Pieretti et al. 2011) and acoustic diversity index (ADI; Villanueva-Rivera et al. 2011) for each recording under each compression type and sample rate treatment combination using the *soundecology* package (Villanueva-Rivera and Pijanowski 2018).

We modelled the effect of audio compression on acoustic indices using mixed-effects linear regression in a Bayesian framework with vague priors (normal(0,1000) for ACI; normal(0,10) for ADI) in the package *brms* (Bürkner 2017). For each acoustic index (ACI, ADI), we built a global model with sample rate and compression treatment as

categorical variables, as well as an interaction between the two and recording name as a random effect. All models were run for three chains with a burn-in of 1,000 and a subsequent 20,000 iterations. We inspected all fixed and random effects for chain convergence, and ensured all rhat values were less than 1.10 and all effective sample size ratios were greater than 0.1. We used 95% credible intervals to determine the presence of between-group differences with 95% confidence on average (Hespanhol et al. 2019).

Results

File compression and reduction in sample rate reduced audio file size anywhere from 72.6% (32,000 Hz WAV) to 6.6% (MP3 96 kbps) of the original size (Table 2). File compression from WAV to MP3 provided the largest file size savings in comparison to sample rate reduction, which was relatively linear.

Community listening

There was an effect of file compression to 96 kbps MP3 on the total abundance of individuals detected during human listening (95% CI = −1.31 to −0.03; Figure 2). There was also a negative effect of file compression to 96 kbps on species richness; however, the 95% CI slightly overlapped zero (95% CI = −0.86 to 0.04). All other 95% CIs strongly overlapped zero.
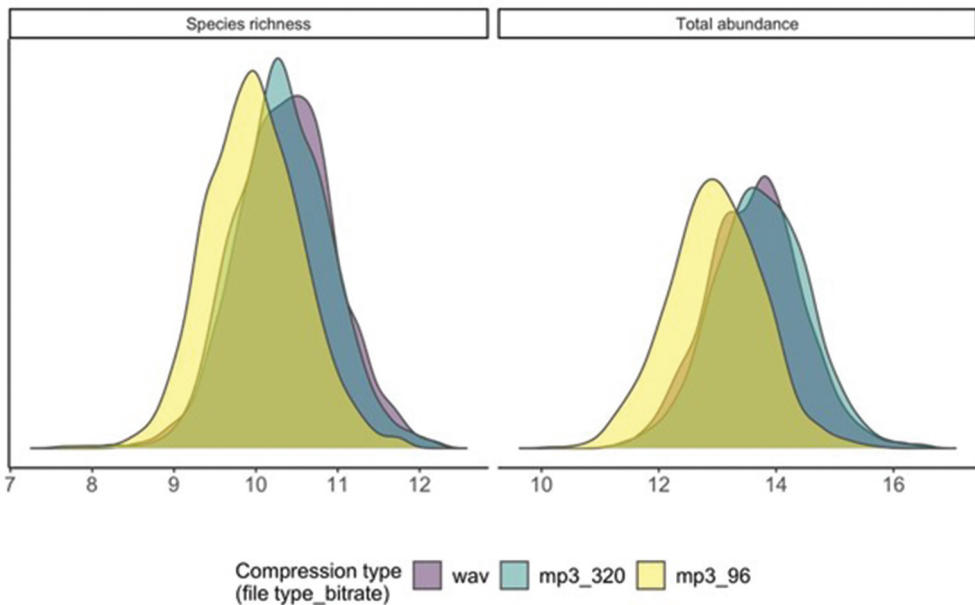

Figure 2. Posterior predictions (*n* = 1000) of species richness and total abundance of individual birds per acoustic recording for community listening under file compression treatments (combination of file type and MP3 bitrate).

Perceptibility

The selected model that best explained the probability of detecting a species or sound contained only a negative effect of distance (Table 3). There was no statistical difference in 83% confidence intervals between the effective detection radius of sounds recorded in WAV and compressed in MP3 (Figure 3).
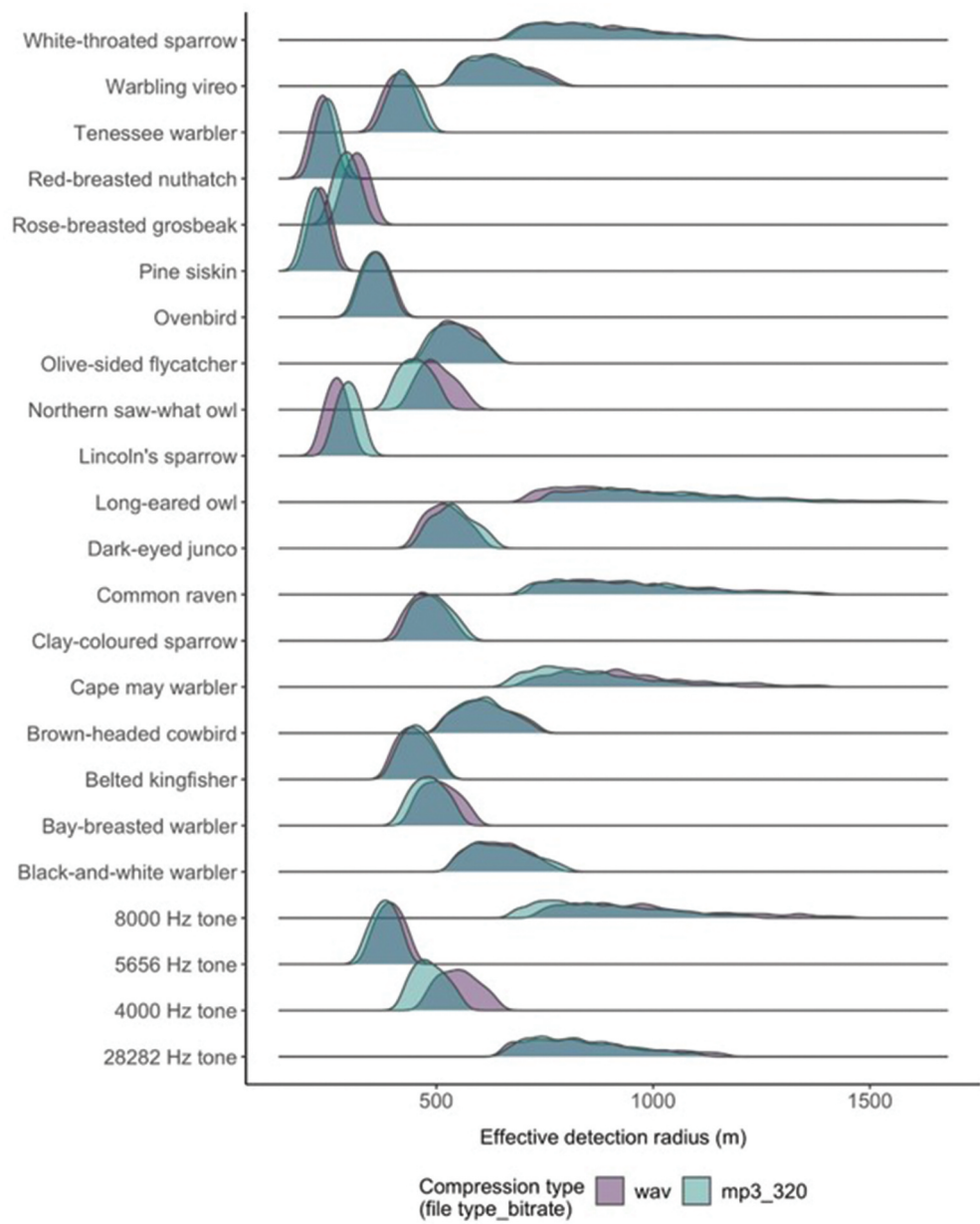

Figure 3. 83% quantile of 1000 bootstraps for effective detection radius of audio recordings of 23 different avian species and tones recorded at various distances for uncompressed and compressed acoustic recordings.

Recognizers

There was an effect of sample rate on precision of the Ovenbird recogniser, with higher precision for 44,100 Hz (beta 95% CI = 0.55–1.04) than the other sample rates (Figure 4). There was also an effect of sample rate on recall of both recognisers, with lower recall for 44,100 Hz (Common Nighthawk beta 95% CI = −0.71 to −0.09; Ovenbird beta 95% CI = −0.60 to −0.33).

There was a strong effect of compression type on recall of both recognisers, with lower recall for the 96 bit rate MP3 recordings (Common Nighthawk beta 95% CI = −1.09 to −0.43; Ovenbird beta 95% CI = −0.61 to −0.33). There was also a significant interaction between 44,100 Hz sample rate and 96 kbps MP3 compression for the Ovenbird recogniser (beta 95% CI = −0.45 to −0.06).

There was an effect of sample rate on recording mean score value with lower values at higher sample rates for both species (Common Nighthawk 32,000 Hz beta 95% CI = −0.12 to −0.01; Common Nighthawk 44,100 Hz beta 95% CI = −0.20 to −0.09; Ovenbird 32,000 Hz beta 95% CI = −0.35 to −0.06, Ovenbird 44,100 Hz beta 95% CI = −0.51 to −0.23). There was also an effect of 96 kbps MP3 compression on recording mean score value (beta 95% CI = −0.31 to −0.20) and an interaction between 96 kbps compression at 32,000 Hz sample rate (beta 95% CI = 0.10 to 0.26) for Common Nighthawk only. All other 95% CIs overlapped zero.
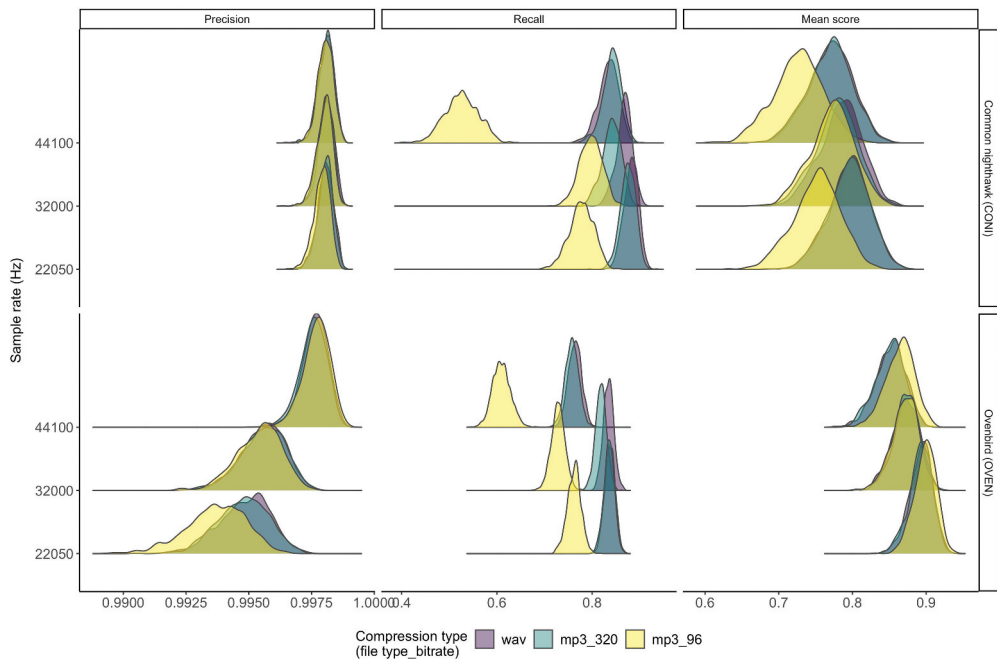

Figure 4. Posterior predictions ($n = 1000$) of precision, recall, and mean score probability per acoustic recording for two single-species deep learning recognisers species under various combinations of sample rate (Hz) and compression type (combination of file type and mp3 bitrate).
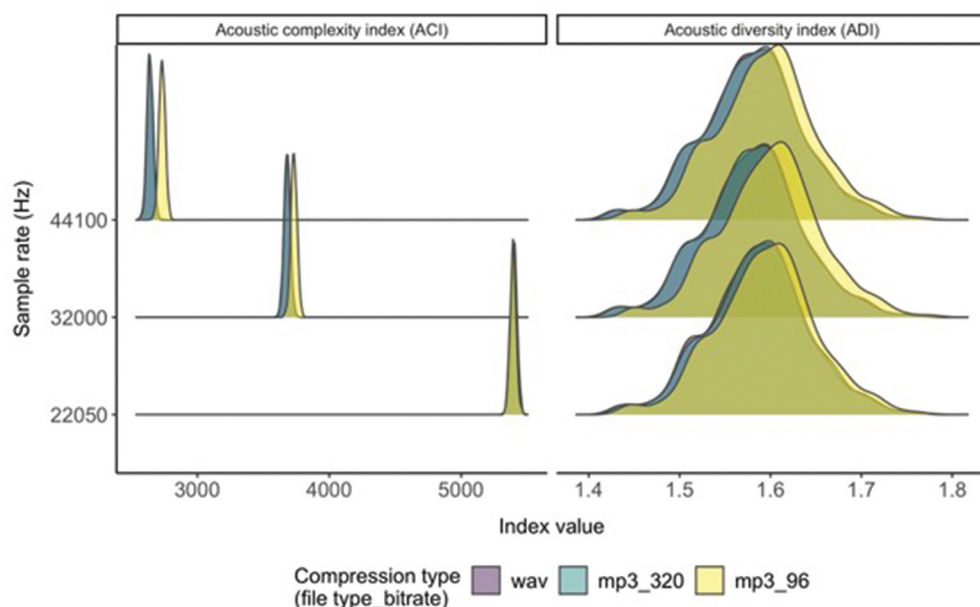
Figure 5. Posterior predictions ($n = 1000$) of acoustic index value for two commonly used indices under various combinations of sample rate (Hz) and compression type (combination of file type and MP3 bitrate) of acoustic recording.

Acoustic indices

There was an effect of sample rate on ACI, with the lowest ACI values for 44,100 Hz and intermediate values for 32,000 Hz (44100 Hz beta 95% CI = −2786.6 to −2733.21; 32000 Hz beta 95% CI = −1746.7 to −1693.2; Figure 5). There was also an interaction between sample rate and the 96 bit rate MP3 compression for ACI, with 96 bit rate compression causing an increase in ACI value, particularly at high sample rates. There was no direct effect of compression type on ACI, but there was a marginal effect of compression type on ADI, with higher ADI values for files with 96 bit rate MP3 compression (MP3_96 beta 95% CI = 0.00 to 0.02). All other 95% CIs overlapped zero.

Discussion

We investigated the effect of compression rate (WAV − 1,411 kbps, and MP3–320 and 96 kbps) and sample rate (41,100 Hz, 32,000 Hz and 22,050 Hz) on community listening, human perceptibility, automated recognition, and acoustic indices. We found no statistical differences in the distance at which different species and sounds were detected; however, both species richness and the abundance of individuals tagged on a recording decreased when audio was compressed to 96 kbps. For recognisers, compression, particularly to 96 kbps, impacted recall of both our focal species, and sample rate affected both recall and precision. Compression had a marginal effect on acoustic indices and there was an effect of sample rate specifically on ACI.

The applications that were most affected by compression were those that rely on the spectrogram. For community listening, a large component of the bird community can be identified by listening alone. However, viewing the spectrogram has been shown to increase annotation accuracy, particularly when there is overlap between species and individuals, improve validation when there is uncertainty in identification, and increase accuracy when estimating abundance of individuals (Ware et al. 2023). While compression to MP3 removes extraneous or redundant information based on the auditory resolution of the human ear, this process results in significant visual deterioration of the spectrogram and thus impacts the ability of an observer to detect signals, particularly those that are fainter or masked by other sounds (Figure 1). We did not, however, find an effect of compression on perceptibility despite a visual component to this process, perhaps because we did not include the MP3 96 kbps compression treatment like in the other applications. Alternatively, the perceptibility application used recordings of single sounds as opposed to complex soundscape recordings like the community listening and recogniser application and so this process may be less susceptible to the loss of information in the spectrogram. We found the strongest effects of compression on automated recognition, likely because the recognisers we used operated directly on the spectrogram. Automated recognition performance is also sensitive to the parameters used to create the spectrogram itself (Knight et al. 2019, refs within). We showed that the highest compression rates were associated with lower confidence score assigned by the recognisers, suggesting that the deterioration of the spectrogram caused by lossy compression results in the removal of elements of the spectrogram that decrease the similarity of signals in audio data to the recogniser training data. Future research should investigate whether there are also effects of training recognisers with compressed data.

The importance of the spectrogram for automated recognition was also emphasised by the effect of sample rate on recall and precision. The positive effect of sample rate on precision was only present for Ovenbird, possibly due to the loss of repetition of the spectrographic pattern that is present in harmonics higher frequencies, even though the CNN was trained with 22.05 kHz data. In comparison, there was no effect of sample rate on Common Nighthawk precision, perhaps because this species has a much simpler acoustic signal and spans a smaller bandwidth. Alternatively, the CNN may simply be overfit to the sample rate of the 22.05 kHz data. Graciarena et al. (2010) previously found no improvement in equal error rate above a frequency range of 13 kHz (26,000 Hz sampling rate); however, they were using recognisers that operated on cepstral coefficients and were therefore likely less sensitive to spectrogram parameters than our full spectrogram CNNs. Contrary to our predictions, the effect of sample rate on recall was negative for both species. Our analysis of score suggested that this relationship was driven by a similarly negative relationship between sample rate and score, whereby because score values were lower, fewer detections exceeded our selected score threshold (0.1) and therefore, recall was lower. We suggest the negative relationship between score and sample rate was because the FFT size of the convolutional neural network (CNN) recogniser was the same across all treatments (128 samples), and so the CNN was operating on spectrograms with coarser frequency resolutions and finer temporal resolutions for higher sample rates. We have previously shown CNN recogniser recall increases with spectrogram frequency resolution, but is highest at intermediate temporal resolution (Knight et al. 2019). Alternatively, it may be because the CNN was trained with

samples recorded at 22,050 Hz, and so the negative effect of sample rate on recall may have been derived by deviation between training and test resolution, not absolute resolution. Future research should endeavour to separate these two processes.

Both of the acoustic indices tested demonstrated effects of sample rate or compression. ADI values, which are calculated by dividing the spectrogram into bins and taking the proportion of the signals in each bin above a power threshold (Villanueva-Rivera et al. 2011), were therefore insensitive to sample rate; however, they were marginally affected by compression, likely because the removal of some information via WAV compression results in lower power values. In contrast, ACI values, which are calculated using the relative difference in intensity in adjacent spectrogram pixels within a frequency bin (Pieretti et al. 2011), were not affected by compression but were strongly affected by sample rate. Our audio recordings primarily contained bird song and ambient sound, which occupy the mid and lower bands of the spectrogram. Higher sample rates increase the frequencies that are recorded, and in these recordings, the unoccupied higher frequency bands result in a lower relative difference in sound intensity and therefore lower ACI scores. Acoustic indices are used in a variety of applications, including event and species identification, as a proxy for species richness, and to summarise the composition of the soundscape and anthropogenic activity (Towsey et al. 2014; Phillips et al. 2018), but the impacts on the interpretation of these applications due to species richness are not well understood. Careful consideration on the effects of sample rate and how ACI and other acoustic indices are interpreted is required until these effects are better understood.

Recommendations for selecting appropriate sample rates are well documented for manual processing of audio data (Sugai et al. 2019a). We therefore did not include sample rate in the community listening and perceptibility applications because it would have increased data processing effort by 3–4 times. Recommendations for automated processing methods are unknown, however, and we showed systematic effects from changes to sample rate in automated processing methods. We suggest mixing of sample rates in analyses is not recommended, although these issues could be potentially addressed at the analysis stage. More importantly, sample rate is a tradeoff when using recognisers, with precision increasing and recall decreasing at higher sampling rates than the training data. Sampling rate choice will depend on how users prioritise precision vs recall in their recogniser outputs and will have to be considered in concert with score threshold choice, which similarly balance precision and recall (Katz et al. 2016; Knight et al. 2017; Knight & Bayne 2019; Priyardarshani et al. 2018). Alternatively, ensemble CNNs that are trained on multiple sample rates of recordings may provide overall improved performance.

When processing community listening data, MP3 compression likely impacts certain types of sounds and signals more than others. Since MP3 compression algorithms prioritise the removal of data that are less likely to be perceived by the human auditory system, high-frequency species, quieter and more distant individuals, and songs/vocalisations that overlap temporally or in frequency are more likely to be filtered out (Jayant et al. 1993). This targeted data removal may therefore decrease detectability for a specific subset of species (e.g. high-frequency warblers, species like aerial insectivores that spend significant time at range from ground-based microphones), which would in turn impact estimates of community composition and richness. Busy recordings with high species richness

may also be more impacted by audio compression due to an increase in over-lapping species and sounds. Likewise, for recogniser data, precision and recall might be differentially affected for different species by compression depending on the same variables mentioned above, as seen here with Common Nighthawk and Ovenbird. Although we found small effect sizes of compression on these various applications, it is possible they may still have downstream effects on the application of that data. For example, a difference of one individual between compressed and uncompressed recordings could have effects on estimates of detection probability and occupancy or abundance (Johansson et al. 2020; Schmidt et al. 2023), and could be particularly problematic if compression type is confounded with other variables in the analysis.

Although compression of audio data has clear benefits for challenges in file storage and processing efficiency of bulky audio files, systematic effects on the identification of wildlife and the significance of those effects to interpretation of statistical analyses should not be ignored. Given these effects of both sample and compression rate on multiple, frequently used approaches for processing wildlife audio data, we recommend the following guidelines and standard practices when collecting, storing, and analysing audio recordings:

(1) Do not combine files of varying sample and/or compression rates or account for this as a random effect statistically in analysis. Both factors introduce systematic variation which could obscure patterns in species occurrence and community composition and cause erroneous interpretation from acoustic indices.

(2) Compress to a lossless audio format such as FLAC (Huang et al. 2014) if possible. This will allow for some data volume savings, but retain important information within the audio, the loss of which we have shown to be important for both human-based and automated approaches for processing data. FLAC reduces data volume by 30–70% (depending on compression rate) and requires minimal processing time when converting from WAV.

(3) Expansion of passive acoustic monitoring globally has increased opportunities for the use of 'big data' in large scale analyses. Carefully consider the costs and benefits of audio compression, both within and across research projects and monitoring programs. Try to adhere to the same parameters as other researchers and collaborators as this ensures easier integration of datasets from different sources.

(4) If recogniser precision is a high priority, use the highest sampling rate possible but ensure the score threshold is adjusted appropriately so recall is not adversely impacted. If recall is a priority, and lower sampling rates are used, be prepared to use additional effort in validation due to lower precision. Note that these recommendations likely only apply to algorithms that operate directly on full spectrograms.

In recent years, the scope and scale of acoustic monitoring programs and the questions and challenges they address have increased significantly (Shonfield and Bayne 2017; Sugai et al. 2019b). An important step in this growth is the development and accessibility of platforms and repositories for storing, processing, and sharing large datasets (Sugai et al. 2019b). Platforms such as WildTrax (https://wildtrax.ca) and

Arbimon (https://arbimon.rfcx.org), both which store audio in FLAC, are examples of repositories where users can store, analyse, and easily share acoustic monitoring data. However, collation of large datasets such as these require adherence to a basic set of criteria for successful data integration and subsequent analysis. Our study contributes to a growing set of community best practices for passive acoustic monitoring that continue to improve the value of large-scale audio datasets for tackling current and future ecological analyses.

Acknowledgements

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Alexander G. MacPhail http://orcid.org/0000-0002-7968-6281
Daniel A. Yip http://orcid.org/0000-0002-3130-0001

Author contributions

AGM, DAY, ECK, RH, MK, JS, EUM, and EMB conceived the ideas and designed the manuscript. AGM, DAY, ECK, RH, MK, JS, and EUM collected the data. DAY and ECK analysed the data. AGM, DAY, and ECK led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

References

Araya-Salas M, Smith-Vidaurre G, Webster M. 2019. Assessing the effect of sound file compression and background noise on measures of acoustic signal structure. Bioacoustics. 28(1):57–73. doi: 10.1080/09524622.2017.1396498.

Arnold TW. 2010. Uninformative parameters and model selection using Akaike's information criterion. J Wildl Manage. 74(6):1175–1178. doi: 10.1111/j.1937-2817.2010.tb01236.x.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw. 67:1–48. doi: 10.18637/jss.v067.i01.

Browning E, Gibb R, Glover-Kapfer P, Jones KE. 2017. Passive acoustic monitoring in ecology and conservation. WWF Conserv Technol Series. 1:1–74.

Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L. 2001. Introduction to distance sampling: estimating abundance of biological populations. 2nd ed. New York, NY, USA: Oxford University Press.

Bürkner PC. 2017. brms: An R package for bayesian multilevel models using Stan. J Stat Soft. 80 (1):1–28. doi: 10.18637/jss.v080.i0.

Gibb R, Browning E, Glover-Kapfer P, Jones KE, Börger L. 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecology Evol. 10(2):169–185. doi: 10.1111/2041-210X.13101.

Graciarena M, Delplanche M, Shriberg M, Stolcke A, Ferrer L. 2010. Acoustic front-end optimization for bird species recognition. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing; 293–296. doi: 10.1109/ICASSP.2010.5495923.

Hespanhol L, Vallio CS, Costa LM, Saragiotto BT. 2019. Understanding and interpreting confidence and credible intervals around effect estimates. Braz J Phys Ther. 4(4):290–301. doi: 10.1016/j.bjpt.2018.12.006.

Huang H, Shu H, Yu R. 2014. Lossless audio compression in the new IEEE standard for advanced audio coding. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP): 6934–6938. doi: 10.1109/ICASSP.2014.6854944.

Jayant N, Johnston J, Safranek R. 1993. Signal compression based on models of human perception. Proceedings of the IEEE. 81:1385–1422.

Johansson O, Samelius G, Wikberg E, Chapron G, Mishra C, Low M. 2020. Identification errors in camera-trap studies result in systematic population overestimation. Sci Rep. 10(1):6393. doi: 10.1038/s41598-020-63367-z.

Kahl S, Wood CM, Eibl M, Klinck H. 2021. BirdNET: a deep learning solution for avian diversity monitoring. Ecol Inform. 61:101236. doi: 10.1016/j.ecoinf.2021.101236.

Katz J, Hafner SD, Donovan T. 2016. Assessment of error rates in acoustic monitoring with the R package monitoR. Bioacoustics. 25:177–196. doi: 10.1080/09524622.2015.1133320.

Knight EC, Bayne EM. 2019. Classification threshold and training data affect the quality and utility of focal species data processed with automated audio-recognition software. Bioacoustics. 28:539–554. doi: 10.1080/09524622.2018.1503971.

Knight EC, Hannah KC, Foley GJ, Scott CD, Brigham RM, Bayne EM. 2017. Recommendations for acoustic recognizer performance assessment and applications to five common automated signal recognition programs. ACE. 12(2):14. doi: 10.5751/ACE-01114-120214.

Knight EC, Solymos P, Scott CD, Bayne EM. 2020. Validation prediction: a flexible protocol to increase efficiency of automated acoustic processing for wildlife research. Ecol Appl. 30:e02140. doi: 10.1002/eap.2140.

Krzywinski M, Altman N. 2013. The meaning of error bars is often misinterpreted, as is the statistical significance of the overlap. Nat Methods. 10:921–922. doi: 10.1038/nmeth.2659.

Phillips YF, Towsey M, Roe P, Radford CA. 2018. Revealing the ecological content of long-duration audio-recordings of the environment through clustering and visualization. PLoS ONE. 13(3):e0193345. doi: 10.1371/journal.pone.0193345.

Pieretti N, Farina A, Morri D. 2011. A new method to infer the singing activity of an avian community: the acoustic complexity index (ACI). Ecological Indicators. 11(3):868–873. doi: 10.1016/j.ecolind.2010.11.005.

Porneluzi P, Van Horn MA, Donovan TM. 2011. Ovenbird (Seiurus aurocapilla), version 2.0. In: Poole AF, editor(s). In the birds of North America. Ithaca, New York (NY): Cornell Lab of Ornithology.

Priyadarshani N, Marsland S, Castro I. 2018. Automated birdsong recognition in complex acoustic environments: a review. J Avian Biol. e01447. doi: 10.1111/jav.01447.

Rempel RS, Hobson KA, Holborn G, van Wilgenburg SL, Elliott J. 2005. Bioacoustic monitoring of forest songbirds: interpreter variability and effects of configuration and digital processing methods in the laboratory. J Field Ornithol. 76:1–11. http://jstor.org/stable/4151255.

Schmidt BR, Cruickshank SS, Bühler C, Bergamini A. 2023. Observers are a key source of detection heterogeneity and biased occupancy estimates in species monitoring. Biol Conserv. 283:110102. doi: 10.1016/j.biocon.2023.110102.

Shonfield J, Bayne EM. 2017. Autonomous recording units in avian ecological research: current use and future applications. Avian Conserv Ecol. 12(1):14. doi: 10.5751/ACE-00974-120114.

Sueur J, Pavoine S, Hamerlynck O, Duvail S, Reby D. 2008. Rapid acoustic survey for biodiversity appraisal. PLoS ONE. 3(12):e4065. doi: 10.1371/journal.pone.0004065.

Sugai LSM, Desjonqueres C, Silva TSF, Llusia D, Pettorelli N, Lecours V. 2019a. A roadmap for survey designs in terrestrial acoustic monitoring. Remote Sens Ecol Conserv. 6(3):220–235. doi: 10.1002/rse2.131.

Sugai LSM, Silva TSF, Ribeiro JW Jr., Llusia D. 2019b. Terrestrial passive acoustic monitoring: review and perspectives. BioScience. 69(1):15–25. doi: 10.1093/biosci/biy147.

The Bioacoustic Unit. 2019. WildTrax acoustic transcription user guide. Version 2.0. Edmonton (AB): University of Alberta and Alberta Biodiversity Monitoring Institute.

Towsey MW, Wimmer J, Williamson I, Roe P. 2014. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. Ecol Inform. 21:110–119. doi: 10.1016/j.ecoinf.2013.11.007.

Truskinger A, Cottman-Fields M, Eichinski P, Towsey M, Roe P. 2014. Practical analysis of big acoustic sensor data for environmental monitoring. 2014 IEEE Fourth International Conference on Big Data and Cloud Computing. p. 91–98. doi: 10.1109/BDCloud.2014.29.

Villanueva-Rivera LJ, Pijanowski BC. 2018. Soundecology: soundscape ecology. R package 1.3.3. https://cran.r-project.org/web/packages/soundecology/index.html.

Villanueva-Rivera LJ, Pijanowski BC, Doucette J, Pekin B. 2011. A primer of acoustic analysis for landscape ecologists. Landsc Ecol. 26(9):1233–1246. doi: 10.1007/s10980-011-9636-9.

Ware L, Mahon CL, McLeod L, Jetté JF. 2023. Artificial intelligence (BirdNET) supplements manual methods to maximize bird species richness from acoustic data sets generated from regional monitoring. Can J Zool E-First. doi: 10.1139/cjz-2023-0044.

Yip DA, Bayne EM, Solymos P, Campbell J, Proppe D. 2017. Sound attenuation in forests and roadside environments: implications for avian point-count surveys. Condor. 119(1):73–84. doi: 10.1650/CONDOR-16-93.1.

Yip DA, Knight EC, Haave-Audet E, Wilson SJ, Charchuk C, Scott CD, Sólymos P, Bayne EM. 2020. Sound level measurements from audio recordings provide objective distance estimates for distance sampling wildlife populations. Remote Sens Ecol Conserv. 6(3):301–315. doi: 10.1002/rse2.118.