Austral ECOLOGY A Journal of ecology in the Southern Hemisphere

*Austral Ecology* (2022) **47**, 440–455

# Accounting for both automated recording unit detection space and signal recognition performance in acoustic surveys: A protocol applied to the cryptic and critically endangered Night Parrot (*Pezoporus occidentalis*)

NICHOLAS P. LESEBERG,*[1,2,3] WILLIAM N. VENABLES,[4] STEPHEN A. MURPHY,[1,2,3] NIGEL A. JACKETT[1,2,3] AND JAMES E. M. WATSON[1,2,3,5]

[1]*School of Earth and Environmental Sciences, The University of Queensland, Brisbane, Queensland, 4072, Australia (Email: n.leseberg@uq.edu.au);* [2]*Green Fire Science, The University of Queensland, Brisbane, Queensland, Australia (Email: n.leseberg@uq.edu.au);* [3]*Research and Recovery of Endangered Species Group, The University of Queensland, Brisbane, Queensland, Australia (Email: n.leseberg@uq.edu.au);* [4]*School of Mathematics and Physics, The University of Queensland, Brisbane, Queensland, Australia; and* [5]*Centre for Biodiversity and Conservation Science, School of Biological Sciences, The University of Queensland, Brisbane, Queensland, Australia*

**Abstract** Research into the suitability of autonomous recording units (ARUs) when surveying for vocal species is increasing. Simultaneously, there has been extensive research into methods for efficiently extracting signals of interest from the acoustic data sets that accrue from the deployment of ARUs. For some species, bioacoustic monitoring supported by computerised signal detection offers the only effective and efficient method for widespread survey. In these circumstances, the detection space of both the ARU and the performance of the signal detection process must be considered concurrently, but typically, these two elements have been considered separately. Here, using the Night Parrot (*Pezoporus occidentalis*) as a case study, we consider both ARU detection space and the signal detection process to develop a robust and repeatable survey protocol for the species. After developing a call recogniser for the Night Parrot, we test its performance on a data set of Night Parrot calls given at a known distance from an array of ARUs. Having established a relationship between ARU type, recogniser performance and distance, we determine the sampling radius of an ARU for a given recogniser score cut-off, and the associated probability of detecting a Night Parrot that calls within that sampling radius. Using these data, we outline how to develop a robust and repeatable survey protocol for the Night Parrot, with a defined probability of detection. This protocol could be adapted for other scenarios where deployment of ARUs is necessary to determine a species' status and distribution.

**Key words:** acoustic monitoring, automated recording unit, bioacoustics, call recogniser, night parrot.

## INTRODUCTION

Shortfalls in our knowledge of the geographical distribution and abundance of rare and threatened species prevent targeted and timely conservation action (Sutherland *et al*. 2004; Hortal *et al*. 2015). If the species is cryptic or difficult to detect, these shortfalls may be acute (Mace *et al*. 2008). Developing effective survey protocols to detect and monitor such species should therefore be a research priority. An effective survey protocol will account for the behaviour and ecology of the target species and, equally importantly, must be robust and repeatable (Sutherland 2006).

*Corresponding author.
Accepted for publication November 2021.

One survey technique increasing in popularity is bioacoustic monitoring (Shonfield & Bayne 2017; Teixeira *et al*. 2019). The availability of commercial autonomous recording units (ARUs) with long-lasting power supplies and large memory capacity has seen bioacoustic monitoring applied widely. ARUs have proven particularly useful for detecting and monitoring species previously considered difficult to detect or assess using conventional survey methods (see, e.g. Lambert & McDonald 2014; Measey *et al*. 2017; Williams *et al*. 2018). There has also been a corresponding increase in research focussed on how to extract signals of interest from, often large, acoustic data sets (Priyadarshani *et al*. 2018). Typically, these two elements of bioacoustic monitoring are considered separately. If acoustic monitoring is to be used for systematic survey, protocols with defined

 doi:10.1111/aec.13128

detection probabilities and sampling areas are necessary (Sugai *et al.* 2020). This requires an improved understanding of the parameters defining detection performance, which include both ARU detection space and call recogniser performance.

Ultimately, the most important parameter to resolve is the probability of detection, which is the product of the probability the species will be detected if it is available for detection, and the probability it is available for detection (Pollock *et al.* 2004). Although influenced by the chosen method of detection, the second part of this equation is largely a behavioural question. If an ARU is set to record in an environment where the species of interest occurs, this will be the probability the species will vocalise in the way necessary for it to be recorded by the ARU. It may vary in response to multiple factors including risk of predation, social context and breeding biology (Teixeira *et al.* 2019). The first part of the equation though is a purely technical question; provided the species does vocalise and the ARU is there to record it, this is the probability that vocalisation will be both recorded by the ARU and extracted from the resulting acoustic data set. Understanding this component of the detection equation requires field testing both the chosen ARU and its associated settings such as sample rate, gain level and microphone type, and the method used to extract the signal of interest from an acoustic data set.

ARU detection space will be driven by technical factors, such as microphone quality and sensitivity, and signal processing capability (Turgeon *et al.* 2017). The signal extraction technique will depend on the signal of interest. Much research has focussed on methods for efficiently extracting relevant calling events from acoustic data sets, including sampling approaches, visual inspection of spectrograms, acoustic indices and the application of signal detection algorithms, termed 'call recognisers' hereafter (Swiston & Mennill 2009; Towsey *et al.* 2014; Joshi *et al.* 2017; Knight & Bayne 2018). Particularly for large data sets, call recognisers are an increasingly popular choice. Call recognisers rely on training an algorithm to detect and evaluate acoustic signals. One of the most commonly used techniques is template-matching, whereby potential signals are scored depending on their similarity to a specific template, with the user adjusting the score threshold at which a match is declared (see, e.g. Charif *et al.* 2010; Katz *et al.* 2016). Although the specific algorithm varies, the process is similar for most recognisers.

When using a call recogniser to extract data from an acoustic data set, it has been shown that a calling event's score is related to the distance that call is made from the ARU (Knight & Bayne 2018). This is expected, as calls made closer to the ARU are likely to have the highest amplitude, exhibit the greatest signal-to-noise ratio and therefore receive a higher score from the recogniser. It follows that the score threshold set to define when a match is declared will also determine the effective sampling radius of the ARU. Raising the score threshold increases the score required for an event to be declared a match, meaning only calls with a higher signal-to-noise ratio, typically those made closer to the ARU, are likely to be extracted, reducing the effective sampling radius. Conversely, lowering the score threshold decreases the score required for an event to be declared a match, and calls with a lower signal-to-noise ratio, such as distant calls, will also be extracted, increasing the effective sampling radius. Reducing the threshold in this way does increase the likelihood of false positive detections and may require manual inspection of declared matches to ensure the integrity of the data.

Research to date has largely focussed on these separate elements, either ARU detection space, or call recogniser performance. Here, using the Night Parrot (*Pezoporus occidentalis*) as a case study, we combine the two elements to determine the probability that a species will be detected given it is available for detection, and the sampling radius associated with that probability. Initially, we compile a data set of Night Parrot calls at known distances from an array of several models of ARU. We then analyse these calls using a call recogniser, confirming that the relationship between recogniser score and distance can be described using an exponential decay curve. Using an extensive database of different call types, we then test this relationship, confirming it is consistent across key call types. Finally, we use this relationship to determine the sampling radius for a range of score thresholds, and the proportion of calls made within that sampling radius that will achieve that score threshold. When combined with the results of research into the species' calling behaviour, these data will support the development of an acoustic survey protocol with a defined probability of detection that can be used to systematically assess the Night Parrot's distribution.

## METHODS

### Study species

The Night Parrot formerly occurred throughout central Australia, before its precipitous decline and virtual disappearance in the late 19th century (Leseberg *et al.* 2021). In 2013, an extant population of birds was discovered in western Queensland (Koch 2013), and the bird has since been detected at several locations in Western Australia (see, e.g. Jackett *et al.* 2017; Mills & Collins 2017; Collins 2021). Research suggests that Night Parrots are relatively sedentary (Murphy, Austin, *et al.* 2017; Murphy, Silcock, *et al.* 2017). They spend the day roosting as dispersed pairs or


doi:10.1111/aec.13128

small groups, in patches of long unburnt *Triodia* spp., a genus of grass widespread throughout central Australia. The birds emerge at dusk and may travel considerable distances to forage on open grasslands and floodplains (Murphy, Silcock, *et al.* 2017). The birds typically return to their roost sites just before dawn.

Night Parrots are predictably vocal, engaging in a brief period of calling most evenings before leaving their roost sites to feed (Murphy, Austin, *et al.* 2017; Murphy, Silcock, *et al.* 2017). Birds occasionally return to their roost sites and call during the night, particularly when breeding, but typically return for another brief period of calling just before dawn. The birds also call occasionally at foraging and drinking sites, and when moving to and from these sites (N. Leseberg, S. Murphy, N. Jackett pers. obs.).

## Research problem

The Night Parrot's sedentary habits and predictable calling behaviour have seen acoustic monitoring prove the most suitable method for detecting and monitoring the species. ARUs have been used to monitor presence at known roost sites, and survey for Night Parrots at prospective roost sites. However, the detection radius and associated probabilities of detection associated with these deployments are not well understood. Furthermore, the large data sets collected require a call recogniser to extract potential Night Parrot calls. While statistics around the precision and recall of one Night Parrot call recogniser have been established (Leseberg *et al.* 2020), how these relate to the detection radius of the ARUs has not been investigated.

These limitations mean robust conclusions about the presence or absence of Night Parrots beyond the immediate vicinity of an ARU are difficult, preventing development of ARU deployment protocols that could provide a probabilistic assessment of the presence or absence of Night Parrots at the landscape scale. To solve this problem, we aimed to determine the probability of detecting a Night Parrot if it was available for detection within a defined distance of an ARU, when using a call recogniser to analyse the resulting sound recordings. To establish these parameters, we (1) constructed a call recogniser for the Night Parrot; (2) determined the relationship between score assigned to each detection by the recogniser and the distance that call was made from the ARU; (3) determined how that relationship varied between ARU types; (4) confirmed that relationship was consistent across call types; and (5) used that relationship to establish the probability of detecting a Night Parrot if it was available for detection, and the detection radius associated with that probability.

## Recording Night Parrot calls

Adult Night Parrots give a variety of calls, categorised broadly as 'whistle', 'bell-like' and 'croak' calls (Leseberg *et al.* 2019). Whistle and bell-like calls are most useful for detection as they are the most commonly heard; there are nine described whistle calls and three described bell-like calls. Of these twelve call types, four whistle and three bell-like calls are commonly heard in the study area. The typical frequency range and duration of these seven call types is given at Table 1. Within these frequency ranges, individual calls of each call type are typically constant in frequency, with a narrow bandwidth of ~100–200 Hz and no modulation.

When recording Night Parrot calls using ARUs, a sampling rate of 24000 Hz is typically used. The upper frequency limit of the Night Parrot's 'croak' call (not relevant to this research) is ~10000 Hz, so this sampling rate avoids aliasing in accordance with the Nyquist–Shannon Sampling Theorem (Landau 1967). In circumstances where long-term monitoring is occurring, a sampling rate of 16000 Hz may be used to reduce the amount of memory required for acoustic data. This sampling rate avoids aliasing for all whistle and bell-like calls, which are the most important for detection and are the focus of this research.

## Recogniser development

We used the R package 'monitoR' (Katz *et al.* 2016; R Core Team 2018) to build a call recogniser for the seven Night Parrot call types most frequently heard in the study area (Table 1). In a comparison with recognisers using machine learning methods and commercially available packages, 'monitoR' performed well (Knight *et al.* 2017) and is relatively easy to use. Using the technique outlined in Katz *et al.* (2016), we constructed a series of binary point templates for the seven call types. Templates are created by clipping an example call from a sound file sampled at 24000 Hz then creating a spectrogram (FFT transformation = Hann window, FFT size = 512, overlap = 0). These spectrogram parameters provide a good balance between the frequency and time resolution necessary for viewing Night Parrot calls, and subsequent processing speed. A selection of cells within the spectrogram is then classified as 'on' or 'off'. 'On' cells are selected to represent the expected region of highest amplitude for the call in the time domain, while 'off' cells are placed strategically where no signal is expected.

**Table 1.** Frequency and duration of specific Night Parrot call types included in the recogniser. The number of templates included for each call type is given

| Call type | Freq. range (kHz) | Duration (s) | Templates |
|---|---|---|---|
| Hollow whistle | ~ 1.9–2.1 | ~ 0.5–0.6 | 2 |
| One-note and two-note trill | ~ 2.2 | Each note: ~ 0.3 | 2 |
| Multi-note whistle | ~ 2.1–3.0 | ~ 0.4–0.8 | 7 |
| Toot | ~ 1.6–1.8 | ~ 0.1 | 2 |
| Ding-de-ding | ~ 2.8–3.8 | Each note: ~ 0.1 | 3 |
| Dink-dink and dink | ~ 2.2–3.5 | Each note: ~ 0.1 | 14 |
| Didit | ~ 2.2 | ~ 0.2 | 1 |

doi:10.1111/aec.13128 

As the seven call types commonly heard at the study site are all distinct, templates were created for each specific call type. To construct a template, a good quality example of that call type was extracted from field data recorded using an ARU, and the procedure described above used to create the template. These templates were then qualitatively tested on a sample of ARU field data containing examples of the call type to ensure they detected known variation in the call type. Some call types are relatively uniform in frequency and duration (e.g. *hollow whistle* and *ding-de-ding*) and required few templates. Other calls which are less uniform, particularly in frequency (e.g. *multi-note whistle* and *dink-dink*), required more templates. Additional templates were created for each call type until qualitative testing on a data set of ARU field recordings from the study area containing 1850 known Night Parrot whistle and bell-like calls suggested the templates could detect most local variation across all call types. This process resulted in 31 different templates (Table 1), which were combined into a template list (Katz *et al.* 2016) and represent the final 'recogniser'. This recogniser was able to achieve maximum recall of 0.94, with an area under the curve (AUC) of 0.76 (Leseberg *et al.* 2020).

## Analysing acoustic data with the recogniser

Before analysis, each sound file is converted to a spectrogram using the same parameters as were used to create the templates. 'monitoR' requires template files and the sound files that will be scanned to have the same sample rate. As the templates for this recogniser were created from clips with a sampling rate of 24000 Hz, any sound files to be analysed that were not recorded at 24000 Hz were downsampled or upsampled to a sampling rate of 24000 Hz before processing. Qualitative testing confirmed that manipulating files in this way had no apparent effect on results. No post-processing, such as band-pass filtering of audio files, occurs before analysis. Each of the 31 templates within the recogniser is then stepped along that spectrogram, and for every step, a similarity score is assigned. The similarity score is the difference between the mean amplitude detected in the template's 'on' and 'off' cells (Katz *et al.* 2016). Hereafter, we refer to this similarity score simply as 'score'. A higher score implies greater similarity between the underlying signal within the sound file and the template file. When plotted against time, this results in a series of peaks, with the recogniser returning a list of these peaks with their associated score. Because the recogniser scans each file with all 31 templates, this may result in multiple peaks at times when more than one template detects the same call. In this circumstance, the recogniser selects the peak with the highest score.

## Establishing the acoustic array

Under the inverse square law, the amplitude of the signal received by an ARU is related to the squared distance of the source of the call from that ARU. This has been demonstrated in practice (Yip *et al.* 2017). Because the score assigned to a detection by the recogniser is directly related to the amplitude of that call, we expect that score will also be correlated with the distance the call is made from the ARU. To determine the relationship between score and distance, we established a 675 m long linear array of ARUs through an area where observations over several months had established that at least one Night Parrot was roosting. This area consisted of patchy *Triodia* hummocks on otherwise open ground, with some scattered shrubs. This is typical of Night Parrot long-term stable roosting habitat elsewhere (Jackett *et al.* 2017; Murphy, Silcock, *et al.* 2017). Three models of popular ARU were tested: Song Meter 3 (SM3) and Song Meter 4 (SM4) (Wildlife Acoustics Inc., Concord, Massachusetts, USA), and Bioacoustic Audio Recorder (BAR) (Frontier Labs, Brisbane, Queensland, Australia). SM3s were placed at both end points of the 675 m array centreline and every 75 m along the centreline (Fig. 1). As there were fewer SM4s and BARs available, they were placed at the end

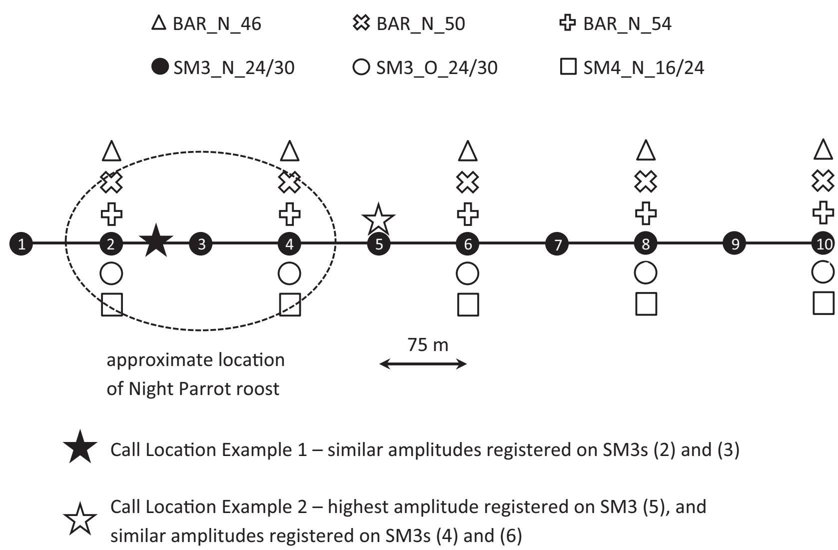


**Fig. 1.** Schematic of ARU array used for determining the relationship between recogniser score and distance. Where grouped, ARUs were all placed within 2 m either side of the array centreline. The code for ARU type gives ARU model (BAR/SM3/SM4), microphone condition (N = New / O = Old) and gain setting (in dB). Examples are also given of the two methods used to establish approximate locations of calls within the array.


doi:10.1111/aec.13128

**Table 2.** Technical parameters of each separate component within the signal processing chain

| | Microphone | | | | | Receiver | Digitisation | Storage |
|---|---|---|---|---|---|---|---|---|
| 'Treatment' | Condition | Directivity | Sensitivity | Signal to Noise Ratio | Pre-amp Gain | Gain | Sample Rate (Hz)/ Bit Depth | Format |
| BAR_N_46 | New | Omni-directional | –28 dB | 80 dB | 20 dB | 46 dB | 48 000 16-bit | .wav file |
| BAR_N_50 | New | Omni-directional | –28 dB | 80 dB | 20 dB | 50 dB | 48 000 16-bit | .wav file |
| BAR_N_54 | New | Omni-directional | –28 dB | 80 dB | 20 dB | 54 dB | 48 000 16-bit | .wav file |
| SM3_N_24 | New | Omni-directional | $-11 \pm 4$ dB | > 68 dB | - | 24 dB | 16 000 16-bit | .wav file |
| SM3_N_30 | New | Omni-directional | $-11 \pm 4$ dB | > 68 dB | - | 30 dB | 16 000 16-bit | .wav file |
| SM3_O_24 | Old | Omni-directional | $-11 \pm 4$ dB | > 68 dB | - | 24 dB | 16 000 16-bit | .wav file |
| SM3_O_30 | Old | Omni-directional | $-11 \pm 4$ dB | > 68 dB | - | 30 dB | 16 000 16-bit | .wav file |
| SM4_N_16 | New | Omni-directional | $-33 \pm 4$ dB | 80 dB | 24 dB | 16 dB | 16 000 16-bit | .wav file |
| SM4_N_24 | New | Omni-directional | $-33 \pm 4$ dB | 80 dB | 24 dB | 24 dB | 16 000 16-bit | .wav file |

point furthest from the known roost, then every 150 m along the array. At locations where the ARUs were grouped, all ARUs were within 2 m of the centreline. ARUs were set with microphones oriented perpendicular to the array centreline. The technical specifications of each ARU are given in Table 2.

Research suggests that microphone condition affects recording quality (Turgeon *et al.* 2017). To test this, two sets of SM3s were deployed, one with new microphones and another with microphones which had spent at least one year permanently deployed in the field and were in apparently poor condition. Microphones used on the SM4s and BARs had been previously deployed for approximately six months and were in good condition.

The gain setting of an ARU determines the input level being received by the ARU from the microphone(s). When recording quiet (low amplitude) sounds, gain can be boosted to increase the input level of those sounds, potentially increasing the ARU's detection radius. However, increasing gain to boost the input level of a signal of interest will also boost the level of any background noise, potentially negating the benefit for processes where signal-to-noise ratio is important, as is the case with template-matching recognisers. While the default gain settings on SM3 and SM4, and the manufacturer recommended gain setting of 46 dB for BAR had been used to make satisfactory field recordings of Night Parrots, we aimed to test whether an increase in gain improved detection radius. SM3 and SM4 are capable of recording in stereo, and the gain for each channel can be set independently. We programmed one channel on each of the SM3s and SM4s to record at the default setting (24 dB and 16 dB respectively), while the other was set at a slightly increased gain (30 dB and 24 dB respectively). BARs have only one microphone, so three sets of BARs were deployed, with gain settings of 46 dB, 50 dB and 54 dB. These gain settings were chosen because testing on Night Parrot calls made close to the ARU suggested the resulting input levels noticeably increased the amplitude of individual calls without distortion. It is important to note that these gain settings are not directly comparable, as the detection space of the ARU and microphone combination depends on other internal specifications that are unique to each model of ARU (see Table 2). Ultimately, with combinations of ARU type, microphone condition and gain setting, we tested nine 'treatments' (see Table 2).

## Collection of acoustic data and call identification

SM3s and SM4s recorded from sunset to sunrise. BARs recorded for 90 mins from sunset and for 120 mins before dawn, the period when most Night Parrot calling activity occurs. ARUs recorded at sampling rates of 16000 Hz (SM3 and SM4) and 48000Hz (BAR). The BAR's microphones apply an 80 Hz high-pass filter. No filtering was applied to SM3 and SM4 data. Acoustic data were saved as 16-bit wav files. All ARUs were left in place for seven nights, recovered and the data downloaded. The recordings from five nights contained significant wind noise. As wind is a variable known to influence ARU detection space (Thomas *et al.* 2020), we discarded all files from these nights to avoid introducing additional variation into the data set. This means our results only apply to recordings from still nights, a trade-off we were willing to accept. Sound files from the remaining two still nights were examined by both listening to and manually scanning spectrograms of each file. Many Night Parrot calls were identified, including some loud calls clearly made close to the array and simultaneously detected on multiple ARUs, and some very faint calls that were likely made some distance from

doi:10.1111/aec.13128 

the array. Of the calls detected, 105 discrete Night Parrot calls that apparently occurred in close proximity to the recording array were selected for further analysis.

## Determination of call location

It was not possible to determine by observation where, relative to the array, the bird giving each of the 105 calls selected for analysis was located; Night Parrots only call at night when they cannot be seen, and the presence of humans in and around roost areas at night is restricted and can affect calling behaviour. Instead, as the amplitude of a sound wave is known to decay consistently with distance squared between the source of the sound and the recorder, we used this principle to try and determine the approximate point from where each of the 105 selected calls were made. The data from the sound files recorded by SM3s with new microphones and default gain setting were examined, and for each call, the mean amplitude in decibels relative to full scale (dB FS) for the fundamental frequency in the time domain was extracted from the original wav file using the 'Plot Spectrum' function in audio-editing public domain software Audacity (version 2.3.0, http://audacity.sourceforge.net/). The sound files from the SM3s with new microphones were used because these were more closely spaced than other ARUs, allowing the relative position of the calls within the array to be determined more accurately. A subset of the 105 calls were identified which registered a similar amplitude (~ $-51.0 \pm 2$ dB FS) on one SM3 and which registered the same or very similar amplitude ($\pm$ 2 dB FS) on an immediately neighbouring SM3. We assumed the bird giving these calls did so from the midpoint on the array between those two SM3s. An example of this call location is labelled as 'Call Location Example 1' in Fig. 1. Similarly, a subset of calls was identified that registered an amplitude on a single SM3 similar to the maximum amplitude across the data set (~ $-42.0 \pm 2$ dB FS), and which also registered very similar amplitudes (but lower than on the single SM3) on both of the two SM3s in the array immediately neighbouring the single SM3. These calls could reasonably be assumed to be given by a bird in the immediate proximity of that centre SM3. These calls were assumed to be given at a point 15 m from that SM3, perpendicular to the array centreline. An example of this call location is labelled as 'Call Location Example 2' in Fig. 1.

After extracting and examining each of the 105 call events from the acoustic data recorded by the SM3s with new microphones, 26 discrete Night Parrot calls were identified that, using the process mentioned, could reasonably be assigned to a specific location relative to the array. Because we assumed that all calls were given either on or very close to the straight line along which the array was placed, it is likely that in most cases, the estimated distance of the calling bird from each ARU in the array is an underestimate.

This method of location determination relies on the decay of the amplitude within the array being consistent; however, numerous variables can affect this assumption. In this case, the area where the array was established was an open stony plain, with only scattered low vegetation, not expected to impede transmission of Night Parrot calls. We also assumed that the direction the bird was facing did not affect the amplitude of the signal being received by the ARU. Because Night Parrots call at night when they cannot be seen there is no way to empirically assess this. A final assumption was that Night Parrots give most of their calls at a relatively constant volume, meaning the decay of each call is similar across the array. Qualitatively, this assumption is reasonable, as different calls given at similar distances are of a similar volume to the human ear (N. Leseberg pers. obs.). As multiple calls were included in subsequent modelling, and scores for each call were registered on multiple ARUs on either side of each call, any error associated with these final assumptions was incorporated into the results of subsequent modelling.

## Extraction and compilation of distance data set

As BARs use a GPS receiver to update their internal clock, it was possible to easily determine where on a specific sound file from any BAR a call detected on another BAR would occur, and manually identify that call or, in the event the call was too faint to be detected aurally or visually from the sound file, the section of the file when the call would have occurred. Because the internal clock for SM3 and SM4 is set by the user, and may wander over time, a combination of characteristics unique to each call or series of calls, and other ambient sounds, was used to calculate the equivalent GPS time of each SM3 and SM4. This enabled the same calls, or at least sections of recording when those calls would have occurred, to be manually identified within the sound files of each SM3 and SM4.

Using this process, the specific time at which each of the 26 calls with associated distance data occurred on each ARU was determined, and a two second section of the recorded wav file clipped that either contained the call, or the time when the call would have occurred. Because nine of these calls occurred outside of the period when the BARs were recording, only 17 call clips were extracted from each BAR. All clips were then analysed using the call recogniser, with the score threshold set to zero. This ensured that a score was registered for each clip, even clips from ARUs too distant for the call to be recorded. For these clips, the score represents the score achieved by ambient noise within the sound file. The resulting data set is subsequently referred to as the 'distance data set'.

## Relationship between score and distance

Because score is directly correlated with amplitude, we expected score to decay consistently with distance in accordance with the inverse square law, provided the atmospheric conditions which influence that decay remain relatively constant. Ambient noise within the sound file still receives a score from the recogniser, albeit low, so we modelled this relationship using a non-linear regression function of the form:


doi:10.1111/aec.13128

$$s = S_a + (R_a - S_a)\exp\{-\exp(l_a)\,d\} + \epsilon.$$

where $s$ is the score observed at a distance $d$ from the calling bird. $S_a$ is the asymptotic value of score (representing the mean score the recogniser assigns ambient noise), and $R_a$ is the conceptual initial value at $d = 0$. $l_a$ is the decay constant. The subscript $a$ denotes the specific combination of ARU model, microphone condition and gain setting. Under this model, the asymptote $S_a$ was allowed to vary for all ARU combinations. The model was fitted using the SSasymp function in statistical software R (R Core Team 2018). To examine the relationship between score and distance, the data for each ARU were plotted and the estimated regression curve added to the plot. Although $R^2$ is not typically a good indication of fit for non-linear models, because in this case the null model of a single mean is a sub-model of the non-linear regression model, we calculated $R^2$ to assess goodness-of-fit (Schabenberger & Pierce 2002).

## Effects of varying gain and microphone condition

To test whether increasing gain changed the relationship between score and distance significantly, after fitting the model to the nine original ARU groups, we condensed these nine groups into four groups: SM3s with old microphones, SM3s with new microphones, BAR and SM4. We then fitted the same model to these four groups, before conducting an analysis of variance for a null hypothesis of no differences between the original model with nine groups and the model with four condensed groups.

Similarly, to test whether using new or old microphones changed the relationship between score and distance significantly, we condensed the four original SM3 groups into two groups representing the different microphone conditions (old and new) and a single group representing all SM3 data. We then fitted the same model to each of these three groups, before conducting an analysis of variance for a null hypothesis of no differences between the model fitted to the old and new microphone groups and the model fitted to the one group representing the entire SM3 data set.

## Relationship between call types

Although Night Parrots have a variety calls, at specific sites, certain call types often dominate (Leseberg *et al.* 2019). Two main call types, *multi-note whistle* and *dink-dink* calls, were detected within the 105 discrete calls detected close to the array. However, the final selection of 26 calls given at known locations only contained variations of *dink-dink* calls, which are treated similarly by the recogniser's templates (see Appendix S1 for spectrograms of these calls). To ensure the relationship established between score and distance using the distance data set holds for other call types not contained within that data set, we compiled a 'call type data set' containing a variety of call types and tested whether the relationships within that data set were consistent with the relationships established in the distance data set.

The call type data set contained 985 Night Parrot calls collected in field recordings taken at the study site. All calls were recorded using SM4s with microphones in good condition, on the default gain settings, and extracted using the same recogniser. Because the call type data set did not have associated distance data, we used the amplitude of each call as a proxy for distance. A relative amplitude for each call within the call type data set was calculated using the R package 'seewave' (Sueur *et al.* 2008), by isolating the call syllable, or the strongest syllable for calls with multiple syllables, and using the function 'meanspec' to extract the mean relative amplitude of the syllable. Amplitude varies consistently with distance in the same way we have demonstrated that score does. We aimed to assess whether the relationship between score and amplitude is similar for different call types, reasoning that if it is, we can safely assume the relationship between score and distance for other call types is similar to that between score and distance determined for the *dink-dink* type calls in the distance data set.

To assess whether the relationships are similar, we grouped the different call types into three broad categories of call type: 'bell-like', 'short whistle' and 'long whistle'. Previous research has demonstrated consistent relationships between score and call type within these groups (Leseberg *et al.* 2020). The call type data set contained 406 bell-like calls, 96 long whistle calls and 483 short whistle calls. We constructed three linear models to test the relationship between score, amplitude and call type: a null model where score was solely a function of amplitude; a model with call type and amplitude as explanatory variables; and a model with amplitude nested within call type as the explanatory variables. We conducted an analysis of variance to compare these models, selecting the model which explained the most variation compared to the null model. We then reviewed the coefficients of the selected model to examine any relationships and determine whether different call types were treated similarly by the recogniser.

## Establishing a survey protocol

Having characterised the relationship between score and distance, we used these results to define the parameters of a survey protocol with a quantified probability that a Night Parrot call given within a specified distance of an ARU would be both recorded by the ARU and extracted using the call recogniser. We achieved this by establishing for a given cut-off score, $c$, what distance from an ARU, $d_c$, a Night Parrot call can be given that will have a probability $p$ of exceeding that cut-off score, therefore solving the equation:

$$\Pr(S > c \mid d_c) = p.$$

where $S$ is the realised score of the call.

To achieve this, we solved for the lower *tolerance* limit of the model which described the relationship between score and distance for the specified recorder type, settings and microphone condition. Unlike solving for a lower *confidence* limit, which would reflect the probability that the mean score for a given distance will exceed the score cut-off, when solving for the lower *tolerance* limit, $p$ reflects the probability that a call given at that distance will exceed the score cut-off. Having defined this relationship, we created a series of tables that specified the effective sampling radius

doi:10.1111/aec.13128 

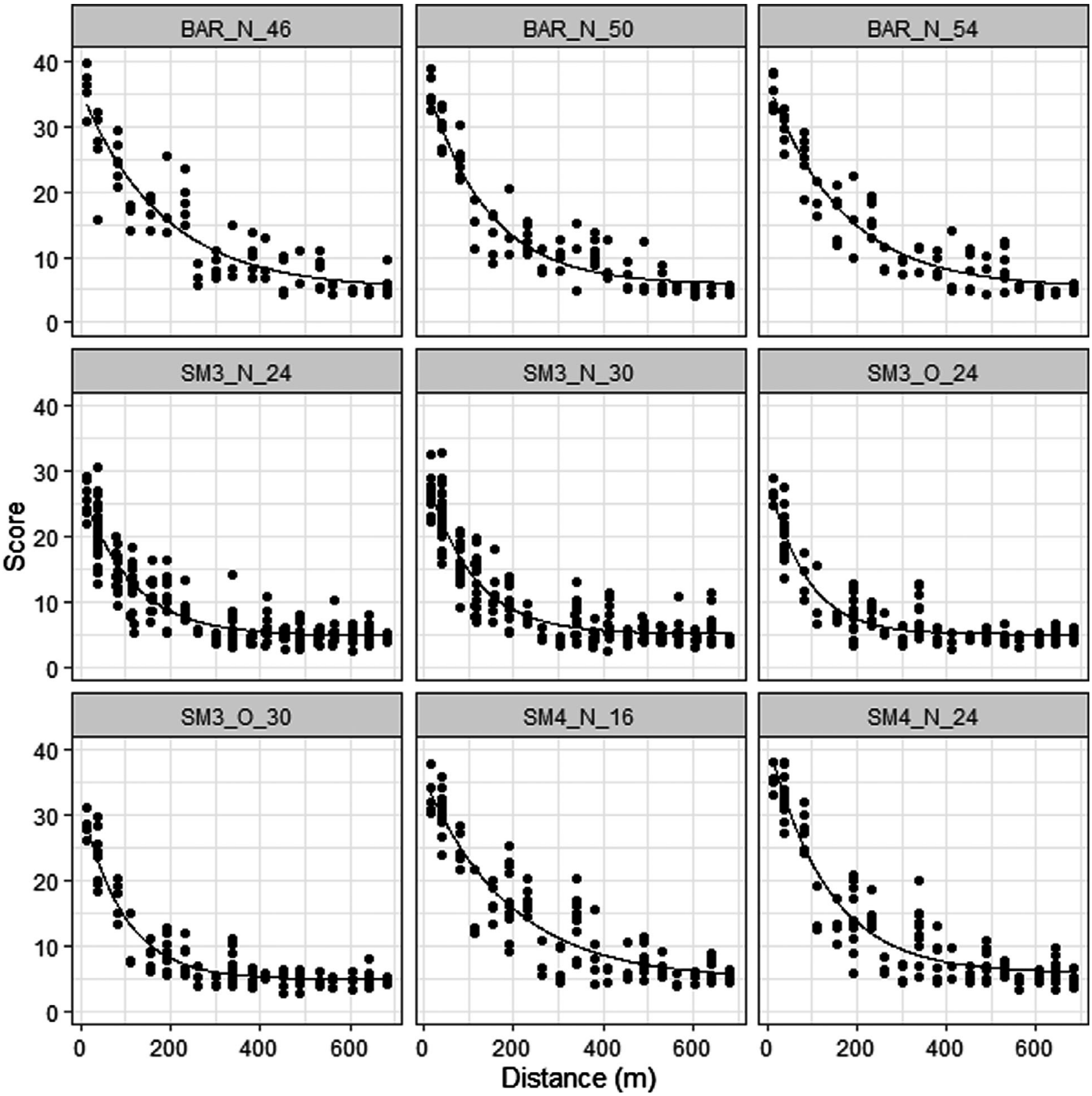


**Fig. 2.** Plots of score against distance for each ARU combination, with the model fitted to each separate ARU, microphone and gain combination. All combinations of BAR and SM4 perform similarly and outperform all combinations of SM3.

for each combination of ARU model, microphone condition and gain setting, for a defined score cut-off, and probability that a call given within that sampling radius would be both recorded and extracted from the data set using the call recogniser.

## RESULTS

### Relationship between score and distance

$R^2$ for the model fitted to each combination of ARU model, microphone condition and gain setting was 0.89, and plots for each ARU combination confirm the model was a good fit (Fig. 2). The broad relationship was similar for each ARU model, gain setting and microphone combination, with score initially decreasing rapidly with increasing distance, before approaching an asymptote at a recogniser score of approximately 5. BAR and SM4 combinations outperformed SM3 combinations; for SM3, calls close to the ARU received lower scores, and score decayed more quickly than for both SM4 and BAR. For an SM4 using the default gain setting, 95% of calls that were given within 214 m of the

 doi:10.1111/aec.13128

ARU received a score ≥ 10. For a BAR using the manufacturer recommended gain setting of 46 dB, this distance was 207 m. For an SM3 using the default gain setting, this distance was 91 m, meaning the detection area for the SM3 using these parameters is less than 20% that of the SM4 or BAR.

### Differences between gain settings and microphone condition

After fitting the model to the four ARU groups (by model and microphone condition), the ANOVA comparing these models with the original nine groups suggested there were significant differences ($F_{15,1216} = 3.31$, $P < 0.001$). While there are statistically significant differences in detection space between gain settings within the groups, plotting the fitted models for each of the nine groups over the data for the condensed groups shows that from a practical point of view, while noticeable for some groups, these differences are neither substantial nor consistent (Fig. 3). This suggests that increasing the gain will not result in any predictable improvement in detection space. This result is expected for a recogniser based on template matching; because monitoR calculates the difference between the mean amplitudes of the 'on' and 'off' cells of a template, increasing gain will also increase the amplitude of the signal in the 'off' cells, negating any benefit of increasing it for the 'on' cells.

Although ANOVA indicates there is no statistically significant difference in detection space between old and new microphones ($F_{3,722} = 2.43$, $P = 0.064$), plotting the fitted models over the pooled data reveals slight differences (Fig. 4). From a practical point of view, these differences are small but noteworthy, suggesting new microphones slightly outperform old microphones for score achieved by the recogniser at a specified distance.

### Relationship between different call types

Compared to the null model with score solely a function of amplitude, the model with amplitude nested within call type explained more variation than the model with amplitude and call type as independent variables ($F_{4,979} = 14.64$, $P < 0.001$). A review of this model's coefficients (Table 3) indicates that intercepts for the short whistle and long whistle calls at an amplitude of 25 dB (chosen because 25 dB is the approximate mean relative amplitude of all calls in the data set) are both slightly greater than for bell-like calls. Further, the slope for both the short whistle and long whistle calls is marginally greater than the slope of bell-like calls. This suggests that both short and long whistle calls actually receive a higher score than a bell-like call of the same amplitude. The relationship appears consistent across the specific call types within each broad group (Fig. 5). These results confirm that short and long whistle calls exhibit the same relationship between score and distance as bell-like calls, meaning these results can be used to define a survey protocol that will be applicable across call types.

### Establishing a survey protocol

Solving the equation defining the probability that a call given within a specified distance will achieve a score greater than the cut-off produces Table 4. This indicates the effective sampling radius for an SM4 using new microphones and the default gain setting. Results for all other ARU, microphone and gain combinations are included at Appendix S2.

## DISCUSSION

If ARUs combined with automated signal detection will be the primary means of surveying for a species, it is critical that the influence of both elements of the probability of detection are understood. To date, most research in this field has focussed on the capabilities of ARUs as a substitute for other survey methods (see, e.g. Digby *et al.* 2013; Lambert & McDonald 2014; Williams *et al.* 2018), the performance of different signal extraction techniques (see, e.g. Swiston & Mennill 2009; Joshi *et al.* 2017; Venier *et al.* 2017) and the relative performance of various automated recognition algorithms for extracting the resulting acoustic data (see, e.g. Dema *et al.* 2017; Knight *et al.* 2017; Priyadarshani *et al.* 2020). The method outlined here demonstrates how both ARU detection space and call recogniser performance must be assessed concurrently to define a robust and repeatable survey protocol for species like the Night Parrot, where the deployment of ARUs followed by analysis using call recognisers offers the only feasible option for widespread survey.

Our pattern of results was largely expected, with score decaying exponentially and predictably with increasing distance from the ARU for all combinations of ARU type, microphone and gain setting. As this research focussed on maximising the area that could be surveyed, ARU spacing was established to better understand the outer limits of the detection space. ARUs were too widely spaced to get a detailed understanding of the relationship between score and distance very close to the ARU (i.e. within ~ 40 m). For another species, or application where precisely understanding detection performance close to the

doi:10.1111/aec.13128 

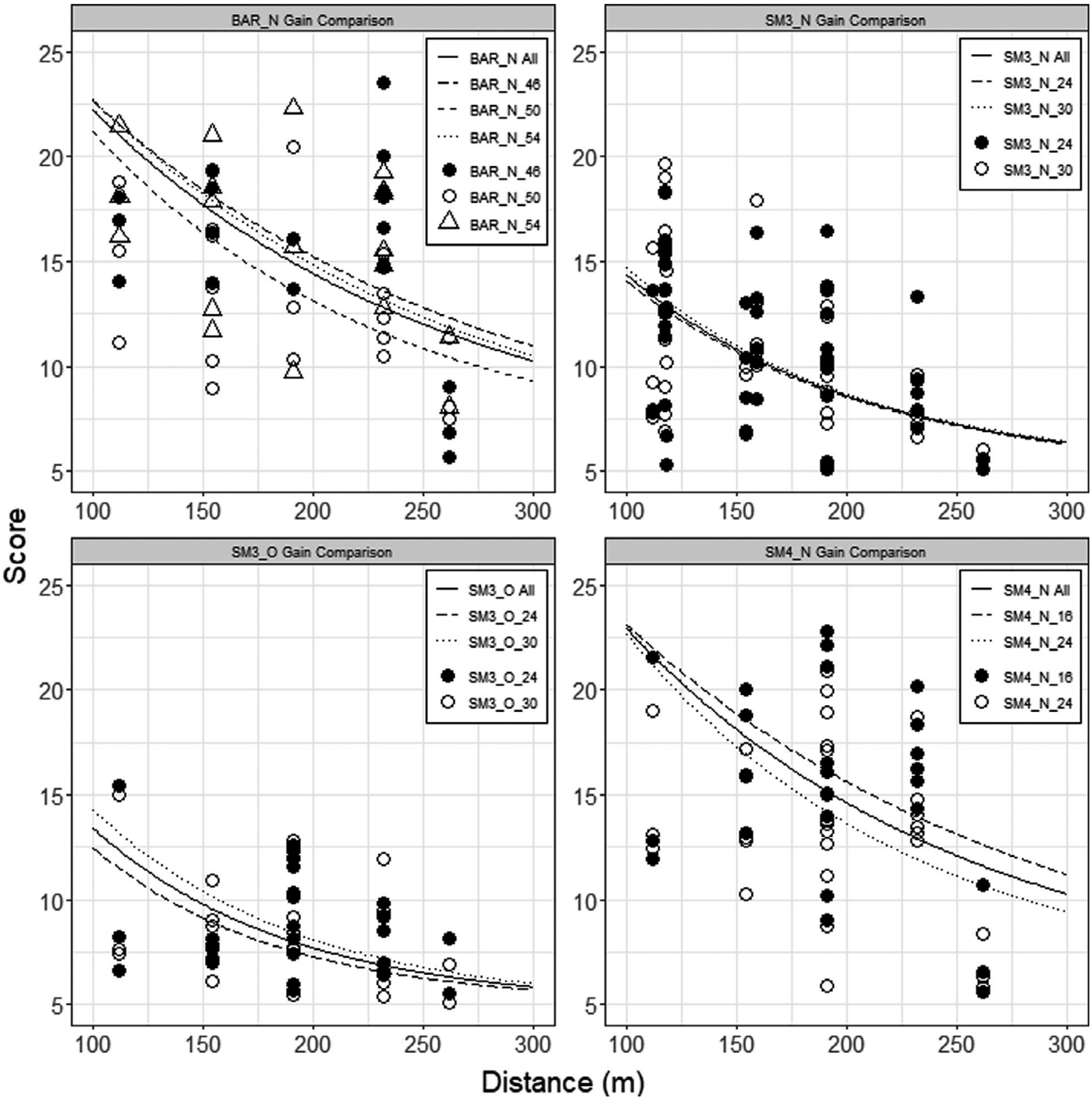


**Fig. 3.** Plots of score against distance for different gain settings within the four broad combinations of ARU model and microphone condition. Only a portion of the x- and y-axes are shown, to highlight the sections of the plot where differences are most obvious. The inconsistency can be seen; for BAR_N, the highest gain setting slightly outperforms the others, whereas for SM4_N, the highest gain setting performs more poorly than the lower setting.

ARU is important, the spacing of ARUs should be considered closely.

Calls made at the same distance from an ARU and extracted using the recogniser achieve a higher score when recorded by SM4 and BAR than when recorded using SM3, and the rate of decay was less for SM4 and BAR than for SM3. This saw both SM4 and BAR achieve a much larger detection space than SM3. These differences are expected given the SM4 and BAR are later generation ARUs than SM3, with improved specifications such as inclusion of a pre-amplifier and better signal-to-noise ratio. Changing gain settings had no practical impact on the results that were achieved by each ARU type, with differences in detection space neither substantial nor consistent. This is probably an artefact of the template-matching algorithm used to build the recogniser, and the spacing of ARUs. As mentioned, ARUs were too widely spaced to really understand the relationship between score and distance close to


doi:10.1111/aec.13128

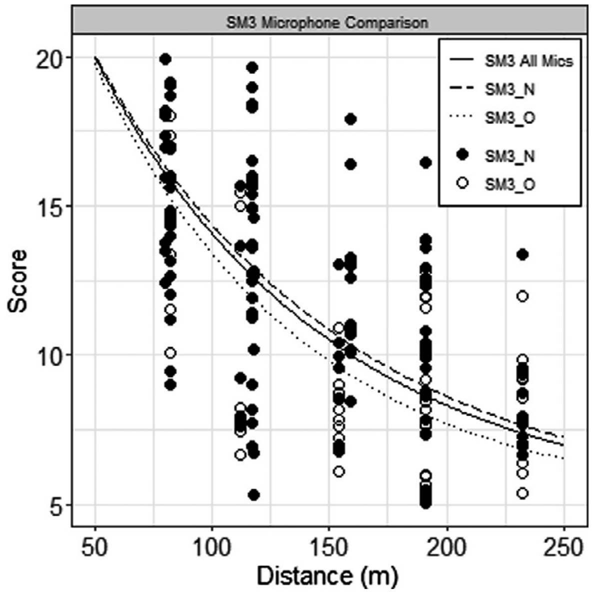


**Fig. 4.** Plot of the model fitted to the entire SM3 data set, and models fitted to only old and new microphone data sets. Machines with new microphones slightly outperform machines with new microphones. Only a portion of the x- and y-axes are shown, to highlight the sections of the plot where differences are most obvious.

**Table 3.** Estimates of the intercept (at a relative amplitude of 25 dB) and slope coefficients for the model incorporating call type and amplitude nested within call type. Of particular note are the higher intercept and slope coefficients for the 'long whistle' and 'short whistle' calls

| Effect | Estimate | SE | *t* value | Pr(>\|*t*\|) |
|---|---|---|---|---|
| call 'bell-like' (at 25 dB) | 20.204 | 0.207 | 97.745 | 0 |
| call 'long whistle' (at 25 dB) | 22.181 | 0.427 | 51.993 | 0 |
| call 'short whistle' (at 25 dB) | 22.239 | 0.189 | 117.674 | 0 |
| call 'bell-like' : amp | 0.859 | 0.019 | 44.635 | 0 |
| call 'long whistle' : amp | 0.869 | 0.037 | 23.446 | 0 |
| call 'short whistle' : amp | 0.896 | 0.022 | 40.663 | 0 |

the ARU, where changes in gain are perhaps more likely to have an impact. As expected, the detection space of the SM3 using old microphones in apparently poor condition was smaller than for an SM3 with new microphones, although this difference was not statistically significant.

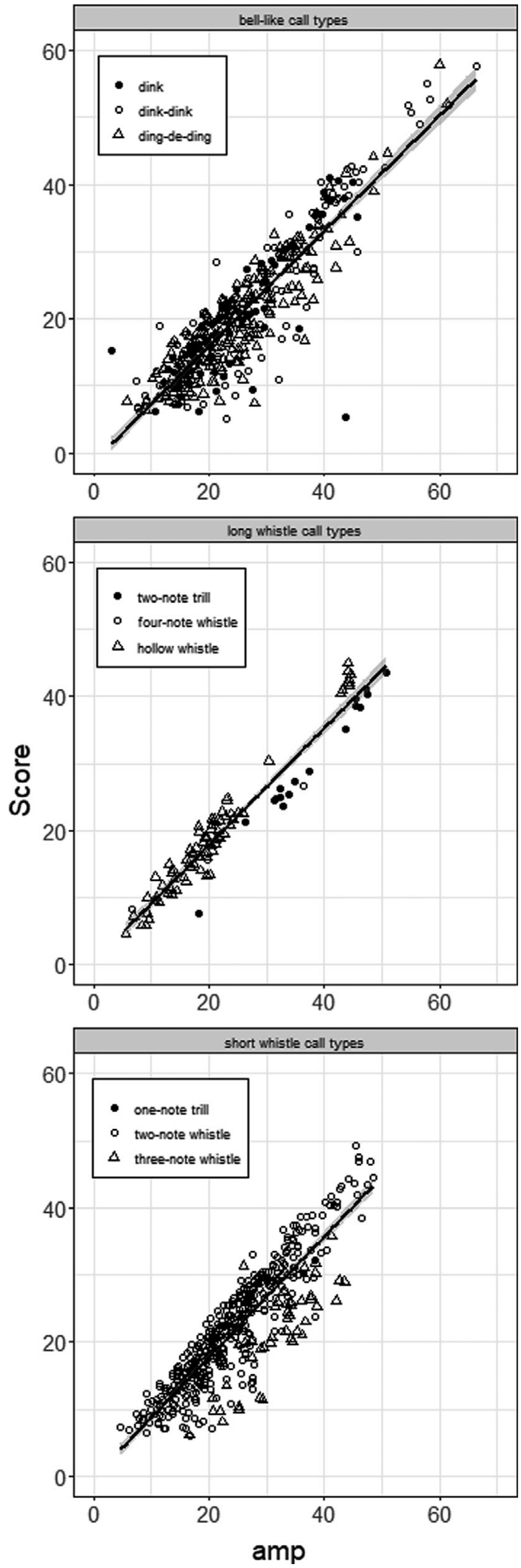


**Fig. 5.** Plots showing the relationship between score and relative amplitude for each call type group. Relationships are similar across groups, with specific call types also displaying similar relationships within the groups.

## Probability of detecting Night Parrots

The overall probability of detection is the product of the probability the species will be detected if it is

doi:10.1111/aec.13128

**Table 4.** Results of solving the equation for detection probability and distance given a specified score cut-off using the SM4 with microphones in good condition and the default gain setting applied (SM4_N_16). For example, if applying a score cut-off of 15, 95% of calls given within 138 m of the ARU will be extracted from the acoustic data set by the recogniser. Alternatively, the probability that a Night Parrot call given within 138 m of the ARU will score ≥ 15 is 95%

Probability call will be detected

| Score Cut-off | 95% | 90% | 85% | 80% | 75% | 70% | 65% | 60% | 55% | 50% |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 214 | 236 | 252 | 266 | 278 | 290 | 302 | 314 | 327 | 340 |
| 11 | 196 | 216 | 230 | 242 | 253 | 264 | 274 | 285 | 295 | 306 |
| 12 | 180 | 198 | 210 | 221 | 231 | 241 | 250 | 259 | 268 | 278 |
| 13 | 165 | 181 | 193 | 203 | 212 | 220 | 228 | 236 | 245 | 253 |
| 14 | 151 | 166 | 177 | 186 | 194 | 201 | 209 | 216 | 223 | 231 |
| 15 | 138 | 152 | 162 | 170 | 178 | 185 | 191 | 198 | 205 | 211 |
| 16 | 125 | 138 | 148 | 156 | 163 | 169 | 175 | 181 | 187 | 194 |
| 17 | 114 | 126 | 135 | 142 | 149 | 155 | 160 | 166 | 172 | 177 |
| 18 | 103 | 115 | 123 | 130 | 136 | 141 | 147 | 152 | 157 | 162 |
| 19 | 93 | 104 | 112 | 118 | 124 | 129 | 134 | 139 | 144 | 148 |
| 20 | 83 | 93 | 101 | 107 | 112 | 117 | 122 | 126 | 131 | 135 |
| 21 | 74 | 84 | 91 | 96 | 101 | 106 | 111 | 115 | 119 | 123 |
| 22 | 65 | 74 | 81 | 86 | 91 | 96 | 100 | 104 | 108 | 112 |
| 23 | 56 | 65 | 72 | 77 | 82 | 86 | 90 | 94 | 97 | 101 |
| 24 | 48 | 57 | 63 | 68 | 72 | 76 | 80 | 84 | 87 | 91 |
| 25 | 40 | 49 | 55 | 59 | 64 | 67 | 71 | 75 | 78 | 81 |

dist (m)
300
200
100

available for detection, and the probability it is available for detection. Our modelling has solved the first part of this equation, allowing the definition of a sampling radius for a specified call recogniser score threshold, and the associated probability of detecting a Night Parrot call, if one calls within that sampling radius. The second part of this equation is a behavioural question and, in the case of the Night Parrot, relates to how likely the species is to call. Research suggests that calling activity in and around long-term stable roost sites is very predictable (Murphy, Austin, *et al.* 2017), and the probability of calls being given during any dusk or dawn calling period is high. If the sampling period encompasses several dusk and dawn calling periods, this probability is close to one (S. Murphy, N. Leseberg unpub. data). The data presented here suggest that robust conclusions about the presence or absence of Night Parrots at prospective long-term stable roost sites are possible using an array of appropriately spaced ARUs deployed for short periods. Understanding the probability of detection away from long-term stable roost sites will require further research into the calling behaviour of Night Parrots when moving within the wider landscape.

## Importance of other recogniser performance parameters

Recogniser performance is typically assessed using two parameters: precision and recall. For a specified score threshold, precision is the proportion of detections returned by a call recogniser that represent actual calls (true positive calls), while recall is the proportion of true positive calls detected by the recogniser that were available for detection within the data set. For all recognisers, there is a trade-off between recall and precision. Setting a high score threshold increases a recogniser's precision, meaning a higher proportion of detections will represent true positives. However, this increases the likelihood of false negatives, actual target signals that do not meet the score threshold, and are not detected, for example quiet or distant calls. This reduces the recogniser's recall, or the proportion of all target signals identified within a data set. Conversely, reducing the score threshold returns more low-scoring target signals as matches, but simultaneously returns more low-scoring non-target signals, or false positives. This increases the recogniser's recall, but also increases the proportion of non-target signals in the resulting data set, thereby decreasing precision.

This research has linked sampling radius to recall for a specified score threshold. Recall is equivalent to the probability that a call will be detected if it is available for detection, the first of the two components of the probability of detection equation outlined earlier. The recogniser's precision will also be affected by changes to the score threshold, but this will not affect the probability of detection. Instead, this will manifest itself in the labour required to manually proof detections returned by the recogniser.

 doi:10.1111/aec.13128

Raising the score threshold will reduce the sampling radius for a specific recall, but will increase the precision of the recogniser, meaning fewer false positives will be returned, requiring less manual proofing of detections. Conversely, reducing the score threshold to increase the sampling radius for a specific recall will decrease the precision of the recogniser, meaning more false positives will be returned, requiring more manual proofing of detections.

## Potential limitations of this research

Our approach relies on several assumptions that are reasonable for this study, but should be reviewed if applying the method more widely. This study used data that were collected from only a few specific sites within an area of only several thousand hectares. Given the similarity between these sites, factors which may affect the models in different circumstances, such as sound transmission and background noise levels, did not vary. It is likely that most Night Parrots will be found in similar habitats and acoustic landscapes. However, when applying the method outlined here to surveys elsewhere, or to another species that may occur in a different acoustic landscape, the influence of these factors should be considered.

In this study, all calls that were incorporated into the distance data set were detected by the recogniser. In practice, this is unrealistic and probably reflects the relatively small number of calls in the data set. All recognisers will have a false negative rate; a selection of calls within the data set that for some reason will not be detected by the recogniser. While care should be taken to minimise this during recogniser development and testing, under some circumstances an increased false negative rate may be unavoidable. This may occur, for example, at a site where call types differ from those used to train the recogniser. Quantifying this effect without a comprehensive test data set is difficult. Ultimately, care should be taken to ensure recogniser performance is understood before applying the methods outlined here in unfamiliar circumstances.

The approach used here is uniquely suited to call recognisers based upon a template-matching algorithm such as ‘monitoR’; because a call’s amplitude is a key component in calculating that call’s score, the established relationship between amplitude and distance is transferable to score. A similar relationship exists between distance and detection probability for acoustic data generally and also for some other recognisers (Knight & Bayne 2018). While it is likely this relationship exists for most recogniser types, the strength of the relationship may not be universal. If this method is applied using another recogniser algorithm, testing will be necessary to determine whether the relationship between recogniser detection score threshold and distance is strong enough to support similar conclusions.

Our method for estimating the distance of calling birds from the ARUs within the array using the relative amplitudes of calls was necessarily imprecise and reflects the unique challenge presented by the Night Parrot. It is nocturnal, extremely cryptic, very rarely seen even by researchers focussed on the species (N. Leseberg, S. Murphy, 2021, pers. obs.), and there is no captive population. Because so few populations are known, ethical considerations preclude unnecessary disturbance at long-term stable roost sites. Given these considerations, the approach applied here was imperfect, but necessary. It is important to note we were more likely to under- than overestimate the distance of a calling bird from the ARU, meaning our calculated sampling radii are also likely to be underestimates and therefore conservative.

For other species where research is less constrained, more accurate methods of distance assessment should be considered. Such methods could include establishing the precise position of the bird within the array via observation or radio tagging, using a sound pressure meter while observing a calling bird to determine the mean amplitude of a call given at a specific distance, or using multiple ARUs with accurate synchronised clocks to triangulate an animal’s position within an array. These results could be used to replicate a calling bird using playback at the appropriate volume within an array, building a more comprehensive test data set. These techniques have been used in other studies seeking to simulate birds calling at a specific distance (see, e.g. Digby *et al.* 2013; Turgeon *et al.* 2017).

We did not quantitatively assess the calibration of the microphones used in this study. While new microphones were assumed to operate within the tolerances specified by the manufacturer, the strength of our conclusions around the performance of the microphones in poor condition is limited. Declining microphone sensitivity over repeated field deployments has been demonstrated (Turgeon *et al.* 2017). However, in our experience, the majority of Night Parrot surveys will be performed by consultants, Indigenous ranger groups and conservation managers unlikely to have the time or expertise to repeatedly test and calibrate ARU microphones. In these circumstances, observing the condition of the microphones, while being aware of the amount of time they have already spent deployed in the field, represents a method to rapidly assess their suitability, hence our use of microphone condition associated with deployment time as a treatment within the study. Ultimately, our results confirm those of other studies that microphone performance will decline over time, and although microphones in apparently

doi:10.1111/aec.13128 

poor condition may still perform adequately, this should be a trigger to confirm microphone performance through calibration and testing.

### How to develop a specific survey protocol?

Applying the results of this research requires consideration of both the target species, a survey's objectives, and the resources available, including funding, and time for both fieldwork and subsequent analysis. As ARUs could be used to survey a range of species, with an enormous variety of vocalisations, the number and spacing of ARUs necessary to collect the initial distance data set will be species and vocalisation specific. For example, spatially accurate surveys of small anurans will require a tighter spacing than general presence or absence surveys of a bird or mammal with a louder or lower frequency vocalisation. These requirements should be considered before designing then testing the initial survey array. Our results establish a significant difference in detection space between different ARU types, which affects the area able to be surveyed, and must be understood. For a small survey, this may be less relevant. For large-scale surveys involving significant investment in field and analysis time, the improved detection space of higher quality ARUs is likely to offset the additional cost. Similarly, our results suggest new microphones outperform older microphones. This difference in detection space will increase over time as microphones are subject to additional exposure through the wear and tear of repeated deployments and long periods in the field. Again, the benefit of ensuring consistent performance through regular testing and, if necessary, replacement of microphones will probably offset the cost of reduced or unpredictable performance. Our research did not find that increasing the gain improved detection space, which could be specific to the scenario and detection algorithm we used. While this may be the case generally, we acknowledge there could be circumstances where adjusting gain is beneficial.

When deciding on score thresholds, the precision-recall trade-off discussed earlier must be considered. Prioritising a high probability of detection while looking to maximise the sampling radius will see a lower score threshold applied, effectively prioritising recall, but reducing precision and therefore necessitating substantial time spent manually proofing detections. Raising the score threshold to improve precision will reduce the number of false positives returned and therefore the amount of manual proofing necessary, but will also reduce the effective sampling radius, requiring more resources to sample the same area. Ultimately, the acceptable sampling radius will be determined by factors such as the importance of extracting all calls from the data set, the area to be sampled and the resources available, including ARUs, field time and analysis time. These should be considered when developing a survey protocol for a specific site. Appendix S3 gives a worked example of how to do this.

The process outlined here generated the data necessary to develop a robust and repeatable survey protocol for the Night Parrot, accounting for both ARU detection space and call recogniser performance, two parameters that have typically been considered separately. For species such as the Night Parrot, whose distribution is poorly known, but for which acoustic monitoring will be the only feasible method of survey, the development of such protocols is critical. Although the rarity and conservation status of the Night Parrot limited our options for collecting some of the required data, the overall process for calculating the required parameters was sound and could be easily improved and applied to other similar species. This process will be of particular use to researchers and field ecologists employing bioacoustic monitoring to better understand the distribution of a species.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the Maiawali people, on whose land the research for this paper was conducted. Bush Heritage Australia and the Australian Government's National Environmental Science Program through the Threatened Species Recovery Hub provided support for this research. NPL received an Australian Government Research Training Program (RTP) Scholarship and additional support through the Max Day Environmental Fellowship, University of Queensland strategic funding, and Birds Queensland. This research was performed under Queensland Government Scientific Purposes Permit WISP17875416 and University of Queensland Animal Ethics Approval Certificate GPEM/392/16.

## AUTHOR CONTRIBUTION

**Nicholas Leseberg:** Conceptualization (lead); Data curation (lead); Formal analysis (equal); Funding acquisition (equal); Methodology (lead); Writing – original draft (lead); Writing – review & editing (lead). **William Venables:** Formal analysis (equal); Methodology (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Stephen Murphy:** Conceptualization (supporting); Formal analysis (supporting); Methodology (supporting); Supervision (supporting); Writing – original draft (supporting); Writing – review &

 doi:10.1111/aec.13128

editing (supporting). **Nigel Jackett:** Conceptualization (supporting); Methodology (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **James Watson:** Conceptualization (supporting); Funding acquisition (equal); Methodology (supporting); Supervision (lead); Writing – original draft (supporting); Writing – review & editing (supporting).

## FUNDING

Bush Heritage Australia, and the Australian Government's National Environmental Science Program through the Threatened Species Recovery Hub provided funding for this research. NL received an Australian Government Research Training Program (RTP) Scholarship, and additional funding through the Max Day Environmental Fellowship, University of Queensland strategic funding and Birds Queensland.

## REFERENCES

Charif R., Waack A. & Strickman L. (2010) *Raven Pro 1.4 User's Manual*. Cornell Lab of Ornithology, Ithaca, NY 25506974

Collins B. (2021) New recordings of critically endangered night parrot music to ears of Kimberley rangers, scientists. ABC Kimberley, www.abc.net.au.

Dema T., Zhang L., Towsey M. *et al.* (2017) An Investigation into Acoustic Anlaysis Methods for Endangered Species Monitoring A Case of Monitoring the Critically Endangered White-bellied Heron in Bhutan. 2017 IEEE 13th International Conference on e-Science, Auckland, New Zealand.

Digby A., Towsey M., Bell B. D. & Teal P. D. (2013) A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods Ecol. Evol.* **4,** 675–83.

Hortal J., de Bello F., Diniz-Filho J. A. F., Lewinsohn T. M., Lobo J. M. & Ladle R. J. (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **46,** 523–49.

Jackett N. A., Greatwich B. R., Swann G. & Boyle A. (2017) A nesting record and vocalisations of the Night Parrot *Pezoporus occidentalis* from the East Murchison, Western Australia. *Australian Field Ornithol.* **34,** 144–50.

Joshi K. A., Mulder R. A. & Rowe K. M. (2017) Comparing manual and automated species recognition in the detection of four common south-east Australian forest birds from digital field recordings. *Emu* **117,** 233–46.

Katz J., Hafner S. D. & Donovan T. (2016) Tools for automated acoustic monitoring within the R package monitoR. *Bioacoustics* **25,** 197–210.

Knight E. C. & Bayne E. M. (2018) Classification threshold and training data affect the quality and utility of focal species data processed with automated audio-recognition software. *Bioacoustics* **28,** 539–54.

Knight E. C., Hannah K. C., Foley G. J., Scott C. D., Brigham R. M. & Bayne E. (2017) Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conserv. Ecol.* **12**(2).

Koch T. (2013) Tweets in the night, a flash of green: is this our most elusive bird? *The Australian* 29 Jun.

Lambert K. T. A. & McDonald P. G. (2014) A low-cost, yet simple and highly repeatable system for acoustically surveying cryptic species. *Austral Ecol.* **39,** 779–85.

Landau H. J. (1967) Sampling, data transmission, and the Nyquist rate. *Proc. IEEE* **55,** 1701–6.

Leseberg N. P., McAllan I. A. W., Murphy S. A. *et al.* (2021) Using anecdotal reports to clarify the distribution and status of a near mythical species: Australia's Night Parrot (Pezoporus occidentalis). *Emu - Austral. Ornithol.* **121**(3), 239–49.

Leseberg N. P., Murphy S. A., Jackett N. A. *et al.* (2019) Descriptions of known vocalisations of the Night Parrot *Pezoporus occidentalis*. *Australian Field Ornithol.* **36,** 79–88.

Leseberg N. P., Venables W. N., Murphy S. A. & Watson J. E. M. (2020) Using intrinsic and contextual information associated with automated signal detections to improve call recognizer performance: A case study using the cryptic and critically endangered Night Parrot Pezoporus occidentalis. *Methods Ecol. Evol.* **11,** 1520–30.

Mace G. M., Collar N. J., Gaston K. J. *et al.* (2008) Quantification of extinction risk: IUCN's system for classifying threatened species. *Conserv. Biol.* **22,** 1424–42.

Measey G. J., Stevenson B. C., Scott T., Altwegg R. & Borchers D. L. (2017) Counting chirps: acoustic monitoring of cryptic frogs. *J. Appl. Ecol.* **54,** 894–902.

Mills V. & Collins B. (2017) Night parrot captured in blurry image in the Great Sandy Desert. Australian Broadcasting Commission, www.abc.net.au.

Murphy S. A., Austin J. J., Murphy R. K. *et al.* (2017) Observations on breeding Night Parrots (*Pezoporus occidentalis*) in western Queensland. *Emu* **117,** 107–13.

Murphy S. A., Silcock J., Murphy R. K., Reid J. R. W. & Austin J. J. (2017) Movements and habitat use of the night parrot *Pezoporus occidentalis* in south-western Queensland. *Austral Ecol.* **42,** 858–68.

Pollock K. H., Marsh H., Bailey L. L., Farnsworth G. L., Simons T. R. & Alldredge M. W. (2004) Separating components of detection probability in abundance estimation: an overview with diverse examples. In: *Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters* (ed W. L. Thompson) pp. 43–58. Island Press, Washington, USA. 43–58

Priyadarshani N., Marsland S. & Castro I. (2018) Automated birdsong recognition in complex acoustic environments: a review. *J. Avian Biol.* **49,** e01447.

Priyadarshani N., Marsland S., Juodakis J., Castro I. & Listanti V. (2020) Wavelet filters for automated recognition of birdsong in long-time field recordings. *Methods Ecol. Evol.* **11,** 403–17.

R Core Team (2018) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Schabenberger O. & Pierce F. J. (2002) *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press, Boca Raton, FL.

Shonfield J. & Bayne E. M. (2017) Autonomous recording units in avian ecological research: current use and future applications. *Avian Conserv. Ecol* **12**(1).

Sueur J., Aubin T. & Simonis C. (2008) Seewave: A free modular tool for sound analysis and synthesis. *Bioacoustics* **18,** 213–26.

 

Sugai L. S. M., Desjonquères C., Silva T. S. F. & Llusia D. (2020) A roadmap for survey designs in terrestrial acoustic monitoring. *Remote Sens. Ecol. Conserv.* **6,** 220–35.

Sutherland W. J. (2006) *Ecological Census Techniques: A Handbook*, 2nd edn. Cambridge University Press, Cambridge.

Sutherland W. J., Pullin A. S., Dolman P. M. & Knight T. M. (2004) The need for evidence-based conservation. *Trends Ecol. Evol.* **19,** 305–8.

Swiston K. A. & Mennill D. J. (2009) Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers. *J. Field Ornithol.* **80,** 42–50.

Teixeira D., Maron M. & van Rensburg B. J. (2019) Bioacoustic monitoring of animal vocal behaviour for conservation. *Conserv. Sci. Pract.* **1**(8)**,** e72.

Thomas A., Speldewinde P., Roberts J. D., Burbidge A. H. & Comer S. (2020) If a bird calls, will we detect it? Factors that can influence the detectability of calls on automated recording units in field conditions. *Emu - Austral Ornithol.* **120,** 239–48.

Towsey M., 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,** 110–9.

Turgeon P. J., Van Wilgenburg S. L. & Drake K. L. (2017) Microphone variability and degradation: implications for monitoring programs employing autonomous recording units. *Avian Conserv. Ecol.* **12,** 84–115.

Venier L. A., Mazerolle M. J., Rodgers A., McIlwrick K. A., Holmes S. & Thompson D. (2017) Comparison of semiautomated bird song recognition with manual detection of recorded bird song samples. *Avian Conserv. Ecol.* **12**(2).

Williams E. M., O'Donnell C. F. J. & Armstrong D. P. (2018) Cost-benefit analysis of acoustic recorders as a solution to sampling challenges experienced monitoring cryptic species. *Ecol. Evol.* **8,** 6839–48.

Yip D. A., Leston L., Bayne E. M., Solymos P. & Grover A. (2017) Experimentally derived detection distances from audio recordings and human observers enable integrated analysis of point count data. *Avian Conserv. Ecol.* **12**(1).

## SUPPORTING INFORMATION

Additional supporting information may/can be found online in the supporting information tab for this article.

**Appendix S1**. Example spectrograms of the *dink* (top) and *dink-dink* (bottom) calls, which comprised the distance dataset.
**Appendix S2**. Detection radius tables for all combinations of ARU model, microphone condition, and gain setting tested in this study. A value of '15' signifies the distance is less than 15 m.
**Appendix S3**. A worked example of how to apply the findings of this research.

 doi:10.1111/aec.13128