Can you hear me now?

A cross-comparison between traditional bird counts and audio recording systems for bird richness estimates

Sara Alves Coutinho
Lincoln University

Abstract

Bird population monitoring is an important indicator of an ecosystems' health. New emerging technologies, such as audio recorders, have the potential to complement or replace traditional bird monitoring methods, but the effectiveness of these new technologies has not yet been widely tested. In this study we compared how the detection of bird species richness varied between traditional five-minute point counts performed by two observers with different levels of expertise and the Department of Conservation AR4 audio recorder. We also compared the effectiveness of two different audio recorders available for sale in New Zealand: the Department of Conservation AR4 and the 2040 Bird Monitor. We found no significant difference between the mean number of species detected by the novice observer, expert observer and the AR4 audio recorder. There was a large overlap between the species detected by the three approaches, but each identified unique species. We found a significant difference in the mean number of species detected by the two audio recorders. We were able to detect four times more bird vocalisations and 50% more bird species with the AR4 when compared to the 2040 Bird Monitor. Our results suggest that audio recorders can be used as a complement or replacement of traditional bird monitoring methods, but they also revealed differences between the performance of two audio recorders. We therefore recommend evaluating the performance of audio recorders before deploying them in the field, and endorse the use of the Department of Conservation AR4 for studies of shorter duration where the main objective is estimating bird species richness.

Introduction

Bird populations are an important indicator of an ecosystems' health (Temple & Wiens, 1989); hence bird monitoring is a powerful tool to provide insights into conservation and policy activities. A widely used method to monitor bird populations is point counts (Ralph, Sauer, & Droege, 1995). In this method observers stay stationary for a set amount of time and record all bird species seen and heard (Greene & Hartley, 2012). Bird identification often relies on bird vocalisations (Ralph et al., 1995), particularly in forest environments (Greene & Hartley, 2012), where birds are more easily heard than seen (Mortimer & Greene, 2017; Scott Brandes, 2008). Point counts have several advantages: they are cheap, easy to run and require no specialised equipment (Greene & Hartley, 2012; Shonfield & Bayne, 2017). This allows counts to be conducted in large numbers (Greene & Hartley, 2012). Point counts are a well-established technique in New Zealand, and five-minute counts have been in use for more than 30 years (Greene & Hartley, 2012). Two downsides of point counts are the need for trained staff to perform them (Scott Brandes, 2008), and the variance between observers. Correct identification relies on the observer's ability to recognise birds and often varies depending on the identifier's skill level, age and hearing capacity (Scott Brandes, 2008).

A revolutionary method made possible by technological advances is the use of automated sound recorders to monitor bird populations. Automated recording units (ARU) can now be used as a replacement or as a complement to point counts. Automated recording units fall into several broad categories: commercial recorders with a timer attached; programmable recording devices, including smartphones; and custom-made single board computers (Scott Brandes, 2008). ARUs have several advantages over traditional bird counting techniques. They are a non-invasive technique (Pérez-Granados, Bota, Giralt, Albarracín, & Traba, 2019) that can be deployed to remote locations and collected at a later stage or left on site; this removes any potential disruption observers might cause to bird species during point counts (Digby, Towsey, Bell, & Teal, 2013; Scott Brandes, 2008). ARUs also provide a permanent and accurate recording of bird vocalisations that can be accessed at any stage for review and replayed as many times as needed, reducing potential user bias on identification (Pérez-Granados et al., 2019). Two of the downsides of using ARUs are the vast amount of data collected and the inherent complexity of manually parsing this data (Wimmer, Towsey, Planitz, Williamson, & Roe, 2013). Some ARUs record continuously, collecting days' worth of data that need to be manually analysed; this process can be fast-tracked by using spectrograms, a visual representation of a sound wave (Digby et al., 2013). Bird identification by using machine learning is under development and looks promising (Wimmer et al., 2013). However, automated detection has its own issues: the methods used rely on large training datasets and are not very effective on birds that don't vocalize often; and automated detection can lead to a large number of false negatives and false positives (Digby et al., 2013).

The infancy of this technology means that research into the efficacy of different ARUs and a direct comparison between them, from both a cost/benefit and detection ability point of view, has been limited. ARUs have been compared to traditional bird counting methods performed by individuals in order to understand the detection performance between field observers and listeners of bird recordings, with inconsistent findings across studies. Past research indicated a slightly lower detection rate for ARUs when compared to field surveys (Venier, Holmes, Holborn, McIlwrick, & Brown, 2012). Another study found a large overlap between species detected by both methods, but with some unique species being detected by each (Leach, Burwell, Ashton, Jones, & Kitching, 2016). More recent studies indicated equivalent results for both methods (Darras et al., 2018). One of the two studies performed in New Zealand indicated that ARUs can be a viable alternative to traditional point counts (Digby et al., 2013), while the other found that the results from both methods produce similar results (Bombaci & Pejchar, 2019). Research comparing different types of ARUs has been sparser. Several studies have attempted this, either by comparing different categories of hardware solutions (Scott Brandes, 2008) or specific models of ARUs. A study comparing different models of ARUs found that analysts of audio recordings detected 10% less species with certain ARU models, although there was more variance between the analysts themselves than the ARUs (Rempel et al., 2013). A recent study found large differences in effectiveness between recorders (Pérez-Granados et al., 2019). Both studies detected no relationship between unit cost and detection performance.

In this study we aim to determine how detection of bird species richness varies in three different scenarios: between point counts performed by two observers with different levels of expertise; between point counts and an ARU; and between two different ARUs. Firstly, we investigated differences in bird identification accuracy rates between a less and a more experienced participant taking part in the same point counts. Secondly, we examined variance of bird species' detection rates between traditional point counts and detection using automated recording units, by analysing the data collected by a human observer and the equivalent data collected by an ARU. Lastly, we compared the detection rates of two ARUs available for sale in New Zealand: the 2040 Bird Monitor (2040, n.d.-b) and the Department of Conservation AR4 Acoustic and Bat recorder (Department of Conservation, n.d.), by analysing the overlapping data collected by both ARUs. We predict that a more experienced participant in traditional point counts will be able to detect more bird species than a less experienced participant; that ARUs may lead to a slightly lower species richness detection than a point count, due to the ARU smaller radius of detection; and that ARUs sold at a similar price point will lead to a similar bird species' detection rate.

Methods

Study area

This study was carried out from 3-7 February 2020, on a beech terrace situated 250m south-east of the Boyle River Outdoor Education Centre (BROEC; 42°31'00.8"S 172°23'02.1"E). The BROEC is situated 15 km south of Lewis Pass, in the northern region of Canterbury, New Zealand. The vegetation of the terrace consists predominantly of mixed-beech forest including red beech (*Fuscopora fusca)*, silver beech (*Lophozonia menziesii*) and mountain beech (*Fuscospora cliffortioides*) at various levels of maturity (Figure 1). The area has a recent history of disturbance, particularly by fire (Department of Conservation, 2006), and is surrounded by kanuka (*Kunzea* sp.) groves to the north and east and grassland to the east. The western side is bordered by State Highway 7.

The area, Poplars Range, is part of the Lewis Pass National Reserve and sits at an altitude of 600 meters, with a landscape characterised by fans and river terraces (Department of Conservation, 2006). The region has a cool, wet climate, with minimum temperatures of -7 °C and maximum of 32 °C (Stewart et al., 1991), and an average rainfall of 1300 mm per year (Department of Conservation, 2006). During the duration of the study temperatures varied between a minimum of 8.6 °C and a maximum of 27 °C. Fourty-four mm of rain occurred during the second day and during the night of the third day. There were also some occasional rain showers, specifically 0.12 mm, during the last two days (NIWA, n.d.).

The bird fauna of the area is particularly diverse, with several declining and endangered native species present, including yellow-crowned parakeet, orange-fronted parakeet, kea, New Zealand falcon and riflemen (Department of Conservation, 2006). There is also a wide range of introduced birds such as blackbird, chaffinch, greenfinch, goldfinch, redpoll, house sparrow and song thrush (Department of Conservation, 2006). Introduced predator species, namely possums, mustelids and rodents represent a threat to native species but a predator control program has been underway since 2013 (Graham, 2019).

Figure 1 – Panorama of part of the study area situated in the beech forest.

Materials and Methods

Site establishment

The area has been the subject of several bird and mammal monitoring projects in the past and we used the same setup as Hoete-Dodd (2013) to provide continuity to those studies. The setup consisted of four transect lines, each 400 m in length and 150 m apart, with five tracking tunnel locations per line. We used the existing tracking tunnel locations as the sampling sites for this study (Figure 2, Appendix A).

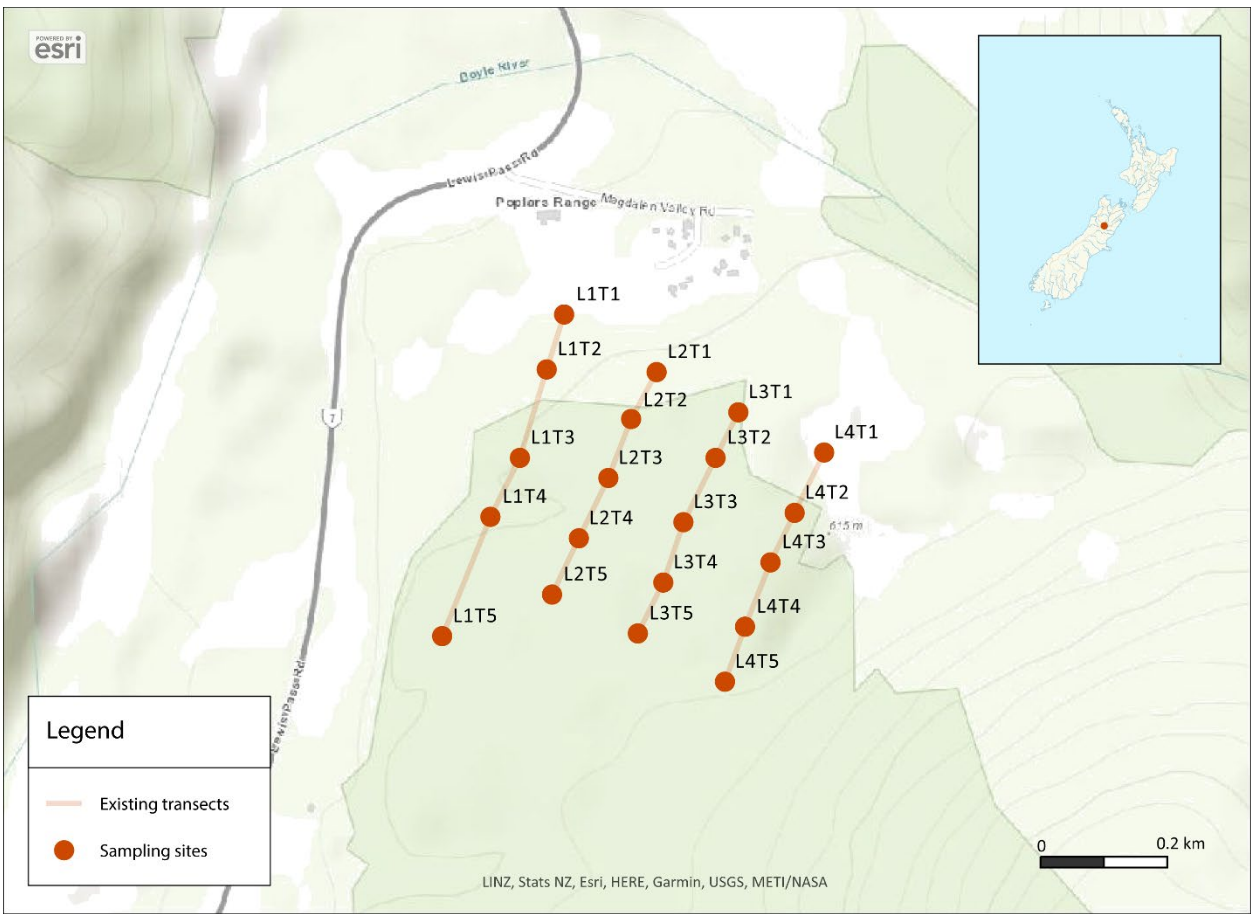

Figure 2 – Map of the study area showing the existing transects and sampling sites.

Point counts

We performed twenty point counts per day at each sampling site from the 4-6 February inclusive, resulting in three counts at each sampling site and a total of sixty counts. We randomized the order of transect lines but performed the counts in a linear order per transect. Count times varied between mid-morning and late afternoon. The point counts followed the standard Department of Conservation protocol for five minute bird counts (Greene & Hartley, 2012), extended to include the minimum and maximum number of individuals present and distance bands.

The more experienced bird identifier took part in half of sixty point-counts and the less experienced birder took part in all of them (Figure 3). Each participant self-assessed their level of expertise.

Figure 3 - The novice birder performing a five-minute bird count in one of the sampling sites. Photo by Uykim Lim.

Automated recording units

We sampled each site using the following automated recording units (Table 1): 20 2040 Bird Monitors (Figure 4) (2040, Christchurch) and 12 Department of Conservation AR4 (Figure 4) (Department of Conservation, Wellington). Given that we had only twelve Department of Conservation (DoC) ARUs for twenty sites, we set them up at site 1, 3 and 5 of each transect line. All twenty replicates had a 2040 Bird Monitor. The ARUs were left on site from late afternoon on 3 February until the morning of 7 February 2020.

Table 1 - Name and specifications of the two ARUs used in the study

ARU Name	**Manufacturer**	**Cost (NZD)**
2040 Bird Monitor	2040	$299 + GST
Department of Conservation AR4	Department of Conservation	$375 + GST* *price on enquiry, cost of current batch

Figure 4 – 2040 Bird Monitor (left) and Department of Conservation AR4 (right) on site.

Each ARU was configured before being taken to each site. We used the default recording settings for the 2040 ARU: 32 random samples per day, each one minute long (2040, n.d.). We originally set up the DoC ARU to record continuously using high-quality settings (protocol of High) but this seemed to trigger an issue that rendered recordings unusable from midnight. We changed DoC ARU settings on the second day to record continuously using high-quality audio recording settings (protocol of High) from 6am to 9pm and low-quality audio recording settings (protocol of Low) from 9pm to 6am (Department of Conservation, n.d.).

We attached the ARUs to trees at a height varying between 1.2 m and 2 m and left them in the same position for the duration of the study. When two ARUs were present on the same site, we attached them to neighbouring trees. We took special care to ensure that the ARUs positioning and proximity didn't interfere with their recording ability and performance.

Due to time limitations, we were unable to run a field trial of the DoC ARUs before using them in this study. This prompted us to change the batteries of the DoC ARU on the third day of the experiment, as we were unsure their charge would last until the end of the study.

Data cleaning

We weren't able to gather complete recordings for some ARUs due to technical issues and potentially user error. We only used ARU recordings that could be matched to the equivalent manual bird count or equivalent ARU of different make (Appendix B).

Comparison between manual bird counts and ARU

We matched the overlapping manual bird counts and the DoC ARU recordings using their timestamps. We then sliced the equivalent five minutes of the DoC ARU recordings and identified the birds present based on their vocalisations, ending up with a total of 18 overlapping counts between the DoC ARU and the human observers.

Comparison between ARUs

We matched the overlapping recordings of both ARUs using their timestamps. We sliced the DoC ARU recordings and extracted the audio portion that matched the one-minute 2040 ARU recordings. During this process it became clear that the recordings didn't match due to the clocks of each device being slightly offset from one another. We found the right offset by converting the 2040 recordings to spectrograms and looking through them for easily recognisable signal, such as distinctive bird calls.

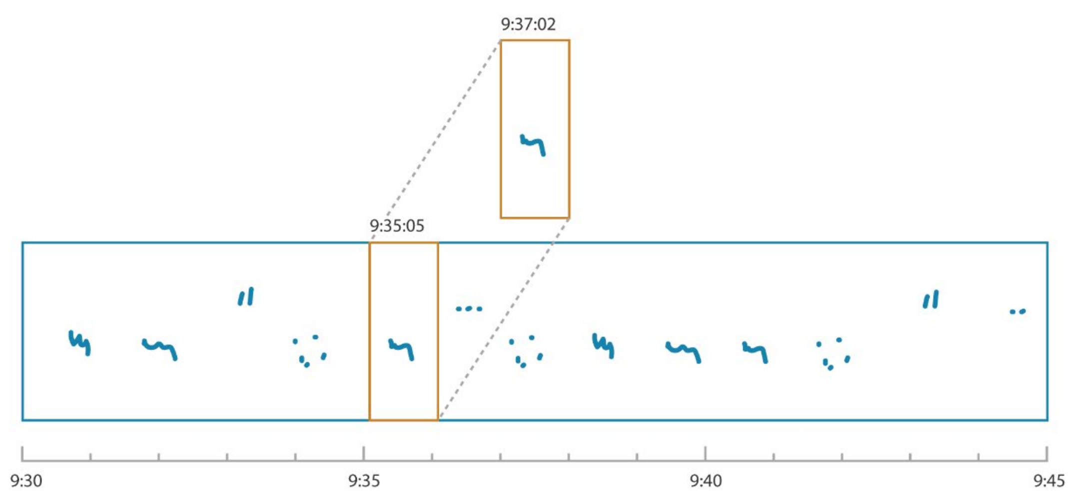

Figure 5 – A visual representation of the process of calculating the clock variance between ARUs by looking at spectrograms of overlapping recordings and finding matches.

We then matched this signal to the corresponding signal in the DoC recording, identified how far off it was from the originally extracted one-minute recording and calculated the clock variance between ARUs (Figure 5) (Appendix B).

Bird vocalisation identification in the audio recordings

We used the software Raven Lite (The Cornell Lab of Ornithology, 2019) to tag bird vocalisations in the audio recordings. We also used the software Adobe Audition (Adobe, 2019) for faster navigation between and within audio files, and to improve the quality of the audio. The improvements included increasing the audio volume, removing noise and reducing low frequencies to remove wind and anthropogenic noise.

We divided the bird species identification into two categories: certain or uncertain. Bird vocalisations tagged as certain were clearly audible and their vocalisation were distinctive enough to identify species with certainty. Uncertain bird vocalisations were often muffled or too distant and weren't distinctive enough for a confident identification. We couldn't identify some bird vocalisations, even with external help, so we tagged those as unknown.

Statistical analysis

Comparison between manual bird counts and ARU

We compared the mean number of species identified by the novice observer and expert observer, and used the equivalent bird species data extracted by listening to the five-minute recordings of the Department of Conservation AR4. We ended up with a total of 18 counts per observer/recorder.

We performed two analyses: one excluding observations where species identification was uncertain and another including all observations, regardless of identification certainty. We ran a one-way repeated measures ANOVA in R (v3.6.2; The R Foundation, 2020) to compare the mean number of species detected per count by the novice, expert and DoC ARU. We analysed significant results and ran post-hoc tests using multiple pairwise paired t-tests, adjusted using the Bonferroni multiple testing correction method.

Comparison between ARUs

We compared the mean number of species identified by 2040 Bird Monitor and the Department of Conservation AR4 by listening to the one-minute recordings of both ARUs that overlapped in time. We then identified all species present based on their vocalisations. We tagged all bird vocalisations in 6 of the 12 replicates, with each replicate having around 130 minutes of audio recordings spread across 4.5 days.

We performed two analyses: one excluding observations where species identification was uncertain and another including all observations, regardless of identification certainty. We ran a Wilcoxon Signed Rank Test in R (v3.6.2) to compare the mean number of species detected by the Department of Conservation AR4 and the 2040 Bird Monitor.

Results

Comparison between novice, expert and ARU

Comparison including only confident detections

There was no significant difference between the mean number of bird species detected by the novice, expert and the Department of Conservation AR4 ARU ($F_{1,21} = 1.423$, $P = 0.254$, $\eta^2_G = 0.031$; Figure 7), when we excluded observations where species identification was uncertain, during the 18 five-minute counts and equivalent five-minute data extracted from the recordings of the AR4 ARU.

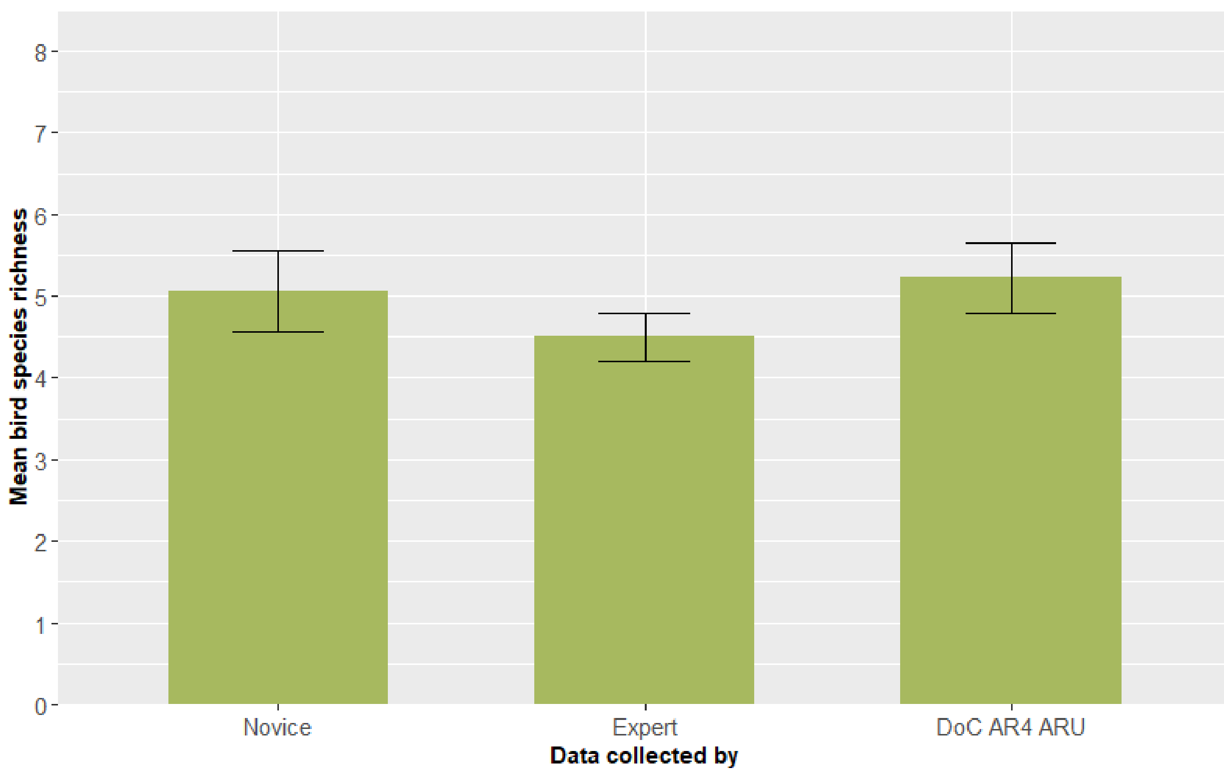

Figure 6 – Mean species richness detected by the novice observer and expert observer during the five-minute counts, and equivalent five-minute data extracted by listening to the recordings of the Department of Conservation AR4. Analysis excludes observation where bird species identification was uncertain. Error bars represent standard error.

Comparison including all observations

There was a significant difference between the mean number of bird species detected by the novice, expert and the Department of Conservation AR4 ARU ($F_{2,34}$ = 6.864, P = 0.003, η^2_G = 0.091; Figure 8) when all observations were included, regardless of their certainty, during the 18 five-minute counts and equivalent five-minute data extracted from the recordings of the AR4 ARU.

Post-hoc analyses with a Bonferroni adjustment revealed that mean number of species detected was significantly different between the novice and the expert (P = 0.021), with the novice detecting more species than the expert; and between the expert and the DoC ARU (P = 0.005), with the DoC ARU detecting more species than the expert. There was no significant difference in the mean number of species detected between the novice and the DoC ARU (P = 0.921) (Figure 8).

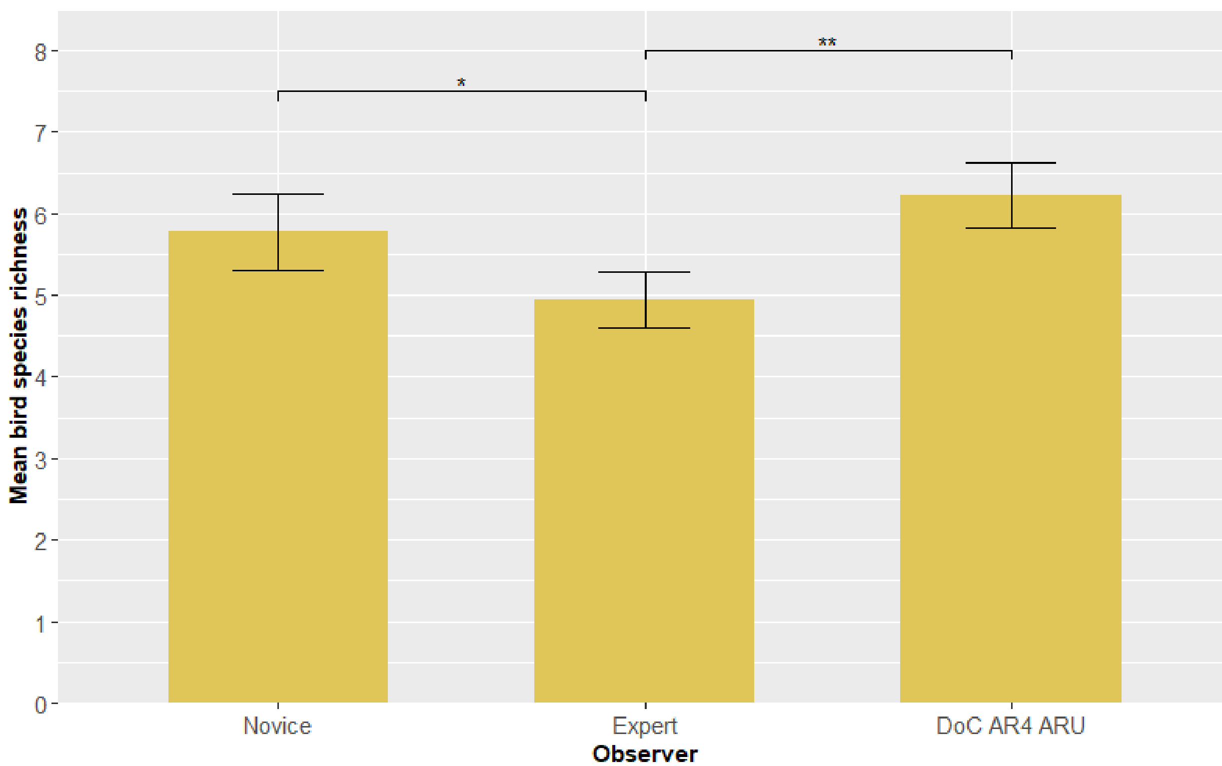

Figure 7 – Mean species richness detected by the novice observer and expert observer during the five-minute counts, and equivalent five-minute data extracted by listening to the recordings of the Department of Conservation AR4. Analysis includes observations note. Horizontal bars represent pairs with significant differences. A single asterisk (*) represents p < 0.05, a double asterisk (**) represents p < 0.01.

There was an outlier in the DoC ARU data that contained observations where bird species identification was both certain and uncertain. We removed that count from the three groups and ran the analysis again. The results still showed a significant difference between the mean number of bird species detected by the novice, expert and ARU ($F_{2,32} = 5.29$, $P = 0.01$, $\eta^2_G = 0.079$). Post-hoc analyses with a Bonferroni adjustment revealed that the mean number of species detected was significantly different between the novice and the expert ($P = 0.043$), with the novice detecting more species than the expert, but this difference was now on the verge of non-significance. There was still a significant difference in the mean number of species detected between the expert and the ARU ($P = 0.009$), with the ARU detecting more species than the expert.

Species identified

A total of 14 species were observed, including uncertain detections, down to 12 species when uncertain observations were excluded. There was a large overlap between the species detected, with a few exceptions: first, greenfinches were confirmed only with certainty by the novice; second, the presence of redpolls was only confirmed with certainty by the Department of Conservation AR4 ARU; and finally the DoC ARU was the only method to detect the presence of tūī (Figure 9).

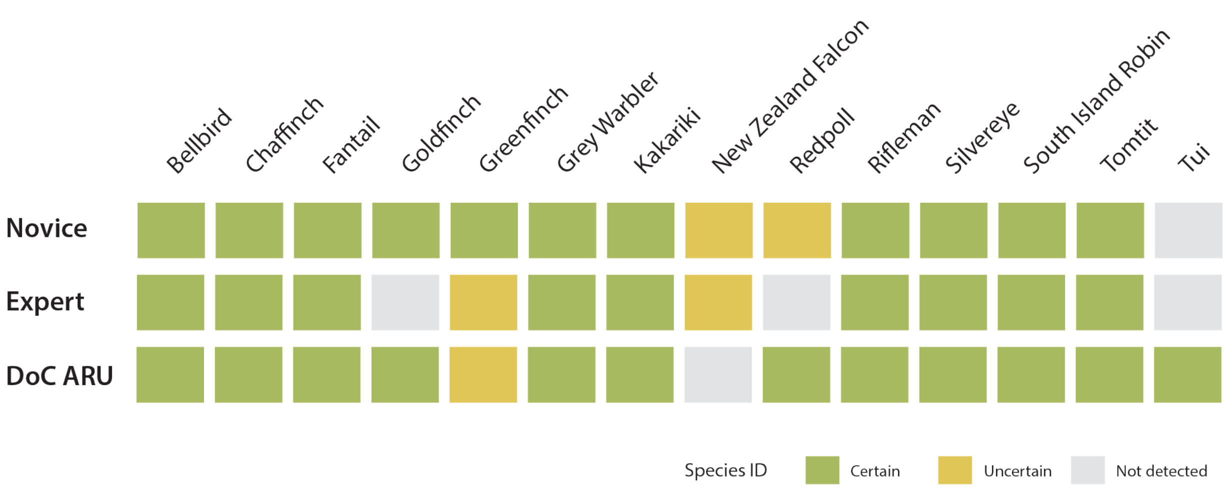

Figure 8 – Total species detected by the novice, expert and DoC ARU.

Species that were tagged as uncertain tended to be detected only on a few occasions. Tūī was recorded by the DoC ARU only once; this identification was crowdsourced from iNaturalist and later confirmed by someone familiar with tūī. Goldfinches were detected only once by the DoC ARU and redpolls were detected three times. Greenfinch, the only uncertain species detected by the DoC ARU, was detected three times.

Greenfinch was detected only once by the novice, but with certainty. The novice also detected the presence of redpolls and the New Zealand falcon only once, and both observations were uncertain. The two uncertain species detected by the expert, greenfinch and New Zealand falcon, were both detected only once.

Comparison between ARUs

Comparison including only confident observations

There was a significant difference between the mean number of bird species detected by the Department of Conservation AR4 and the 2040 Bird Monitor (V = 0, P = 0.03501; Figure 11), when we excluded observations where species identification was uncertain. The data for this analysis was extracted from the overlapping one-minute audio recordings of six replicates across both ARUs.

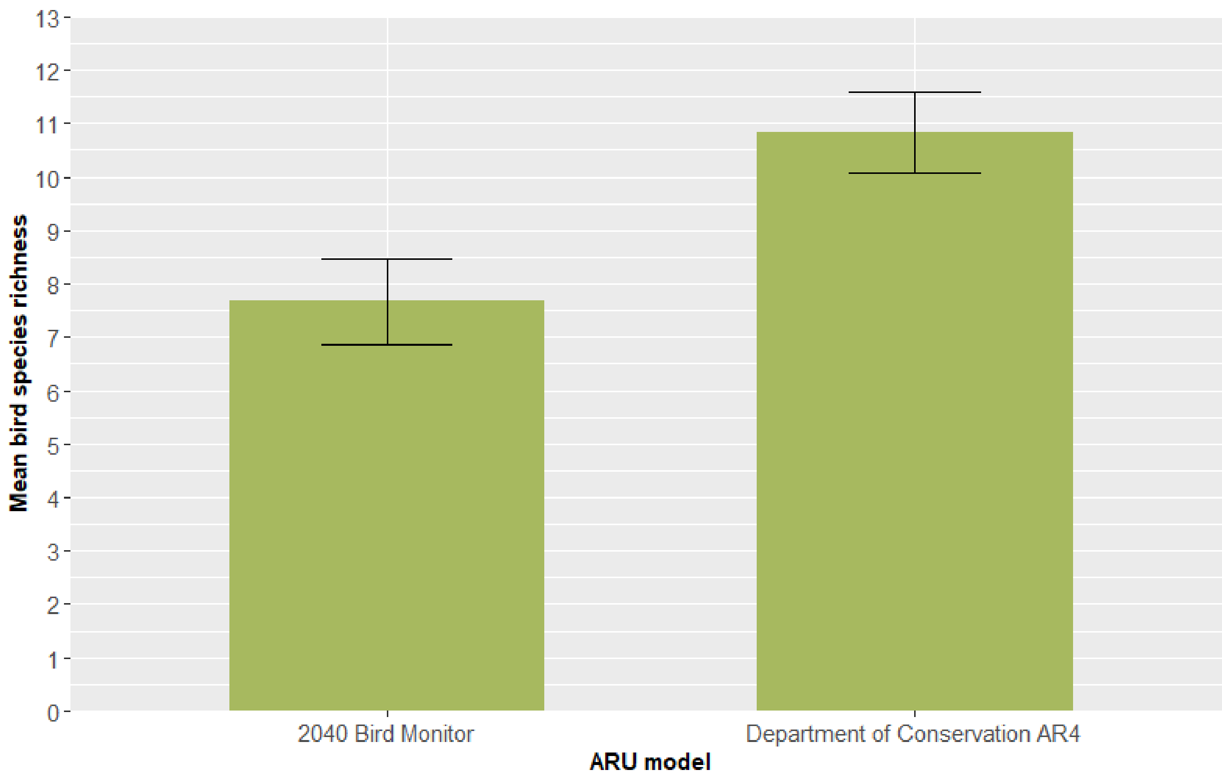

Figure 9 – Mean species richness detected by listening to the one-minute overlapping recordings of both ARUs and identifying all bird species present. Analysis excludes observation where bird species identification was uncertain. Error bars represent standard error.

The number of species detected by the 2040 ARUs in one of the replicates (L4T1) was much lower (4 species) than the number of species detected in all other replicates, making it an outlier. Conversely, the number of species detected by the DoC ARU in the same replicate was one of the highest (13 species). First, we looked at the total number of detections across all replicates (Figure 12). We then compared the total number of bird vocalisations detected by both ARUs to understand if there was a relationship between the number of vocalisations detected and the number of species detected per replicate. We found that the 2040 ARU had the lowest number of bird vocalisation and species detected in that particular replicate (L4T1) but this wasn't the case with the DoC ARU (Figure 13).

We removed the data from that replicate and ran the analysis again. The mean number of species detected was now not significant (V = 0, P = 0.05676), although this value was on the verge of significance.

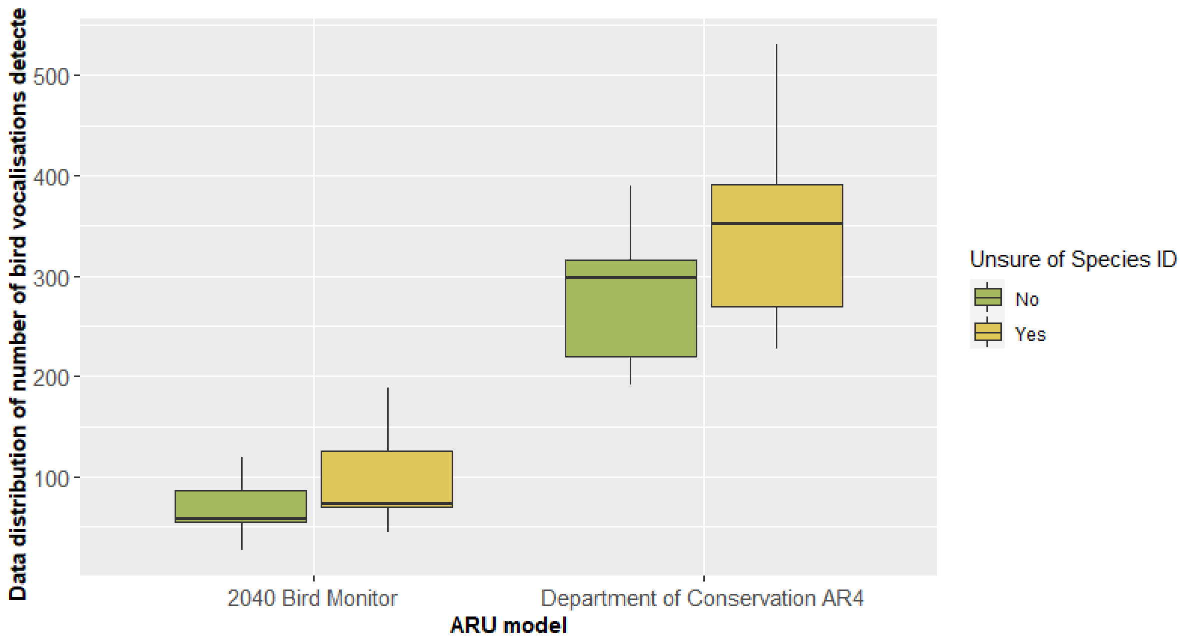

Figure 10 - Distribution of number of bird vocalisations detected by the 2040 Bird Monitor and the Department of Conservation AR4.

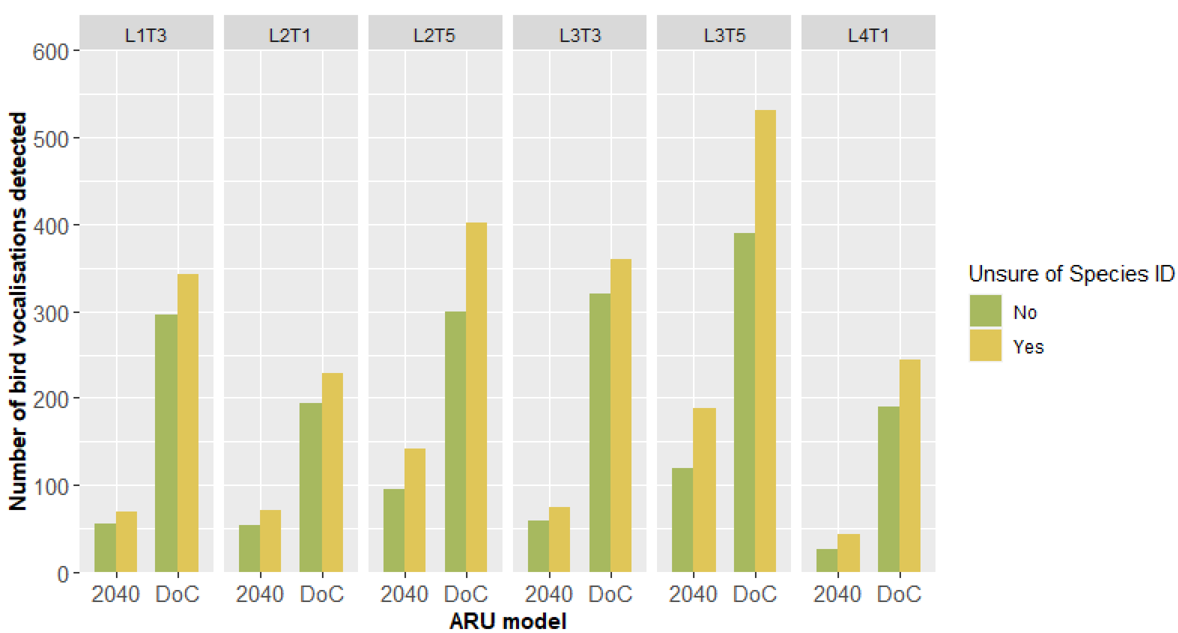

Figure 11 - Number of bird vocalisations detected by the 2040 Bird Monitor and the Department of Conservation AR, grouped by replicate.

Comparison including all observations

There was a significant difference between the mean number of bird species detected by the Department of Conservation AR4 ARU and the 2040 Bird Monitor (V = 0, P = 0.03401; Figure 12) when all observations were included, regardless of their certainty. The data for this analysis was extracted from the overlapping one-minute audio recordings of six replicates across both ARUs.

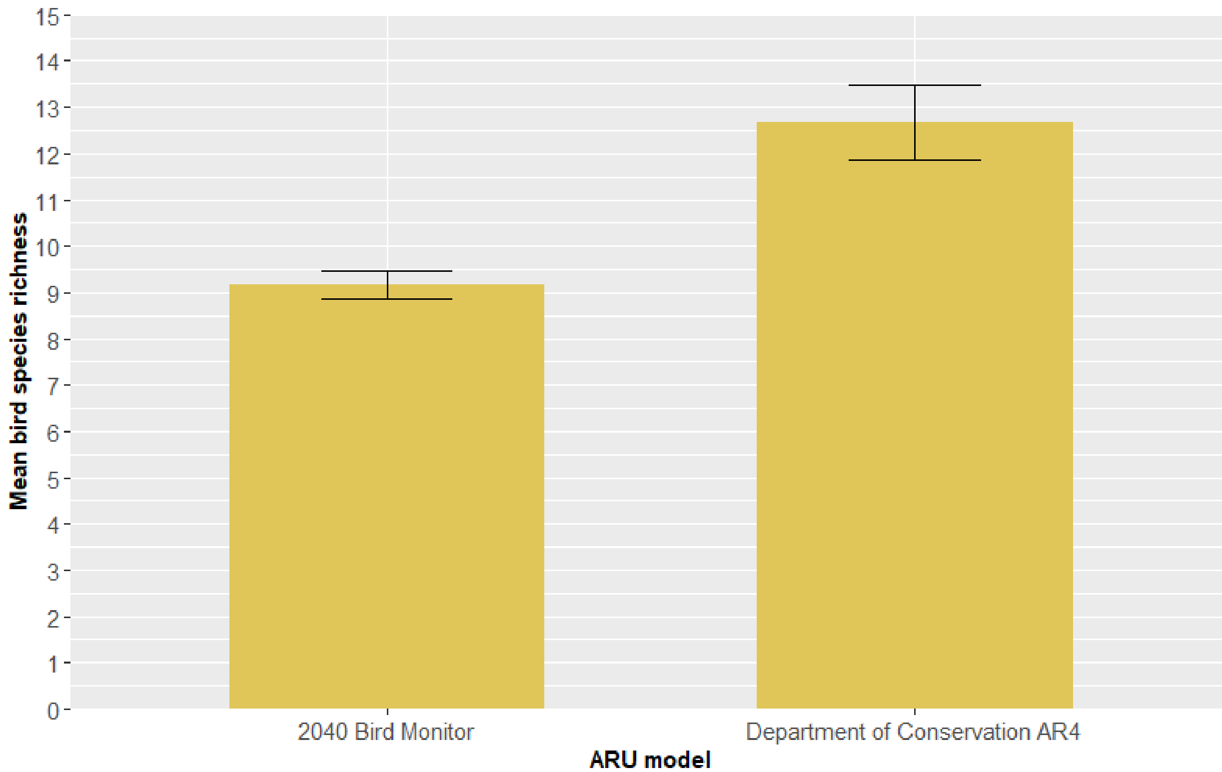

Figure 12 – Mean species richness detected by listening to the one-minute overlapping recordings of both ARUs and identifying all bird species present. Analysis excludes observation where bird species identification was both certain and uncertain. Error bars represent standard error.

Species identified

A total of 19 species were identified, including uncertain detections, down to 15 species when uncertain observations were excluded. We were able to detect considerably more species with Department of Conservation AR4 ARU than with the 2040 Bird Monitor. We detected 19 species with the DoC AR4 and 12 species with the 2040 Bird Monitor, including both certain and uncertain identifications. We detected 15 species with the DoC AR4 and 9 species with the 2040 ARU when we excluded species where identification was uncertain (Figure 13).

Figure 13 – Total species detected by each ARU.

We were able to confirm the presence of all four species of finch in New Zealand (greenfinch, goldfinch, redpoll and chaffinch) with the DoC ARU; these species were recorded by the 2040 ARU but we were not certain of their identification. We were also only able to detect blackbird, brown creeper, redpoll and kea, albeit without certainty, with the DoC ARU. Finally, we were only able to identify long-tailed cuckoo and New Zealand falcon with the DoC ARU; those were confident observations that we were not detected at all in the recordings of the 2040 ARU.

Uncertain species were detected a maximum of five times by both recorders. Interestingly, some species, such as New Zealand Falcon and long-tailed cuckoo were detected only once, but their vocalisations were clear and distinctive enough to identify them.

Discussion

Comparison between novice, expert and ARU

We found no difference between the mean number of species detected between the novice observer, expert observer and the Department of Conservation AR4 ARU when only species where identification was certain were considered. There was a significant difference between the novice and the expert and between the expert and the Department of Conservation AR4 ARU when uncertain species were included in the analysis. Both analyses used the data collected during the 18 five-minute counts and equivalent five-minute bird data extracted from the audio recordings of the AR4 ARU. Given that uncertain species were only detected a few times at most, it is likely that an increase in the number of counts would also increase the probability of detecting those species with certainty. There was a large overlap between species detected by different observers, but each observer identified unique species, supporting the results of Leach, Burwell, Ashton, Jones, & Kitching (2016).

Surprisingly, the ARU detected the greatest number of species and the expert observer detected the least. This might be due to several factors: first, the level of expertise of the human observers was self-assessed and based largely on the years of experience. A study by Mortimer, Greene, & Mortimer (2019) found that less experienced observers tended to underestimate their ability. The study suggestion of using a computer-based quiz to evaluate the observers' expertise should be employed in future studies to accurately assess each observer. Second, the expert observer may have been more conservative in the species identification than the novice observer. Third, hearing acuity might vary between observers and lead to different results.

The slightly higher species detection of the ARU can likely be explained by several factors: first, sound recordings can be replayed as many times as needed when species confirmation is uncertain. Second, recordings can be edited to increase the volume of certain sounds or reduce background noise, making species identification easier. Third, identification of unsure species was crowdsourced by using iNaturalist and didn't rely on a single observer.

The ARU recordings were also tagged by the novice observer; while there's a danger that this might introduce some biases in the analysis, we'd like to note that the results obtained by the novice and the ARU were different. There was indeed a large overlap in the species found, but each method identified unique species and the levels of confidence in species' identification varied. The recordings were also tagged after spending a considerable number of hours tagging the dataset used for the ARU-ARU comparison and this likely provided the novice with extra familiarity with bird vocalisations.

Comparison between ARUs

Our results show that there is a significant difference in species richness detection between ARUs, confirming the findings of two previous studies (Pérez-Granados et al., 2019; Rempel et al., 2013). It is worth noting that for our analyses we only used data that was directly comparable. We analysed the bird data extracted from the overlapping one-minute audio recordings of both ARUs, across six replicates where we had recordings for the full duration of the study. The Department of Conservation AR4 detected on average around 33% more species than the 2040 Bird Monitor per replicate, and the total number of species detected by the DoC ARU was almost 50% higher. Both ARUs are sold at a similar price point but their bird detection efficacy is considerably different.

We were able to detect four times more vocalizations with the DoC AR4 than with the 2040 Bird Monitor. This is likely due to the DoC ARU having better hardware and a wider detection range than the 2040 ARU. There are many factors that may impact the performance of ARUs, as Rempel et al. (2013) highlighted previously. The directionality of the microphones may have been a factor affecting performance, but neither manufacturer provides information about the directionality of their microphones and we didn't test them in the field.

Outlier

We removed one replicate for the comparison between ARUs, even though species identification was certain, as it was an outlier. This replicate had the lowest number of species and bird vocalisations detected by the 2040 ARU, but this was not the case for the DoC ARU. Interestingly, this same replicate was not an outlier when we looked at the data for all observations, regardless of detection certainty. The location of this replicate, an area of kanuka grove, might explain these results. The kanuka grove is more exposed to the elements than the beech forest and could indicate that the 2040 ARU is not as effective in open areas or that there was equipment malfunction.

Reliability

We weren't able to collect recordings for the full period of this study in three of the twelve DoC AR4 ARUs. This might have been because none of us were familiar with the DoC ARU before the study. We were unsure of what recording settings to use, battery life duration and storage space capacity. Conversely, we had used the 2040 Bird Monitors in a previous field trial and only lost recordings in one of them.

Detection of uncommon species

We were able to detect a species rarely recorded in the area with the AR4 ARU: a long-tailed cuckoo. This species was detected at night, highlighting the usefulness of audio recordings to detect rare birds.

Bird call tagging process

Time expenditure

It took us, on average, around 1.5 hours to analyse two hours of the 2040 ARU audio recordings and around 4.5 hours to analyse the matching DoC AR4 recordings. This ratio between times is consistent with the vocalization detection ratio between ARUs – we detected four times more vocalizations with the DoC AR4 than with the 2040 ARU. Stopping and starting the audio playback to tag bird vocalisations was the main reason for the increase of the time expenditure.

For the comparison between ARUs, we spent 9 hours tagging the matching one-minute 2040 ARU audio recordings and 27 hours tagging the DoC AR4 audio recordings. We spent 7 hours tagging the DoC AR4 five-minute recordings equivalent to the manual bird counts. Although these time efforts are equivalent to the time spent in traditional point counts in the field, we believe they cannot be directly compared as the nature of the work is vastly different. We found that there was only so much bird tagging we could do in a day before we started making mistakes or skipping bird vocalisations by accident. We therefore recommend that future studies are conservative when estimating how much time they need to allocate for record tagging.

Tooling

We found the tools available to tag bird vocalizations lacking and look forward to new developments in this area. Although Raven Pro is a well-established program (Wimmer et al., 2013) and commonly used in bird studies (Bombaci & Pejchar, 2019; Cook & Hartley, 2018; Digby et al., 2013; Rempel et al., 2013), we ended up using Adobe Audition as the main tool to play, edit and skip audio recordings. The main pain point identified with Raven Pro was the inability to jump to a specific location in an audio recording without having to play the whole recording. Adding this feature to Raven Pro would make the tagging process considerably faster.

Crowdsourcing of identifications

One of the advantages of ARUs, as mentioned by other studies (Rempel et al., 2013; Tegeler, Morrison, & Szewczak, 2012) is the ability to replay recordings and have more than one observer analysing the recording. We crowdsourced the identification of uncertain bird vocalisations by using iNaturalist. Almost two thirds of the vocalisations were able to be identified (two or more people agreed with the identification). This highlights how citizen science platforms and crowdsourcing can greatly benefit scientists around the world.

Limitations

We performed two-thirds of the counts used to compare the novice observer, expert observer and ARU during rainy and slightly windy weather. Wind and rain impair the ability of both humans (Greene & Hartley, 2012; Ralph et al., 1995) and audio recorders (Wimmer et al., 2013) to detect birds, and also affect the behaviour of birds (Greene & Hartley, 2012). We therefore recommend repeating these counts under different weather conditions.

Unfortunately, we didn’t have access to the Department of Conservation AR4 ARU before the experiment and as a result we had partial loss of data in three of the twelve AR4 ARUs. This once again highlights the need for field tests before deploying ARUs, as mentioned by Pérez-Granados, Bota, Giralt, Albarracín, & Traba (2019).

During our analysis we discovered that the clocks of ARUs didn’t match and drifted at a different rate. There was as much as almost a minute and a half variance in clock time between ARUs left in the same location. We adjusted the variation by looking at spectrograms and finding distinctive bird calls that could be used as a reference point, but this was a manual and time-consuming process. This issue might be mitigated by using an external service to resync the clocks of ARUs on a regular basis. Both the DoC AR4 and the 2040 Bird Monitor can be connected to a GPS network, and the 2040 Bird Monitor can be connected to a cellular network. Both options are a viable option for syncing the ARUs’ clocks but come at the cost of reduced battery life.

Due to clocks drifting we cannot also guarantee that the manual bird counts overlap the ARU recordings in their totality. Given the very similar number of species detected this was likely not a major factor in the results, but we’d recommend the method used by Van Wilgenburg, Sólymos, Kardynal, & Frey (2017) in future studies, where human observers say out loud when they begin and end a count so this information can be used later on to match manual to ARU counts.

Conclusion

Our results provide supporting evidence for the effectiveness of ARUs as a complement or replacement of field surveys performed by humans, particularly when evaluating bird species richness. Our results also revealed significant differences between the effectiveness of two ARU models available for sale in New Zealand. We therefore endorse the recommendation of Pérez-Granados et al. (2019) to test the performance of different ARUs before committing to using a specific model in long term conservation activities and studies. Based on our research we recommend the use of the Department of Conservation AR4 for studies of shorter duration where the main objective is estimating bird species richness.

Acknowledgements

I would like to thank the following people and organizations: my SCIE393 colleagues and staff for helping with the setup and removal of the bird monitors and for their company during bird counts; Myles MacKintosh for bringing all the extra necessary gear; 2040 for providing their bird monitors free of charge in exchange for setting them up; Lincoln University for providing the Department of Conservation AR4 ARUs; The Cacophony Project for their help troubleshooting a faulty 2040 Bird Monitor, in particular Menno Finlay-Smits and Cameron Ryan-Pears; the iNaturalist community for their help identifying unknown or dubious bird vocalizations, in particular Dave Holland, Finn Davey (skipperdogman), Lloyd Esler, britt_fb and hedgehog111; Tim Curran and Jon Sullivan for sane-checking my statistical analysis; Jon Sullivan for taking part in the bird counts as the expert observer; Julie Hardie for reviewing my report draft; Jonathan Hardie for his help automating many of the tedious processes, particularly downloading and extracting matching bird counts, reviewing my drafts and for moral support. To all, thank you.

Research was conducted under the Department of Conservation permit: 64058-RES to Tim Curran.

References

2040. (n.d.-a). *2040 Bird Monitor Setup Instructions* (pp. 1–35). pp. 1–35. Retrieved from https://cdn.shopify.com/s/files/1/0016/0622/1873/files/Bird_Monitor_Set_up_instructions_v4.pdf?176

2040. (n.d.-b). *Automatic Bird Monitor*. Retrieved from https://www.2040.co.nz/collections/cacophonometer-bird-monitoring/products/cacophonometer-wifi-bird-monitoring

Adobe. (2019). Adobe Audition. A professional audio workstation. Retrieved from https://www.adobe.com/nz/products/audition.html

Bombaci, S. P., & Pejchar, L. (2019). Using paired acoustic sampling to enhance population monitoring of New Zealand's forest birds. *New Zealand Journal of Ecology*, *43*(1). https://doi.org/10.20417/nzjecol.43.9

Cook, A., & Hartley, S. (2018). Efficient sampling of avian acoustic recordings: Intermittent subsamples improve estimates of single species prevalence and total species richness. *Avian Conservation and Ecology*, *13*(1). https://doi.org/10.5751/ACE-01221-130121

Darras, K., Batáry, P., Furnas, B., Celis-Murillo, A., Van Wilgenburg, S. L., Mulyani, Y. A., & Tscharntke, T. (2018). Comparing the sampling performance of sound recorders versus point counts in bird surveys: A meta-analysis. *Journal of Applied Ecology*, *55*(6), 2575–2586. https://doi.org/10.1111/1365-2664.13229

Department of Conservation. (n.d.). *AR4 Acoustic and Bat Recorder*. Retrieved from https://ftp.doc.govt.nz/public/folder/CpR1cRv_cE_rqb9ua5WRTg/electronics/Acoustic Recorders/AR4 Instructions _2.pdf

Department of Conservation. (2006). *The Poplars Pastoral Lease: Conservation Resources Report*. Retrieved from www.linz.govt.nz/system/files_force/media/crown-property-attachments/tenure-review/ig/the-poplars-con-res.pdf?download=1&download=1

Digby, A., Towsey, M., Bell, B. D., & Teal, P. D. (2013). A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*, *4*(7), 675–683. https://doi.org/10.1111/2041-210X.12060

Graham, B. J. (2019). *Boom and Bust: Rodent Abundance Correlates with Beech Tree Masts in the Lewis Pass*. Unpublished student research report, Lincoln University.

Greene, T., & Hartley, L. (2012). Incomplete counts: five-minute bird counts. *Department of Conservation*, pp. 1–22. Retrieved from https://www.doc.govt.nz/globalassets/documents/science-and-technical/inventory-monitoring/im-toolbox-birds-incomplete-five-min-counts.pdf

Hoete-Dodd, V. (2013). *Species richness and relative abundance indices of mammalian predators at Boyle River, North Canterbury, New Zealand*. Unpublished student research report, Lincoln University.

Leach, E. C., Burwell, C. J., Ashton, L. A., Jones, D. N., & Kitching, R. L. (2016). Comparison of point counts and automated acoustic monitoring: detecting birds in a rainforest biodiversity survey. *Emu*, *116*(3), 305–309. https://doi.org/10.1071/MU15097

Mortimer, J. A. J., & Greene, T. C. (2017). Investigating bird call identification uncertainty using data from processed audio recordings. *New Zealand Journal of Ecology*, *41*(1), 126–133. https://doi.org/10.20417/nzjecol.41.10

Mortimer, J. A. J., Greene, T. C., & Mortimer, S. P. (2019). Assessing bird vocalisation identification accuracy using a computer-based quiz. *New Zealand Journal of Zoology*, *46*(3), 201–224. https://doi.org/10.1080/03014223.2018.1536068

NIWA. (n.d.). The National Climate Database. Retrieved from https://cliflo.niwa.co.nz/

Pérez-Granados, C., Bota, G., Giralt, D., Albarracín, J., & Traba, J. (2019). Cost-Effectiveness Assessment of Five Audio Recording Systems for Wildlife Monitoring: Differences between Recording Distances and Singing Direction. *Ardeola*, *66*(2), 311. https://doi.org/10.13157/arla.66.2.2019.ra4

Ralph, C. J., Sauer, J. R., & Droege, S. (1995). Monitoring Bird Populations by Point Counts. In *General Technical Report PSW-GTR-149*. https://doi.org/https://doi.org/10.2737/PSW-GTR-149

Rempel, R. S., Francis, C. M., Robinson, J. N., Rempel, R. S., Francis, C. M., Robinson, J. N., & Campbell, M. (2013). Comparison of audio recording system performance for detecting and monitoring songbirds. *Journal of Field Ornithology*, *84*(1), 86–97. https://doi.org/10.1111/jofo.12008

Scott Brandes, T. (2008). Automated sound recording and analysis techniques for bird surveys and conservation. *Bird Conservation International*, *18*, S163–S173. https://doi.org/10.1017/S0959270908000415

Shonfield, J., & Bayne, E. M. (2017). Autonomous recording units in avian ecological research: current use and future applications. *Avian Conservation and Ecology*, *12*(1), 14. https://doi.org/10.5751/ace-00974-120114

Stewart, G. H., Rose, A. B., Veblen, T. T., Journal, S., Dec, N., Glenn, H., … Thomas, T. (1991). Forest Development in Canopy Gaps in Old-Growth Beech (Nothofagus) Forest, New Zealand. *Journal of Vegetation Science*, *2*(5), 679–690.

Tegeler, A. K., Morrison, M. L., & Szewczak, J. M. (2012). Using extended-duration audio recordings to survey avian species. *Wildlife Society Bulletin*, *36*(1), 21–29. https://doi.org/10.1002/wsb.112

Temple, S., & Wiens, J. a. (1989). Bird populations and environmental changes - can birds be bio-indicators? *American Birds*, *43*(2), 260–270.

The Cornell Lab of Ornithology. (2019). Software - Cornell Lab of Ornithology. Retrieved from http://ravensoundsoftware.com/software/

The R Foundation. (2020). The R Project for Statistical Computing. Retrieved May 21, 2020, from https://www.r-project.org/

Van Wilgenburg, S. L., Sólymos, P., Kardynal, K. J., & Frey, M. D. (2017). Paired sampling standardizes point count data from humans and acoustic recorders. *Avian Conservation and Ecology*, *12*(1). https://doi.org/10.5751/ACE-00975-120113

Venier, L. A., Holmes, S. B., Holborn, G. W., McIlwrick, K. A., & Brown, G. (2012). Evaluation of an automated recording device for monitoring forest birds. *Wildlife Society Bulletin*, *36*(1), 30–39. https://doi.org/10.1002/wsb.88

Wimmer, J., Towsey, M., Planitz, B., Williamson, I., & Roe, P. (2013). Analysing environmental acoustic data through collaboration and automation. *Future Generation Computer Systems*, *29*(2), 560–568. https://doi.org/10.1016/j.future.2012.03.004

Appendices

Appendix A: Sampling sites

Table 1 - Geographic information for the twenty sampling sites.

Site	Northing NZTM2000	Easting NZTM2000	Latitude WSG 84	Longitude WSG 84
L1T1	1549456	5292481	-42.51838813	172.38467564
L1T2	1549428	5292395	-42.51916076	172.38432717
L1T3	1549387	5292256	-42.52040982	172.38381574
L1T4	1549342	5292163	-42.52124437	172.38325968
L1T5	1549266	5291975	-42.52293238	172.38231777
L2T1	1549602	5292393	-42.51919013	172.38644518
L2T2	1549563	5292317	-42.51987200	172.38596371
L2T3	1549526	5292225	-42.52069807	172.38550515
L2T4	1549481	5292130	-42.52155064	172.38494893
L2T5	1549440	5292040	-42.52235844	172.38444183
L3T1	1549731	5292330	-42.51976587	172.38801002
L3T2	1549696	5292257	-42.52042099	172.38757752
L3T3	1549645	5292157	-42.52131820	172.38694784
L3T4	1549614	5292062	-42.52217169	172.38656207
L3T5	1549576	5291981	-42.52289864	172.38609231
L4T1	1549867	5292268	-42.52033304	172.38966019
L4T2	1549821	5292172	-42.52119457	172.38909177
L4T3	1549784	5292093	-42.52190359	172.38863439
L4T4	1549745	5291993	-42.52280159	172.38815081
L4T5	1549713	5291906	-42.52358297	172.38775357

Appendix B: ARU status and offset

Table 2 – Status of each ARU per sampling site and temporal offset of the Department of Conservation AR4 clock, in seconds, when compared to the 2040 Bird Monitor clock.

Site	2040 Bird Monitor ARU status	Department of Conservation AR4 ARU status	ARU clock offset
L1T1	OK	Missing last two days of recordings	34
L1T2	OK	N/A	
L1T3	OK	OK	-29
L1T4	OK	N/A	
L1T5	OK	OK	-34
L2T1	OK	OK	-65
L2T2	OK	N/A	
L2T3	OK	Missing first day of recordings	-85
L2T4	OK	N/A	
L2T5	OK	OK	-69
L3T1	OK	OK	11
L3T2	OK	N/A	
L3T3	OK	OK	-78
L3T4	OK	N/A	
L3T5	OK	OK	-82
L4T1	OK	OK	-59
L4T2	OK	N/A	
L4T3	Water damage to microphone	OK	Unknown
L4T4	OK	N/A	
L4T5	Lost most recordings during upload	Missing last three days of recordings	Unknown