

## A practical comparison of manual and autonomous methods for acoustic monitoring

Andrew Digby<sup>1\*</sup>, Michael Towsey<sup>2</sup>, Ben D. Bell<sup>1</sup> and Paul D. Teal<sup>3</sup>

<sup>1</sup>School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand; <sup>2</sup>Science and Engineering Faculty, Queensland University of Technology, Brisbane, QLD, Australia; and <sup>3</sup>School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand

### Summary

1. Autonomous acoustic recorders are widely available and can provide a highly efficient method of species monitoring, especially when coupled with software to automate data processing. However, the adoption of these techniques is restricted by a lack of direct comparisons with existing manual field surveys.

2. We assessed the performance of autonomous methods by comparing manual and automated examination of acoustic recordings with a field-listening survey, using commercially available autonomous recorders and custom call detection and classification software. We compared the detection capability, time requirements, areal coverage and weather condition bias of these three methods using an established call monitoring programme for a nocturnal bird, the little spotted kiwi (*Apteryx owenii*).

3. The autonomous recorder methods had very high precision (>98%) and required <3% of the time needed for the field survey. They were less sensitive, with visual spectrogram inspection recovering 80% of the total calls detected and automated call detection 40%, although this recall increased with signal strength. The areal coverage of the spectrogram inspection and automatic detection methods were 85% and 42% of the field survey. The methods using autonomous recorders were more adversely affected by wind and did not show a positive association between ground moisture and call rates that was apparent from the field counts. However, all methods produced the same results for the most important conservation information from the survey: the annual change in calling activity.

4. Autonomous monitoring techniques incur different biases to manual surveys and so can yield different ecological conclusions if sampling is not adjusted accordingly. Nevertheless, the sensitivity, robustness and high accuracy of automated acoustic methods demonstrate that they offer a suitable and extremely efficient alternative to field observer point counts for species monitoring.

**Key-words:** acoustic monitoring, automated animal call recognition, bioacoustics, census, research techniques

### Introduction

#### CALL COUNTS

Point call counts with field observers are one of the most common methods of measuring abundance or presence or absence of birds (Rosenstock *et al.* 2002). Such counts are prone to several well-known biases. Observers vary greatly in their ability to detect and classify calls due to differences in age, experience and hearing (Emlen & Dejong 1992; Rosenstock *et al.* 2002; Hutto & Stutzman 2009), and errors become more prevalent with high call rates (Hutto & Stutzman 2009). The presence of observers can also affect vocal activity through disturbance of the study species. Counts usually take place during fine weather to maximise detections and minimise variability, but can still be biased by weather conditions (Bas *et al.* 2008).

Since counts are usually short, they are prone to temporal biases arising from daily or seasonal activity variation (Diefenbach *et al.* 2007; Bas *et al.* 2008). This can lead to behavioural information being missed or population sizes significantly underestimated (Bridges & Dorcas 2000). In addition to these biases, field counts require substantial observer effort, which can be expensive, particularly when trained observers are required to monitor in remote areas.

#### AUTONOMOUS RECORDERS

Autonomous acoustic recorders offer an alternative not subject to many of the biases of field surveys. By not requiring the presence of an observer, they can be less expensive (Charif 2008), can be used in inaccessible or inhospitable habitats (Hutto & Stutzman 2009), avoid bias from subject disturbance (Alldredge *et al.* 2007) and minimise temporal bias through extended sampling. Acoustic recorders can provide reliable data more rapidly than human-based survey techniques

\*Correspondence author. E-mail: andrew.digby@vuw.ac.nz

(Parker 1991; Riede 1993, 1998) and can perform better at sampling bird communities than skilled observers using audio-visual field point counts (Celis-Murillo *et al.* 2009; Kirschel *et al.* 2011). Recorders provide a permanent data record for re-analysis by independent observers (Swiston & Mennill 2009) and combined in microphone arrays (Mennill *et al.* 2012) can yield spatial and behavioural information that is not available to an observer in the field (Fitzsimmons *et al.* 2008; Mennill & Vehrencamp 2008; Kirschel *et al.* 2011).

A principal obstacle to the widespread use of autonomous acoustic recorders is that they can require substantial expert time to process the recordings by visually inspecting spectrograms. Although this can be substantially less time than required for field point counts (Charif 2008), autonomous surveys sample over much longer time periods and so often yield very large data volumes.

Automatic species detection and classification software provide an efficient method of processing such large data sets. A variety of methods are available (Brandes 2008), with some accessible through dedicated software packages (Charif 2008). These methods present two main issues. One is that classification methods such as hidden Markov Models and neural networks rely on large training data sets, which can be restrictive for species with low call rates (Acevedo *et al.* 2009; Towsey *et al.* 2012). The second is the difficulty in minimising the number of false-negative and false-positive errors (Waddle *et al.* 2009; Bardeli *et al.* 2010).

A further barrier preventing wide-scale adoption of autonomous acoustic methods for conservation management is that recorders usually have a lower sensitivity than a human listener (Hutto & Stutzman 2009). This can be remedied by modified sampling, but necessitates that their efficacy be assessed against existing monitoring methods. This has not been undertaken for most species and is rarely performed in the context of established monitoring programmes.

#### KIWI MONITORING

Of particular benefit for cryptic vocal species, point call counts are an established tool for monitoring kiwi (Apterygidae). Call surveys have been used in kiwi conservation management to detect previously unknown remnant populations at risk of extinction (Jolly & Colbourne 1991) and to measure response to predator control or advocacy (Robertson *et al.* 1993; Pierce & Westbrooke 2003; Robertson & Colbourne 2004). They are widely used to assess kiwi distribution and population trends throughout New Zealand (Robertson *et al.* 2003, 2005; Colbourne 2006). The vast majority of surveys use field counts, with observers listening for 1–2 h on up to eight nights per year at designated sites (Robertson *et al.* 2003). As a result, these surveys are prone to temporal sampling bias, with population estimates dependent on the density of listening stations, the length of counts and the number of consecutive nights on which surveys take place (Colbourne & Kleinpaste 1984; Miles *et al.* 1997). Call rates can also vary with weather conditions (Pierce & Westbrooke 2003) and moon phase (Miles *et al.* 1997). Autonomous recorders have recently been used for kiwi

surveys (S. Cockburn, pers. comm.), but no published results exist of evaluation against existing count methods.

#### THIS STUDY

In this article, we compare manual and automated examination of autonomous acoustic recordings with a field observer survey, using an established kiwi call count programme as a test case. We assess method sensitivity and reliability from a conservation monitoring perspective and measure variation in performance due to different ambient conditions. By application to a real monitoring situation, these results provide an applied and quantitative assessment of the efficacy of automated methods for conservation monitoring of vocal species. This article also provides the first report on the automated recognition of kiwi calls.

### Materials and methods

#### CALL COUNTS

Since 2002, kiwi call counts have been conducted annually between January and March at Zealandia, a 225-ha mammalian predator-free fenced wildlife reserve in Wellington, New Zealand. In 2000–2001, 40 little spotted kiwi (*Apteryx owenii*; LSK) were translocated to the sanctuary from nearby Kapiti Island, and a 2010 census estimated the population size at *c.* 100 (H. Robertson, pers. comm.). Counts follow the protocol described in the national Kiwi Call Count Scheme (Robertson *et al.* 2003). From two sites in the reserve, two or three observers record the sex, estimated distance and magnetic bearing of all kiwi calls heard for an hour from 30 min after sunset. Weather conditions are recorded (Table 1), and surveys only take place on nights with low or moderate background noise. Established procedure at Zealandia required that the observers were varied as much as possible between counts.

#### CALL RECORDING

On 53 nights from 17 January–16 March 2011 and 18 January–30 March 2012 an autonomous recorder (Song Meter 2; Wildlife Acoustics Inc., Concord, MA, USA) was positioned near a kiwi count site in Zealandia. This site was on top of a 10-m high viewing platform above the tree canopy, on a spur in the central part of the reserve (S41.29776 E174.74432). Due to security considerations and the likely impact of wind noise on the microphones, the recorder was placed next to the platform at 2 m above ground level. The recorder was therefore approximately 10 m below and 5 m away horizontally from the

**Table 1.** Environmental variables recorded by the observers during each call count

Variable	Values
Temperature	Cold, Mild, Warm
Rain	Nil, Light, Moderate
Wind	Calm, Light, Moderate, Strong
Cloud cover	Clear, Partly cloudy, Overcast
Light	Light, Dark, Black
Noise	None, Slight, Moderate
Ground condition	Dry, Damp, Wet

observers. To assess whether this distance affected the likelihood of detecting kiwi calls, for 27 nights in 2012, a microphone connected to one channel was positioned in the tree canopy 2 m horizontally from the platform edge. By comparing calls detected by each channel, the suitability of the ground-level listening position for comparison with the call counts could be assessed.

Recordings were conducted for 70 min each night, starting 5 min before and ending 5 min after each count. Recordings were made in stereo (microphone frequency response = 20–20 000 Hz) and were digitised at 16 kHz, 16-bit precision. A hardware gain of +48 dB was applied to each channel.

## SPECTROGRAPHIC ANALYSIS

The recordings from the count were analysed spectrographically by an experienced observer (AD). Spectrograms were viewed in RAVEN PRO 1.4 (Charif *et al.* 2010) with 512-sample Hann window, 31.3 Hz spectral resolution. Files were viewed in 15 min segments, and to improve detectability of LSK calls against background noise, spectrograms were smoothed with ten-sample averaging. Each LSK call was marked and annotated with its sex and a subjective signal-to-noise quality score, ranging from 1 (very high signal-to-noise) to 10 (call barely visible or audible). Detection was primarily conducted from visual inspection of the spectrograms, but aural confirmation was used for very faint calls.

## AUTOMATED CALL DETECTION

Automated detection of LSK calls was performed using custom software written in C#. Although a commercial detector (capable of recognising multiple species) might have been employed, customised recognisers can offer better performance for autonomous monitoring in acoustically unconstrained environments (Bardeli *et al.* 2010; Towsey *et al.* 2012). We approached the problem of LSK call detection in three steps: (i) identification of putative LSK calls; (ii) feature extraction from putative calls; and (iii) classification of the putative calls using a decision tree.

## PUTATIVE CALL DETECTION

LSK calls consist of a long series of repeated rising syllables or elements that persist for 20–40 s (Digby *et al.* 2013). The sexes call in largely nonoverlapping frequency ranges (Digby *et al.* 2013); hence, it is possible to search for male and female calls in separate bands (2.2–3.2 and 1.3–2.2 kHz, respectively) using the same algorithm.

Due to the comparatively long duration of an LSK call, the signals were down-sampled to 17 640 samples per second and spectrograms were prepared from nonoverlapping windows (frame size) of 2048 samples. The spectral values were converted to decibels followed by noise reduction as described in Towsey *et al.* (2012). The frequency bands of male and female calls were processed separately. The average decibel value was calculated for each window within the appropriate frequency band. Putative LSK calls were identified by searching for a pattern of repeated energy peaks (corresponding to LSK call syllables) within the acoustic energy profile of the frequency band. An autocorrelation coefficient was calculated for lengths of 32 frames (3.715 s duration) in 1 s steps. A portion of the spectrogram was marked as a putative call if it met the following conditions: the acoustic periodicity was in the range 0.3–1.8 s; the autocorrelation coefficient exceeded a threshold of 0.2; the autocorrelation exceeded the threshold for not <7 s and not >70 s.

Note that the autocorrelation technique did not work well unless preceded by noise reduction in the spectrogram. To avoid duplicate detections of calls containing pauses, putative calls were merged if they were within 5 s of another event in the same frequency band. True LSK calls that were undetected in this first step were treated as false negatives.

## FEATURE EXTRACTION

Several other bird species have calls with a similar repetition of elements in the same frequency band. We therefore extracted a total of five features from each putative call event: (i) the change in syllable periodicity over duration of call; (ii) the degree of frequency modulation within each syllable; (iii) the consistency of amplitude across the syllables; (iv) correlated acoustic energy outside the bandwidth; and (v) a weighted combination of the four feature scores.

**1. Period score.** The syllable rate of true LSK calls declines (period increases) through the duration of a call. A periodicity score was assigned to each putative call depending on the duration (in seconds) for which the period continuously increased. The score was normalised in [0,1] by setting 0 and 20 s as minimum and maximum durations, respectively.

**2. Chirp score.** The syllables of a true LSK call consist of a rapidly rising chirp, that is, the frequency bin having maximum energy increases over consecutive frames. For each chirp, the frame ( $n$ ) having maximum energy was located, and the frequency bin in frames  $n - 1$  and  $n + 1$  having maximum energy was determined. The frequency difference (in Hz) between the two was the chirp score:  $\text{Chirp Score} = F_{\max}(n + 1) - F_{\max}(n - 1)$ . The score was normalised in [0,1] by setting 0 and 100 Hz as minimum and maximum chirp scores, respectively.

**3. Consistency score.** The syllables of a true LSK call have consistent amplitude from beginning to end. A consistency score was calculated from the frames in each chirp having maximum energy. The consistency score equalled the average absolute difference in energy from one chirp maximum to the next divided by the energy of the maximum chirp. The latter step normalises the consistency score.

**4. Bandwidth score.** The syllables of a true LSK call lie within a fixed bandwidth. However, there were other birds (such as kaka, *Nestor meridionalis*, an endemic parrot) in the same recordings whose calls consist of syllables repeated at the same rate, but whose bandwidth is greater and crosses that of the LSK calls. To resolve these calls, we calculated a bandwidth score which declines as correlated acoustic energy spills outside the LSK frequency band. A buffer of 150 Hz was excluded above and below the LSK frequency band to allow for uncertainty in its extremes. The correlation coefficients for acoustic activity above and below the LSK frequency band were added and normalised in [0,1].

**5. Combined score.** Finally, we calculated a fifth feature as the weighted combination of the above normalised feature scores. Weights for the features were  $w_{\text{Period score}} = 0.25$ ,  $w_{\text{Chirp score}} = 0.25$ ,  $w_{\text{Consistency score}} = 0.20$ ,  $w_{\text{Bandwidth score}} = 0.3$ .

## CALL CLASSIFICATION

Following feature extraction, putative LSK calls were classified by a C5.0 decision tree classifier (version 2.09; RuleQuest Research, St Ives, NSW, Australia). The decision tree was trained using 22 three-hour recordings with low background noise from the study site, randomly selected over a range of dates. LSK calls were identified in these recordings by spectrographic inspection. Calls with the lowest quality score or which overlapped other biotic or abiotic sources were removed from the training data on the principle that a classifier is better trained on unambiguous examples.



The training data contained 459 true LSK calls (368 male, 91 female). The call detector identified 3411 putative calls, of which 338 were true kiwi calls and 3073 were not. This training set was input to C5.0, with each instance consisting of a feature vector having five values and its true classification (kiwi or not kiwi). Different methods of building the decision tree were evaluated using 10 repeats of 10-fold cross-validation and compared using their recall and precision. Recall is defined as  $TP/(TP + FN)$  and precision as  $TP/(TP + FP)$ , where TP = true positives, FN = false negatives and FP = false positives. The final classifier used ten decision trees combined by adaptive boosting and no softening (C5.0 files provided in online supplement). It had a mean error rate of 0.9% and a cross-validated recall and precision of 93.4% and 97.1%, respectively. This recall value on training data was higher than would be expected in operational conditions due to overlearning (the phenomenon whereby a classifier learns feature combinations that are specific to the training data but which are not relevant in the operational environment). In addition, the training data contained only unambiguous calls (quality score of at least nine) to assist learning. This decision tree was used to classify 'unseen' putative calls from the autonomous recordings made during the field survey.

#### COMPARING COUNTS

The calls recorded by the call count observers (field monitoring) were matched with those identified from spectrogram inspection and from the automatic detection and classification software. For each method, detections were marked either as true positives, false positives or false negatives (Fig. 1). For LSK calls which were detected by spectrogram inspection but missed by the automatic method, the likely reason for the false negative was noted. The field counts were taken to be the truth, except for sexing calls, and in cases when a call was recorded as close or medium distance (<250 m) by the observers, no LSK call and a call from another species was apparent in the spectrogram at the time. Species at the study site that can be confused acoustically with LSK are ruru (*Ninox novaeseelandiae*), a native owl and kaka. In these cases, and when the sex was recorded incorrectly, the field detection was

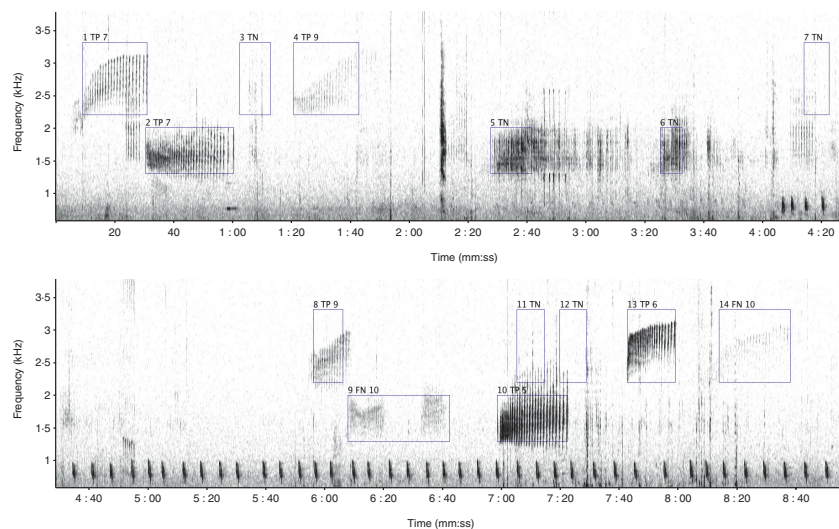
marked as a false positive. This approach unavoidably underestimates the number of false positives and overestimates the true positives for the field counts, since distant calls not visible on spectrograms could not be verified.

#### RECORDER DETECTION LIMITS

The distance limits at which the LSK calls could be detected from the autonomous recordings were determined by broadcasting kiwi calls at realistic amplitude at varying distances from the recorder. The appropriate speaker volume was determined by playing a LSK call recorded at known close range (<5 m) at different volume settings and recording it from the same distance with the same equipment as used to make the initial recording (SD722 recorder, Sound Devices, Reedsburg, WI, USA with Telinga Twin Science microphone and Telinga 53 cm foldable parabola). The appropriate volume setting was that for which the rerecorded power spectra most closely matched that of the original call over the peak energy range. With appropriate volume settings determined for each sex, male and female LSK calls were played at 100 m intervals at distances ranging from 100 to 1200 m from the Song Meter 2 recorder. Calls were broadcast using a handheld recorder (model PCM-M10P, Sony, Tokyo, Japan) and 45-W loudspeaker (model MA-101, MIPRO, Chiayi, Taiwan). Tests were conducted at night in low background noise conditions. The distance limits were determined by manually and automatically scanning the spectrograms from the autonomous recorder and ascertaining the distance threshold to which the known calls were detected.

#### COUNT VARIATION WITH AMBIENT CONDITIONS

Generalised linear models were used to assess variation in detection with ambient conditions, allowing for an annual trend. A different model was calculated for each detection method (field monitoring, spectrogram inspection and automatic call detection). The response variable was the number of calls detected during each count, and explanatory variables were year (2011 or 2012) and environmental



**Fig. 1.** A section of a spectrogram from the autonomous recorder, showing detections from the automatic software. The spectrogram is split into two contiguous lines. Annotations above each detection give an event number, a verification of the detection (TP = true positive, FN = false negative, TN = true negative) and the call quality (1–10) for kiwi calls. Events 1, 4, 8, 13 and 14 are male LSK calls, and events 2, 9 and 10 are female LSK calls. The faintest kiwi calls (9 and 14) are not detected by the automatic software, but were by the other methods. Events 3, 5 and 6 are duck calls. The repeated signals in the lower line at c. 0.8 kHz are from an owl (*Ninox novaeseelandiae*).

conditions recorded by the observers, except wind direction (Table 1). Poisson models with a log link were fitted using function GLM in R (version 2.15; R Development Core Team 2011). Observations from 52 nights were used; one count that was abandoned due to poor weather was removed. Full models were checked for overdispersion and adequacy (Zuur *et al.* 2009). Model selection followed an informatic-theoretic approach (Anderson *et al.* 2000), with models fitted for all possible combinations of explanatory variables without interactions. These were ranked by corrected Akaike Information Criteria (AICc; Hurvich & Tsai 1989), and a 95% confidence model set was selected based on cumulative Akaike weight. Inferences for the three detection methods were then based on the weighted average model and the relative variable importance calculated from this set.

## Results

### TIME EXPENDITURE

Spectrogram inspection took 2–5 min per hour of recording, depending on the call rate. The detection and classification software took approximately 2 min per recording hour on a four-core 2.4 GHz computer. Replacement of recorder memory cards and batteries was required about every 6 weeks. Including these resupply times, the total time required for the methods using autonomous recorders was less than one person-hour per week. The field counts required at least 28 person-hours per week, with two people spending 1 h listening and 1 h walking to and from the site for each night. This is a conservative estimate and does not include additional time in processing the hand-written count results and organising observers.

### COMPARISON BETWEEN GROUND-LEVEL AND CANOPY SITES

A comparison of the ground-level and canopy recording sites using spectrogram inspection yielded 348 calls (291 M, 57 F) from the canopy microphone and 345 (284 M, 61 F) from the microphone near ground level. The median difference in the number of calls detected at each site for each count was not significantly different from zero (Wilcoxon paired signed-rank test,  $V = 94.5$ ,  $P = 0.702$ ). The ground level microphone site was therefore comparable with the observer listening position, and so recordings from this site were used for spectrogram inspection.

### DETECTION DISTANCE LIMIT

The simulated kiwi calls of both sexes were reliably detected from spectrogram inspection to at least 400 m. Some calls were discerned up to 600 m, but beyond 400 m were very faint on the spectrograms and not evident for all repeats at that distance. The detection software consistently located calls of both sexes to 300 m and some up to 400 m.

### DETECTION PERFORMANCE

A total of 900 calls (721 male, 171 female, 8 sex unknown) were detected by all three methods over 52.2 h on 53 evenings.

Recall of this total ranged from 40% for the automatic detection to 94% for the field survey (Table 2). Manual and automatic spectrogram analysis detected 85% and 42%, respectively, of the calls heard by the human listeners. The autonomous recording methods detected proportionally more female calls: spectrogram inspection recovered 93% of female calls and 84% of male calls heard by the listeners, and the detection software found 47% of the female calls and 42% of male calls from the field survey. The observers missed 6% of the total calls and misclassified 4% of the calls they detected. The automatic detector found one very faint call which was missed by both the other methods; otherwise, false negatives from spectrogram inspection were also false negatives for the automatic method.

The detection distance limit of at least 400 m for spectrogram inspection equates to a minimum areal coverage of *c.* 50 ha. Assuming a uniform density of calling kiwi, the recall values indicate that the field and automatic surveys therefore sampled calls from areas of *c.* 59 and 25 ha, respectively. This latter figure agrees with the areal coverage of 28 ha calculated from the 300 m detection limit for automatic detection.

As expected, calls not detected from spectrogram inspection were those at further distances: 85% of those missed were marked as >250 m by the observers. All of these false negatives were due to low signal strength, either not visible in the spectrograms or audible in the recordings or only detected upon very detailed re-inspection. Of the 177 false negatives, 44 (25%) were detected in the spectrograms upon closer examination.

The performance of the automatic detector improved markedly with increasing call quality, with a recall of more than 80% for calls with a quality score up to seven (Fig. 2). Over half (54%) of the automatic method false negatives that were visible in the spectrograms were attributed to low signal strength. Other common reasons for calls being missed by the automatic detector were signal degradation effects due to acoustic interference from other species (23% of false negatives), wind gusts (9%), reverberation (5%) or overlapping LSK calls (5%). A small number of false negatives were also caused by LSK calls that were very short (<7 s; 1.6% of false negatives) or contained gaps or unusual spectral structure (1%).

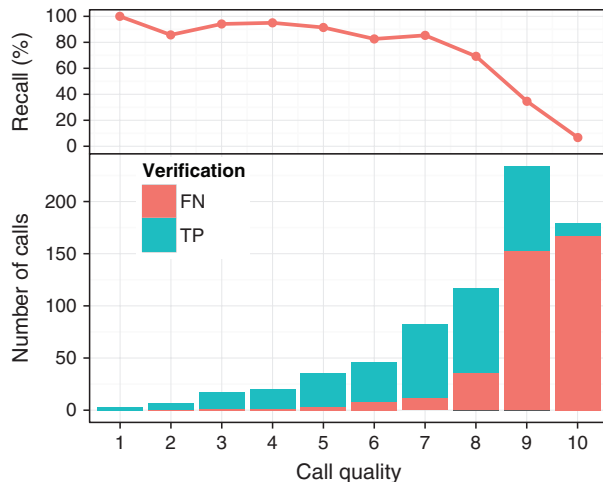
### VARIATION WITH YEAR AND AMBIENT CONDITIONS

The inferences from the average generalised linear models differed between detection methods. Ground moisture, cloud cover and background noise significantly affected calls detected by the field observers, whereas only wind strength affected spectrogram inspection, and noise and wind strength influenced automatic detection (Fig. 3 and Table S1). However, the sign of the effects on calls detected was the same for all models, with counts reduced with increasing wind strength, noise and cloud cover, and dry ground. The most important variables influencing detection were ground conditions for field counts and wind strength for the autonomous recording methods (Fig. 4).

**Table 2.** Detection scores of the three methods, with true positives (TP), false positives (FP), false negatives (FN), recall (R) and precision (P), shown for the total of 900 calls

Method	TP	FP	FN	R (%)	P (%)
Field survey	849	36	51	94.3	95.9
Spectrogram inspection	723	0	177	80.3	100.0
Automatic detection	358	7	542	39.8	98.1

Note that the field survey false positives are a lower estimate, since for distant calls accuracy could not be verified from spectrograms.



**Fig. 2.** The recall and number of training calls that were either detected (TP) or not detected (FN) by the automatic software, as a function of subjective spectrographic call quality (10 is low quality). Bars are stacked, so total bar height shows the number of calls for each quality score. The majority of calls not detected are those in the two lowest quality bins (9 and 10).

## Discussion

As expected, kiwi call counts conducted by observers in the field were more sensitive than those from autonomous recorders. Misclassification and false-negative rates for the field counts were low, confirming that surveys using trained volunteers, many with no previous experience of bird counts, can yield accurate results. While autonomous recorders were less sensitive, their areal coverage was comparable with field surveys and adequate for many applications. For example, assuming an LSK territory size of 3 ha for a dense population (Colbourne & Robertson 1997; Holzapfel *et al.* 2008), spectrogram inspection was able to monitor approximately 16 territories, compared with 19 for the field count. The fully automatic survey sampled about half the number of territories, as a result of its inability to detect the fainter, lower quality calls. Assuming a uniform density of calling kiwi, the number of calls will increase with the square of distance, but the amplitude of those calls will decrease by the same ratio. A sensitivity-limited sample will therefore contain a high proportion of lower quality calls, and the area sampled will decrease strongly with sensitivity. This skew to lower quality calls is evident in Fig. 2, which also shows that the automatic method had a constantly high

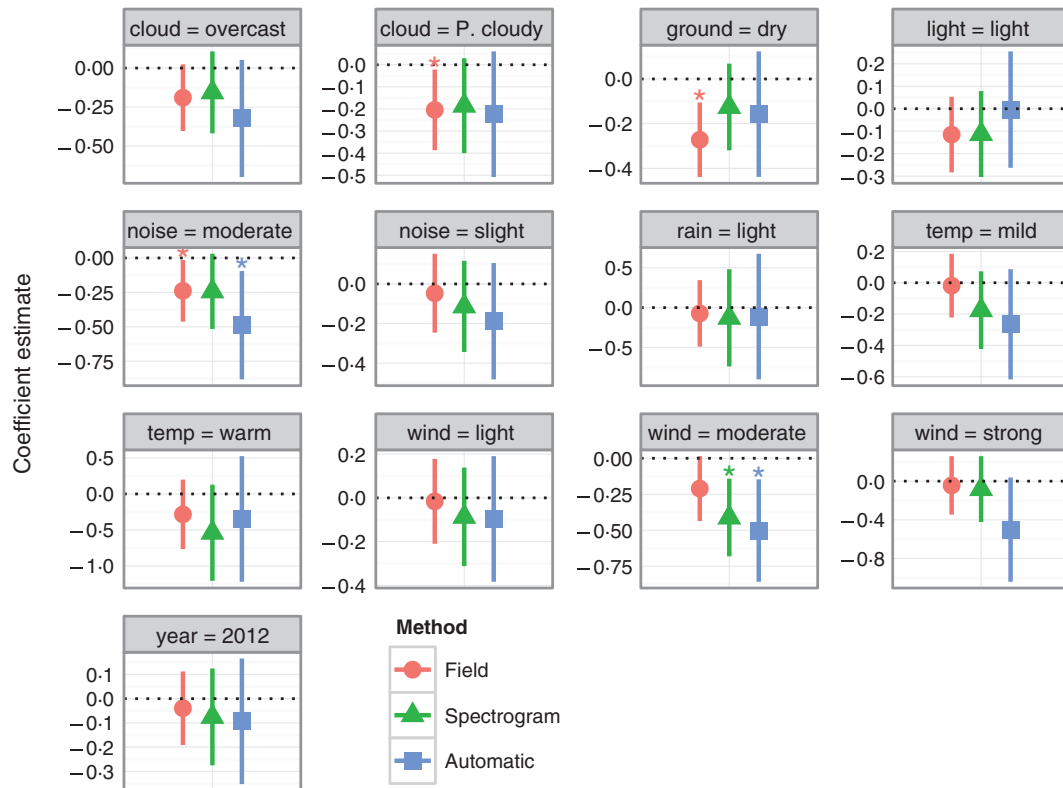
recall (>80%) up to a quality of seven, but after that recall declined sharply. This is because these lower quality calls would mostly be beyond the 300 m reliable detection limit of the automatic method. A larger area could therefore be surveyed simply using more recorders, spaced according to this distance limit. Although the automatic detection recall was relatively low, the very few false positives suggest that it was unaffected by noise conditions, which often cause problems for automatic signal detection and classification (Bardeli *et al.* 2010).

It should be noted that the performance of the field survey is likely overestimated in this study, since distant detections could not be confirmed by the recorders. Conversely, the recall from spectrogram inspection was a lower limit, since very faint calls that were missed during spectrogram scanning were seen upon re-inspection and would likely have been detected with more careful examination. However, these caveats would make at most only minor adjustments to the performance of each method and do not alter the overall conclusions.

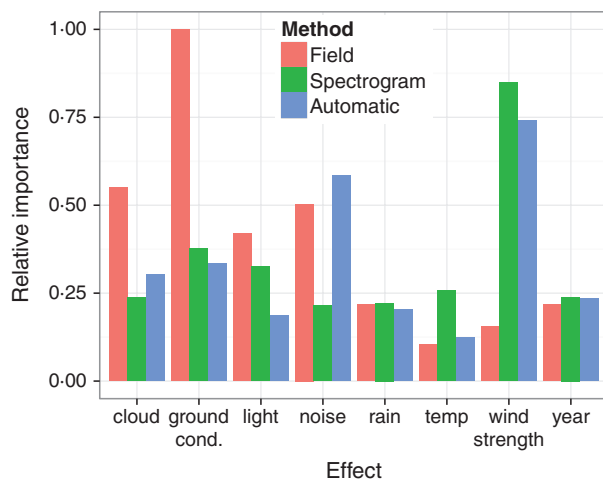
Automatic detection software is ideally suited to species such as LSK that have low call variability. For animals with a wider repertoire, such as song birds, or for detection of multiple species, a larger training set is necessary and detector performance is likely to be reduced (Bardeli *et al.* 2010). A further restriction is that the detection software in this study was often unable to resolve calls that overlapped with other sources, either conspecifics or other species. This would limit its usefulness in measuring abundance in areas of high population density or where there are high numbers of acoustically competing species. However, field counts are also prone to such bias (Hutto & Stutzman 2009), which is caused by either the insufficiency of human auditory ability in temporally or spectrally resolving simultaneous calls or the requirement that observers record call details as they occur. The latter also affects the quality of observations made, since call parameters were occasionally omitted or inaccurate during periods of high calling activity. It is significant that spectrogram inspection tended to perform better than the field observers during high call rates.

The principal advantage of using autonomous recorders is the gain in efficiency, with the spectrogram inspection around 30 times more efficient than traditional counts. These substantial gains in productivity offered by recorders could be utilised by sampling over longer periods each night and at other times of year to reduce temporal biases. The only significant time input required for the fully automated method is in resupplying recorders and data management, and a nightly single point survey could be achieved with a time expenditure of only about an hour per month. For remote areas or in determining species presence or absence, this method is extremely attractive.

The automatic recorder surveys also offer a major benefit in reduction in temporal and observer bias that can affect field surveys. The automatic detection method is fully repeatable as it requires no subjective input at all and so is well suited for assessing variation in counts between sites or time periods. Minimal disturbance of the target species and a permanent data record are further benefits of the autonomous methods.



**Fig. 3.** Effect coefficient estimates and 95% confidence intervals for the average models determining the influence of year and environmental conditions on counts from each detection method. Effects which significantly impact counts are those for which the coefficient confidence range does not intersect zero and are marked by an asterisk.



**Fig. 4.** The relative importance of the year and environmental variables in the average model for each method (field survey, spectrogram inspection and automatic detection).

A disadvantage of autonomous recorders is that each is restricted to a single fixed location. In contrast, human observers can move around the survey area to make further observations if necessary. Field-based observers can also discern between signals from different directions to improve abundance estimates or determine territories, which is not possible with a single-channel recorder. The simple use of autonomous

recorders in this experiment is therefore limited compared with field counts when estimating numbers of individuals or territories. However, stereo recorders can yield directional information from the difference in signal arrival time at each microphone (Benesty *et al.* 2008). With spatially separated microphones, species densities can be estimated from relative signal intensities (Dawson & Efford 2009), and synchronised microphone arrays allow accurate determination of caller position (Collier *et al.* 2010; Mennill *et al.* 2012). Acoustic recorders therefore offer the potential to provide population and behavioural information beyond that available to human listeners (Fitzsimmons *et al.* 2008; Mennill & Vehrencamp 2008; Kirschel *et al.* 2011).

The generalised linear models suggested that wind noise affected the autonomous detections more than the field survey. This may have been due to the observers being slightly above the canopy and so surrounded by fewer scattering surfaces and sources of noise in windy conditions than the microphones, which were below canopy level. Also, unlike the static microphones, the observers were able to move to minimise the effect of wind noise. This reflects an inherent restriction of acoustic recorders, which are adversely affected by wind noise on microphones and should be placed in a sheltered position.

The significant impact of ground conditions on the field counts, but not on the autonomous counts, is a more serious concern. Pierce & Westbrooke (2003) reported an increase in call rate with increasing ground moisture index for brown kiwi



(*Apteryx mantelli*), suggesting that this is a real effect. The inability of the autonomous methods to detect a significant influence may have been a result of their reduced sensitivity. This led to fewer counts and wider effect confidence intervals and also reduced areal sampling that may have affected the ground moisture dependency if that effect was nonuniform throughout the study site. Additionally, higher call rates under damp or wet ground conditions may have reduced the recall of the fully automated method. However, the ground moisture coefficient estimates for the autonomous methods are similar to the field counts, albeit nonsignificant. A longer sampling period is necessary to sufficiently reduce confidence intervals for the autonomous methods. This would be easily performed given their efficiency, and it should be noted that the short recording period of 1 h per night is artificially low, used only to provide direct comparison with the field survey.

The year effect would of main interest in most conservation applications and shows no difference between the three methods. It is slightly surprising that no call increase is apparent between years, given that the LSK population in Zealandia is growing at c. 10% per annum (H. Robertson, pers. comm.). However, counts taken over many years would be required to measure such a trend, and call rates do not always follow population changes (Robertson *et al.* 2003).

## RECOMMENDATIONS

This study has evaluated the use of autonomous recorders for point call counts, applied to a typical, volunteer-run monitoring programme, rather than a one-off research experiment. These results demonstrate that while trained lay observers provide accurate surveys, autonomous recorders offer a viable alternative. Spectrogram inspection can yield comparable coverage with field observers for considerably less effort, and fully automated methods offer robust alternatives that are well suited to longer-term projects and those in remote or unmonitored areas. Our modelling results show that autonomous methods can provide similar ecological inferences to field counts, but that care must be taken to ensure sufficient temporal sampling to adjust for the different biases these are subject to.

The detection ability of our autonomous recording methods depended upon signal strength, with recall dropping substantially for the faintest calls. We demonstrated this using a subjective measure of call spectrogram quality, but wider adoption of autonomous techniques would be aided by measuring performance against a quantitative signal-to-noise measure. A plot of recall against signal-to-noise, or ideally the distance to which a species can be reliably detected, would provide a useful definition of the detection limits of a survey method that would guide the spacing of recorders.

Field call counts can offer significant benefits for conservation advocacy and community engagement. We recommend that autonomous recorders do not replace these, but are utilised to increase the spatial and temporal coverage of existing call count regimes. Acoustic recorders should also be used during call counts to verify the accuracy of obser-

vations (Hutto & Stutzman 2009), particularly during periods of high call rates. Autonomous recorders would also be suitable for areas where existing surveys are not established and are particularly appropriate for determining presence or absence of a species in remote areas where field call counts are costly and inefficient.

## Acknowledgements

The authors appreciate very helpful comments from Dr Alex Kirschel and another anonymous reviewer. We are very grateful for the support and assistance of the staff and volunteers at Zealandia, particularly Raewyn Empson and Erin Daldry, who organised the kiwi call counts and shared survey data. The effort of the Zealandia volunteers who contributed to this study through their part in the kiwi counts is greatly appreciated. Andrew Digby was financially supported by a Victoria University of Wellington (VUW) Doctoral Assistantship and Submission Scholarship and by funding from the VUW Centre for Biodiversity and Restoration Ecology, the VUW School of Engineering and Computer Science and the Bank of New Zealand Save the Kiwi Trust.

## Conflict of interest

Other than having purchased their equipment for research purposes, the authors have no relationship with the company Wildlife Acoustics Ltd.

## References

- Acevedo, M.A., Corrada-Bravo, C.J., Corrada-Bravo, H., Villanueva-Rivera, L.J. & Aide, T.M. (2009) Automated classification of bird and amphibian calls using machine learning: a comparison of methods. *Ecological Informatics*, **4**, 206–214.
- Allredge, M.W., Pollock, K., Simons, T., Collazo, J., Shriner, S. & Johnson, D. (2007) Time-of-detection method for estimating abundance from point-count surveys. *The Auk*, **124**, 653–664.
- Anderson, D.R., Burnham, K.P. & Thompson, W.L. (2000) Null hypothesis testing: problems, prevalence, and an alternative. *The Journal of Wildlife Management*, **64**, 912–923.
- Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K.H. & Frommolt, K.H. (2010) Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters*, **31**, 1524–1534.
- Bas, Y., Devictor, V., Moussus, J. & Jiguet, F. (2008) Accounting for weather and time-of-day parameters when analysing count data from monitoring programs. *Biodiversity and Conservation*, **17**, 3403–3416.
- Benesty, J., Chen, J. & Huang, Y. (2008) *Microphone Array Signal Processing*, 1st edn. Springer-Verlag, Berlin.
- Brandes, S.T. (2008) Automated sound recording and analysis techniques for bird surveys and conservation. *Bird Conservation International*, **18**, S163–S173.
- Bridges, A. & Dorcas, M. (2000) Temporal variation in anuran calling behavior: implications for surveys and monitoring programs. *Copeia*, **2000**, 587–592.
- Celis-Murillo, A., Deppe, J.L. & Allen, M.F. (2009) Using soundscape recordings to estimate bird species abundance, richness, and composition. *Journal of Field Ornithology*, **80**, 64–78.
- Charif, R. (2008) *Automated detection of Cerulean Warbler songs using XBAT data template detector software*. Technical report, The Cornell Lab of Ornithology, Ithaca, New York, USA.
- Charif, R., Strickmann, L. & Waack, A. (2010) *Raven Pro 1.4 User's Manual*. The Cornell Lab of Ornithology, Ithaca, New York, USA.
- Colbourne, R.M. (2006) *Kiwi call scheme*. Technical report, Department of Conservation, Wellington, New Zealand.
- Colbourne, R.M. & Kleinpaste, R. (1984) North Island brown kiwi vocalisations and their use in censusing populations. *Notornis*, **31**, 191–201.
- Colbourne, R.M. & Robertson, H. (1997) Successful translocations of little spotted kiwi (*Apteryx owenii*) between offshore islands of New Zealand. *Notornis*, **44**, 253–258.
- Collier, T.C., Kirschel, A.N.G. & Taylor, C.E. (2010) Acoustic localization of antbirds in a Mexican rainforest using a wireless sensor network. *The Journal of the Acoustical Society of America*, **128**, 182.
- Dawson, D.K. & Eford, M.G. (2009) Bird population density estimated from acoustic signals. *Journal of Applied Ecology*, **46**, 1201–1209.
- Diefenbach, D., Marshall, M. & Mattice, J. (2007) Incorporating availability for detection in estimates of bird abundance. *The Auk*, **124**, 96–106.



- Digby, A., Bell, B.D. & Teal, P.D. (2013) Vocal cooperation between the sexes in Little Spotted Kiwi *Apteryx owenii*. *Ibis*, **55**, 229–245.
- Emlen, J. & Dejong, M. (1992) Counting birds: the problem of variable hearing abilities. *Journal of Field Ornithology*, **63**, 26–31.
- Fitzsimmons, L., Foote, J. & Ratcliffe, L. (2008) Frequency matching, overlapping and movement behaviour in diurnal countersinging interactions of black-capped chickadees. *Animal Behaviour*, **75**, 1913–1920.
- Holzapfel, S., Robertson, H., McLennan, J., Sporle, W., Hackwell, K. & Impey, M. (2008) *Kiwi (Apteryx spp.)* Recovery Plan. Threatened Species Recovery Plan 60. Department of Conservation, Wellington, New Zealand.
- Hurvich, C.M. & Tsai, C.-L. (1989) Regression and time series model selection in small samples. *Biometrika*, **76**, 297–307.
- Hutto, R.L. & Stutzman, R.J. (2009) Humans versus autonomous recording units: a comparison of point-count results. *Journal of Field Ornithology*, **80**, 387–398.
- Jolly, J. & Colbourne, R.M. (1991) Translocations of the little spotted kiwi (*Apteryx owenii*) between offshore islands of New Zealand. *Journal of the Royal Society of New Zealand*, **21**, 143–149.
- Kirschel, A.N.G., Cody, M.L., Harlow, Z.T., Promponas, V.J., Vallejo, E.E. & Taylor, C.E. (2011) Territorial dynamics of Mexican ant-thrushes *Formicarius moniliger* revealed by individual recognition of their songs. *Ibis*, **153**, 255–268.
- Mennill, D. & Vehrencamp, S. (2008) Context-dependent functions of avian duets revealed by microphone-array recordings and multispeaker playback. *Current Biology*, **18**, 1314–1319.
- Mennill, D.J., Battiston, M., Wilson, D.R., Foote, J.R., & Doucet, S.M. (2012) Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour. *Methods in Ecology and Evolution*, **3**, 704–712.
- Miles, J., Potter, M. & Fordham, R. (1997) Northern brown kiwi (*Apteryx australis mantelli*) in Tongariro National Park and Tongariro Forest – ecology and threats. *Science for Conservation*, **51**, 1–23.
- Parker, T.A. (1991) On the use of tape recorders in avifaunal surveys. *The Auk*, **108**, 443–444.
- Pierce, R. & Westbrooke, I. (2003) Call count responses of North Island brown kiwi to different levels of predator control in Northland, New Zealand. *Biological Conservation*, **109**, 175–180.
- R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. Version 2.15. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org>. Downloaded 02 April 2012.
- Riede, K. (1993) Monitoring biodiversity: analysis of Amazonian rainforest sounds. *Ambio*, **22**, 546.
- Riede, K. (1998) Acoustic monitoring of Orthoptera and its potential for conservation. *Journal of Insect Conservation*, **2**, 217–223.
- Robertson, H.A. & Colbourne, R.M. (2004) Survival of little spotted kiwi (*Apteryx owenii*) on Kapiti Island. *Notornis*, **51**, 161–163.
- Robertson, H.A., Colbourne, R.M. & Nieuwland, F. (1993) Survival of little spotted kiwi and other forest birds exposed to brodifacoum rat poison on Red Mercury Island. *Notornis*, **40**, 253–262.
- Robertson, H., Colbourne, R.M. & Olsen, D. (2003) *Kiwi (Apteryx spp.)* best practice manual. Technical report, Department of Conservation, Wellington, New Zealand.
- Robertson, H.A., McLennan, J., Colbourne, R.M. & McCann, T.J. (2005) Population status of great spotted kiwi (*Apteryx haastii*) near Saxon Hut, Heaphy Track, New Zealand. *Notornis*, **52**, 27–33.
- Rosenstock, S.S., Anderson, D.R., Giesen, K.M., Leukering, T., Carter, M.F. & Thompson, F. III (2002) Landbird counting techniques: current practices and an alternative. *The Auk*, **119**, 46–53.
- 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. *Journal of Field Ornithology*, **80**, 42–50.
- Towsey, M., Planitz, B., Nantes, A., Wimmer, J. & Roe, P. (2012) A toolbox for animal call recognition. *Bioacoustics*, **21**, 107–125.
- Waddle, J., Thigpen, T.F. & Glorioso, B.M. (2009) Efficacy of automatic vocalization recognition software for anuran monitoring. *Herpetological Conservation and Biology*, **4**, 384–388.
- Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A. & Smith, G.M. (2009) *Mixed Effects Models and Extensions in Ecology with R*, 1st edn. Springer, New York, USA.

Received 24 January 2013; accepted 2 April 2013

Handling Editor: Luca Giuggioli

## Supporting Information

Additional Supporting Information may be found in the online version of this article.

**Appendix S1.** Zip archive containing the C5.0 (version 2.09) files for the classifier used in the automatic detection process.

**Table S1.** Table of coefficients of generalised linear models relating year and environmental conditions to calls detected for each of the three survey methods.