CENTRE FOR
INVASIVE SPECIES SOLUTIONS

AUTOMATED DETECTION: TRIGGERING SMARTER, FASTER, BETTER RESPONSE TO INCURSIONS

FINAL REPORTS FOR PROJECT P01-T-002

AUTHORS

Susan Campbell,
Kyle Armstrong,
David Barnard
and David Lucas

OCTOBER 2022

Prepared for the Centre for Invasive Species Solutions

INVASIVES.COM.AU

The Centre for Invasive Species Solutions gratefully acknowledges the financial and in-kind contributions made by its members, associate members, and partners.

We are a not-for-profit, member-based organisation formed to address the impact of invasive plants and animals across Australia.

We acknowledge the Traditional Custodians of the lands on which we meet and work and pay our respects to Elders — past, present and emerging.

We acknowledge all Aboriginal and Torres Strait Islander peoples and their continuing connection to country, culture and community.

The Centre for Invasive Species Solutions is governed and managed by Invasive Animals Limited.

CITATION

This report should be cited as: Campbell S, Armstrong K, Barnard D, Lucas D and Yuki Konishi (2020). *Automated Detection: Triggering Smarter, Faster, Better Response to Incursions: Final Reports for Project P01-T-002*. Report for the Centre for Invasive Species Solutions.

invasives.com.au

ISBN e-Book 978-1-925727-98-2

ISBN Print 978-1-925727-99-9

ACKNOWLEDGEMENT OF PROJECT PARTNERS

The *Automated Detection: Triggering Smarter, Faster, Better Response to Incursions* project was led by the Western Australian Agriculture Management Authority in partnership with David Barnard trading as DKB Solutions Australia and Specialised Zoological.

The project was funded by the Australian Government Department of Agriculture, Fisheries and Forestry with in-kind support from Western Australia Department of Primary Industries and Regional Development.

Cover designed by Hodge Environmental, hodgeenvironmental.com.au

Cover image: Automated detection. Credit Western Australian Department of Primary Industries and Regional Development.

AUTOMATED DETECTION: TRIGGERING SMARTER, FASTER, BETTER RESPONSE TO INCURSIONS

FINAL PROJECT REPORT FOR P01-T-002

Susan Campbell,[1] Kyle Armstrong,[2] David Barnard,[3] David Lucas,[4]

[1] Department of Primary Industries and Regional Development, Albany WA 6330, Australia, susan.campbell@dpird.wa.gov.au

[2] Specialised Zoological, Adelaide SA 5061, Australia, kyle.n.armstrong@gmail.com

[3] DKB Solutions, Perth WA 6001, Australia, david.barnard@dkbsolutions.com.au

[4] Lynxemi Pty Ltd, #03-05 Singapore S387298, david.lucas@gmail.com

Contents

Figures

INTRODUCTION

EARLY DETECTION IS VITAL FOR BIOSECURITY RESPONSES

Invasive species are those that establish and proliferate in areas outside their natural range after human introduction (Kolar and Lodge 2001; Lodge et al. 2006; Perrings et al. 2002). Increases in trade (both legal and illegal), tourism, intensification of agriculture and deliberate releases all greatly exacerbate the impact of invasive species on global economic, environmental, social and cultural assets (Bradshaw et al. 2021; Perrings et al. 2002). Invasive species can be nearly impossible to eradicate once established in previously unoccupied areas. Early detection of incursions is therefore critical if control actions are to be both effective and affordable (Larson et al. 2020). Technology that assists with early detection can help provide managers with the best chance of initiating a rapid response.

Numerous methods are currently being applied to invasive species surveillance, including:

- conducting environmental sampling for species-specific DNA (e.g. CISS project P01-I-004; Larson et al. 2020; McDonald et al. 2020; Ruppert et al. 2019)
- monitoring electronic trade in wildlife (Stringham et al. 2020)
- application of lures to enhance trapping success (e.g. starlings, Campbell et al. 2012; European wasp [currently unpublished]; Asian black-spined toad, Caley et al. 2022)
- capitalising on the surveillance potential of citizen scientists (e.g. Atlas of Living Australia's DIGIVOL platform; the Western Kimberley rubber vine Aquila project; WA Department of Primary Industries and Regional Development's MyPestGuide app)
- using remote sensing devices, including cameras, acoustic detectors and LiDAR [light detection and ranging] sensors (reviewed in Juanes 2018).

ACOUSTIC DETECTION CAN IMPROVE THE FOOTPRINT OF PEST SURVEILLANCE

Many animals in both terrestrial and aquatic habitats generate unique sounds (including ultrasonic sounds) that can be recorded on acoustic recording unit (ARU) sensors (e.g. Shonfield and Bayne 2017; Barber-Meyer et al. 2020; Sousa-Lima et al. 2013; Britzke et al. 2013; Walters et al. 2012). A sensor is a device that detects or measures a physical or biological property, permitting responses such as records, alerts or control.

Acoustic recording devices have expanded the utility of simple surveys at different scales so that data can be recorded, digitised, saved, scanned, compared to known libraries and easily mapped (Juanes 2018). Such sensors permit an efficient, non-invasive and taxonomically broad means to study wildlife populations (including incursions of invasive species) and monitor wildlife community responses to environmental change and anthropogenic influences (reviewed in Gibb et al. 2019). To date, ARUs have predominantly been applied to conservation/biodiversity assessment programs – for example, detecting the presence of rare, cryptic species (Armstrong et al. 2021) – with minimal application to the detection of invasive species. Notable exceptions (reviewed in Juanes 2018) come from Australia where ARUs have been applied with mixed success to detect cane toads (*Rhinella marina*) at incursion fronts (Tingley et al. 2017), and acoustic lures have been trialled on both cane toad and Asian black-spined toad (*Duttaphrynus melanostictus*) traps to try to improve trapping efficacy (Muller et al. 2020; Caley et al. 2022).

Automated acoustic surveillance can potentially increase the spatial and temporal footprint of surveillance. However, one of the key challenges facing the integration of ARUs into large-scale invasive species management programs is the ease with which operators can effectively manage the ensuing ‘big data’. Reliable automated data analysis is required to free operators from the burden of manually verifying thousands of hours of sound recordings. However, it is important to emphasise that whilst surveillance for new biological invasions can be transformed by changes in both how the environment is monitored and who is doing the monitoring (Caley et al. 2022), surveillance is only valuable if it is linked to management responses that can rapidly prevent the spread and/or impact(s) of invasive species (Larson et al. 2020).

KEEPING WESTERN AUSTRALIA STARLING-FREE

Starlings (*Sturnus vulgaris*) are an aggressive generalist species identified by the International Union for Conservation of Nature and Natural Resources as one of the 100 worst invasive species globally. Starlings feed on arthropods and other ground-dwelling organisms, grain crops and horticultural produce. They consume and spoil livestock feed, foul stock water points and livestock with excrement, are vectors for disease, foul and damage public and private amenities, and compete with native fauna for resources (particularly nesting hollows). Approximately 1 million km^2 in WA is considered highly suitable habitat for starlings (Kirkpatrick 2008). An area this size could support up to 12.5 million starlings, potentially causing losses of more than $176 million annually to WA agriculture (Campbell et al. 2016).

The WA starling management program aims to prevent starlings from establishing, and is unique in terms of longevity (> 50 years of near-continuous management), scale (> 10,000 km^2) and outcome (starlings are not established in WA). Despite these successes, modelling (Anderson 2009, 2017; Campbell et al. 2015) indicates that additional spatial and temporal surveillance is required to keep WA starling-free.

OBJECTIVES

The key research objectives were to:

- Develop and demonstrate a remote acoustic surveillance, detection and reporting solution, using starlings as an initial case study.
- Develop and demonstrate the application of the remote acoustic detection solution for additional, high-priority invasive pest animal(s).
- Communicate outcomes and promote end-user uptake of the technology.

METHODS

INTEGRATING AUTOMATED ACOUSTIC RECORDING UNITS INTO LANDSCAPE-SCALE INVASIVE ANIMAL CONTROL PROGRAMS

Adoption of ARUs into ongoing wildlife management programs requires that the technology be:

- durable
- reliable
- accurate
- upgradable
- accompanied by automated data analysis.

DEVELOPING A COMPREHENSIVE STARLING CALL REFERENCE LIBRARY

We drew upon three primary sources of reference data to build the starling call reference library for this project:

1. In 2011, 50 ARUs (SongMeter2, Wildlife Acoustics) were deployed throughout habitat at high risk of starling incursion on the south coast of WA. They were programmed to record for several hours every day at dawn and dusk, and for 5–10 minute intervals every hour (Campbell et al. 2013). A subset of the resulting 27 terabytes of recordings assisted the evolution of algorithm development.
2. During 2011 and 2012, high-quality reference calls from starlings were collected from several locations in SA using a handheld digital recorder (Nagra BB+) and directional microphone (Campbell et al. 2013). These annotated calls were incorporated into the training of the current starling convolutional neural network (CNN) algorithm.
3. New SA recordings were collected from Adelaide city, Cape Jervis and several other locations on the Fleurieu Peninsula using a Bioacoustics Audio Recorder (BAR) (Frontier Labs), a Rode NTG-2 shotgun microphone connected to a Zoom H1n digital audio recorder, and the CM4 prototype (all recordings taken at a sampling rate of 48 kHz).

TRAINING THE MODEL TO RECOGNISE STARLING CALLS

Using our comprehensive reference library of starling calls, we trained a one-dimensional CNN to recognise three main call types. The model available for broadscale deployment was trained with 10,553 calls comprising the three starling call types (n = 6,377) and other signals (n = 4,176) from field test sites in SA, plus 8,329 non-target signals from WA (95.5% accuracy, 95.7% precision, 95.4% recall).

The starling CNN has been trained to recognise three call types: two variations on a ‘buzz’ call sequence (Figures 1a and b) and a shrill descending whistle (Figure 1c).

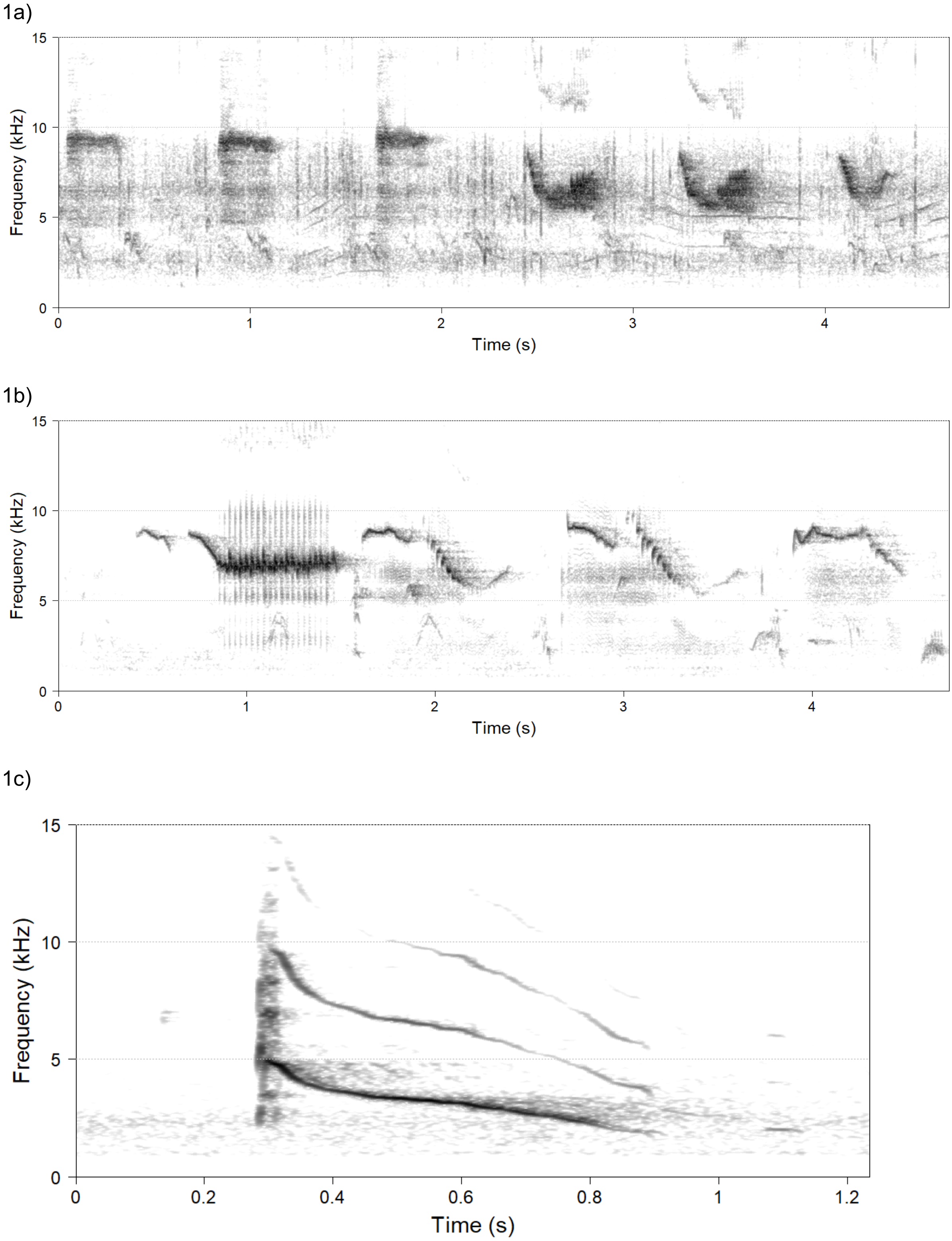


Figure 1. Spectrograms (time versus frequency) illustrating the three call types detected by the starling CNN; two variations on a starling 'buzz' call sequence (a,b) and also a shrill descending whistle (c).

SOFTWARE RESOURCES

CREATING AND TRAINING THE STARLING-DETECTION ALGORITHM

To develop the bioacoustic recording and processing system, we first tested several different analysis approaches before settling on the development of the CNN-based system. A Masters of Engineering student project (University of Adelaide) developed systems for the recording and classification of starling calls. They used Python and bash scripts, feature extraction with mel-frequency cepstral coefficients, tested filtering and noise-removal options and used K-Nearest Neighbour to allocate target detections. These were all running within a Docker container on a balenaFin 1.1 (https://www.balena.io/fin/) microcomputer that was remotely controllable through the balenaCloud resource (https://www.balena.io/cloud/).

We developed a feature extraction and classification analysis pipeline using functions from several bioacoustics-related packages in the [R] statistical computing language (*bioacoustics*, Marchal et al. 2021; *monitoR*, Hafner and Katz 2018; *warbleR*, Araya-Salas and Smith-Vidaurre 2017). Classification was tested using linear Discriminant Function Analysis and random forest walks. Initial tests with both these pipelines demonstrated that the rate of false-positive detections was likely to be prohibitively large for the intended application.

The CNN-based pipeline was also first developed and tested on both an NVIDIA Jetson Nano Developer Kit (V3) (https://developer.nvidia.com/embedded/jetson-nano-developer-kit) and a Raspberry Pi 4 model B (https://www.raspberrypi.com/products/raspberry-pi-4-model-b/). Both these hardware devices were successful in running the CNN pipeline steps from recording through to classification, but ultimately the Raspberry Pi Compute Module 4 and IO board (https://www.raspberrypi.com/products/compute-module-4-io-board/) were chosen on the basis of power management and communications.

Three main resources were produced for the CNN analysis pipeline by Thomas Rowntree, Senior Research Engineer at the Australian Institute of Machine Learning, University of Adelaide. The first ('Training Code') is a custom routine written in Python, used to train the CNN model with both starling calls and non-target signals of various types.

The second ('Runtime Code') is a signal classification pipeline within a Docker container environment, used to apply the trained CNN model to field recordings on the microcomputer. It ran a pipeline comprising the following steps from recording to storing the outputs of signal classification:

- **Sound detection**: Sound is taken from the microphone and analogue-to-digital converter and is transferred into computer memory, at a sampling rate of 48 kHz.
- **Feature extraction and CNN model classification**: The CNN model is applied to a 5-minute segment of this audio stream in real time using mel-frequency cepstral coefficients for feature extraction. A vector of probability scores reflecting the degree of model to signal match across this audio stream is produced.
- **Segmentation**: Signals that match the CNN model with a probability score above a chosen threshold are identified as starlings. The threshold score is set by the user based on performance testing results. For performance testing, all signals with a probability score of 0.50 and above were saved, and for field tests currently being undertaken in WA, a probability score of 0.96 is used to achieve the best compromise between the rate of false-positive signal rejection and incorrect target signal rejection.
- **Saving for validation**: Segmented signals are saved at the midpoint within a buffered length of audio stream to a total of five seconds. These five-second snippets are saved as 16-bit 48 kHz WAV sound files, with filenames encoding information about snippet date and start time, and the probability score for the detection.

- **Saving for further development**: Each 5-minute segment of audio stream analysed in real time is saved as a date/time-stamped 16-bit 48 kHz WAV file, for archival purposes. This allowed performance assessment of the various CNN model versions and will ultimately be excluded in the most mature pipeline to be deployed widely.
- **Logging and housekeeping**: A log file is produced of all recordings, and WAV recordings accumulate in an 'outputs for validation' directory.

The third CNN resource ('Desktop Runtime Code') was a modified version of the Runtime code, designed to run on a desktop computer and take archived field recordings on an external drive as the input. This was used to process the 2011 recordings made with Song Meters in WA, and to provide a resource of likely false-positive signal types that would be present in the WA soundscape. This resource was used to re-train the CNN model again to reduce the false positives requiring manual validation.

We trained the CNN model and then performance evaluated the model in an iterative process, beginning with the development based on signals from Adelaide city, followed by more intensive field testing at a site at Cape Jervis with an early prototype of the chosen Compute Module 4 (CM4) hardware, and resulting in full deployment at two sites in WA.

The model versions include:

- **Model0**: First model produced with legacy and newly collected Adelaide city reference calls. Initial testing within Adelaide city.
- **Model1**: Improved model based on field recordings from Cape Jervis, Clayton Bay and Adelaide city. Evaluated via field test 1 at Cape Jervis using a CM4 hardware prototype.
- **Model2**: Improved model based on verified output from first field test (based on further examples of labelled true and false positives). Evaluated in second field test using a CM4 hardware prototype and deployed in WA at two tower sites.
- **Model3**: Model with additional training to reject false positives (derived from Model2 when run over historical Song Meter data with the Desktop Runtime code).
- **Model4+**: Future model(s) to be re-trained to reject false positives accumulated in the broadscale deployment in WA.

ADDING EXTRA FUNCTIONS TO THE FIELD DEVICE

Several additional scripts were developed to ensure the final field device was fully automated, and could be remotely queried and updated without end users opening the device in the field. A management controller enabled the following to be undertaken:

- waking of the main central processing unit (CPU) as scheduled (e.g. relative to sunset/sunrise times)
- gathering of solar data on an hourly basis (total energy and battery status) and reporting this data to the main CPU on request
- operating the external indicator LEDs
- listening for Bluetooth connections and commands
- instructing the main CPU to enable or disable wi-fi (providing security and power savings)
- handling failure scenarios.

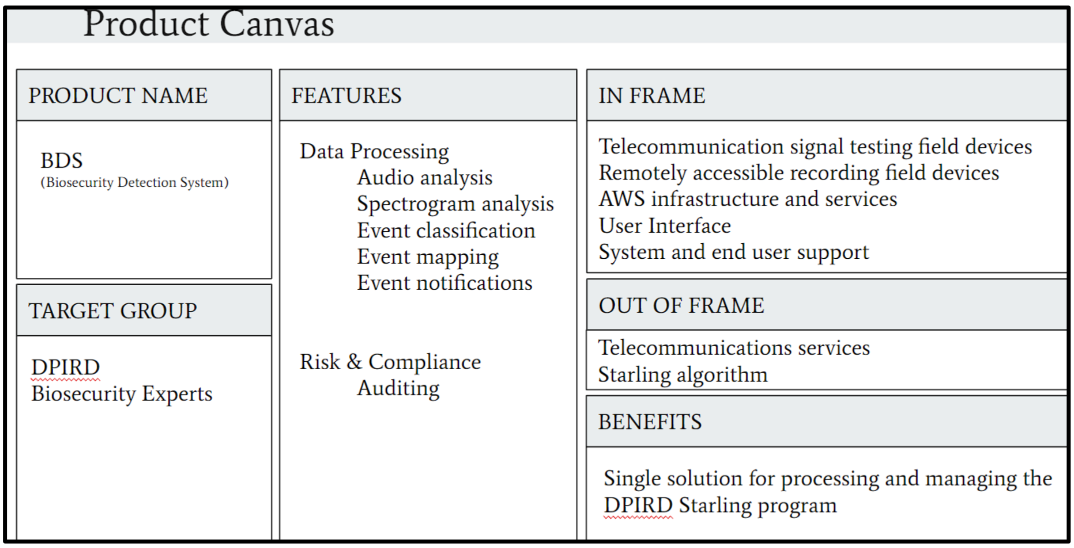
The main controller was developed with the following functionality:

- upon waking, retrieve location and synchronise time
- start the Long-Term Evolution (LTE) communications and monitoring system
- set the time on the management controller to ensure synchronisation
- retrieve any update configuration from the cloud
- set the next wake-up time with the management controller
- retrieve solar data from the management controller
- report wake-up and solar data to the cloud
- start the Docker container housing the detection system
- await detection messages and attempt to send them to the cloud (or queue them for later if initial delivery fails)
- stop the detector as scheduled
- report sleep state to the cloud
- turn off communications and shut down.

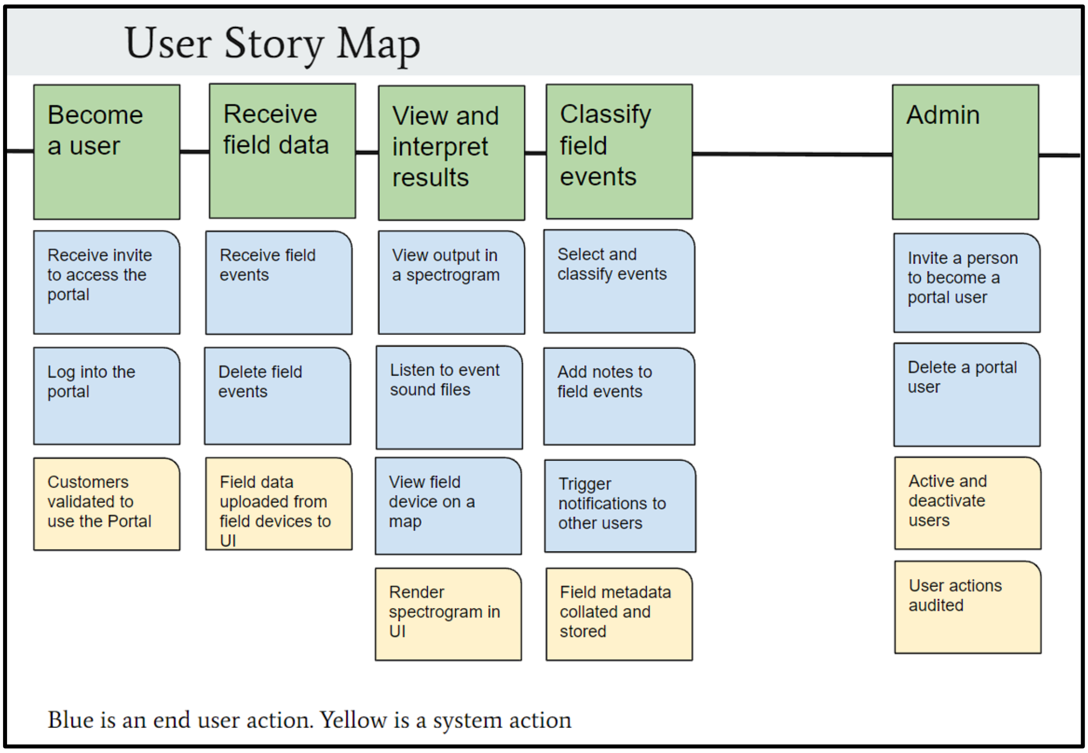
DEVELOPING A USER INTERFACE THAT MEETS PROGRAM AND USER NEEDS

The user interface (Detect-It) went through several development iterations until we built an interface that was going to meet the program's needs. This involved starling-program discovery sessions between DKB Solutions and the starling team. Development packs for each phase were generated from these sessions. The development packs produced an agreed product canvas (Figure 2a), personas for end users, a user story map (Figure 2b), and the user interface functions and acceptance criteria to support them (Figure 2c). Software development was conducted using agile software development methodologies, resulting in a minimum viable product that then underwent user acceptance testing over a short time frame. Additional functionality was then added and tested to achieve all in-scope items.

2a)



2b)



2c)

View and interpret results

Story	Acceptance criteria
Listen to event sound files	* Events can be selected and sound files played/paused
View output in a spectrogram	* Events can be selected and spectrograms can be viewed
View field device on a map	* See where the field device is on a map * The user can zoom in and out of the map

Figure 2. Schematic representation of the early consultation and planning that facilitated development of the online end-user application 'Detect-It'.

ON-BOARD PROCESSING AND REMOTE COMMUNICATIONS ELIMINATE THE NEED FOR REGULAR SITE VISITS

The device design dramatically reduces the amount of data required to be transferred off the device and negates further processing. The design uses edge computing architecture that exploits large computer power processes and stores data in the field before sending data to back-end services. Via a custom user interface, program managers can view high-detection-probability events sent from devices via long-term evolution (LTE) Cat-M1 low-bandwidth communications to cloud services.

Both the main unit and the battery box are waterproofed ABS boxes, with additional waterproofing added during deployment (with epoxy resin coating around external connector points and box seals). The dual microphones are waterproofed, omnidirectional condensers with a signal-to-noise ratio of 80 dB at 1 kHz and a sensitivity of –28 db ± 3 dB at 1 kHz (only one is used currently – selectable with redundancy).

The computing module is a Rasp Pi CM4 with eMMC, rather than an SD card, for improved reliability. The device uses a 20 Ah LiFePo4 battery, which is recharged via a solar regulator connected to a 100 W lightweight PET solar panel. Capacity refinements can be made from energy studies of the devices. There is a low-power custom management controller (always on) that collects the energy data and manages the wake and sleep cycles of the main processor. An operator in the field can connect to this controller via Bluetooth on a mobile device to operate it outside normal cycles, to avoid opening the casing (e.g. users can remotely wake the main CPU and engage the wi-fi and other communication elements).

We used both an online tool to investigate signal strength (based on antennae location and propagation information) plus a custom network scanner device to measure signal strength of the LTM Cat-M1 network onsite. Both these approaches enabled us to select sites with suitable coverage to ensure effective communication of detections from field devices. We applied the information from the Australian Communications and Media Authority web portal to predetermine antenna height, power levels, gain, relevant beam width and azimuth. The Yagi antennae (15 element, 14 dBi, 700 MHz, Band 28) attached to a 2-m pole atop the main mast was then directed towards the most effective communications tower. The main 4.2-m galvanised mast was cemented 0.8 m into the ground and anchored by four guy wires, each rated to 1 tonne of pressure via hydraulically driven anchors. The device also contains a GPS for both time synchronisation and locating itself on every wake.

TESTING THE ACOUSTIC RECORDING UNITS' COMMUNICATIONS, ALGORITHM PERFORMANCE AND MICROPHONE SENSITIVITY

Once the towers were installed at both sites, we conducted playback experiments to test onboard communications, algorithm performance and microphone sensitivity. Example recordings of both the buzz and whistle starling call sequences were loaded onto a microSD card and projected at 70–90 dB (measured at 1 m with a Lutron SL-4011 sound-level meter) towards the ARU from a portable Bluetooth speaker (Xeneo) attached to the end of a 4-m telescopic pole (to simulate the height of a starling calling from a small tree). Recordings were played every 10 m up to 150 m along five evenly spaced radial transects around the ARU. To assess the playback results, we remotely downloaded approximately 2.5 hours of field recording for the period that the playback experiments occurred, and manually annotated all discernible starling calls. For further validation, we reviewed the detections communicated from the ARU to the cloud server in real time.

RESULTS

REFINING THE CNN DETECTION ALGORITHM

A highly accurate and precise bioacoustics-based starling call recording and recognition system has been produced and field-tested. Key to this was the design and implementation by Thomas Rowntree (formally, Australian Institute of Machine Learning). The detection model was produced by the one-dimensional CNN method and is embedded in a routine that records signals from the microphone, classifies signals within the audio stream in real time, and outputs results and example snippets of putative target signals ready for communication back to base. Specifically, there were five core software resources created during the project:

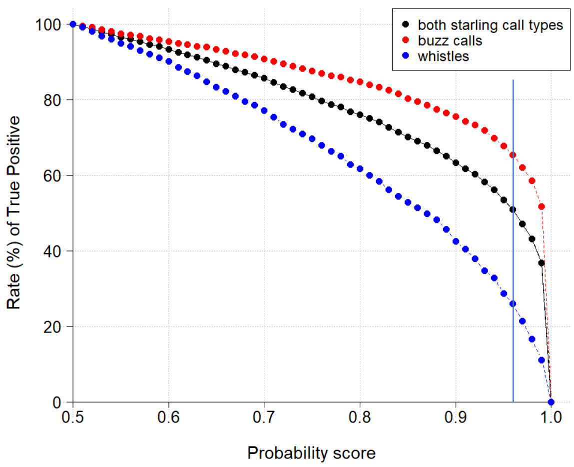
- **Training code**: a routine to train a new CNN model on a desktop computer.
- **Runtime code**: Python and shell scripts to create a Docker environment, record an audio stream, run the CNN model over the audio stream, and save five-second sound file snippets containing putative signals. Five-minute files were also saved in field tests to allow performance evaluation.
- **Desktop Runtime code**: a subset of the Runtime code that enables running the CNN model over bulk data fed into a desktop computer. It enabled collation of false positives from archived WA Song Meter recordings.
- **Other**: additional scripting to handle power management and communications between the CM4 and cloud-based server.
- **Detect-It**: a cloud-based environment enabling efficient visualisation and validation of potential targets collected from all field devices.

The CNN model development was iterative, starting at Model0 and reaching Model3 by the end of this project. The performance of each CNN model version was evaluated with field-collected data, then each model was refined using the outputs from each evaluation. For training, field-collected data was mainly derived from Adelaide city and several sites on the Fleurieu Peninsula for Model0, predominantly from Cape Jervis for Model1 and Model2, and from the 2011 WA Song Meter recordings for Model3.

LOW DETECTION AND HIGH FALSE-POSITIVE RATES FOR FIRST ITERATION OF STARLING ALGORITHM

During training, we calculated the accuracy, precision and recall of Model1. In a three-week field trial at Cape Jervis, SA, we further evaluated its performance by manually verifying putative detections (correct identifications and incorrect false-positive identifications). Both the full recording from the field test, and the putative detections (five-second files) were used for manual verification. The true-positive identification rates expected for buzz, whistle and both call types combined is depicted in Figure 3a, with the rate of true positives given for each probability score (level of confidence in the result). For Model1, 50.9% of true starling calls are detected, with 49.1% of true starling calls missed at a probability threshold of 0.96. For this first iteration of the CNN, the false-positive rate at a probability threshold of 0.96 was unacceptably high, with 22.2% of all returned snippets labelled incorrectly as starling (Figure 3b).

3a)



3b)

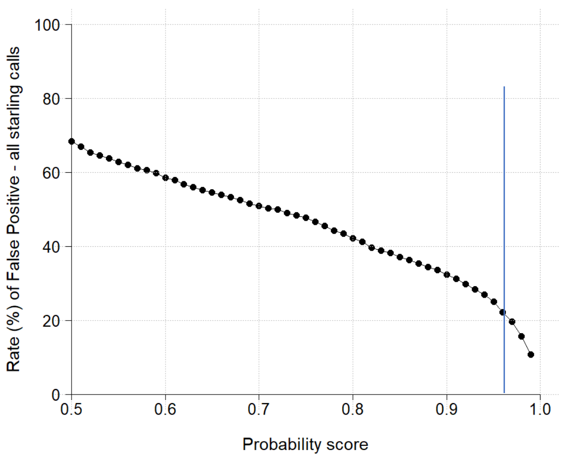
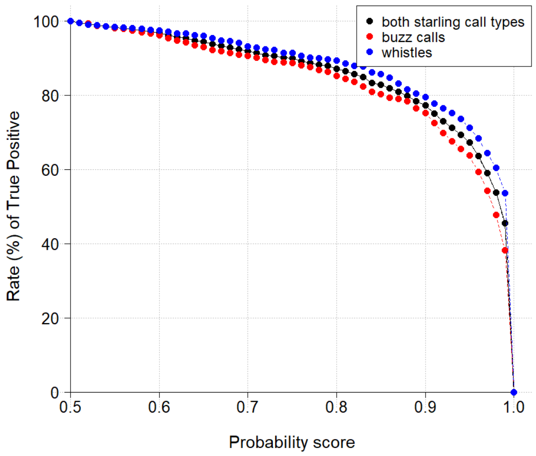


Figure 3. Plots of true positive (a) and false positive (b) rates for starling CNN Model1. Probability score (confidence) on the x-axis and rate as a percentage of total true positives recovered (a) and percentage of returned detections (snippets) that were incorrect (i.e. not starling) (b). Vertical blue line indicates 0.96 Probability threshold.

PROBABILITY OF DETECTION INCREASES WITH MODEL 2 AFTER RETRAINING WITH CALLS THAT MODEL1 FAILED TO IDENTIFY.

To improve model performance, we manually checked all recordings from the first field test for starling calls that were missed (i.e. labelled with a probability score of 0.50 or less by Model1). These missed calls were re-labelled as positives and fed back for re-training. The resulting Model2 detected a much greater proportion of starling calls (63.6% detected at probability threshold of 0.96). Detection rate of the whistle call type was greatly improved from Model1. These increases in performance by Model2 (Figure 4a) were gained without jeopardising the false-positive rate, which was also reduced to 4.2% at a 0.96 threshold rate (Figure 4b). The 36.4% of true starling calls missed by Model2 are typically very faint, low-quality examples; the model performs very well on good-quality repeated calls (Figure 5). Model2 is now installed on the two detection towers in the field and performs with 95.5% accuracy, 95.7% precision and 95.4% recall.

4a)



4b)

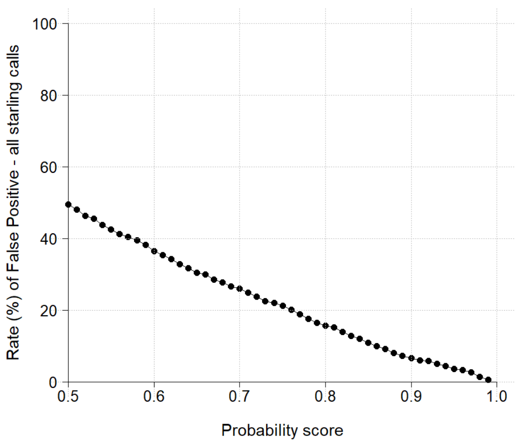
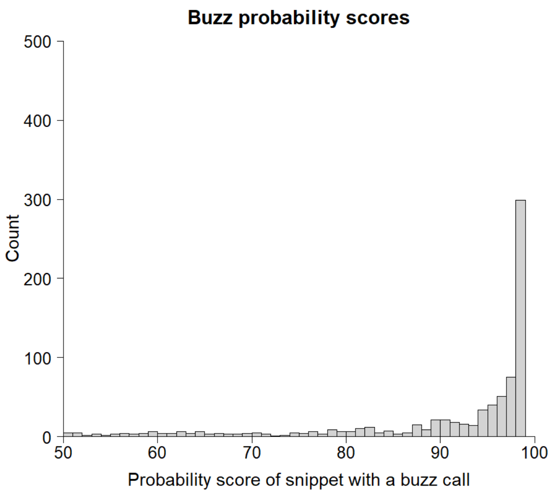


Figure 4. Plots of true positive (a) and false positive (b) rates of detection for starling CNN Model2. Probability score (confidence) on the x-axis and rate as a percentage of total true positives recovered (a) and percentage of returned detections (snippets) that were incorrect (i.e. not starling) (b). Vertical blue line indicates 0.96 Probability threshold.

5a)



5b)

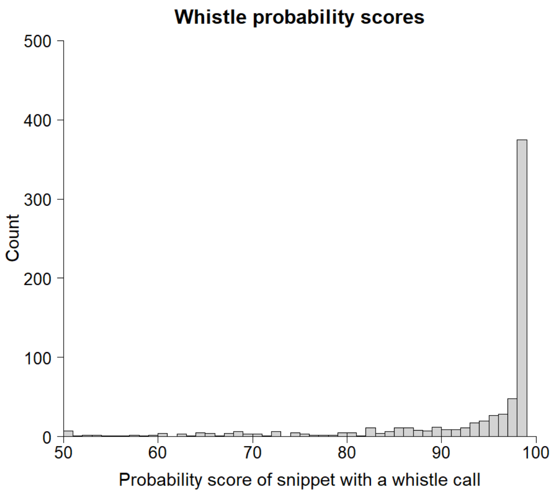


Figure 5. Cumulative plots illustrating the proportion of starling calls detected for a given probability threshold (x-axis) for buzz call types (a) and whistle call type (b). Most starling calls are recognised with a high probability. The small proportion of calls with low probability scores (left hand side tail of plots) are those calls that have low signal quality relative to background noise, for example those generated by birds calling too far from the microphone, or unfavourable background noise during high wind.

MODEL 3 WAS RE-TRAINED WITH NON-TARGET SIGNALS FROM ARCHIVED WA RECORDINGS

After installing field listening towers with Model2, we developed a Desktop Runtime code to process the terabytes of field recordings collected on Song Meter recorders (Wildlife Acoustics) in WA in 2011 (Campbell et al. 2013). Model3 has been retrained with 8,329 non-target signals from the starling management program area on the south coast of WA. Further analysis is required to quantify the improvement in accuracy, precision and recall for Model3.

STARLINGS MODEL TRAINING CODE WAS USED AS A BLUEPRINT FOR CALL DETECTION OF OTHER SPECIES

The first version of a similar bioacoustics-based detection system for the Asian black-spined toad has been trained on recordings provided by a global network of collaborators. These include recordings made in India, Indonesia, Madagascar, and Singapore. The output model (Model_0) is ready for a small-scale field test in a habitat with challenges such as other frog calls, signal sources and overlapping calls of the target species.

The training code developed to train the starling CNN model was applied to a new dataset of 561 Asian black-spined toad target calls and 523 non-target call signatures, resulting in the generation of a Model_0 for Asian black-spined toad acoustic detection. These calls consist of long call train bursts of variable duration (approximately 12 seconds long) that consist of repeated pulses spanning the frequency range 0.5–4.0 kHz (Figure 6).

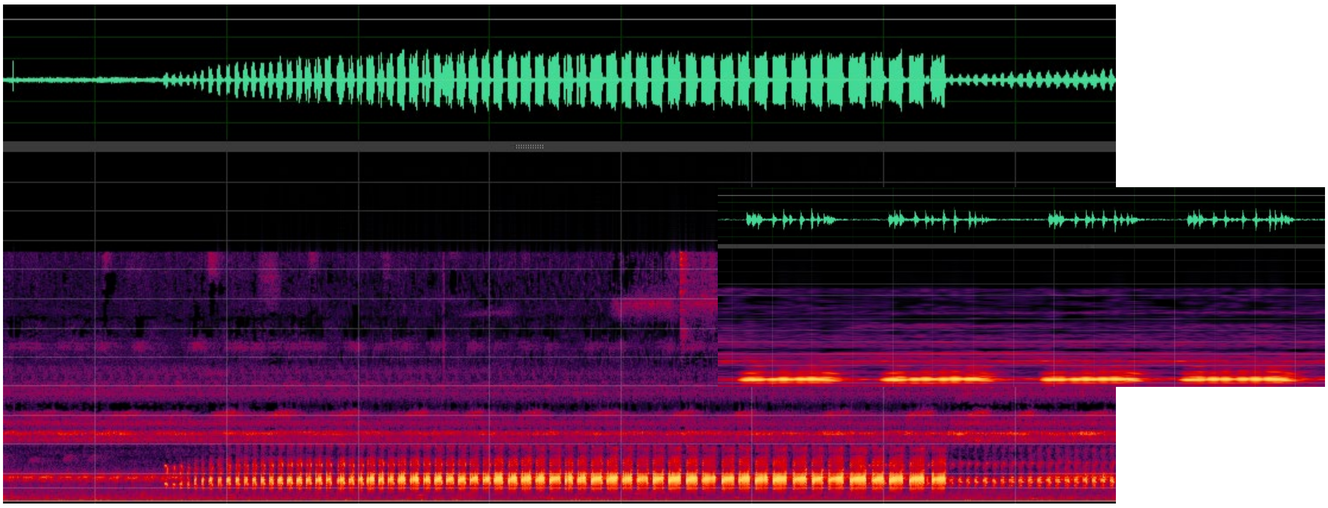


Figure 6. Waveform (top) and spectrogram of an example Asian black-spined toad call with time on the x-axis and frequency (not displayed, but between 0.5 – 4kHz) on the y-axis. Inset image illustrates detail of the four pulses within the long 12-second call train.

AUTOMATED STARLING ACOUSTIC RECORDING UNITS CAN DETECT CALLS UP TO 70 M AND RECORD CALLS AT 150 M AWAY

Two permanent ARU towers for starling surveillance were successfully installed at Bremer Bay (April 2022) and Gibson (May 2022) in WA. Both sites have a history of starling incursions, and the towers will provide additional ongoing starling surveillance for the Department of Primary Industries and Regional Development's starling management program. The towers (Figure 8a–c) were operational from the day of installation, with proven end-to-end communication of acoustic signals to cloud servers over the Telstra LTE Cat-M1 network and were available via the Detect-It user interface (Figure 7). Two additional towers have been commissioned and will be permanently installed at strategic locations selected for the suitability of starling habitat and history of known incursions.

Preliminary investigation of playback recordings indicate that the devices are capable of recording starling calls at a distance of (and potentially beyond) 150 m. Whilst the playback files could be manually verified via listening and viewing of spectrograms (RavenPro, Cornell Lab) of the full recordings downloaded from the ARU, the maximum distance that the onboard algorithm was able to positively identify the starling playback calls was approximately 70 m. Beyond 70 m, we suggest that the signal-to-noise ratio is insufficient and the CNN is unable to confidently identify the calls (with a probability of > 0.96). Further investigation into the influence of both biotic (e.g. other bird calls) and abiotic variables (e.g. wind, rain or anthropogenic noises) is required to provide a quantitative assessment of the 'detection footprint' for each ARU. It is likely that this footprint will be variable, but we have a high level of confidence that our system will communicate a positive detection for all high signal-to-noise ratio starling calls.

Evidence of noise clipping was present when the devices were set to the highest gain. This was reduced by setting a normalised gain. Other induced noise was detected in some circumstances once the device was operational in remote locations, which were not evident during prototype development and testing in both Singapore and Adelaide. Future iterations of the detectors will remedy this. Lastly, although the south-west of WA seems to be well serviced by Telstra's 700 MHz network, Nullarbor coverage is much more limited and could impact where the current system can deliver additional starling surveillance. This information can be confirmed through the Australian Communications and Media Authority web portal and will change as additional Telstra towers are constructed.

DETECT-IT USER INTERFACE

Every positive detection communicated over Telstra's LTE Cat-M1 network is displayed as a single line via the Detect-It user interface with associated metadata including Event ID, Device ID, Confidence level and Date/Time Stamp (Figure 7a). Users can hover over and enlarge the thumbnail spectrogram image from this landing page for a quick visual review of the spectrogram for each detection, and action (verify) each detection if the outcome (starling/not starling) is very clear at this point. Alternatively, users can further investigate individual detections by viewing the full spectrogram, playing back the five-second audio, downloading the audio, locating the detection on a map, annotating and sharing detections (Figure 7b). Buffering the call signal identified by the algorithm within a five-second window of surrounding audio provides the end user a much clearer context for the detection, with a more realistic listening experience to help assess the true source of the signal.

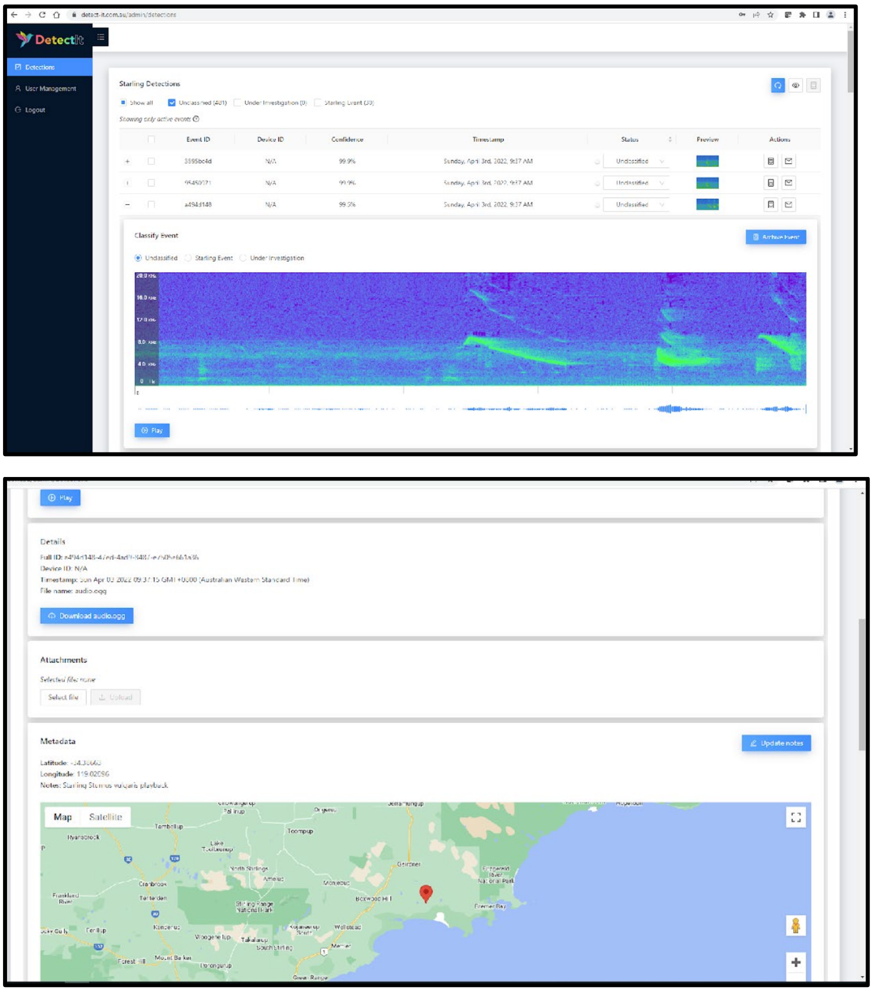
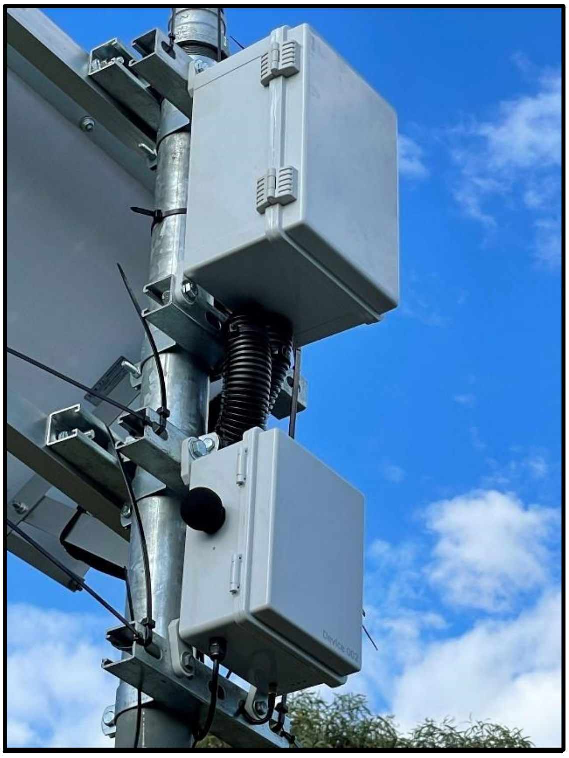


Figure 7. Screen shots from 'Detect-It', user interface for viewing, verifying, listening to, annotating, mapping, archiving and sharing putative detections returned from the CNN onboard the starling field ARUs

8a)



8b)

8c)

Figure 8. a) Automated recording unit (ARU) with two omni-directional microphones positioned beneath Lithium-Ion solar charged battery. Split conduit piping surrounds exposed cabling to provide protection from wildlife. b) Four guy wires ratcheted to one tonne of pressure hold the 6m galvanised central mast (c) which supports the ARU, battery, solar panel and Yagi antennae.

APPLICATION

EXPANDING THE FOOTPRINT OF TRADITIONAL SURVEILLANCE

By deploying fully automated, remote, passive acoustic surveillance technology to expand the footprint of starling surveillance in remote locations, we have demonstrated how early acoustic detection could facilitate an effective control response.

Starling numbers in WA are predicted to grow exponentially, with slow initial increases followed by a rapid increase over a 10-year period (Anderson 2017). Enhancing the current management regime by increasing surveillance will help keep starling numbers low and help prevent this pest species from establishing in WA. The ARU technology we developed and tested is one way the program may cost-effectively increase starling surveillance, noting that an effective control response is still required to remove any detected starlings. The project team will participate in an online workshop (Tuesday 8th November with Department of Primary Industries and Regional Development staff and management involved with the starling program to discuss options for a large-scale rollout of ARU surveillance technology under its ongoing program.

DISCUSSION

THE NEXT STEPS FOR THE PROJECT TEAM

A draft commercialisation and utilisation strategy has been prepared for CISS outlining how the products of this project could potentially be used by other end users, and which presents a potential pathway (the detection hub/wildlife monitoring portal) to apply the technology in other programs.

We also intend to:

- Investigate the influence of abiotic variables on detection radius with statistical modelling.
- Re-engage with the international Asian black-spined toad research community to enhance availability of test files.
- Further develop the commercialisation strategy.
- Continue verifying false-positive signals from Bremer Bay and Gibson towers.
- Investigate the source and possible fix of ‘noise’ returned intermittently from towers.
- Present findings at Queensland University of Technology’s bioacoustics symposium (November 2022).
- Present findings at Australasian Wildlife Management Society (December 2022) New Zealand conference.
- Facilitate collaboration with Department of Biodiversity, Conservation and Attractions to re-train CNN with the existing Western ground parrot call reference library.
- Prepare grant applications to support rollout of passive acoustic surveillance technology for starlings ‘at scale’ throughout the south coast of WA.
- Continue collaborations and prepare proposals to support integration of passive acoustic surveillance technology into a broader ‘detection hub’ along with thermal and visual artificial intelligence technology.
- Continue sharing the innovative technology we have developed with the international research community and identify additional end users.

STARLING CNN MODEL HAS BROADSCALE POTENTIAL FOR OTHER SPECIES OF INTEREST

Re-training the CNN with a reference library of Asian black-spined toad calls has demonstrated the broader applicability of this technology to other species. Sharing the project resources allows both conservation and biosecurity programs to benefit from the application of automated acoustic surveillance, enabling earlier detections and rapid responses. Proving the suitability of our CNN algorithm for other species via the Desktop Runtime code is the first step in sharing the project resources. Subsequently, both biosecurity (e.g. Asian black-spined toad) and conservation (e.g. Western ground parrot) programs could integrate their species-specific algorithm with the hardware and communication solutions developed under this project.

The Invasive Animal Ltd. board has approved the licensing of the Desktop Runtime code developed in this project to SA Department of Environment and Water and WA Department of Biodiversity, Conservation and Attractions for non-commercial purposes. Some adjustments may be required for the design of other species’ ARUs (e.g. placement of microphones for ground-dwelling species). Specific challenges to communication of data from alternative placements would require planning but are not insurmountable.

WHAT HAVE WE LEARNED?

REMOTE, FULLY AUTOMATED ACOUSTIC DETECTION IS POSSIBLE

Our solution supports and improves field monitoring programs in what can be considered a step-change, when compared with current monitoring methods. We have enabled:

- autonomous field units operating year-round
- near-real-time notifications of high probability starling events with call processing via the onboard algorithm, call storage and communication to transfer call event data
- one user interface displaying each call event with time, location, audio playback and spectrogram analysis
- the latest field data available to end users without the burden of retrieving historical data from the field.

SECURE, PERMANENT INFRASTRUCTURE IS EASY TO USE

Specialised hydraulic anchoring equipment can safely secure galvanised masts to support solar powered field units year-round. Each fully autonomous field unit runs wake/sleep, battery level, communication checks and data-transfer protocols. Field visits to update or inspect units are no longer required with remote software updates. Data stored securely in the cloud places the solution at the forefront of technology and support with Amazon Web Services (AWS) MQTT messaging and the AWS core.

ACCURATE, PRECISE AND RELIABLE AUTOMATED DATA ANALYSIS IS POSSIBLE

Machine learning approaches have proven successful for developing an automated starling detection algorithm with high performance and, encouragingly, few false-positive detections. The training script developed from this project enables other users to re-train the CNN framework with calls from other species of interest, allowing both conservation and biosecurity programs to efficiently analyse vast volumes of surveillance data.

REFERENCES

*Note: bold references indicate project outputs. Other CISS and IA CRC publications are marked with an **

Anderson D (2009) *Modelling the probability of starling (*Sturnus vulgaris*) eradication in Western Australia*, Landcare Research Contract Report: LC0910-061.

Anderson D (2017) *Surveillance effort required to control starlings along the southern coast of Western Australia*, Landcare Research Report prepared for the Department of Agriculture and Food, Western Australia, June 2017.

Araya-Salas M and Smith-Vidaurre G (2017) 'warbleR: an R package to streamline analysis of animal acoustic signals', *Ecology and Evolution*, 8(2):184–191, doi:10.1111/2041-210X.12624.

Armstrong KN, Broken-Brow J, Hoye G, Ford G, Thomas M and Corben C (2021) 'Effective detection and identification of sheath-tailed bats of Australian forests and woodlands', *Australian Journal of Zoology* 68(6):346–363, doi:10.1071/ZO20044.

Barber-Meyer SM, Palacios V, Marti-Domken B and Schmidt LJ (2020) 'Testing a new passive acoustic recording unit to monitor wolves', *Wildlife Society Bulletin,* 44(3): 590–598, doi:10.1002/wsb.1117.

Bradshaw CJ, Hoskins AJ, Haubrock PJ, Cuthbert RN, Diagne C, Leroy B, Andrews L, Page B, Cassey P, Sheppard AW and Courchamp F (2021) 'Detailed assessment of the reported economic costs of invasive species in Australia', *NeoBiota*, 67:511–550.

Britzke ER, Gillam EH and Murray KL (2013) 'Current state of understanding of ultrasonic detectors for the study of bat ecology', *Acta Theriologica*, 58(2):109–117, doi:10.1007/s13364-013-0131-3.

Bylemans J, Gleeson DM, Hardy CM and Furlan E (2018) 'Toward an ecoregion scale evaluation of eDNA metabarcoding primers: a case study for the freshwater fish biodiversity of the Murray–Darling Basin (Australia)', *Ecology and Evolution*, 8(17):8697–8712, doi:10.1002/ece3.4387.*

Caley P, Campbell S, Csurhes S and P Cassey (2022) *Development of integrated passive and active surveillance tools and networks*, Draft Final Report prepared for Centre for Invasive Species Solutions, project P01-I-003. *

Caley P, Welvaert M and Barry SC (2020) 'Crowd surveillance: estimating citizen science reporting probabilities for insects of biosecurity concern', *Journal of Pest Science*, *93*(1):543–550, doi:10.1007/s10340-019-01115-7. *

Campbell S, Cook S, Mortimer L, Palmer G, Sinclair R and Woolnough AP (2012). 'To catch a starling: testing the effectiveness of different trap and lure types', *Wildlife Research*, 39(3):183–191.

Campbell S, Parsons S, Woolnough A and Obolonkin V (2013) *Identifying corridors, refuges and entry points for starlings,* Sturnus vulgaris, *at an invasion front* [PDF 1.8 MB], report to the Australian Government Department of Agriculture, Fisheries and Forestry, Invasive Species Science Group.

Campbell S, Roberts EJ, Craemer R, Pacioni C, Rollins L and Woolnough AP (2016) 'Assessing the economic benefits of starling detection and control to Western Australia', *Australasian Journal of Environmental Management*, *23*(1):81–99. *

Gibb R, Browning E, Glover-Kapfer P and Jones KE (2019) 'Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring', *Methods in Ecology and Evolution*, 10(2):169–185.

Hafner S and Katz J (2018) ‘monitoR: Acoustic template detection in R. R package version 1.0.7’, Vermont Cooperative Fish and Wildlife Research Unit Spreadsheet and R Project.

Juanes F (2018) ‘Visual and acoustic sensors for early detection of biological invasions: Current uses and future potential’, *Journal for nature conservation*, 42:7–11.

Kirkpatrick W and Massam M (2008) *Risk assessment for Australia for Common Starling (*Sturnus vulgaris*)*, Department of Agriculture and Food, Western Australia.

Kolar CS and Lodge DM (2001) ‘Progress in invasion biology: predicting invaders’, *Trends in ecology & evolution*, 16(4):199–204.

Lamont TA, Chapuis L, Williams B, Dines S, Gridley T, Frainer G, Fearey J, Maulana PB, Prasetya ME, Jompa J, Smith DJ and Simpson SD (2022) ‘HydroMoth: Testing a prototype low-cost acoustic recorder for aquatic environments’, *Remote Sensing in Ecology and Conservation*, 8(3):362–378, doi:10.1002/rse2.249.

Larson ER, Graham BM, Achury R, Coon JJ, Daniels MK, Gambrell DK, Jonasen KL, King GD, LaRacuente N, Perrin-Stowe TI, Reed EM, Rice CJ, Ruzi SA, Thairu MW, Wilson JC and Suarez AV (2020) ‘From eDNA to citizen science: emerging tools for the early detection of invasive species’, *Frontiers in Ecology and the Environment*, 18(4):194–202, doi:10.1002/fee.2162.

Lehmann GUC, Frommolt KH, Lehmann AW and Riede K (2014) ‘Baseline data for automated acoustic monitoring of Orthoptera in a Mediterranean landscape, the Hymettos, Greece’, *Journal of Insect Conservation*, 18(5):909–925, doi:10.1007/s10841-014-9700-2.

Lodge DM, Williams S, MacIsaac HJ, Hayes KR, Leung B, Reichard S, Mack RN, Moyle PB, Smith M, Andow DA and Carlton JT (2006) ‘Biological invasions: recommendations for US policy and management’, *Ecological applications*, 16(6):2035–2054.

Marchal J, Fabianek F, Scott C, Corben C, Riggs D, Wilson P and Biserkov J (2021) ‘Bioacoustics: analyse audio recordings and automatically extract animal vocalizations’, R package version 0.2.5, https://CRAN.R-project.org/package=bioacoustics.

McDonald JI, Wellington CM, Coupland GT, Pedersen D, Kitchen B, Bridgwood SD, Hewitt M, Duggan R and Abdo DA (2020) ‘A united front against marine invaders: developing a cost-effective marine biosecurity surveillance partnership between government and industry’, *Journal of Applied Ecology*, 57(1):77–84, doi:10.1111/1365-2664.13557.

Muller BJ, Schuman MJ and Johnson SA (2020) ‘Efficacy of acoustic traps for cane toad (*Rhinella marina*) management in Florida’, *Florida Scientist*, 83(1):42–53.

Perrings C, Williamson M, Barbier EB, Delfino D, Dalmazzone S, Shogren J, Simmons P and Watkinson A (2002) ‘Biological invasion risks and the public good: an economic perspective’, *Conservation Ecology*, 6(1).

Ruppert KM, Kline RJ, Rahman MD (2019) ‘Past, present and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA’, *Global Ecology and Conservation* 17:e00547, doi:10.1016/j.gecco.2019.e00547.

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

Sousa-Lima RS, Norris TF, Oswald JN and Fernandes DP (2013) ‘A review and inventory of fixed autonomous recorders for passive acoustic monitoring of marine mammals’, *Aquatic Mammals,* 39(1):23-53, doi:10.1578/AM.39.2.2013.205.

Stringham OC, Toomes A, Kanishka AM, Mitchell L, Heinrich S, Ross JV and Cassey P (2021) ‘A guide to using the internet to monitor and quantify the wildlife trade’, *Conservation Biology*, 35(4):1130–1139, doi:10.1111/cobi.13675.

Thomas ML, Gunawardene N, Horton K, Williams A, O'Connor S, McKirdy S and van der Merwe J (2017) 'Many eyes on the ground: citizen science is an effective early detection tool for biosecurity', *Biological Invasions*, 19:2751–2765, doi:10.1007/s10530-017-1481-6.

Tingley R, Ward-Fear G, Schwarzkopf L, Greenlees MJ, Phillips BL, Brown G, Clulow S, Webb J, Capon R, Sheppard A, Strive T, Tizard M and Shine R (2017) 'New weapons in the Toad Toolkit: a review of methods to control and mitigate the biodiversity impacts of invasive cane toads (*Rhinella marina*)', *The Quarterly Review of Biology*, 92(2):123–149, doi:10.1086/692167.

Towsey M, Planitz B, Nantes A, Wimmer J, Roe P (2012) 'A toolbox for animal call recognition', *Bioacoustics*, 21(2):107–125, doi:10.1080/09524622.2011.648753.

Voigt CC, Russo D, Runkel V and Goerlitz HR (2021) 'Limitations of acoustic monitoring at wind turbines to evaluate fatality risk of bats', *Mammal Review*, 51(4): 559–570, doi:10.1111/mam.12248.

Walters CL, Freeman R, Collen A, Dietz C, Fenton MB, Jones G, Obrist MK, Puechmaille SJ, Satler T, Seimers BM, Parsons S and Jones KE (2012) 'A continental-scale tool for acoustic identification of European bats', *Journal of Applied Ecology*, 49(5):1064–1074, doi:10.1111/j.1365-2664.2012.02182.x.

APPENDIX 1. SPECIALISED ZOOLOGICAL DETAILED FINAL REPORT

Specialised Zoological

Development and performance evaluation of the starling sentinel acoustic detector

Prepared for **Dr Susan Campbell**
Department of Primary Industries and Regional Development, Western Australia

Prepared by **Dr Kyle Armstrong** and **Yuki Konishi**
Specialised Zoological ABN 92 265 437 422
Tel 0404 423 264
kyle.n.armstrong@gmail.com

Version **5 October 2022**
SZ project reference **SZ530**

This report should be included as an appendix in any larger submission to Government, and cited as:

Specialised Zoological (2022). Development and performance evaluation of the starling sentinel acoustic detector. Unpublished report by Specialised Zoological for Dr Susan Campbell, Department of Primary Industries and Regional Development, Western Australia, version 5 October 2022, project reference SZ530.

Table of contents

Summary

This final report[1] presents details of the development and testing of a one-dimensional Convolutional Neural Network (CNN) model trained to recognise buzz and whistle calls of a pest bird, the European (or common) starling *Sturnus vulgaris*. The model forms part of a bioacoustic recording and analysis pipeline conducted onboard an acoustic 'sentinel' planned for deployment in remote areas of southern Western Australia to detect and allow expedient response to starling incursions from South Australia. The output detection CNN model is suitable for broad-scale field deployment and further iterative updates.

Reference calls of starlings used to build the CNN model came predominantly from recordings at two main locations in Adelaide city, and five locations on the Fleurieu Peninsula south of Adelaide. The main source of calls and the site for field testing was at a residence at Cape Jervis where recorders could be placed close to where a resident flock of starlings settled regularly in nearby trees, drank from a garden pond, and bred in nest boxes.

A full recording and analysis pipeline was produced by Thomas Rowntree (Senior Research Engineer, Australian Institute of Machine Learning, University of Adelaide) that runs in a Docker container within a version of Linux. It is suitable for the chip architecture and fully functional on the chosen hardware device—a 64-bit Raspberry Pi Compute Module 4 on the Raspberry Pi IO board and a custom Analogue to Digital Converter from Lynxemi Pte Ltd to facilitate sound recordings. The code resources produced include a Training code routine to produce ('train') CNN models; a Runtime code routine to classify signals, with the CNN model embedded within a bioacoustic recording and signal processing pipeline running within a Docker container; and a Desktop Runtime code routine that allows classification of signals that have been recorded previously and archived on disk.

The CNN model was improved incrementally through various iterations by training it with starling calls (two types of buzz call, and a descending whistle), non-target signals from South Australia, and non-target signals from relevant soundscapes in Western Australia. The performance of two major iterations of the model (Model1, Model2) was assessed following the deployment of a prototype Raspberry Pi Compute Module 4 at the site at Cape Jervis that had resident starlings. Model performance was assessed in two different ways: as metrics of Accuracy, Precision and Recall; and as calculated rates of True Positives, False Negatives and False Positives derived from manual inspection and validation of identifications in the both the 5-second WAV snippets containing putative True Positives and those that were missed but available in archived 5-minute WAV files comprising a continuous recording of the test period. A guide to understanding how these performance measures are calculated is provided in a glossary and an appendix.

A significant improvement was observed when the validated and labelled output from Model1 was used to train Model2. The detection rate for starling calls increased overall, but particularly for whistles, and the rate of False Positive detections was reduced significantly.

The metrics of Accuracy, Precision and Recall available from the training process were all relatively high for both Model1 and Model2 were relatively high (Model1: 97%, 97%, 98%,

[1] In support of Centre for Invasive Species Solutions Project No. P01-T-003 'Automated Detection: Triggering Smarter, Faster, Better Response to Incursion'

respectively; Model2: 95.5%, 95.6%, 95.4%, respectively). However, recalculation of Precision and Recall from the validated output of the two field tests provided a good estimate of performance in natural recording situations. For a user-chosen threshold probability value of 0.96, Model2 misses around 36.4% of starling calls with an associated probability of a match to the model of 0.5 or greater, though the signals that are missed are often of low amplitude or are obscured by other signals. It typically produces a list of putative detections where c. 3.3% are False Positive detections. This compromise between levels of True Positive rate and False Positive rate is considered practical for initial deployment of the model at two sites in Western Australia.

Further training was undertaken to produce Model3, which has had both the input of validated and labelled examples from Field Test 2 of Model2, plus the 8,329 False Positive examples derived from applying Model2 to recordings made across the South Coast of Western Australia in 2011 on Song Meter (Wildlife Acoustics) recorders. The performance of this Model3 has not yet been evaluated, but it is available for broad scale deployment, testing, and further refinement.

In summary, Model3 available for broader deployment has been trained on at least 6,969 starling calls (plus many more as part of the development of Model1), plus 3,493 False Positive signals from South Australia and 8,329 False Positive signals from Western Australia.

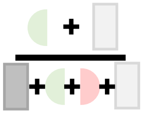
In conclusion, the significant contribution of thousands of examples of starling buzz and whistle calls, and an iterative process of retraining models with labelled examples of True and False Positives recorded on field tests has produced a deep learning CNN model that is suitable for deployment as part of a Passive Acoustic Surveillance in southern Western Australia. It can be improved further in the future based on accumulated resources and experience resulting from a broader deployment beyond the current two test sites in Western Australia. The coding resources are also suitable for the development of models and Passive Acoustic Surveillance systems for other vocalising species, and a Model0 for the Asian Black-spined Toad was produced as a proof of concept.

Glossary

There are terms within this report that have specific meanings, as defined in this section.

Accuracy—The fraction of predictions the model got correct (number of correct predictions over total number of samples). Calculated by (see **Appendix 1** for explanation of the illustration):

$$\textit{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$



Call—Any signal that was produced by a species of bird, or a signal not from a bird that is mistaken for a bird call. Here the word is used in a general sense, and can include only some or all of the syllables and elements.

CNN—The one-dimensional Convolutional Neural Network method chosen ultimately for application in this project. It is one-dimensional because the input is a sound wave, and during analysis the ‘kernel’ (a filter used to extract features) slides along the signal in one dimension (along the time axis; or in the frequency domain: Cheuk et al. 2020). A sound wave displayed as time series data is actually in two dimensions (time by amplitude), but it is the behaviour of the kernel that gives the CNN its dimensionality. Two-dimensional CNNs are common because image classification has many applications, and it requires the kernel to slide over a flat image (e.g. a digital photograph) in two dimensions (width by height).

Detection model (‘the model’, ‘the CNN model’, ‘the algorithm’)—A trained CNN classification system that is optimised for making identifications of starling calls. This term is used generally, and the term “Model#” (where # is a number) is used to specify a particular iteration of the CNN model that was applied and tested.

Detection probability—A value between 0 and 1, or in percent, representing the probability of detection for target call types based on fitting of the CNN model.

Detection threshold probability value (‘threshold’)—A probability value above which calls are attributed to the target species, as the basis for a species identification.

False Negative (FN)—A missed detection. Calls from the target species that were not detected by the model, given a chosen detection threshold probability value. During this development phase of the project, there are two types of false negative. ‘FN1’ is anything with a detection probability of 0.5 or more and below an arbitrary threshold value, which are represented in 5-second WAV snippets that are output for investigator validation. These were used in model performance evaluations and were the basis of estimates of Accuracy, Precision and Recall. ‘FN2’ is anything that was missed by the model and had a detection probability of less than 0.5. Examples of these calls are only observable in the 5-minute WAV calls recorded, and do not contribute to estimates of Accuracy, Precision and Recall. Rather, the rate of missed detection in 5-minute WAVs is compiled and reported separately. See also the Confusion Matrix in **Appendix 1**.

False Positive (FP)—The incorrect attribution of an identification of 'starling' to a call. The source of the call is either another bird, or else the 'call' is a portion of the background noise having characteristics that fit the model with high probability. The rate of False Positive detections of target calls is the proportion of non-target signals with a probability score above the set detection threshold probability value. False Positives are confirmed by manual inspection of the call in a spectrogram by a person experienced with examining starling calls. See also the Confusion Matrix in **Appendix 1**.

Model development—Developing a model or algorithm using reference calls from the target species, other signals that an investigator thinks might be, or observes to be, attributed incorrectly to the target species, and long background recordings that do not contain calls from the target species. The iterative process of model development in a 'train' coding routine uses three types of dataset:

Training dataset: The sample of data used to fit the model. This is typically most of the available reference call recordings.

Validation dataset: The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while the machine learning engineer fine-tunes model hyperparameters. The model processes this data as part of an evaluation, but it does not learn from it. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration. The validation dataset is typically a small proportion of the reference call recordings (e.g., 15%; also called the development dataset).

Test dataset: A separate sample of data used to provide an unbiased evaluation of a final model fit on the training dataset. It is only used once a model is completely trained. The test dataset contains selected examples that encompasses the variation that is likely to be faced by the model in a real deployment.

Precision—The proportion of identified calls that are from starlings (total correct predictions over the total number of predictions). The inverse is the False Positive rate. Calculated by (see **Appendix 1** for explanation of the illustration):

$$\textit{Precision} = \frac{TP}{TP+FP}$$

Recall—The proportion of starling calls that are identified (total positive predictions over all positive samples). Calculated by (see **Appendix 1** for explanation of the illustration):

$$\textit{Recall} = \frac{TP}{TP+FN}$$

Single Board Computer (SBC)—a complete computer built on a single circuit board, with microprocessor/s, pre-determined amount of RAM, input/output (I/O) and other features required of a functional computer, but with no expansion slots for peripherals. One of the most widely known SBCs is the Raspberry Pi range that was developed for education, and is now popular with hobbyists and 'makers'. Some, such as the NVIDIA Jetson Nano are built for specialised applications that require greater use of GPUs. A wide range of other expansion boards ('hardware attached on top' or 'HATs') with specialised functions are available. For example, in this project, digital audio conversion of sound from a microphone is handled by a custom expansion board. The functionality of SBCs can also be split into two separate boards, one that handles I/O and HATs, and another containing CPU and RAM that functions as a 'compute module'.

True Negative (TN)—The correct rejection of an identification of 'starling' to a call. The rate of True Negative detections is the proportion of non-target signals with a probability score below the set detection threshold probability value. True Negatives are confirmed by manual inspection of the call in a spectrogram by a person experienced with examining starling calls. In practice, this number is extremely large because there are many signals that are the basis for small peaks in probability that are well below the threshold. See also the Confusion Matrix in **Appendix 1**.

True Positive (TP)—The correct attribution of an identification of 'starling' to a call. The rate of True Positive detections of target signals is the proportion of target signals with a probability score above the set detection threshold probability value. True Positives are confirmed by manual inspection of the call in a spectrogram by a person experienced with examining starling calls. See also the Confusion Matrix in **Appendix 1**.

Target signal—in this project, one of the two primary call types, or a syllable, from the target species that the model was trained on.

> Whistle—a target call type; a long descending whistle of c. 500 milliseconds beginning anywhere between c. 6 and 9 kHz, and descending to c. 2.5 kHz (example in **Figure 1**). Sometimes a second whistle follows the first with no inter-call period, with each possibly emitted by two separate individuals.
>
> Buzz sequence—a target call type, of which there are two main variants; any of the buzz-like calls that span the frequency range between c. 6 and 10 kHz (example in **Figure 1**).

Target species—in this project the European starling (common starling) *Sturnus vulgaris.*

1 Introduction

1.1 Background

Surveillance of remote and large expanses of Australia for wildlife species of interest requires either enormous sustained and intensive field survey effort, or a technological solution that can reduce the effort and increase the effectiveness of species detection. In environmental impact studies, targets of interest are typically threatened species listed under State, Territory or Commonwealth legislation. The movements of invasive species are also of interest in a biosecurity context. The European (or common) starling *Sturnus vulgaris* is one such introduced bird species that is a Declared pest animal under the Western Australian *Biosecurity and Agriculture Management Act 2007* and associated Regulations, and presents particular challenges for control in this state where it has not yet established.

Starlings are reasonably common and widespread across most of eastern Australia (Atlas of Living Australia[2]), and most incursions into Western Australia (WA) are from individuals that disperse frequently across the Nullarbor from South Australia (SA). Such incursions represent a significant threat to WA's agricultural, public amenity and biodiversity assets (Campbell et al. 2015). Starlings cause damage to high-value fruit crops such as cherries and grapes, affect intensive cattle, pig and poultry production by consuming and spoiling feed, spread weeds through their droppings and regurgitations, are a nuisance to residents through their noise, soiling of garden trees and outdoor furniture, vehicles, and footpaths by droppings, cause damage and disease risk from nesting in buildings, and compete aggressively with native species for nest sites (review in Campbell et al. 2015; DPIRD 2018). If starlings are left to establish in WA unchecked, the costs are likely to be in the tens of millions of dollars, so it is more economical to effect control (Campbell et al. 2015).

Incursions of individuals and small flocks of starlings have been controlled by the WA Government since 1971. The control programme aims to prevent the establishment and subsequent increase in starling numbers within WA, mainly by halting the persistent incursions from SA. Numerous actions have been conducted as part of the control programme, including the widespread installation of wooden nest boxes in swamps that may have helped concentrate breeding and thus assisting control effort (Campbell et al. 2012a), and trapping using caged live starlings as a lure (Campbell et al. 2012b). The cost and ethical basis of the latter has been questioned. Further, while starling numbers continue to be suppressed, Campbell et al. (2015) stated that further improvements in the efficiency of starling detection and control are required.

[2] https://www.ala.org.au/

1.2 Aims and scope

The present project seeks to build an acoustics-based surveillance ('sentinel') system in areas west of the Nullarbor Plain where incursions of starlings occur to further improve the efficiency with which these birds can be detected across large areas of mostly uninhabited land.

The effort is supported by the Commonwealth Government's Centre for Invasive Species Solutions Project No. P01-T-003 'Automated Detection: Triggering Smarter, Faster, Better Response to Incursion'.

The scope of the involvement of Specialised Zoological in the project extended to the following objectives:

1. Develop a reference call library for starlings that will support the development of a machine learning based bioacoustic detection system.

2. Develop, with the collaborative technical assistance of Thomas Rowntree (Senior Research Engineer, Australian Institute of Machine learning, University of Adelaide), an acoustics-based starling call detection and identification software system that is suitable for operation on a Single Board Computer (SBC) forming the core component of an integrated Passive Acoustic Surveillance (PAS) hardware solution for starling detection.

3. Refine this system to optimise the rate of true detections and minimise the rate of incorrect identifications through field testing of machine learning models loaded onto PAS units deployed in South Australia.

4. Using similar coding tools, develop an automated acoustics-based detection and identification system for at least one additional invasive species, the Asian Black-spined Toad.

To guide the development of the software component of the PAS, a 'Scenario' was described that outlined the needs of the project and anticipated constraints (**Appendix 2**). The present document builds on the progress reports of Specialised Zoological (2020, 2022).

2 Choosing an analysis method for a Passive Acoustic Surveillance bioacoustic recorder

There are many tools available for the analysis of bioacoustic signals. Commercial software has found widespread use over many years but, more recently, open-source packages in computing languages such as Python and [R] have seen much development—both for sound analysis and signal processing in general, and for the application of many 'machine learning' methods. When choosing amongst these tools for the current application, there were numerous considerations, but there are two constraints imposed by the hardware that limit the range of tools available:

> The first is whether these packages will run on the type of processing chips that are on the more capable (higher specification) Single Board Computers (SBCs). Most of these small computing devices have a 'system on a chip' (SoC) with an integrated ARM-compatible central processing unit (CPU). Some software packages are not compatible with this kind of 'chip architecture' (AArch32, AArch64), at least without further development.
>
> The second is that some classification methods are computationally intensive, which might place too great a demand on either the processor (CPU or GPU) or the power supply. Thus, the starling detection software needed to be composed of tools that will operate both efficiently and effectively on a relatively small computing device that can run without mains power.

Since most SBCs can run a version of Debian-based Linux, finding a way to run Windows-based software popular for analysing bird calls was not possible or desirable, especially because such software lacks the flexibility for integration into a larger automated acoustic recording, signal processing and classification system. Examples of such software include RAVEN PRO (Cornell Laboratory of Ornithology, Cornell University [3]) and SOUNDID [4]. Given that many biologists now use the [R] statistical computing language for a very broad range of analyses, there are also popular packages for bioacoustic analysis. In the present context, the real value of considering these software programmes is to understand what methods can been incorporated into a full recording and analysis system. While the methods might be different, they have similar overall components:

1. pre-processing of sound—filtering, de-noising, background subtraction;
2. segmentation of signals—defining putative calls or syllables in an audio stream;
3. feature extraction—deriving a set of measurements or output from filters; and
4. classification—allocation of signals to a predefined set of groups or clusters.

An early step in the workflow facilitated by RAVEN PRO is the segmentation of signals from background with a customised Band Limited Energy Detector (BLED, or a 'blob' detector). Similar threshold-based extraction of signals from the background soundscape is a necessary first step before feature extraction in SOUNDID (Jinnai 2018). Segregation in ANABAT INSIGHT

[3] https://ravensoundsoftware.com/
[4] http://www.soundid.net/SoundID/Software_Home.html

software for bat call analysis uses a zero-crossings threshold (Titley Scientific 2020). In [R] language packages such as WARBLER 'blobs' are segmented with a function that uses a threshold together with other optional constraints (Araya-Salas and Smith-Vidaurre 2017; using *auto_detec()*). Segmentation can also be conducted using a set of 'templates' of a target signal, and calculating a correlation score between these templates and each defined time bin in a recording. This is implemented in the [R] package MONITOR (Hafner and Katz 2018). The correlation scores across the recording can be considered as a series of peaks, and signals corresponding to peaks above a threshold score can be allocated to the target signal category.

Feature extraction in RAVEN PRO is a user's choice of up to 82 measurements derived from an area in the time-frequency domain defined by the BLED rectangular window. Compiled measurements can be exported for classification elsewhere. A similar system is available in the [R] package BIOACOUSTICS (Marchal et al. 2021). In SOUNDID, the signal of interest is decomposed with the Linear Predictive Coding (LPC) method to obtain a representation of the spectral envelope (rather than the more commonly used fast-Fourier Transform, FFT). The shape of this envelope is compared with representations from reference signals with the acoustically-appropriate Geometric Distance metric (Jinnai et al. 2009, 2010, 2012) as a way of classifying the signal of interest to the categories of target or non-target. ANABAT INSIGHT relies on measurements taken from a trend of the echolocation chirp shape in the time-frequency domain, and uses filters to subset ranges of these measurements within a Decision Tree process to allocate species names. In WARBLER, fundamental or dominant frequency contours as a time series can be applied to segmented signals, and sets of measurements can be taken; or else summary variables derived from sets of Mel Frequency Cepstral Coefficients (MFCCs) are saved as input for a range of classification options. Common machine learning classification methods used in bioacoustics include Discriminant Function Analysis, Gaussian Mixed Models, Hidden Markov Models, K-Nearest Neighbours, Random Forest Walks and Support Vector Machines (e.g., Agranat 2009, 2012, 2016; reviews in Priyadarshani et al. 2018 and Das et al. 2020; Clink and Klinck 2021; Marchal et al. 2021).

The methods above have generally been developed for specific applications and then cross-applied to other tasks (e.g., MFCC and LPC from speech recognition), which, it has been argued, has contributed to stagnation in the field of audio event recognition (Humphrey et al. 2013). The best performance appears to be gained with newer 'deep learning' approaches, rather than 'hand-crafted' methods. Deep learning is a class of machine learning algorithms that are based on artificial neural networks and use multiple layers to progressively extract higher-level features from raw input. In deep learning, a simple, general transform is applied to the input data, and the network then learns a feature representation and performs classification.

A demonstration of the level of improvement that deep learning methods have brought to bioacoustics was evident in the BirdCLEF challenge, a code competition posed by the Cornell Lab of Ornithology via Kaggle[5] to automate the acoustic identification of birds in soundscape recordings. When deep learning was applied, mean average precision (MAP) scores rose from a maximum of 0.45 using to 0.69 (Kiskin et al. 2020). The application of deep learning for bird call recognition and identification has since taken two main pathways: one that classifies based on audio streams (e.g., following wavelet transformation or Short-time Fourier Transformation

[5] https://www.kaggle.com/c/birdclef-2021/overview

(=sliding window Discrete Fourier Transform); Kiskin et al. 2020; Cheuk et al. 2020), and another where the audio stream is first converted into a spectrogram image showing a representation of signals in the time-frequency domain. Perhaps the best known and largest project based on images of spectrograms is the Cornell Lab of Ornithology's BirdNET[6] project that uses a model architecture derived from the family of residual networks (ResNets) (Kahl et al. 2021). It allows the annotation and identification of bird species from North America and Europe, and its performance for audio event detection is equivalent to that for object detection, with a MAP score of 0.79 for single-species recordings.

Deep learning has been applied previously, not only to bird call identification in general, but specifically for starlings in the context of pest management. Dolezel et al. (2019) describe an evolved system that takes sound recordings as input, and uses a Convolutional Neural Network (CNN) as a decision-making tool based on spectrographic images of the sound recordings. CNN is a class of artificial neural network most commonly applied to analyse images. Dolezel et al. (2019) recorded monaural sound at a sampling rate 44.1 kHz, filtered for noise and silence, normalised for energy amplitude, and divided the audio stream into 3-second segments. These were converted to a three-dimensional spectrogram, represented as normalised frequency (x) by samples (y) by power over frequency (z). The performance of five CNN architectures was evaluated on a recording of 50 minutes in total, where 6 minutes was of starling calls, and on spectrograms of different resolution (number of pixels). The best results (as measured by Accuracy, Precision and Recall; *see Glossary*) demonstrated good results with the LeNet CNN architecture and 150 x 150 pixel spectrographic image size.

In their field implementation, their goal was to localise the source of starlings within a vineyard and then deter them from feeding. Their system consisted of a central processing device receiving input from multiple microphone sources spread throughout the vineyard, and an acoustics-based deterrent (or 'scarer') (Dolezel et al. 2015, 2016). The deployment requirements and goals of the work by Dolezel et al. (2015, 2016, 2019) are different to those for the present starlings detection system. They were able to take advantage of mains power, and needed a device that could process multiple inputs (the NVIDIA Jetson "AI platform for autonomous machines"[7] combined with the Intel Movidius Visual Processing Unit Neural Compute Stick[8]). They also sought to identify the starling from a wide range of call types, rather than just a few of the more common call types. Despite these differences to the present application in Western Australia, their experience informed both hardware and software choice for the deployment situation of a PAS unit trained to detect starling calls in remote areas of Australia.

---

[6] https://birdnet.cornell.edu/
[7] https://www.nvidia.com/en-au/autonomous-machines/embedded-systems/
[8] https://www.intel.com/content/www/us/en/developer/tools/neural-compute-stick/overview.html

3 Bird calls

3.1 Calls of starlings

An acoustics-based PAS system for starlings in the Western Australian environment needs to be able to detect call types that migrating starlings might emit when in groups or as lone individuals. Starlings have an extensive repertoire of calls, and the detection system needs to be trained to recognise one or more common call types that could be made by starlings dispersing outside their established range.

The vocal repertoire of starlings in their original natural range has been the subject of numerous behavioural studies that attribute function (e.g., to song: Eens et al. 1989, 1990, 1991, 1993). Male starlings emit complex songs that can last over a minute and comprise over 90 short song types that extend between 0.16 to 2.4 seconds each (Eens et al. 1989). The elements within song types can be described as whistles, clicks, rattles, squeaks and screeches (citations in Eens et al. 1989).

Various websites that provide bird species information profiles have also described call types and the situations in which calls are made. The Cornell Lab of Ornithology[9] mentions warbles, harsh trills, chatter, metallic chips, scream-like calls, bill clacks, and smooth liquid sounds. A purr-like call is given when a bird takes flight, and a rattle is given when an individual joins a flock on the ground. Females also sing, and starlings also imitate other bird species (e.g., Hindmarsh 1984). The species profile on the website of BirdLife Australia[10] mentions that birds do not call while in the spectacular murmuration flocks but are incredibly vocal once they all alight at a roost.

There is an abundance of example recordings on the global online repositories for bird call recordings at Xeno Canto[11] (1,389 foreground recordings as of 2022-02-14) and the Macaulay Library[12] (1,816 recordings). Inspecting these examples when tasked with choosing a few that are thought likely to be emitted by dispersing individuals outside their known range as either individuals or flocks presents as a formidable challenge. In this project, we needed to predict which of these many call types might be made outside their natural range without having the benefit of actual recordings from a starling in exactly the situation where it needs to be detected. Previous work on an earlier version of a deep learning model trained to recognise starlings was based on buzz and whistle call types (Campbell et al. 2013).

[9] https://www.allaboutbirds.org/guide/European_Starling/sounds
[10] https://birdlife.org.au/bird-profile/common-starling
[11] https://xeno-canto.org/
[12] https://search.macaulaylibrary.org

3.2 Making reference call recordings

Only a small proportion of the variety of call types that can be made by starlings could be expected to be useful in our classification system. Thus, the approach to target call selection involved consideration of several practical actions and solutions:

1. Make as many recordings of starlings as possible in the Adelaide area and surrounds to understand what types of calls are heard commonly in areas of South Australia that might be part of the source of dispersing individuals in Western Australia.
2. Record starlings when in flocks, especially when roosting or on the ground.
3. Compare candidate calls of starlings with calls of relevant Western Australian species available on Xeno Canto to identify possible sources of False Positive identifications.
4. Accept that the detection model built will have a degree of limitation in detecting starlings because of the variation that the model is built upon, but allow for future iterative improvement over time as awareness of relevant call type variation is discovered, including during this development phase.
5. Rely on a chosen target call as an initial ‘probe’ to efficiently discover the presence of other call types made within a few seconds of the target.

Acoustic recorders (Bioacoustic Audio Recorders, BARs; Frontier Labs, Brisbane; 48 kHz sampling frequency) were placed at two residential addresses in suburbs of Adelaide city (Unley, Mitchell Park), one in Mt Barker, a property in McLaren Flat, two properties in Clayton Bay, and one property in Cape Jervis. Recording sessions at each location were at least one week, and the Adelaide city and Cape Jervis recording periods totalled several months. From these recordings, and general ‘birdwatching’ activity within Adelaide city where recordings were made from starlings under direct observation (using a Røde NTG-2 shotgun microphone connected to a Zoom H1n digital audio recorder; 48 kHz sampling frequency), it was apparent that ‘whistle’ and ‘buzz’ type calls were relatively common (**Figure 1**). Some call types could not be attributed with confidence to starlings, for example some higher frequency descending whistles that might have derived instead from the blackbird. Other call types were attributed unambiguously to starlings under direct observation, but did not appear to be common outside one location (e.g., the ‘wow’ call; **Figure 1**).

It was especially difficult to obtain quality recordings from animals in flocks. While starlings could be observed easily throughout the Fleurieu Peninsula, birds in flocks were timid and difficult to approach. The alternative to approaching flocks was to place recorders in areas where starlings congregated regularly. This also proved challenging, but a small airstrip for model planes in Clayton Bay regularly had starlings feeding on an open area of ground. Despite the informed placement of two BAR recorders, the resulting signal quality of the recordings made over two unattended recording sessions was low. Thus, calls of birds in flocks do not yet contribute to the detection system.

Most quality recordings of starling calls came from suburbs in Adelaide city and a residence in Cape Jervis, which were also the sites where testing of the detection model and recording hardware was undertaken (development of Model1 and Model2; Field Test 1 of Model1; Field Test 2 of Model2). The model was trained on three main call types: the ‘whistle’ and two types of ‘buzz’ sequence (**Figure 1**).

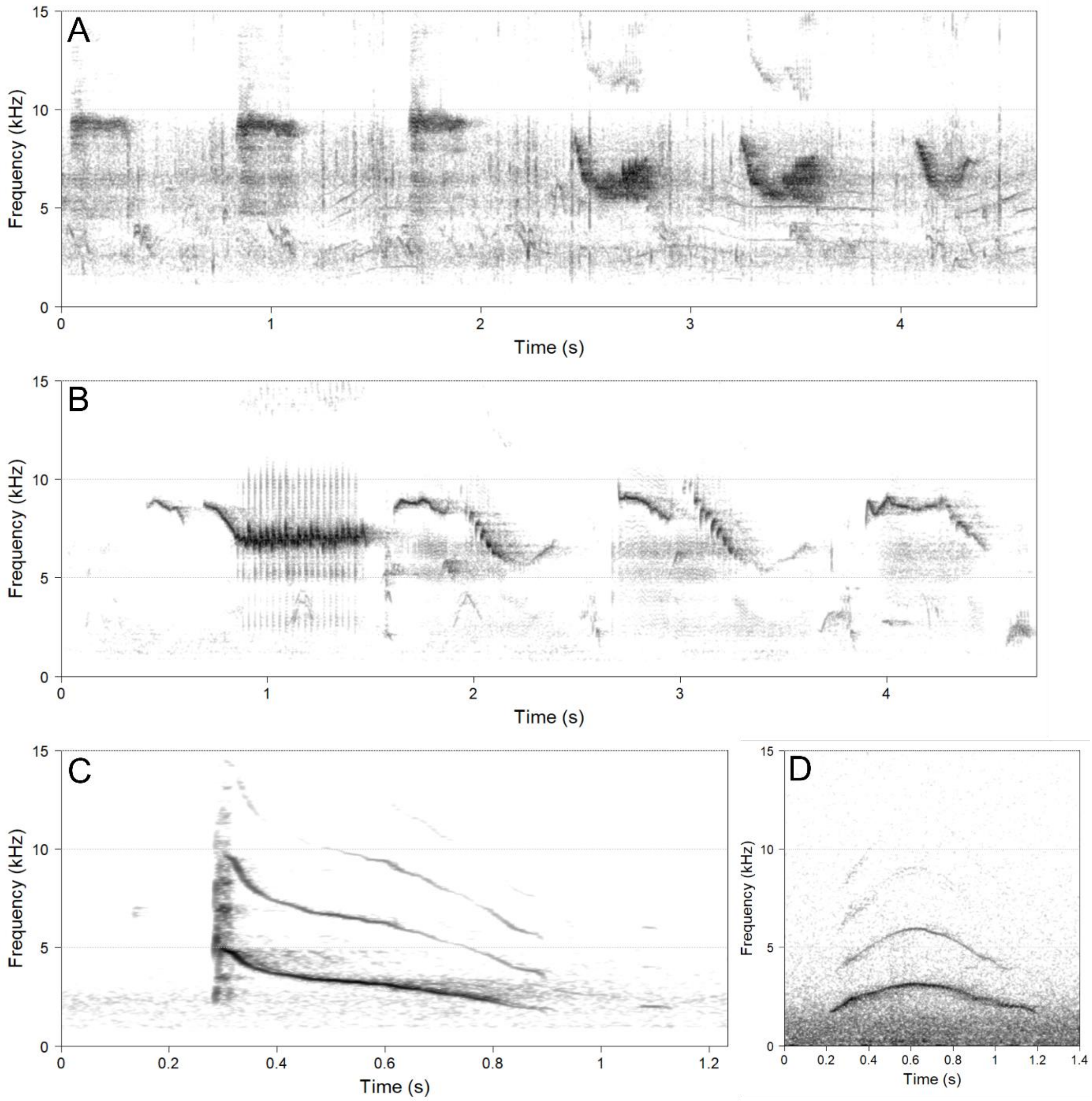


Figure 1. Examples of the target call types from starlings that were used to train the CNN model (**A**: buzz sequence type 1 containing high frequency and u-shaped buzz syllables; **B**: buzz sequence type 2; **C**: multi-harmonic whistle c. 600 milliseconds in duration; **D**: a ‘wow’ call with a duration of c. 1 second, which was not used for training because it was only detected in the Unley recording area).

3.3 Birds of the Western Australian deployment area

To identify whether the whistle and buzz call types that appeared to be most common in Adelaide and the Fleurieu Peninsula had a strong similarity in shape and frequency characteristics to species found in the Western Australian deployment areas, a list of bird species was derived from two sources: the Atlas of Living Australia[13] (a general area bounded by latitudes –30.1 and –34.7 S and longitudes 118.9 and 129.0 E; downloaded 2020-05-20; **Appendix 3**). Calls of these species were then examined on the Xeno Canto database. While it is difficult to predict what an artificial neural network model will mistakenly attribute an identification of starling to, at least the more obvious candidates could be identified. The list in **Appendix 3** has not been checked for current taxonomic accuracy, and detailed acoustic comparisons have not been made to identify potential sources of False Positive.

The intention of making this compilation was to inform generally about the potential for False Positives to occur before field testing. Most bird species in Western Australia, that is 10 of the 173 species listed, appear not to have a perceived likelihood of producing a call type that will result in a mistaken attribution.

The actual signal types that appear as False Positives during testing of Model2 in Field Test 2 and subsequent deployments will only be discovered by application of the model. False Positives may also be discovered by the application of Model2 in a modified runtime script to background recordings provided by Susan Campbell from previous deployments of Wildlife Acoustics Song Meter SM2 recorders in the Western Australian deployment area (Campbell et al. 2013). Having a list of candidate species might help to identify the source of these False Positives. Previously, calls of the Australian Raven and New Holland Honeyeater were common sources of False Positive detections in the previous effort for applying machine learning to starling call detection (Campbell et al. 2013).

[13] https://www.ala.org.au/

4 Early considerations and prototypes

The development of the bioacoustic recording and processing system progressed through non-systematic testing with different analysis approaches and hardware options. Most attention was given to the development of the one-dimensional CNN system (see *Glossary*) written by Thomas Rowntree, but other activities facilitated by Specialised Zoological informed some minor aspects of this.

1. A full recording pipeline was developed through student projects at the University of Adelaide, which supported five Master of Engineering (Electronic) candidates in the School of Electrical and Electronic Engineering in 2019 and 2020. They developed systems for the recording and classification of bat and bird calls using Python and bash scripts, feature extraction with MFCCs, tested filtering and noise removal options and used K-Nearest Neighbour to allocate target detections (e.g., Methra 2019; Xia 2020). A key challenge for the students was to embed their bioacoustic system into an operating system capable of running on an SBC. Two main hardware solutions were successfully implemented, in versions of both Raspbian OS[14] and Balena OS[15] on both the Raspberry Pi 4 Model B[16] and the balenaFin 1.1[17] containing a Raspberry Pi Compute Module 3[18]. The system was ultimately deployed within a Docker[19] container that could be updated and accessed via the balenaCloud[20] system that supports deployment of identical software across a fleet of devices.

2. An NVIDIA Jetson Nano Developer Kit (V3)[21] and a Raspberry Pi 4 Model B were also trialled in the early stage of the project. The Jetson Nano device is a GPU-accelerated computing platform running NVIDIA CUDA-X (a collection of 40+ acceleration libraries) in a version of the Ubuntu Linux operating system. In this project, it was used as an exemplar SBC by Thomas Rowntree to test early versions of the recording and processing system that incorporated the CNN model runtime code (powered mains supply and using a USB-connected conferencing microphone). The Jetson Nano has a maximum reported power consumption of 5 Watts, and while optimised for AI, was considered to be excessively powerful and inefficient in power consumption for the application. When the Raspberry Pi Compute Module 4[22] and IO board[23] became available, testing was transferred to a prototype supplied by the project collaborators David Barnard at DKB Solutions Pty Ltd and David Lucas at Lynxemi Pte Ltd.

14 https://www.raspberrypi.com/software/operating-systems/
15 https://www.balena.io/os/
16 https://www.raspberrypi.com/products/raspberry-pi-4-model-b/
17 https://www.balena.io/fin/
18 https://www.raspberrypi.com/products/compute-module-3-plus/
19 https://www.docker.com/
20 https://www.balena.io/cloud/
21 https://developer.nvidia.com/embedded/jetson-nano-developer-kit
22 https://www.raspberrypi.com/products/compute-module-4/?variant=raspberry-pi-cm4001000
23 https://www.raspberrypi.com/products/compute-module-4-io-board/

3. Consideration was also given to microphone zones of detection. With the goal of maximising the detection of starlings within the vicinity of a PAS system, a relevant consideration became whether to make mono or stereo omni-directional recordings. Having two channels increases a zone of detection around a microphone, but doubles the amount of processing used. A solution offered to increase the zone of detection but restrict recording to a single channel was the Compression Zone Microphone (CZM; Riverforks Research Corporation, USA[24]; described by patent US 6,681,023 B1). This device allows radially directional sound to be collected, and it comprises a parabolic-shaped zone in cross-section within a waveguide formed by two circular disks positioned one above the other; and with a microphone positioned in the middle of lower disk to pick up sound from the central position. It effectively collects sound from all cardinal directions, and its size and shape are optimised for the frequency range of ambient sound, including bird calls. Some initial testing was undertaken, and consideration may be given to incorporating it into the ongoing DPIRD starlings eradication programme once the core device has proven functional.

[24] http://www.riverforks.com/

5 Overview of the coding resources produced to create and apply the CNN model

Ultimately the deep learning CNN model approach appeared to have the greatest potential for bird call classification embedded in an acoustic recording and analysis pipeline on a Raspberry Pi computer.

Three main resources were produced for the CNN analysis pipeline:

Training code—a routine to train a new CNN model on a desktop computer using starling calls, non-target signals of various types and background soundscape without the calls of starlings within it.

Runtime code—Python and shell scripts to create a Docker environment, record an audio stream, run the CNN model over the audio stream, and then save 5-second sound file snippets containing putative target signals. Five-minute files over complete 24-hour daily recording periods were also saved to allow performance evaluation in field tests.

Desktop Runtime code—a subset of the Runtime code that allows running the CNN model over existing large recording datasets fed into a desktop computer from an external hard drive.

6 Steps in the training and development of the CNN model

6.1 The process of developing the CNN model in 15 steps

The CNN model was trained and then its performance was evaluated in an iterative process that began with the development using signals from Adelaide city, through to more intensive field testing at a site at Cape Jervis with an early prototype of the chosen Raspberry Pi Compute Module 4 hardware, and finally to a full deployment for testing at two sites in Western Australia.

The following is a summary of the entire process:

1. **Collect reference calls from starlings** and background soundscape recordings from Adelaide city and the Fleurieu Peninsula (see section *3.2 Making reference call recordings*).
2. **Write the Training code routine to create a CNN model** and make iterative improvements of it.
3. **Write the Runtime code routine** to implement a multi-step recording and classification pipeline suitable for deployment on the chosen SBC.
4. **Embed this Runtime routine in a Docker container** as part of a larger on-board acoustic recording and analysis system suitable for an SBC.
5. **Create a CNN model using reference calls of starlings**, False Positives and background recordings ('Model0'; see *Glossary* for 'model development'). The set of reference call recordings used ('Batch1') comprised 155 whistle type calls and 21 buzz type calls (see *Glossary*; see section *3.2 Making reference call recordings*) from starlings, and 157 WAV files with anticipated False Positives (from the New Holland Honeyeater, Common Blackbird and other signals). Model0 was used to test the runtime code on the Jetson Nano device.
6. **Further training of Model0** to produce a version ready for field testing ('Model1'). A second set of reference call recordings ('Batch2') comprising 291 buzz-type calls, 217 whistle-type calls, and 6 hours of background collected mainly at Cape Jervis was used to produce Model1. This model was loaded onto a Raspberry Pi Compute Module 4 and IO board (hardware 'Prototype1'; also referred to as the 'CM4') provided by David Lucas and then run continuously for 53 days in the backyard of a residence at Mitchell Park in Adelaide. The device made over 100,000 detections, which were saved to a hard drive as 5-second WAV snippets. Most of these detections were False Positives. A significant proportion of these were hand-labelled and added back to the training data, and iterated further on the design of Model1 to reduce the False Positive rate. This model was loaded then back onto the CM4 to collect more detections. Numerous iterations were undertaken in this way to further refine Model1.
7. **Undertake a performance evaluation of Model1** using measures of Accuracy, Precision and Recall (see *Glossary*). A test set of recordings was maintained to help validate models during training. This balanced test set was compiled from a random excerpt from the full set of training data, which included 210 positive samples and 210

negative samples (birds, cars, aeroplanes, dogs). Given that the 'real world' is likely to be extremely unbalanced, with most sound being composed of non-target signals, the reported Precision will be lower in the real world due to the increased opportunity for the model to encounter False Positives. Thus, 'real world tests' were required to give a better understanding of the usefulness of the model.

8. **Undertake 'Field Test 1' on Model1 at Cape Jervis** in a natural environment over 26 days of recordings in October–November 2021 with the system loaded onto the CM4 hardware prototype (see section *8.1 Site characteristics of the Cape Jervis study site*).

9. **Assess the performance of Model1** on the field test by compiling the results in the framework of a Confusion Matrix. The Field Test 1 recordings were examined in Adobe Audition version 22.0. For 15 of these daily recordings (2021-10-24 to 2021-11-07), the number of 5-second WAV snippets containing at least one example of a buzz or whistle was scored as one, and files without starling calls were scored as zero. When the source signal responsible for the probability score was obvious, this was noted. The final matrix also included a column with the probability score, as derived from the WAV filename. The outcomes are reported in section *9 Results of the performance evaluations.*

10. **Further train the model to produce 'Model2'** using a total of 6,414 5-second WAV snippets of manually validated starling calls, False Positives, and missed detections (probability < 0.5; False Negative type 2, FN2).

11. **Undertake 'Field Test 2' at Cape Jervis on Model2**, and then test its performance in the same way as Model1 by recalculating rates in the framework of a Confusion Matrix. The field test was conducted over 21 days in March–April 2022. The outcome of a performance evaluation of Model2 is reported in section *9 Results of the performance evaluations.*

12. **Deploy Model2 at two sites in Western Australia** as part of a full deployment test with all hardware and communications components [*not reported on in detail here*].

13. **Use the Desktop Runtime coding routine and Model2** to process the 2011 recordings made with Song Meters in Western Australia and provide a resource of False Positive signal types that would be present in the Western Australian soundscape.

14. **Further train the model to produce 'Model3'** using compiled WAV snippet files from Field Test 2 and those generated using the Desktop Runtime code. These signals were used to retrain the CNN model specifically to reduce the False Positives requiring manual validation.

15. **Deploy Model3 on a broad scale**, and further assess the performance of this model to produce future iterations as required. Model3 was not assessed for Accuracy, Precision and Recall, and has not yet been assessed in a field test.

6.2 Summary of the models produced

A summary of the models produced for each iteration includes:

Model0—the first model produced with legacy and newly-collected reference calls from Adelaide city. It was produced with 176 calls of starlings and 157 examples of False Positive signals. Initial testing and improvement with calls from Adelaide city.

Model1—an improved model based on field recordings from Cape Jervis, Clayton Bay and Adelaide city. It was produced initially with 508 calls of starlings and 6 hours of background recordings containing only False Positive signals; and then further trained across many iterations with thousands more examples of starling calls from a single location in Adelaide city. It was evaluated in Field Test 1 at Cape Jervis using the CM4 hardware prototype.

Model2—an improved model that incorporated validated output (labelled True Positives and False Positives) from Field Test1. It was produced with 3,811 calls of starlings with probabilities >0.50, 991 calls of starlings with probabilities <0.50 (FN2) and 2,603 examples of False Positive signals that were the output from Field Test 1 of Model1. It was evaluated in Field Test 2 at Cape Jervis using the CM4 hardware prototype. It has also been deployed in WA at two tower sites, but a performance evaluation in this environment has not yet been undertaken.

Model3—added further training using 1,483 starling calls and 733 False Positives derived from Field Test 2 on Model2, as well as 8,329 False Positives derived from running the Desktop Runtime code over legacy recordings (those from Song Meters deployed in Western Australia in 2011).

Model4+ —future model(s) to be retrained to reject False Positives accumulated in the broadscale deployment in WA.

In summary, Model3 available for broader deployment has been trained on at least 6,969 starling calls (plus many more as part of the development of Model1), plus 3,493 False Positive signals from South Australia and 8,329 False Positive signals from Western Australia.

7 Steps in the on-board acoustic recording and analysis system

The later models of the deep learning classification system were embedded into a larger system that begins with sound recording, and ends at the point where accumulated sound files with putative target signals are available on the device for investigator validation. A schematic representation of the core bioacoustic recording and analysis system designed for the starling PAS system describes six steps (below; also **Figure 2**). This Runtime routine was placed in a private repository on Github and made available for further development by project collaborators responsible for hardware the development component.

1. **Sound detection**—Sound is taken from the microphone and Analogue to Digital Converter and is transferred into computer memory, at a sampling rate of 48 kHz.
2. **Feature extraction and CNN Model classification**—A 5-minute segment of this audio stream is processed in real time using MFCC for feature extraction, and a vector of probability scores reflecting the degree of match of the model to signals across this audio stream is produced.
3. **Segmentation**—Signals that match the CNN model with a probability score above the chosen threshold are identified as starlings. The threshold score is set by the user. For performance testing, all signals with a probability score of 0.50 and above were saved. For field tests currently being undertaken at two sites in Western Australia a probability score of 0.96 was used to achieve the best compromise between the rate of False Positive signal rejection and incorrect rejection of the target signal.
4. **Saving for validation**—Segmented signals are saved at the midpoint within a buffered length of audio stream to a total of five seconds. These '5-second snippets' are saved as 16 bit 48 kHz WAV files, with filenames having the following components separated by underscores:
   - DateYYYYMMDD_
   - DayoftheWeek_
   - StartTimeHHMMSS_
   - CountOfDetectionsPer5MinFile_
   - 'time'_TimeWithin5MinFileSeconds_
   - 'prob'_ProbabilityScore.wav
5. **Saving for further development**—Each 5-minute segment of audio stream analysed in real time is saved as a date-time-stamped 16 bit 48 kHz WAV file, for 'archival purposes'. This allowed performance assessment of signals of the various CNN model versions, and will ultimately be excluded in the most mature pipeline to be deployed widely. In the first hardware prototype, all 288 5-minute periods within each 24-hour period were saved, so no sunrise or sunset-based start and stop to the recordings was implemented.
6. **Logging**—A log file is produced of all recordings, and WAV recordings accumulate in the following example directory structure:

   /detections/20211007_Thu/recordings
   /detections/20211007_Thu/snippets

   These are stored ready for compression and transmission to the Detect-it system web app developed by DKB Solutions that allows browsing of snippets and other functionality.

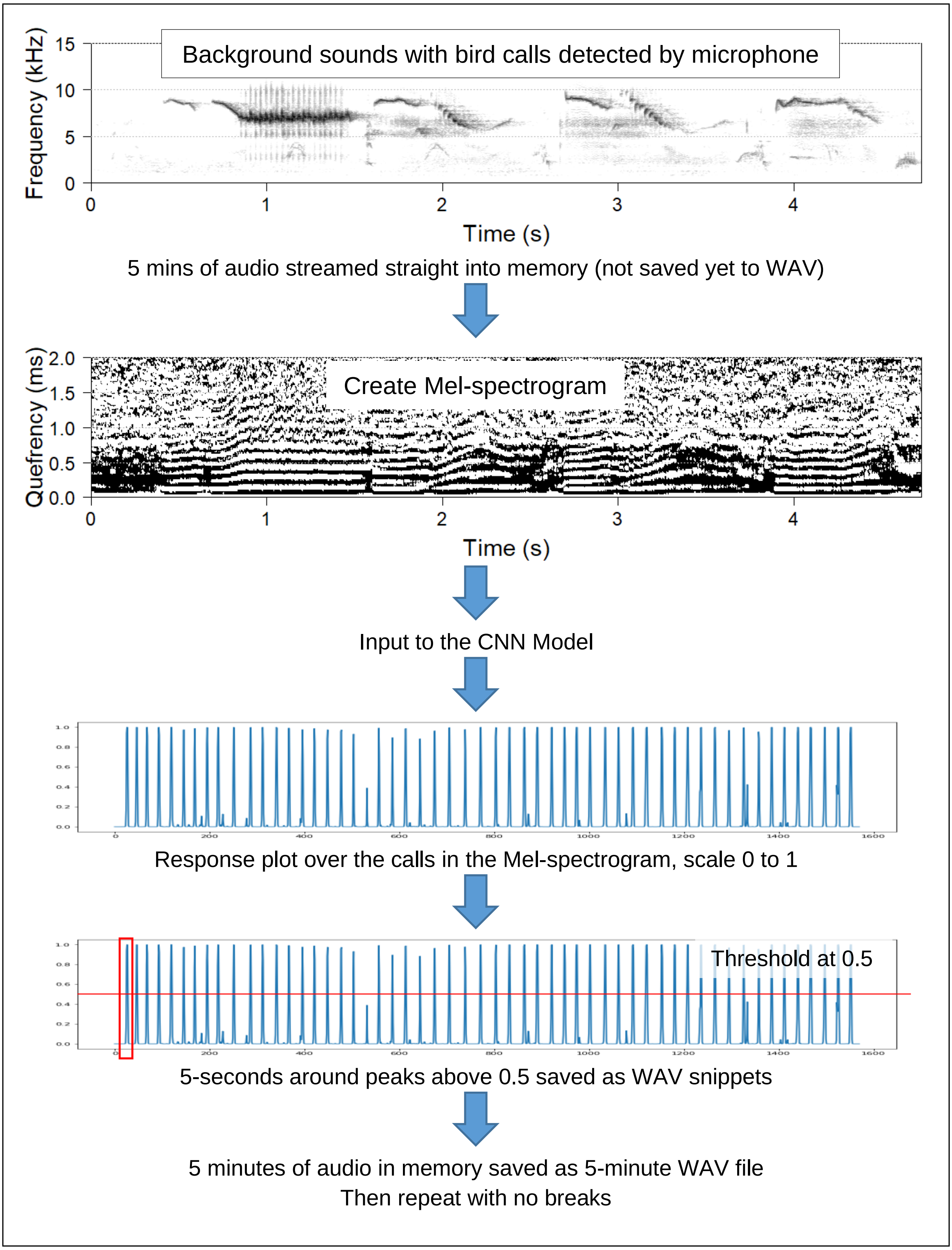


Figure 2. Illustration of what happens on the prototype recording device.

8 The Cape Jervis study site

8.1 Site characteristics of the Cape Jervis study site

Extensive recordings to collect reference calls were made across a total of nine months with two BAR units between November 2019 and June 2021; and across two field tests with the prototype Raspberry Pi CM4 recorder in October–November 2021 and March–April 2022.

The equipment was placed in the garden of a property on Sorata Street, Cape Jervis, at the tip of the Fleurieu Peninsula, South Australia (**Figure 3**). This property is within 100 metres of the ocean shoreline and is fronted by revegetated heathland. The rear of the property, where the device was located, contains a small garden of native plants and has numerous characteristics that made it a good choice for the field tests:

- Within a rural near-coastal vegetation community and containing a bird assemblage that are likely to be relatively similar while still being relatively accessible from Adelaide city;
- Regularly visited by starlings, appear as residents;
- Contains nest boxes that starlings were using during the test;
- Has overhead powerlines upon which starlings sit, either singly or in groups;
- Larger trees in adjacent properties where starlings often congregate are within auditory detection range of an observer;
- Relatively shielded from the wind;
- Presence of a small water fountain used as a water source by starlings;
- Relatively unvisited by people that would cause starlings to vacate or avoid the garden;

Figure 3. **Top**: A frequent sight of a flock of starlings that sits on powerlines above the native garden at the Cape Jervis residence; **Bottom**: A view of the CM4 recorder overlooking the garden, with a funnel enclosing the microphone (a Little Raven perches nearby).

8.2 Deployment and performance of the Raspberry Pi CM4 hardware prototype on field tests at Cape Jervis

For both field tests, the Raspberry Pi CM4 prototype was set in an open area of the garden of a residence at Cape Jervis. A 240 Volt power supply was provided from a connection in an adjacent shed (extension power cord was routed in a protective enclosure). The device and all electrical power connectors were placed together in a plastic lunchbox, which was then placed into a waterproof bag. The microphone on the supplied extension cable was attached to a pole c. 1 metre from the ground. It was enclosed in the funnel-like top of a soft drink container to protect it from rain, and directed horizontally at an area where starling calls were expected to be relatively frequent. While this funnel would have narrowed the zone of sound reception, it has no significant influence on the outcome of the type of analyses performed herein. The device was observed to be functioning normally at the end of both field tests.

The field tests encompassed the following periods:

- Field Test 1: a total of 26 days in the period 2021-10-07 – 2021-11-14.

- Field Test 2: a total of 21 days in the period 2022-03-14 – 2022-04-03.

The quality of signals on the last day of recording was equivalent to that on the first day, as assessed by casual inspection of WAV files in a spectrogram in Adobe Audition version 22.0 software, suggesting no damage had occurred to the microphone from weather conditions.

Large portions of some days had ‘choppy’ signal with innumerable short periods of signal loss, which was eventually traced to either poor wire connections in the microphone or a buffering issue (**Figure 4**). Remarkably however, the model still matched with a relatively high probability to many signals with periods of silence breaking into starling calls.

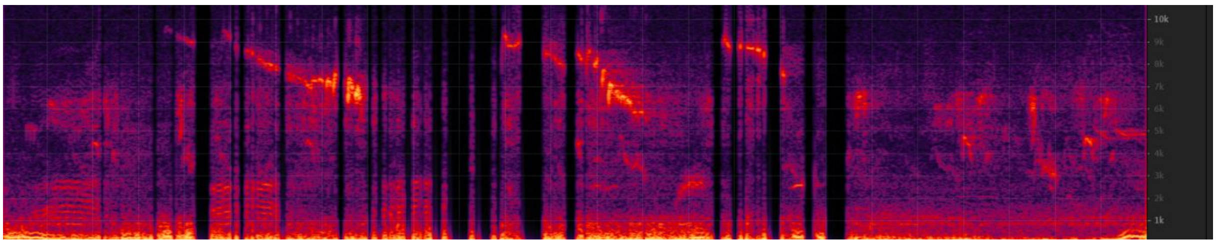


Figure 4. An example of a moderately ‘choppy’ recording segment in a 5-second WAV snippet, with 0.99 probability of a match to starling call.

8.3 Bird assemblage at Cape Jervis

Being able to identify all bird species in the field test study site was important for understanding the contributing sources to false positive signals. Given the broad similarity of the habitat ('near-coastal rural southern Australia') at the study site, there will be some degree of overlap in this bird assemblage with that of the planned deployment areas in Western Australia.

A total of 26 terrestrial bird species (i.e., excluding waterbirds) were noted on the recordings or by casual observation during visits to the field test site at Cape Jervis (**Table 1**). This is undoubtedly an incomplete list given that many other species with a range that includes the Fleurieu Peninsula are likely to visit on occasion. A total of 14 of these species are found in the Western Australian deployment areas and surrounds (**Appendix 3**), so they have effectively been a relevant part of the performance testing of the detection model.

Table 1. List of bird species detected in recordings or by casual observation during visits at the field test site at Cape Jervis (asterisk denotes an introduced species; WA: Y indicates present in the deployment area in Western Australia).

Common name	*Genus species*	WA
Nankeen Kestrel	*Falco cenchroides*	Y
Common Bronzewing	*Phaps chalcoptera*	Y
Galah	*Eolophus roseicapilla*	Y
Rainbow Lorikeet	*Trichoglossus haematodus*	Y
Musk Lorikeet	*Glossopsitta concinna*	
Purple-crowned Lorikeet	*Parvipsitta porphyrocephala*	Y
Adelaide Rosella	*Platycercus elegans subadelaidae*	
Crimson Rosella	*Platycercus elegans elegans*	
Superb Fairy-Wren	*Malurus cyaneus*	
White-browed Scrubwren	*Sericornis frontalis*	
Little Wattlebird	*Anthochaera chrysoptera*	
Red Wattlebird	*Anthochaera carunculata*	Y
Noisy Miner	*Manorina melanocephala*	
Singing Honeyeater	*Gavicalis virescens*	Y
Yellow-plumed Honeyeater	*Ptilotula ornata*	Y
New Holland Honeyeater	*Phylidonyris novaehollandiae*	Y
Grey Fantail	*Rhipidura fuliginosa*	Y
Grey Shrike-Thrush	*Colluricincla harmonica*	Y
Magpie-Lark	*Grallina cyanoleuca*	Y
Australian Magpie	*Gymnorhina tibicen*	Y
Australian Raven/Little Raven	*Corvus coronoides/C. mellori*	Y
Eurasian Skylark	*Alauda arvensis*	
Red-browed Finch	*Neochmia temporalis*	
*House Sparrow	*Passer domesticus*	
*Common Blackbird	*Turdus merula*	
*European Starling	*Sturnus vulgaris*	

9 Results of the performance evaluations

9.1 Accuracy, Precision and Recall

The initial performance evaluation was made with the test sets of signals. For Model1, a random and balanced selection of 210 starling calls and 210 False Positive signals comprising sounds from other bird species, cars, aeroplanes and dogs was used. For Model2, a total of 1,720 non-training WAV files was used, with equal proportions of positive and negative signals.

The values of Accuracy, Precision and Recall were all relatively high at the conclusion of training (**Table 2**; see *Glossary* and **Appendix 1** for an explanation of these metrics). Accuracy shows the percentage of correct predictions (both correctly identified starling calls and correctly rejected False Positives) out of all predictions made. Precision is the percentage of correctly identified starling calls out of all signals attributed to starlings. Recall is the percentage of all starling calls detected in the set of recordings. The decrease in these values from Model1 to Model2 reflects the greater amount of variation in signal types that have been included in the training process.

Table 2. Percentages of Accuracy, Precision and Recall for Model1 and Model2.

Model	Accuracy	Precision	Recall
Model1	97	97	98
Model2	95.5	95.7	95.4

Plots of Precision and Recall against threshold probability scores between 0.5 to 1.0 show a significant change in slope at 97–98% (**Figure 5**; as derived from values taken from TensorBoard[25], for Model1 only), indicating that the ‘best’ Precision and Recall is obtained for probabilities above 97%.

The Precision versus Recall curve shows the trade-off between precision and recall for different threshold values above 50% (0.5 probability). In an ideal model, values of Precision would stay high (close to 100) for increasing values of recall (thus maximising the area under the curve). The curve represents the trade-off between False Positives and False Negatives. A perfect classifier would maintain the maximum Precision for all thresholds—i.e. it will not make any mistakes. A very poor classification model will give maximum recovery of all targets but also give a high output of False Positives. In the outcome from Model1, the relationship between the two is equivalent indicating that as the threshold is increased, the trade-off is equivalent between finding starling calls and rejecting non-target signals. Thus, the model needs to both improve its recognition of target signals when they are present, and improve rejection of non-target signals.

Further plotting to include Model2, and based on the data compiled from field tests, shows a marked improvement when Model1 is retrained with additional examples (**Figure 6**). For Model2, values of Precision and Recall are higher at all probabilities, and the Precision-Recall plot has an overall shape that is closer to that expected for a high performing model. Both Precision and Recall are higher, indicating that the model is finding more starling calls, and rejecting more False Positives.

25 https://www.tensorflow.org/tensorboard

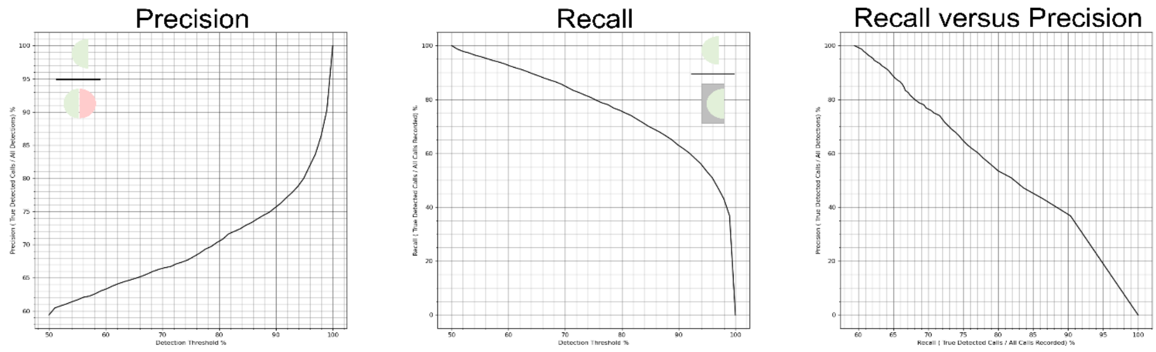


Figure 5. Plots of Precision and Recall for Model1, for probability values above 0.5 (50%), as derived from values taken from TensorBoard used to monitor model development during training.

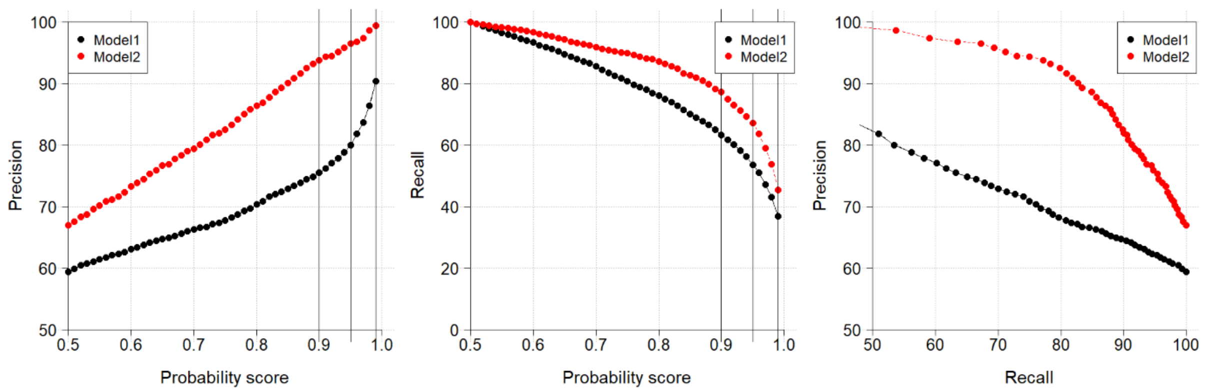


Figure 6. Plots of Precision and Recall for Model1 and Model2 derived from data calculated from the two field tests, for probability values above 0.5 (50%). Vertical lines are at probabilities of 0.90, 0.95 and 0.99.

9.2 Outcomes of field tests

Calculating and optimising values of Accuracy, Precision and Recall are a standard way of assessing how well deep learning models perform. However, simpler metrics can also be easier to understand for non-specialists that would like to know, in the present case, how many starling calls are being missed, and whether the number of incorrect identifications can be minimised to a level that is manageable for regular, manual validation. Therefore, presenting components of the Confusion Matrix (**Appendix 1**; see also the *Glossary*), which are used to calculate Accuracy, Precision and Recall, can make the outcomes of performance testing more generally comprehensible. The outcomes of the two field tests are therefore summarised in terms of these Confusion Matrix components.

Key to understanding the performance evaluation metrics of the models tested at Cape Jervis are the following:

1. All signals with a probability of 0.5 and above participated in deriving the various performance metrics, with some of these signals being produced by starlings and the remainder coming from other sources.
2. The threshold (see *Glossary*) is not set at one value, but instead the change in performance is summarised at different values of this threshold.
3. Signals with a probability score to a match of the model below 0.5 are effectively discarded and not used in all except one analysis to derive a rate of 'FN2' (see *Glossary*); and some of these discarded signals are starling calls.

The recordings made as part of field tests were scored as '1' for WAVs containing starling calls, and WAVs without starling calls were scored as '0'. When the source of a non-target signal responsible for the probability score was obvious, this was noted. The final matrix also included a column with the probability score, as derived from the WAV filename.

An [R] script was written to derive the numbers of True Positive, False Negative and False Positive from this matrix comprising outcomes from 6,414 5-second snippet files for Model1, and 2,216 snippets for Model2.

A plot of the distribution of probability scores showed that there was a conspicuous difference in the frequency of probability scores between buzz calls and whistles for Model1 (**Figure 7**). Most probability scores for the buzz type calls have a probability score of 0.99. For whistles, there is an almost even distribution of probability scores between 0.50 and 0.90, and only a relatively small increase after 0.90. This difference in performance for the two call types was highly statistically significant (Welch Two Sample t-test: $t = 20.27$, $df = 2562.5$, p-value = 2.2e-16; **Table 3**).

When Model2 tested in Field Test 2, there was a significantly greater number of whistle calls at higher probability values than for buzzes (Welch Two Sample t-test: $t = -3.16$, $df = 1481$, p-value = 0.0016; **Table 3**; **Figure 7**), indicating that training had resulted in a large improvement in the detection of whistles.

Table 3. Summary of statistics for the probability scores of two target call types (Mean ± Standard Error, Range, Standard Deviation, Median).

	buzz	whistle
Model1	91.69 ± 0.25	82.26 ± 0.39
	50 – 99	50 – 99
	12.46	14.71
	99	86
Model2	91.25 ± 0.43	93.12 ± 0.41
	50 – 99	50 – 99
	12.06	10.81
	97	99

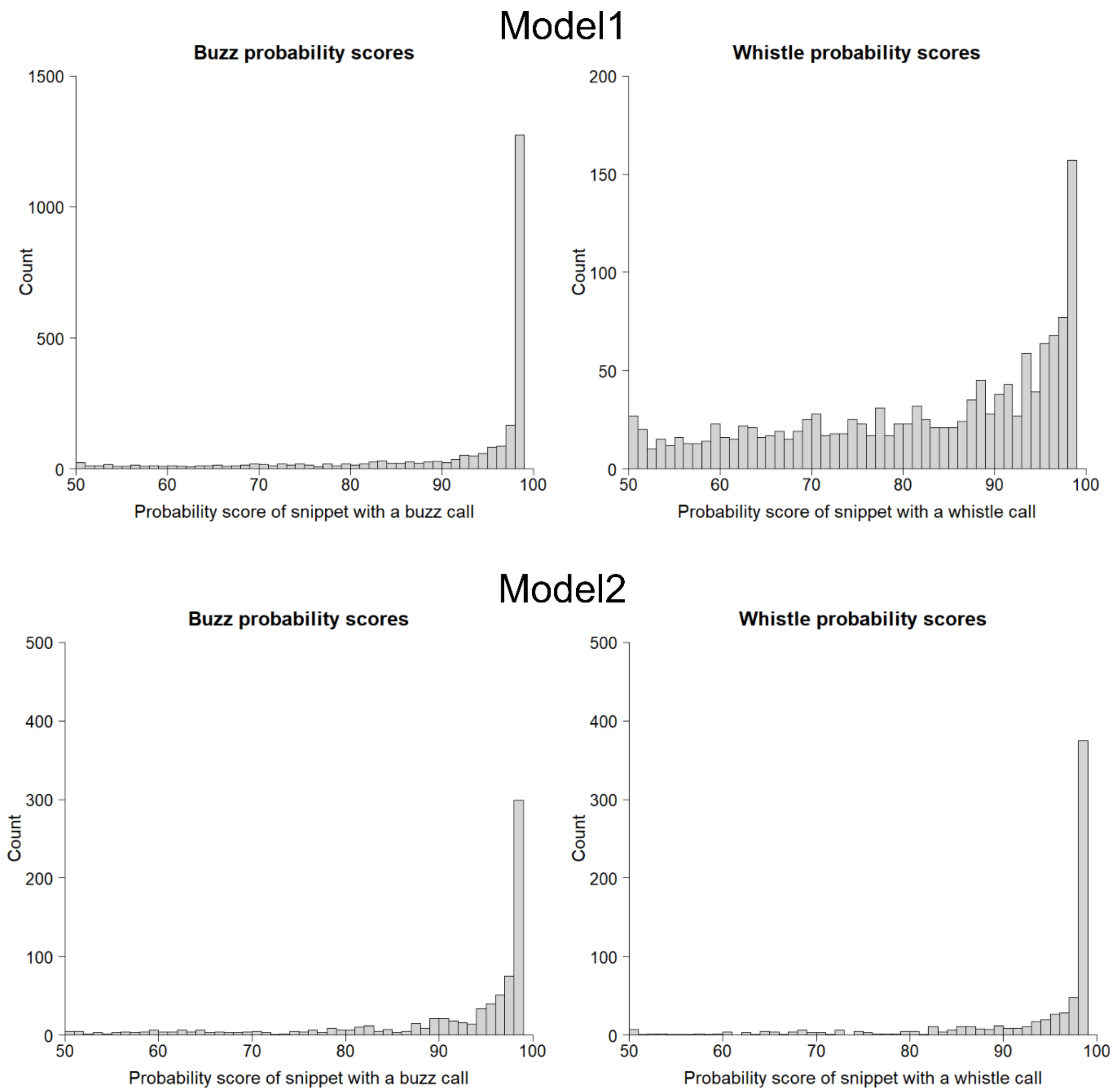


Figure 7. Frequency distribution of probability scores associated with starling buzz and whistle calls, showing a marked improvement in the detection of starlings from whistles with the development of Model2.

Model improvement is also clearly evident when cumulative rates of True Positive and False Positive are plotted against probability scores for both models (**Figure 8**). For the True Positive plot, this a ‘perfect’ rate of detection whereby all examples are recovered at the 0.50 probability threshold. As the threshold is raised, the number of detections gradually decreases. Ideally, the rate of True Positive should approach 100% for all probability values, indicating that a maximum number of starling calls will be detected at the highest probability values. For the False Positive plot, this shows how the proportion of False Positive examples in the dataset is reduced with increasing probability threshold value—i.e., non-starling signals are more likely to have lower probability values assigned to them by the model. Thus, higher probability values have less chance of returning an incorrect identification.

For Model1, there was more strongly declining percentage of True Positives with increasing probability for whistles, illustrating that the detection of whistles was relatively poor compared to the detection of buzz calls. For both call types combined, the rate of False Positive detections was relatively high, with an inflection point around 0.95 indicative of a potentially useful threshold point.

There is a clear improvement in three aspects of performance for Model2. First, more whistle calls are being detected with a higher probability threshold. Second, there is an overall improvement for both starling call types in terms of the proportion of calls being detected with a higher probability threshold. Third, the number of False Positive identifications has reduced significantly overall, and to the point where even relatively low probability threshold values will return minimal numbers of incorrect identifications (e.g., 6.6% of identifications will be incorrect for a probability threshold of 0.90 for Model2, versus 32.4% for Model1; **Table 4**).

A summary of True Positive, False Negative (the inverse of TP), and False Positive values for probability threshold values of 0.90 and above is presented in **Table 4**. From inspection of these values, the improvement in the detection of whistles, and the rejection of False Positives are the most conspicuous changes with the development of Model2. A threshold of 0.96 has been chosen for the field test in Western Australia. While this will miss an estimated 36.4% of starling calls (buzzes and whistles combined), it is likely to give a False Positive rate of only 3.3%. This is important since a higher False Positive rate will result in too many snippets requiring validation when the system is deployed more broadly with many more units.

Key to understanding the significance of the False Negative rate of 36.4% is a consideration of the quality of calls and the effective detection range of the recording device. During inspections of the snippets as part of the performance evaluation, it was noted that many of the lower probability detections of the buzz call were from lower quality recordings. In some of these, the calls were overlain by other signals in the same frequency band, or else they were of relatively low amplitude, and in others the recording was chopped by brief periods of silence that interrupted the call. It is relatively straightforward for a human investigator to distinguish starling calls that are very faint, or that overlap with other signals. The real challenge for identifying starlings will therefore most likely come from signals that are of low amplitude relative to background levels (signals that are emitted off-axis, or towards the limit of the detection range of the microphone), or interrupted in some way. However, given that Model2 recognises a broad range of variation for each of the call types that it was trained on, the expectation is that if birds move closer to the microphone and keep calling, then detection is likely to result if just one good quality example can be recorded.

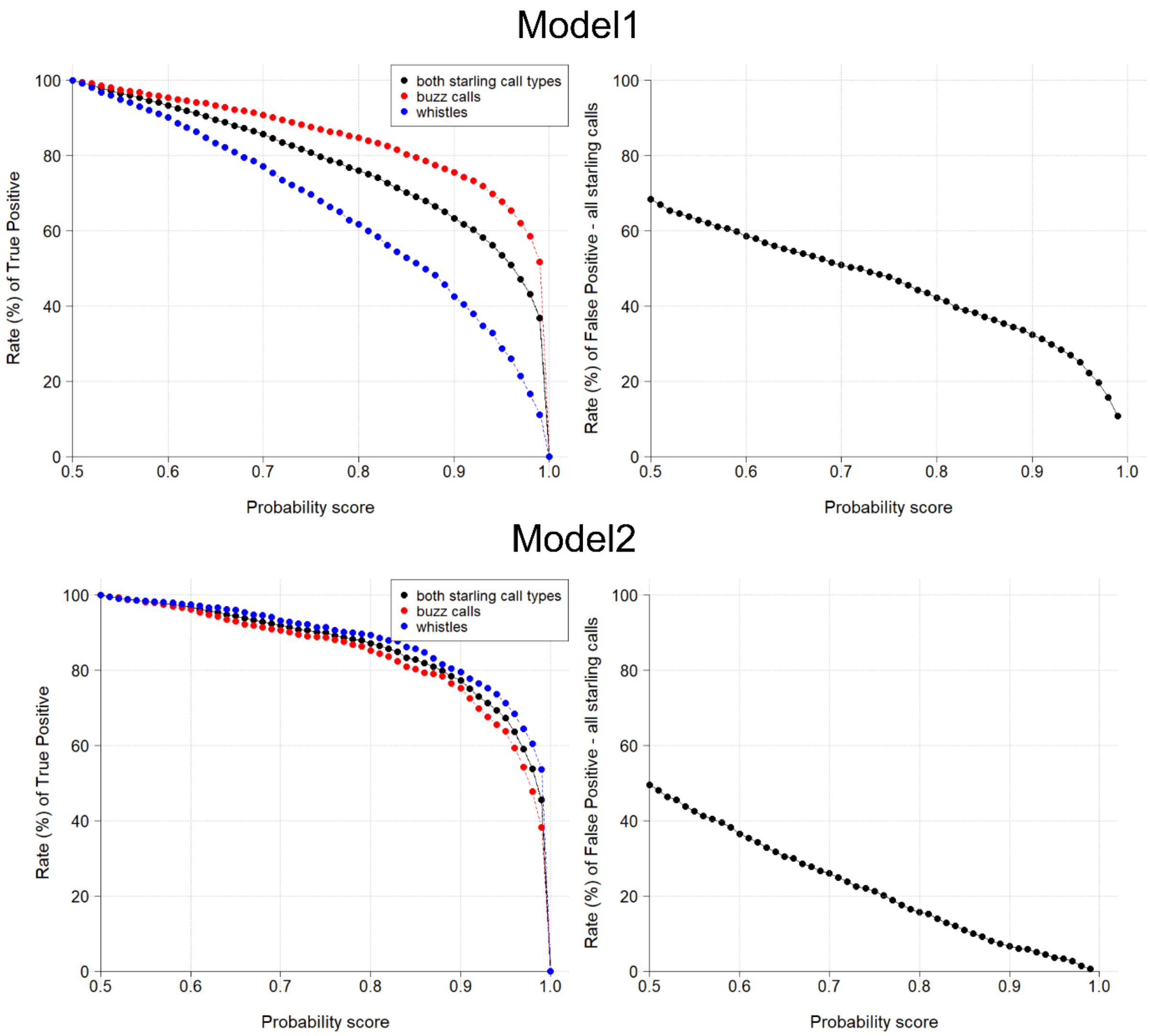


Figure 8. Rates of True Positive and False Positive detections of starling calls for the two models, across probability values from 0.5 to 1.0.

Table 4. Summary of True Positive, False Negative and False Positive rates (as percentages) at threshold scores 0.90 – 0.99 for Model1 and Model2. Orange highlight indicates the rates for a threshold of 0.96, which has been chosen for the field tests of two units in Western Australia.

		Model1			**Model2**	
Threshold value	**Buzz**	**Whistle**	**Both**	**Buzz**	**Whistle**	**Both**
TP—True Positives (rate of correct identifications)						
0.90 and above	75.4	42.5	63.2	75.2	79.4	77.2
0.91 and above	74.2	40.5	61.8	72.5	77.7	75.0
0.92 and above	73.3	37.8	60.2	69.9	76.4	73.0
0.93 and above	71.9	34.8	58.2	67.6	75.1	71.1
0.94 and above	69.8	32.9	56.2	65.5	73.6	69.3
0.95 and above	67.7	28.7	53.5	63.7	71.1	67.2
0.96 and above	65.4	25.9	51.0	59.4	68.3	63.6
0.97 and above	62.0	21.4	47.2	54.3	64.4	59.1
0.98 and above	58.5	16.6	43.1	47.8	60.4	53.7
0.99 and above	51.7	11.1	36.8	38.2	53.6	45.4
FN—False Negatives (rate of incorrect rejections)						
0.90 and above	24.6	57.5	36.8	24.8	20.6	22.8
0.91 and above	25.8	59.5	38.2	27.5	22.3	25.0
0.92 and above	26.7	62.2	39.8	30.1	23.6	27.0
0.93 and above	28.1	65.2	41.8	32.4	24.9	28.9
0.94 and above	30.2	67.1	43.8	34.5	26.4	30.7
0.95 and above	32.3	71.3	46.5	36.3	28.9	32.8
0.96 and above	34.6	74.1	49.0	40.6	31.7	36.4
0.97 and above	38.0	78.6	52.8	45.7	35.6	40.9
0.98 and above	41.5	83.4	56.9	52.2	39.6	46.3
0.99 and above	48.3	88.9	63.2	61.8	46.4	54.6
FP—False Positives (rate of incorrect identifications)						
0.90 and above	—	—	32.4	—	—	6.6
0.91 and above	—	—	31.2	—	—	5.9
0.92 and above	—	—	29.7	—	—	5.8
0.93 and above	—	—	28.4	—	—	5.1
0.94 and above	—	—	26.9	—	—	4.4
0.95 and above	—	—	25.0	—	—	3.6
0.96 and above	—	—	22.2	—	—	3.3
0.97 and above	—	—	19.6	—	—	2.7
0.98 and above	—	—	15.8	—	—	1.4
0.99 and above	—	—	10.7	—	—	0.6

9.3 Sources of False Positive detections

The possible source of False Positive detections in 6,414 5-second WAV snippets resulting from the deployment of Model1 in Field Test 1 was scored for all files that did not contain starling calls. It was not usually possible to decide which of the many signals had provided a fit to the model, but in some cases the source was obvious because it was the dominant or only signal present in the centre of the snippet.

The two most frequent sources of False Positives were from the Grey Fantail and the Common Skylark (**Table 5**). These species could occur in the deployment area in Western Australia, so it was considered important to train Model2 to reject these calls. Upon inspection of their call types in a spectrogram, they were observed to have a very different structure and should be separable from the calls of starlings in the validation process. However, as the goal of development was to reduce the number of False Positives contributing to the task of validation, it was ensured that there were numerous examples of these species amongst the 4,270 False Positive snippets used for training Model2.

The contribution of House Sparrows to the overall rate of False Positive was probably underestimated, but this species is absent from the Western Australian deployment areas and does not represent a significant issue for the model. It was also well-represented in the set of False Positive snippets.

Table 5. Source of various False Positive detections.

Source	No. 5-second snippets	Percent
Australian Raven/Little Raven	19	0.4
Common Blackbird	2	0.0
Grey Fantail	255	6.0
House Sparrow	57	1.3
New Holland Honeyeater	11	0.3
Common Skylark	298	7.0
unknown	3,628	85.0
Total 5-second snippets with FPs	**4,270**	

9.4 Estimating the prevalence of FN2: False Negatives with probabilities below 0.5

Further consideration was given to starling calls contained in the 5-minute WAV files that had not been detected—i.e., they had not met the 0.50 probability threshold for detection. These calls may have been missed because they were of relatively low amplitude, they overlapped with other signals in the soundscape, or else they represented variation that the model had not been trained to recognise. In this report, this proportion of False Negatives has been designated 'FN2' (represented by the blue square areas in **Appendix 1**). Model2 was explicitly trained to further recognise a broader range of variation in starling calls, and these were extracted from the 5-minute recordings made as part of Field Test 1 on Model1. The rate of FN2 was estimated for each model from both Field Test 1 and Field Test 2.

For Model1, three periods on days 2021-11-01 and 2021-11-03 were used to estimate the rate of FN2. While only a total of 675 minutes was examined, a trend was certainly evident from this amount of recording data. The 5-second WAV data contents were first snipped out of the 5-minute WAVs so that signals with a probability score of 0.5 or greater would not be included in the total. This was undertaken using a custom [R] script, and resulted in 5-minute WAV files with one or more 5-second periods of silence. Each WAV file was opened in Adobe Audition 22.0 and the number of syllables from starling buzz and whistle type calls was counted. It was not possible to estimate how many 5-second snippets might have resulted from a sequence of several buzz syllables, so direct comparisons between the number of 5-second snippets with one or more starling syllables and the number of syllables counted from 5-minute WAVs are not possible. The process was repeated with 1,345 minutes on days 2022-03-20 - 2022-03-21 using the 5-minute WAVs from Field Test 2.

The FN2 rate at which buzzes and whistle syllables were missed was relatively high for Model1 (buzz: 45.7%; whistle: 65.1%; **Table 6**). After retraining to produce Model2, this rate decreased slightly, but especially for whistles (buzz: 40.5%; whistle: 43.3%; **Table 6**). In conclusion, it appears that Model2 is somewhat more successful at recognising a relatively greater amount of starling call variation, especially for whistles. Thus, Model2 returned a higher rate of overall detection because more starling calls are being recognised at higher probability values.

Table 6. Number of missed starling call syllables from 675 minutes of recordings in November 2021 from Field Test 1, and 1,345 minutes in March 2022 from Field Test 2 at Cape Jervis—compared with the number of all call types that were detected in the 5-second snippets.

Date	Period	No. buzz missed	No. buzz detected	No. whistle missed	No. whistle detected
Field Test 1					
2021-11-01	05:54 – 09:59 (245 mins)	329	469	102	54
2021-11-01	13:44 – 20:24 (400 mins)	273	238	101	46
2021-11-03	06:09 – 06:39 (30 mins)	133	166	53	37
Totals		735	873	256	137
	Precent missed Model1	45.7%		65.1%	
Field Test 2					
2022-03-20	11:24 – 20:04 (520 mins)	106	253	140	165
2022-03-21	06:19 – 20:04 (825 mins)	191	184	104	154
Totals		297	437	244	319
	Precent missed Model2	40.5%		43.3%	

10 Creating Model3 using recordings from Western Australia

Almost all of the signals used to train Model0, Model1 and Model2 have been recorded in South Australia, either from Adelaide city, north of the city, or various locations on the Fleurieu Peninsula. Thus, the most derived Model2 that has been deployed at the two test sites in Western Australia is naive to the Western Australian soundscape. To ensure that this does not lead to an excessive number of False Positives that require validation in Detect-It, Model2 was retrained to Model3 using not only the validated signals from the performance evaluation of Field Test 2, but also signals from Western Australia.

Exposure of the model to the Western Australian soundscape was achieved using a modified version of the Runtime code that applies the model to the audio stream. A Desktop Runtime code routine was derived that could apply the model to a resource of previously recorded WAV files residing on an external hard drive. Using the Desktop Runtime code, Model2 was applied to field recordings collected from the starling management program area on the South Coast of Western Australia in 2011 on Song Meter (Wildlife Acoustics) recorders (Campbell et al. 2013). This produced 8,329 5-second snippets containing non-target signals, which were used as labelled negative data to train Model3. Further analysis from field testing is required to quantify the improvement in Precision and Recall for Model3.

11 Using the Training code to create a Model0 for the Asian Black-spined Toad

A key stated goal of project P01-T-003 'Automated Detection: Triggering Smarter, Faster, Better Response to Incursion' was to use the same code resources to develop an automated acoustics-based detection and identification system for at least one additional invasive species, the Asian Black-spined Toad (ABST).

Recordings were provided by a global network of collaborators. These include recordings made in India, Indonesia, Singapore and Madagascar. Some of the calls are of relatively low amplitude, and others dominate the 'foreground' of recordings. The call of the consists of long call train bursts of variable duration (c. 12 or more seconds long) that consist of repeated multi-pulse calls spanning the frequency range 0.5 – 4 kHz (**Figure 9**).

The first version (Model0) of a similar bioacoustics-based detection system for the Asian Black-spined Toad was trained with the Training code resource of this project using 561 calls of the target species, and 523 signals from non-target sources. It is ready for a small-scale field test in a habitat where it will be challenged with the calls of other frogs and signal sources, as well as overlapping calls of the target species.

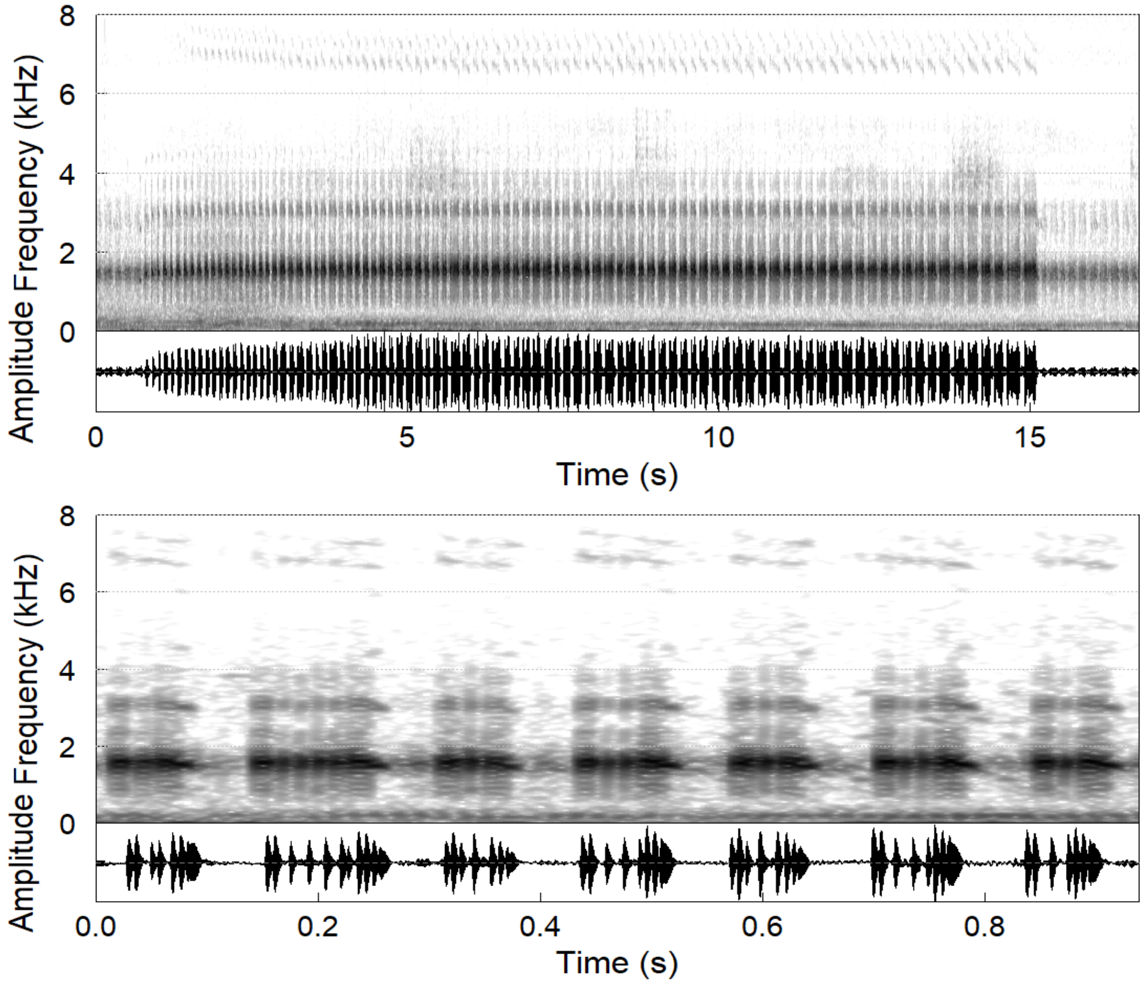


Figure 9. **Top**: Waveform and spectrogram of an example Asian Black-spined Toad 14.4 second call train of pulses. **Bottom**: Further detail of seven multi-pulse calls within the long call train.

12 Conclusions

1. A reference call library for starlings was compiled, containing both target and non-target signals (6,969 examples of starling calls and 11,822 examples of False Positive signals; in total across all model iterations).

2. The contribution of thousands of examples of starling buzz and whistle calls, and an iterative process of retraining models with labelled examples of True and False Positives recorded on two field tests has produced a deep learning CNN model that forms the core component of bioacoustics recording and analysis system for starling detection.

3. The Runtime code containing the CNN model (Model2) and associated bioacoustic recording and signal processing steps has been integrated into a fully-featured Passive Acoustic Surveillance (PAS) hardware solution that is now being tested at two sites in Western Australia.

4. The model has also been trained with 8,329 signals from the Western Australian soundscape that have the potential to elevate rates of False Positive detections to levels where validation would no longer be expedient. This newly trained Model3 is ready for field testing.

5. The coding resources are also suitable for the development of models and Passive Acoustic Surveillance systems for other vocalising species, and a Model0 for the invasive Asian Black-spined Toad was produced as a proof of concept.

References

Agranat, I. (2009). Automatically identifying animal species from their vocalizations. White paper available at: https://www.wildlifeacoustics.com/uploads/publications/Automatically-Identifying-Animal-Species-from-their-Vocalizations.pdf

Agranat, I. (2012). Bat species identification from zero crossing and full spectrum echolocation calls using HMMs, Fisher Scores, Unsupervised Clustering and Balanced Winnow Pairwise Classifiers. White paper available at: http://www.wildlifeacoustics.com/downloads/batid.pdf

Agranat, I. (2016). Kaleidoscope Pro 4.1 Auto ID for bats. White paper available at: http://www.wildlifeacoustics.com/downloads/batid4.pdf

Araya-Salas M. and Smith-Vidaurre, G. (2017). WARBLER: an r package to streamline analysis of animal acoustic signals. https://dx.doi.org/10.1111/2041-210X.12624

Campbell, S., Cook, S., Mortimer, L., Palmer, G., Sinclair, R. and Woolnough, A.P. (2012a). To catch a starling: testing the effectiveness of different trap and lure types. *Wildlife Research* 39: 183–191. https://doi.org/10.1071/WR11115

Campbell, S., Powell, C., Parr, R., Rose, K., Martin, G. and Woolnough, A. (2012b). Can artificial nest-cavities be used as a management tool to assist the control of Common Starlings (*Sturnus vulgaris*)? *Emu – Austral Ornithology* 112: 255–260. https://doi.org/10.1071/MU11050

Campbell, S., Parson, S., Woolnough, A. and Obolonkin, V. (2013). Identifying corridors, refuges and entry points for starlings, *Sturnus vulgaris*, at an invasion front. Final Report prepared for the Bureau of Rural Sciences, Department of Agriculture, Fisheries and Forestry, Australian Government

Campbell, S., Roberts, E.J., Craemer, R., Pacioni, C., Rollins, L. and Woolnough, A.P. (2015). Assessing the economic benefits of starling detection and control to Western Australia. *Australasian Journal of Environmental Management* 23: 81–99. http://dx.doi.org/10.1080/14486563.2015.1028486

Cheuk, K.W., Anderson, H., Agres, K. and Herremans, D. (2020). NNAUDIO: An on-the-fly GPU audio to spectrogram conversion toolbox using 1d convolutional neural networks. *IEEE Access* 8: 161981–162003. http://dx.doi.org/10.1109/ACCESS.2020.3019084

Clink, D.J. and Holger Klinck, H. (2021). GIBBONR: An R package for the detection and classification of acoustic signals using machine learning. R package version 0.0.0.9000. https://doi.org/10.48550/arXiv.1906.02572

Das, N., Mondal, A., Chaki, J., Padhy, N. and Dey, N. (2020). Machine learning models for bird species recognition based on vocalization: a succinct review. pp.117–124 In: *Information Technology and Intelligent Transportation Systems* (eds. L.C. Jain et al.), IOS Press. https://doi.org/10.3233/FAIA200052

Dolezel, P., Rozsival, P. and Mariska, M. (2015). European starling detection in agricultural areas using multilayer artificial neural network. pp. 471–474 In: 25th International

Conference Radioelektronika, April 2015, IEEE. https://doi.org/10.1109/RADIOELEK.2015.7129029

Dolezel, P., Skrabanek, P. and Gago, L. (2016). Pattern recognition neural network as a tool for pest birds detection. pp. 1–6 In: 2016 IEEE Symposium Series on Computational Intelligence (SSCI), December 2016, IEEE. https://doi.org/10.1109/SSCI.2016.7849988

Dolezel, P., Stursa, D. and Honc, D. (2019). Convolutional Neural Network for sound processing – study of deployed application. pp. 1–5 In: 29th International Conference Radioelektronika, April 2019, IEEE. https://doi.org/10.1109/RADIOELEK.2019.8733479

DPIRD (2018). Biosecurity alert: Common starling. Department of Primary Industries and Regional Development. Webpage last updated: 26 July 2018, URL: https://www.agric.wa.gov.au/birds/biosecurity-alert-common-starling?page=0%2C0

Eens, M., Pinxten, R. and Verheyen, R.F. (1989). Temporal and sequential organization of song bouts in the starling. *Ardea* 77: 75–86. https://avibirds.com/wp-content/uploads/pdf/spreeuw12.pdf

Eens, M., Pinxten, R. and Verheyen, R.F. (1990). On the function of singing and wing-waving in the European starling *Sturnus vulgaris*. *Bird Study* 37: 48–52. https://doi.org/10.1080/00063659009477038

Eens, M. and Pinxten, R., and Verheyen, R.F. (1991). Male song as a cue for mate choice in the European starling. *Behaviour* 116: 210–238. https://doi.org/10.1163/156853991X00049

Eens, M., Pinxten, R. and Verheyen, R.F. (1993). Function of the song and song repertoire in the European starling (*Sturnus vulgaris*): an aviary experiment. *Behaviour* 125: 51–66. https://doi.org/10.1163/156853993X00182

Hafner S and Katz J (2018). MONITOR: Acoustic template detection in R. R package version 1.0.7, http://www.uvm.edu/rsenr/vtcfwru/R/?Page=monitoR/monitoR.htm

Hindmarsh, A.M. (1984). Vocal mimicry in starlings. *Behaviour* 90: 302–324. https://www.jstor.org/stable/4534373

Humphrey, E.J., Bello, J.P. and LeCun, Y., 2013. Feature learning and deep architectures: New directions for music informatics. *Journal of Intelligent Information Systems* 41: 461–481. https://doi.org/10.1007/s10844-013-0248-5

Jinnai, M., Tsuge, S., Kuroiwa, S., Ren, F. and Fukumi, M. (2009). New similarity scale to measure the difference in like patterns with noise. *International Journal of Advanced Intelligence* 1: 59–88. http://aia-i.com/ijai/contents.html

Jinnai, M., Tsuge, S., Kuroiwa, S., and Fukumi, M. (2010). A new Geometric Distance method to remove pseudo difference in shapes. *International Journal of Advanced Intelligence* 2: 119–144. http://aia-i.com/ijai/contents.html

Jinnai, M., Boucher, N., Fukumi, M. and Taylor, H. (2012). A new optimization method of the geometric distance in an automatic recognition system for bird vocalisations. In: *Acoustics 2012*. URL: https://hal.archives-ouvertes.fr/hal-00810639/

Jinnai, M. (2018). A new segmentation method of bird vocalisations recorded in the distance. Proceedings of the 2018 Autumn (September) Meeting of the Acoustical Society of Japan, pp. 1397–1400. https://www.geometricdistance.com/_userdata/pdf/0087_3-1-5.pdf

Kahl, S., Wood, C.M., Eibl, M. and Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics* 61: 101236. https://doi.org/10.1016/j.ecoinf.2021.101236

Kiskin, I., Zilli, D., Li, Y., Sinka, M., Willis, K. and Roberts, S. (2020). Bioacoustic detection with wavelet-conditioned convolutional neural networks. *Neural Computing and Applications* 32: 915–927. https://doi.org/10.1007/s00521-018-3626-7

Methra, G. (2019). Customised and flexible devices for acoustics-based biological surveys. ELEC ENG 7077 Master of Engineering (Electronic) thesis, School of Electrical and Electronic Engineering, the University of Adelaide.

Marchal, J., Fabianek, F. and Scott, C. (2021). BIOACOUSTICS: analyse audio recordings and automatically extract animal vocalizations. R package version 0.2.5. https://CRAN.R-project.org/package=bioacoustics

Priyadarshani, N., Marsland, S. and Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: a review. *Journal of Avian Biology* 49: e01447. https://doi.org/10.1111/jav.01447

Specialised Zoological (2020). Update Report 2020-06-25 on progress of acoustic identification system for an acoustic sentinel. Unpublished report by Specialised Zoological for Susan Campbell and David Barnard, 25 June 2020, project reference SZ530.

Specialised Zoological (2022). Development and performance evaluation of the starling sentinel acoustic detector: field test of Hardware Prototype 1 and call detection Model 1 in South Australia, Spring 2021. Unpublished report by Specialised Zoological for Dr Susan Campbell, Western Australian Department of Primary Industries and Regional Development, 25 March 2022 Draft for comment, project reference SZ530.

Titley Scientific (2020). Anabat Insight user manual. Version 2.2, 2020-08-12, Titley Scientific Pty Ltd. URL: https://www.titley-scientific.com/au/downloads/user-manuals

Xia, S. (2020). Single Board Computers for professional applications in biological surveys. ELEC ENG 7077 Master of Engineering (Electronic) thesis, School of Electrical and Electronic Engineering, the University of Adelaide.

Acknowledgements

I am sincerely grateful to Thomas Rowntree (ABN 31 122 011 042) for his technical contribution to this work, and for his troubleshooting and testing effort. While I communicated Susan's direction on pipeline requirements, Thomas selected the machine learning method and wrote the code for the entire recording, signal processing and classification system, as well as other aspects related to having this run within a Docker container on a Raspberry Pi computing device. I also acknowledge the latitude given to Thomas by Professor Ian Reid of the Australian Institute of Machine Learning to contribute to this project outside of his formal employment, and for his technical advice.

Sincere thanks to David Lucas at Lynxemi Pte Ltd and David Barnard at DKB Solutions Pty Ltd who assisted with advice and provided the first prototype hardware device.

I am also very grateful to the people that kindly offered their properties as recording sites, especially those I was able to visit ultimately: Pat and Peter Smytherman, Peter Mirtschin, Alan Hancox, Ros Morgan, Robin Donaldson; and especially Carolyn Schultz and Michael Rumsewicz who cheerfully tolerated my frequent visits to their property at Cape Jervis.

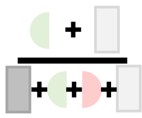
This work has been supported by two grants from the Centre for Invasive Species Solutions: Project No. P01-T-002 'Automated Detection: Triggering Smarter, Faster, Better Response to Incursion'; and Project No. P01-I-003 'Development of integrated passive and active surveillance tools and networks'.

Appendices

Appendix 1. Explanation of performance metrics and Confusion Matrix components.

As part of the performance evaluation of how well a trained artificial neural network model performs, there are several measures that are used to report its accuracy. These include Accuracy, Precision, and Recall (see *Glossary*), and are calculated based on the components of a general Confusion matrix (**Table S1-1**).

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$



Precision = $\frac{TP}{TP+FP}$

The proportion of ID'ed calls that are from starlings.
The inverse is the False Positive rate.

Recall = $\frac{TP}{TP+FN}$

The proportion of starling calls that are ID'ed.

Table S1-1. Confusion matrix showing a summary of the four possible identification outcomes from running the model on acoustic recordings, and then validating the outcome. 5-second snippet files are saved if the signal is accepted by the model at any threshold probability score of 0.5 or greater.

		Predicted condition (testing and validation outcome)	
		Predicted Positive (PP) Accepted as starling	**Predicted Negative (PN)** Rejected as starling
Actual condition	**Positive (P)** Starling call	**TP—True Positive** Correct identification SNIPPET SAVED Validated from 5-sec snippets	**FN—False Negative** Incorrect rejection, a 'miss' Type II error SNIPPET SAVED FN1: Validated from 5-sec snippet FN2: Validated from 5-min files
	Negative (N) Other signal	**FP—False Positive** Incorrect identification Type I error 'false alarm' SNIPPET SAVED Validated from 5-sec snippets	**TN—True Negative** Correct rejection SNIPPET SAVED Not easily validated from field recordings

If we can imagine a box full of starling calls on one side (closed circles on a dark grey background) and signals attributable to other sources on the other side (open circles on a light grey background) (**Figure S1-1**). All of these starling and non-starling signals have a probability of fit to the model of 0.5 and above, and are represented in 5-second snippets.

The model acts as a circular ‘cookie cutter’, cutting out signals that it thinks are from starlings from this matrix of target and non-target signals. The size of the circle relative to the square represents a set threshold. Some of these are correct identifications of starling calls (True Positive on a green background), and some of them are incorrect identifications of signals attributable to other sources (False Positive on a red background).

By retraining the model with more validated data, the circular portion should shift to the left, and expand to correctly label more starling calls (**Figure S1-2**).

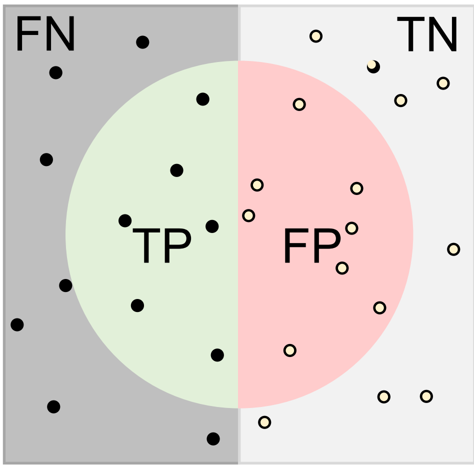


Figure S1-1. Schematic diagram of the acoustic detection process showing regions containing True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) detections (*see further details in Table S1-1*).

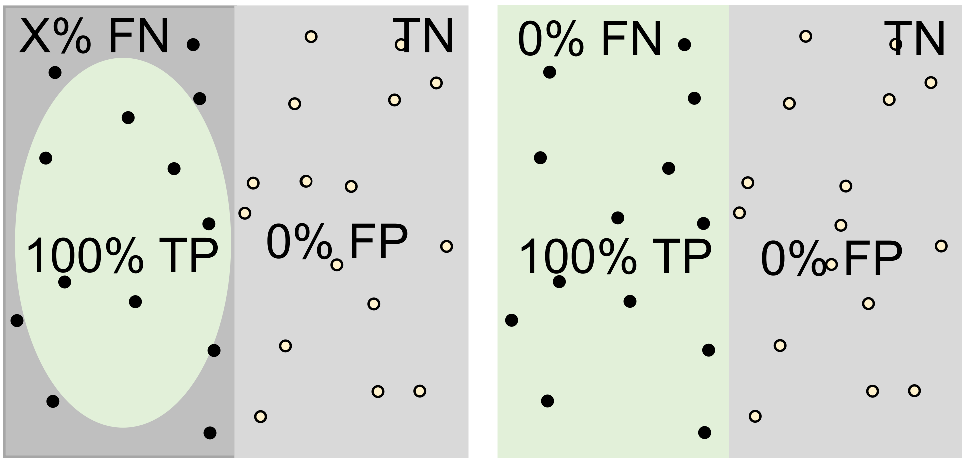


Figure S1-2. Two aspirational conditions. On the left, there is perfect Precision at a particular threshold probability, with only starling calls detected, but some starling calls are missed because the threshold is set too high. On the right, there is perfect Recall where the model detects every starling call within range of the microphone without mistakenly identifying signals from other sources.

What is not shown are all bird calls that have probability scores of below 0.5 (as represented in blue; **Figure S1-3**). In reality, the peaks below 50% contain many more signals. The majority of these will be True Negatives, but there is likely to be some starling calls as well (FN2; see *Glossary*). Perhaps the source of these starling calls is further away so that call quality is lower, or else the shape of the call is too different and beyond the experience of the model.

Improving the model by further training will reduce the number of False Positives (**Figure S1-4 left**). But also, by including a larger amount of natural variation in the calls, the model will find more starling calls in both the blue and grey areas by allocating them a higher probability score (**Figure S1-4 right**).

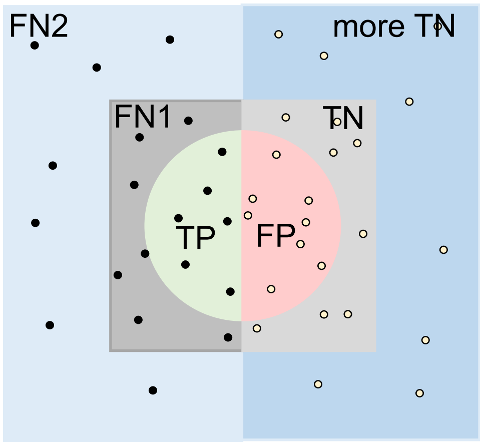


Figure S1-3. The true extent of missed detections as represented in the blue areas, all of which have a probability score of less than 0.5. The grey square encompasses all signals with a probability score of 0.5 and above, and the circle the threshold value.

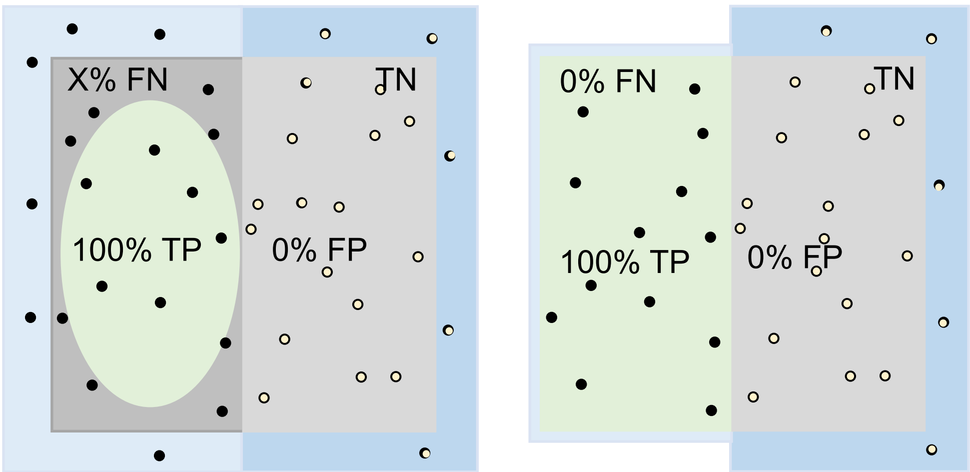


Figure S1-4. Two aspirational conditions. On the left, there is perfect Precision at a particular threshold probability, with only starling calls detected, but some starling calls are missed because the threshold is set too high, and there are still starling calls with probability scores below 0.5 being missed. On the right, there is perfect Recall where the model detects every starling call within range of the microphone without mistakenly identifying signals from other sources.

Appendix 2. The Scenario for guidance in the development of the software and hardware components (provided by Dr Susan Campbell; see also Specialised Zoological 2020).

The design of a custom detection system needs to be informed in the first instance by the description of a typical scenario. In the scenario are details of the deployment of equipment, the availability of networks for communication, and the requires of the investigator waiting to receive data from the devices.

The starling scenario in Western Australia

Purpose

- Expand our area of surveillance
- Extend our areas of surveillance into inhospitable areas
- Facilitate a rapid response to detected incursions
  - Ideally = real-time notification
  - Practically = within 24-48hrs is OK

How many units?

The high risk area currently consists of ~22,000km2 of the south coast region of WA with several hundred mapped swamp areas. Starlings are drawn towards these swamps because of their high requirement for water and the dead trees often associated with swamps provide their preferred nesting habitat.

- “Ideally” we would have units at all of these swamps (assuming detection radius per unit of ~300m, would typically require between 2-8 units / swamp.
- “Practically”, not all areas have adequate 4G coverage (nb: Telstra set to decommission 3G in 2024). It would be relatively easy to identify a priority list of areas for acoustic surveillance.
- For proof-concept: I’d like up to six sites (more than three), with average of ~4 units per site = 24-30 units depending on swamp size and detection range.
- For operational surveillance: More than 12 sites would be preferred, 48-60 units in total.

Recording schedule?

Typically starlings flock-up in winter (when their numbers allow it) and roost communally at night and feed associatively with livestock during the day (noting they’re never far from fresh water). In early spring through to early autumn they pair up and show nest site fidelity throughout the season. Typically breed twice, but three times in good year. Nesting sites (in WA) are in dead trees in swamps, but as a rule, starlings are very flexible and will nest elsewhere if competition for resources is high.

- Therefore, most benefit derived from ARUs recording from ~start Sept through to ~end March. I have had most success recording starlings at dawn, but calls do become mixed in with b/ground of general dawn chorus. General observation is that starlings are not the first bird in the chorus, typically beginning their calling close to, or just after, sunrise.
- Recommend starting recording at ~5min before civil sunrise and ideally continuing until dusk with on-board processing on the fly only retaining important information.
- Practically: detecting a starling in WA is a big deal, so missing a detection has substantial consequences. We're trying to prevent, rather than cure, incursions, therefore I need to be convinced and confident that on-board processing doesn't miss true positives. Trade-off is that large amounts of data (and false positives) may generated and require transmission over network.
- Therefore, perhaps don't record all day, may be just an hour or so at dawn.

Technical requirements

If a starling call is detected, I need to be able to place it in context, therefore I require a bit of recording (audio and the sonogram) before and after the detection to manually review.

- Ideally – the whole recording
- Practically - ~5sec before and ~15sec after may suffice.

Currently, the call signatures we've focussed on (descending whistle and electric buzz (or variation on whistle…see Bureau of Rural sciences final report) are quite high freq (providing opportunity to filter out low freq noise like vehicles and livestock), therefore selected 44kHz sampling rate. Consequence is increase in size of data files. On an SM2, stereo recording at 44.1kHz: 1hour30min recording = 930,267KB; 25min = 258,431 KB.

Ideally, I'd like a hardy, solar powered unit that listens all day and processes data on-board. When a positive detection is returned, unit stores that days file and remotely sends the detection +/- 20sec of audio via network to 'Detect-it', I receive a notification and I can go in and review audio+songogram. If any one unit is not in 4G range, then it has the capability of sending positive detections to neighbouring units that can then relay data to a base station for same day transmission of any detections over the network. If processing has to occur off the unit (either cloud-based or other), then it needs to be cost-effective to transmit (daily) recordings from units to the point of processing, for this solution I would assume units no longer record all day, but just at dawn.

Appendix 3. List of bird species recorded in the Western Australian deployment area.

Taken from the Atlas of Living Australia as being within the area bounded by latitudes -30.1 and -34.7 S and longitudes 118.9 and 129.0 E (downloaded 2022-02-15; FP: predicted potential (p) for False Positive starling identifications based on examination of spectrograms in the Xeno Canto database; CJ: recorded at Cape Jervis—*see Table 1*).

Family	Genus	species	Common name	FP	CJ
Acanthizidae	*Acanthiza*	*apicalis*	Inland Thornbill	no	
Acanthizidae	*Acanthiza*	*chrysorrhoa*	Yellow-tail	no	
Acanthizidae	*Acanthiza*	*inornata*	Western Thornbill	no	
Acanthizidae	*Acanthiza*	*iredalei*	Slender-billed Thornbill	no	
Acanthizidae	*Acanthiza*	*uropygialis*	Chestnut-rumped Thornbill	p	
Acanthizidae	*Acanthiza*	*robustirostris*	Slaty-backed Thornbill	no	
Acanthizidae	*Aphelocephala*	*leucopsis*	Southern Whiteface	p	
Acanthizidae	*Calamanthus*	*campestris*	Rufous Fieldwren	no	
Acanthizidae	*Calamanthus*	*cautus*	Shy Heathwren	no	
Acanthizidae	*Gerygone*	*fusca*	Western Gerygone	no	
Acanthizidae	*Pyrrholaemus*	*brunneus*	Redthroat	no	
Acanthizidae	*Smicrornis*	*brevirostris*	Weebill	no	
Accipitridae	*Accipiter*	*fasciatus*	Brown Goshawk	no	
Accipitridae	*Accipiter*	*cirrocephalus*	Collared Sparrowhawk	no	
Accipitridae	*Aquila*	*audax*	Wedge-tailed Eagle	no	
Accipitridae	*Circus*	*approximans*	Kahu	no	
Accipitridae	*Circus*	*assimilis*	Spotted Harrier	no	
Accipitridae	*Elanus*	*axillaris*	Black-shouldered Kite	no	
Accipitridae	*Elanus*	*scriptus*	Letter-winged Kite	no	
Accipitridae	*Haliaeetus*	*leucogaster*	white-bellied sea-eagle	no	
Accipitridae	*Haliastur*	*sphenurus*	Whistling Kite	p	
Accipitridae	*Hamirostra*	*melanosternon*	Black-breasted Buzzard	no	
Accipitridae	*Hieraaetus*	*morphnoides*	Little Eagle	no	
Accipitridae	*Lophoictinia*	*isura*	Square-tailed Kite	no	
Acrocephalidae	*Acrocephalus*	*australis*	Australian Reed Warbler	no	
Aegothelidae	*Aegotheles*	*cristatus*	Australian Owlet-nightjar	p	
Alcedinidae	*Dacelo*	*novaeguineae*	Kookaburra	no	
Alcedinidae	*Todiramphus*	*pyrrhopygius*	Red-backed Kingfisher	no	
Alcedinidae	*Todiramphus*	*sanctus*	Sacred Kingfisher	no	
Apodidae	*Apus*	*pacificus*	Fork-tailed Swift	no	
Artamidae	*Artamus*	*cinereus*	Black-faced Woodswallow	no	
Artamidae	*Artamus*	*cyanopterus*	Dusky Woodswallow	no	
Artamidae	*Artamus*	*minor*	Little Woodswallow	no	
Artamidae	*Artamus*	*personatus*	Masked Woodswallow	no	
Artamidae	*Artamus*	*superciliosus*	White-browed Woodswallow	no	
Artamidae	*Cracticus*	*nigrogularis*	Pied Butcherbird	no	
Artamidae	*Cracticus*	*torquatus*	Grey Butcherbird	no	

Family	Genus	species	Common name	FP	CJ
Artamidae	*Gymnorhina*	*tibicen*	Australian Magpie	no	CJ
Artamidae	*Strepera*	*versicolor*	Grey Currawong	no	
Atrichornithidae	*Atrichornis*	*clamosus*	Noisy Scrub-bird	no	
Burhinidae	*Burhinus*	*grallarius*	Bush Stone-curlew	no	
Cacatuidae	*Cacatua*	*pastinator*	Western Corella	no	
Cacatuidae	*Cacatua*	*sanguinea*	Little Corella	no	
Cacatuidae	*Calyptorhynchus*	*banksii*	Red-tailed Black Cockatoo	no	
Cacatuidae	*Calyptorhynchus*	*baudinii*	Long-billed Black-cockatoo	no	
Cacatuidae	*Calyptorhynchus*	*latirostris*	Carnaby's Black-cockatoo	no	
Cacatuidae	*Eolophus*	*roseicapilla*	Galah	no	CJ
Cacatuidae	*Lophochroa*	*leadbeateri*	Major Mitchell's Cockatoo	no	
Cacatuidae	*Nymphicus*	*hollandicus*	Cockatiel	no	
Campephagidae	*Coracina*	*novaehollandiae*	Black-faced Cuckoo-shrike	no	
Campephagidae	*Coracina*	*maxima*	Ground Cuckoo-shrike	no	
Caprimulgidae	*Eurostopodus*	*argus*	Spotted Nightjar	no	
Casuariidae	*Dromaius*	*novaehollandiae*	Emu	no	
Charadriidae	*Vanellus*	*miles*	Masked Lapwing	no	
Charadriidae	*Vanellus*	*tricolor*	Banded Lapwing	no	
Climacteridae	*Climacteris*	*rufa*	Rufous Treecreeper	no	
Climacteridae	*Climacteris*	*affinis*	White-browed Treecreeper	no	
Columbidae	*Columba*	*livia*	Rock Dove	no	
Columbidae	*Geopelia*	*cuneata*	Diamond Dove	no	
Columbidae	*Geopelia*	*striata*	Peaceful Dove	no	
Columbidae	*Ocyphaps*	*lophotes*	Crested Pigeon	no	
Columbidae	*Phaps*	*chalcoptera*	Common Bronzewing	no	CJ
Columbidae	*Phaps*	*elegans*	Brush Bronzewing	no	
Columbidae	*Streptopelia*	*chinensis*	Spotted Turtle-dove	no	
Columbidae	*Streptopelia*	*senegalensis*	Laughing Turtle-dove	no	
Corvidae	*Corvus*	*bennetti*	Little Crow	no	CJ
Corvidae	*Corvus*	*coronoides*	Australian Raven	no	CJ
Corvidae	*Corvus*	*orru*	Torresian Crow	no	
Cuculidae	*Cacomantis*	*flabelliformis*	Fan-tailed Cuckoo	no	
Cuculidae	*Cacomantis*	*pallidus*	Pallid Cuckoo	no	
Cuculidae	*Chalcites*	*basalis*	Horsfield's Bronze-cuckoo	no	
Cuculidae	*Chalcites*	*lucidus*	Shining Bronze-cuckoo	no	
Cuculidae	*Chalcites*	*osculans*	Black-eared Cuckoo	no	
Cuculidae	*Chrysococcyx*	*lucidus*	Shining Cuckoo	no	
Dasyornithidae	*Dasyornis*	*longirostris*	Western Bristlebird	no	
Estrildidae	*Stagonopleura*	*oculata*	Red-eared Firetail	p	
Estrildidae	*Taeniopygia*	*guttata*	Zebra Finch	no	
Falconidae	*Falco*	*longipennis*	Australian Hobby	no	
Falconidae	*Falco*	*hypoleucos*	Grey Falcon	no	
Falconidae	*Falco*	*peregrinus*	Peregrine Falcon	no	
Falconidae	*Falco*	*subniger*	Black Falcon	no	

Family	*Genus*	*species*	Common name	FP	CJ
Falconidae	*Falco*	*berigora*	Brown Falcon	no	
Falconidae	*Falco*	*cenchroides*	Nankeen Kestrel	no	CJ
Hirundinidae	*Cheramoeca*	*leucosterna*	White-backed Swallow	no	
Hirundinidae	*Hirundo*	*neoxena*	Welcome Swallow	no	
Hirundinidae	*Petrochelidon*	*nigricans*	Tree Martin	no	
Hirundinidae	*Petrochelidon*	*ariel*	Fairy Martin	no	
Maluridae	*Amytornis*	*striatus*	Striated Grasswren	no	
Maluridae	*Malurus*	*assimilis*	Purple-back Fairy-wren	no	
Maluridae	*Malurus*	*elegans*	Red-winged Fairy-wren	no	
Maluridae	*Malurus*	*pulcherrimus*	Blue-breasted Fairy-wren	no	
Maluridae	*Malurus*	*splendens*	Splendid Fairy-wren	no	
Maluridae	*Malurus*	*leucopterus*	White-winged Fairy-wren	no	
Maluridae	*Stipiturus*	*malachurus*	Southern Emu-wren	no	
Megaluridae	*Cincloramphus*	*cruralis*	Brown Songlark	no	
Megaluridae	*Cincloramphus*	*mathewsi*	Rufous Songlark	no	
Megaluridae	*Megalurus*	*gramineus*	Little Grassbird	no	
Megapodiidae	*Leipoa*	*ocellata*	Malleefowl	no	
Meliphagidae	*Acanthagenys*	*rufogularis*	Spiny-cheeked Honeyeater	no	
Meliphagidae	*Acanthorhynchus*	*superciliosus*	Western Spinebill	no	
Meliphagidae	*Anthochaera*	*lunulata*	Western Wattlebird	no	
Meliphagidae	*Anthochaera*	*carunculata*	Red Wattlebird	no	CJ
Meliphagidae	*Certhionyx*	*variegatus*	Pied Honeyeater	no	
Meliphagidae	*Epthianura*	*aurifrons*	Orange Chat	no	
Meliphagidae	*Epthianura*	*albifrons*	White-fronted Chat	no	
Meliphagidae	*Epthianura*	*tricolor*	Crimson Chat	p	
Meliphagidae	*Gavicalis*	*virescens*	Singing Honeyeater	no	CJ
Meliphagidae	*Gliciphila*	*melanops*	Tawny-crowned Honeyeater	no	
Meliphagidae	*Lichenostomus*	*cratitius*	Purple-gaped Honeyeater	no	
Meliphagidae	*Lichmera*	*indistincta*	Brown Honeyeater	no	
Meliphagidae	*Manorina*	*flavigula*	Yellow-throated Miner	no	
Meliphagidae	*Melithreptus*	*brevirostris*	Brown-headed Honeyeater	no	
Meliphagidae	*Melithreptus*	*chloropsis*	Swan River Honeyeater	no	
Meliphagidae	*Nesoptilotis*	*leucotis*	White-eared Honeyeater	no	
Meliphagidae	*Phylidonyris*	*niger*	White-cheeked Honeyeater	no	
Meliphagidae	*Phylidonyris*	*novaehollandiae*	New Holland Honeyeater	no	CJ
Meliphagidae	*Ptilotula*	*ornata*	Yellow-plumed Honeyeater	no	CJ
Meliphagidae	*Ptilotula*	*plumula*	Grey-fronted Honeyeater	no	
Meliphagidae	*Purnella*	*albifrons*	White-fronted Honeyeater	no	
Meliphagidae	*Sugomel*	*niger*	Black Honeyeater	no	
Meropidae	*Merops*	*ornatus*	Rainbow Bee-eater	no	
Monarchidae	*Grallina*	*cyanoleuca*	Magpie-lark	no	CJ
Monarchidae	*Myiagra*	*cyanoleuca*	Satin Flycatcher	p	
Monarchidae	*Myiagra*	*inquieta*	Restless Flycatcher	no	
Nectariniidae	*Dicaeum*	*hirundinaceum*	Mistletoebird	no	

Family	Genus	species	Common name	FP	CJ
Neosittidae	*Daphoenositta*	*chrysoptera*	Varied Sittella	no	
Oreoididae	*Oreoica*	*gutturalis*	Crested Bellbird	no	
Otididae	*Ardeotis*	*australis*	Wild Turkey	no	
Pachycephalidae	*Colluricincla*	*harmonica*	Grey Shrike-thrush	no	CJ
Pachycephalidae	*Falcunculus*	*frontatus*	Crested Shrike-tit	no	
Pachycephalidae	*Pachycephala*	*rufiventris*	Rufous Whistler	no	
Pachycephalidae	*Pachycephala*	*occidentalis*	Western Whistler	no	
Pachycephalidae	*Pachycephala*	*inornata*	Gilbert's Whistler	no	
Pardalotidae	*Pardalotus*	*striatus*	Striated Pardalote	no	
Pardalotidae	*Pardalotus*	*punctatus*	Spotted Pardalote	no	
Petroicidae	*Drymodes*	*brunneopygia*	Southern Scrub-robin	no	
Petroicidae	*Eopsaltria*	*griseogularis*	Western Yellow Robin	no	
Petroicidae	*Eopsaltria*	*georgiana*	White-breasted Robin	no	
Petroicidae	*Melanodryas*	*cucullata*	Hooded Robin	no	
Petroicidae	*Microeca*	*fascinans*	Jacky Winter	no	
Petroicidae	*Petroica*	*boodang*	Scarlet Robin	no	
Petroicidae	*Petroica*	*goodenovii*	Red-capped Robin	no	
Phasianidae	*Coturnix*	*pectoralis*	Grey Quail	no	
Phasianidae	*Coturnix*	*ypsilophora*	Swamp Quail	no	
Podargidae	*Podargus*	*strigoides*	Tawny Frogmouth	no	
Pomatostomidae	*Pomatostomus*	*superciliosus*	White-browed Babbler	no	
Psittacidae	*Barnardius*	*zonarius*	Australian Ringneck	no	
Psittacidae	*Melopsittacus*	*undulatus*	Budgerigar	no	
Psittacidae	*Neophema*	*elegans*	Elegant Parrot	no	
Psittacidae	*Neophema*	*petrophila*	Rock Parrot	no	
Psittacidae	*Neophema*	*splendida*	Scarlet-chested Parrot	no	
Psittacidae	*Northiella*	*narethae*	Naretha Parrot	no	
Psittacidae	*Parvipsitta*	*porphyrocephala*	Purple-crowned Lorikeet	p	CJ
Psittacidae	*Platycercus*	*icterotis*	Western Rosella	no	
Psittacidae	*Polytelis*	*anthopeplus*	Regent Parrot	no	
Psittacidae	*Psephotus*	*varius*	Mulga Parrot	no	
Psittacidae	*Purpureicephalus*	*spurius*	Red-capped Parrot	no	
Psittacidae	*Trichoglossus*	*haematodus*	Rainbow Lorikeet	p	CJ
Psophodidae	*Cinclosoma*	*clarum*	Copperback Quail-thrush	no	
Psophodidae	*Cinclosoma*	*castaneothorax*	Chestnut-breasted Quail-thrush	p	
Psophodidae	*Cinclosoma*	*cinnamomeum*	Cinnamon Quail-thrush	no	
Psophodidae	*Psophodes*	*nigrogularis*	Western Whipbird	no	
Rhipiduridae	*Rhipidura*	*albiscapa*	Grey Fantail	no	CJ
Rhipiduridae	*Rhipidura*	*leucophrys*	Willie Wagtail	no	
Strigidae	*Ninox*	*connivens*	Barking Owl	no	
Strigidae	*Ninox*	*novaeseelandiae*	Southern Boobook	no	
Sturnidae	*Acridotheres*	*tristis*	Common Myna	no	
Timaliidae	*Zosterops*	*lateralis*	Silvereye	no	
Turnicidae	*Turnix*	*velox*	Little Button-quail	no	

Family	***Genus***	***species***	**Common name**	**FP**	**CJ**
Turnicidae	*Turnix*	*varius*	Painted Button-quail	no	
Tytonidae	*Tyto*	*novaehollandiae*	Masked Owl	no	
Tytonidae	*Tyto*	*javanica*	Eastern Barn Owl	no	
Tytonidae	*Tyto*	*alba*	Barn Owl	no	

Centre for Invasive Species Solutions
Building 22, University of Canberra
University Drive South, BRUCE ACT 2617
T 02 6201 2887
E communications@invasives.com.au