Scaling Acoustic Data Analysis through Collaboration and Automation

Jason Wimmer, Michael Towsey, Birgit Planitz, Paul Roe and Ian Williamson
Microsoft QUT eResearch Centre
Queensland University of Technology
Brisbane, Australia
{j.wimmer; m.towsey; b.planitz; p.roe; i.williamson}@qut.edu.au

***Abstract*—Monitoring and assessing environmental health is becoming increasingly important as human activity and climate change place greater pressure on global biodiversity. Acoustic sensors provide the ability to collect data passively, objectively and continuously across large areas for extended periods of time. While these factors make acoustic sensors attractive as autonomous data collectors, there are significant issues associated with large-scale data manipulation and analysis. We present our current research into techniques for analysing large volumes of acoustic data effectively and efficiently. We provide an overview of a novel online acoustic environmental workbench and discuss a number of approaches to scaling analysis of acoustic data; collaboration, manual, automatic and human-in-the loop analysis.**

Keywords- global climate change; sensors; acoustic sensing; acoustic analysis; data analysis

I. INTRODUCTION

Monitoring and assessing environmental health is becoming increasingly important as human activity and climate change place greater pressures on global biodiversity. Protecting biodiversity and developing effective conservation strategies requires a thorough understanding of natural systems, the relationship between organisms and their environment and the effects of climate change. This understanding is traditionally derived from field observations using manual methods such as fauna and vegetation surveys.

While manual methods will continue to play an important role in conservation, they are limited in their ability to monitor the effects of environmental change over large spatiotemporal scales. It is in this context that providing scientists with technology and tools to rapidly collect and analyse environmental data on a large scale is critical [1, 2]. Sensors are one of tools being utilised increasingly to extend the reach of scientists engaged in environmental monitoring [3-5]. They can be relatively inexpensive, provide a continuous, in-situ monitoring capability over large areas and record data over long periods of time. A wide range of sensors are available to monitor numerous aspects of the environment. This research focuses on the use of acoustic sensors and the analysis of associate acoustic sensor data.

There are numerous roles for acoustic sensors in ecology, conservation biology and wildlife management research. These include:

- Measures of species richness: detecting and measuring the number of different species in a given area [6-8];
- Measures of species abundance: detecting and measuring the size of specific species populations in a given area [9, 10];
- Localisation: detecting specific vocalisations/acoustic events and determining the spatial origin of the call [11, 12];
- Measures of ecosystem health: generalised or relative measures of ecosystem health in the context of a system or specific species [13, 14]

Acoustic sensing provides ecologists with the capability to massively scale ecological observations, both temporally and spatially. For example, traditional avian point counts may involve trained ecologists making 10 minute observations at dawn, noon and dusk over a period of five days at a single site. At 2.5 hours, the total observation time for a short term manual survey is a fraction of the potential 120 hours of a continuous automated acoustic sensor recording over the same period of time at the same site. At long term scales, even scheduled recordings (e.g. five minute recordings every 30 minutes) provide ecologists with significantly more data than manually collected long term survey data.

Conservation biologists and ecologists are increasingly turning to sensor technology to assist their work in the field [1, 3-5]. Sensors provide an effective means to accumulate data at large scales and high resolutions [15]. Acoustic sensors have been employed in this capacity for some time, both in marine and terrestrial environments [14, 16-18]. They can be used to collect data passively, objectively and continuously across large areas for extended periods of time. While these factors make acoustic sensors attractive as autonomous data collectors, large scale data collection presents its own problems:

- Data volume – acoustic sensors generate vast quantities of raw data which must be stored, analysed and summarized.
- Complex analysis requirements – recognising the vocalisations of individual animals or species against a background of general noise and other vocalisations is a complex and challenging task.

This paper describes a novel online Acoustic Environmental Workbench which addresses both the problems of data deluge and complex analysis through collaboration and human-in-the-loop semi-automation. The workbench is a web-based application which includes integrated data upload, storage, management, playback, interactive analysis and annotation tools which enable users to work collaboratively to scale acoustic analysis tasks.

In Part II of this paper we outline the basic architecture of our system. In Part III we describe our analysis techniques and Part IV proposes future work.

II. Online Environmental Workbench

Part of our ongoing research has been to compare the effectiveness of acoustic sensors with traditional manual fauna survey methods. This work identified the need to provide ecologists with a framework which facilitates close interaction with their data and the ability to work collaboratively with other scientists. Working with ecologists we have identified the following core functionality:

- Data upload and storage.
- Data organisation and structure.
- Recording playback and visualisation.
- Recording analysis and annotation.
- Automated generation of metadata using Ecological Metadata Language (EML).

We describe these core functions in turn.

A. Data Upload and Storage

Format flexibility and centralised access were identified as core functionality to enable recordings from a variety of devices, in many formats to be uploaded and accessed via the internet. To achieve this, the acoustic workbench provides web-based access to recordings collected from a variety of sources including, but not limited to, networked sensors and standalone data loggers. Acoustic data in MP3 or WAV format may be uploaded from any digital recording device capable of generating files in these formats.

This centralised approach provides a number of advantages:

- Online access and collaboration: multiple users have access to the same data and same analysis tools, enabling users to collaborate on analysis tasks.
- Data retention: all raw data is stored and retained to allow future analysis as techniques improve and to enable long term comparisons of historical data.
- Data security and backup: all data is stored securely with regular backups and recovery facilities to prevent data loss.
- Data provenance and context retention (metadata): key experimental design details are retained to ensure accurate comparisons between datasets.

B. Data Organisation and Structure

The acoustic workbench allows users to browse and manipulate data in a logical, structured manner. Acoustic data is presented to users based on a hierarchical model of Projects, Sites, Sensors and Recordings.

Projects represent the top level of the hierarchy. Each project consists of a collection of Sites. A project may represent an individual experiment or series of experiments. Sites are physical locations (identified by GPS coordinates), with sensors deployed at each site. Sensors are physical recording devices whose details are stored to ensure retention of experimental design details. Recordings are the raw acoustic data collected from sensor devices in the field and uploaded to the website.

Users are granted role-based permissions on a Project basis. These control the level of access to project data and analysis tasks. Access levels include:

- None (default): user has no access to any data or functionality in the project.
- Read Only: user can view/play acoustic data, but cannot annotate spectra, cannot upload data and cannot perform analysis tasks.
- Full: user can view/play acoustic data, can annotate spectra, can upload data and can perform analysis tasks.

C. Recording Playback and Visualisation

Recordings can be played online using a custom audio playback tool developed for the workbench. The playback tool displays a spectrogram which allows the user to visualise and hear audio simultaneously. Long recordings are split into four minute segments which are loaded dynamically as the player reaches the end of each segment. For example, a continuous 24 hour recording is divided into 360 four-minute segments. This allows the user to start listening without waiting for the entire recording to download.

Users have sequential or random access to the contents of a recording using the player’s navigation tools. This provides the ability to scan recordings rapidly or to locate specific times of interest, for example dawn and dusk. In addition, several recordings may be selected at once to create a 'playlist' of audio to play or to assign to analysis tasks. These playlists are generated using a search tool, which provides the capability to filter recordings based on time and date recorded, project, site and on tags which have been annotated in any recording.

D. Recording Analysis and Annotation

One of the key goals of this research is to devise analysis tools to scan large volumes of acoustic data and identify distinct species. The results of this analysis can then be used to generate a list of species with vocalisations in the recordings. Apart from the obvious challenges of data volume and arbitrary noise sources, animal calls show great variation both within and between species. There can also be significant regional variation in many species.

To overcome these issues the workbench provides a flexible approach to data analysis. Users can work alone or in collaboration. They can annotate recordings manually, run fully automated tools or interact with the system in a semi-automated fashion.

Manual analysis requires users to identify vocalisations aurally and/or visually and to annotate spectrograms directly. Annotation involves drawing a rectangular marquee around the spectral content of a call as it appears in a spectrogram. See Fig. 1 for an example of a spectrogram with annotations. The user then assigns an identifying tag to the marquee. The marquee identifies the upper and lower frequency bound and the start and end times of the call.

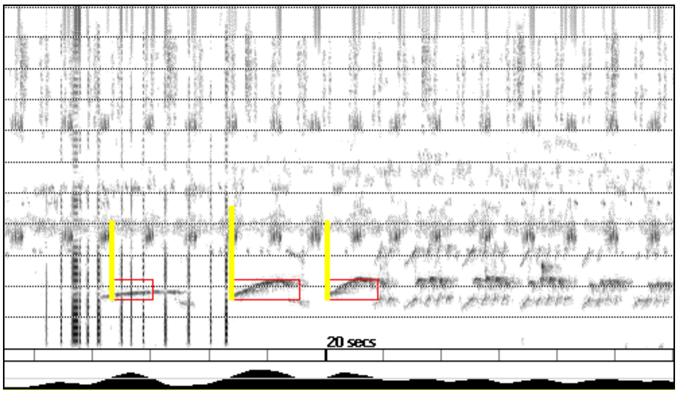

Figure 1. Marquees identifying three Bush Stone Curlew (*Burhinus grallarius*) syllables in a series of calls which are masked by noise and various insect vocalisations.

Users can search the acoustic database for all recordings containing an identifying tag. These recordings can be assigned to a playlist or downloaded as a report in CSV format, listing the tagged events in the recordings.

E. Automated Generation of EML

We use Ecological Metadata Language (EML) [19] to facilitate incorporation of ecological projects undertaken using our acoustic workbench into the Long Term Ecological Research (LTER) metadata catalogues. EML defines, standardises and formalises metadata considered essential for the adequate description of ecological data. EML defines metadata attributes such as geographic and temporal coverage, taxonomic coverage and data types. Maintaining and standardising metadata associated with ecological experiments retains the context of the data and enables reinterpretation and re-analysis of the data for long term research.

III. Online Analysis Techniques

There are significant challenges associated with analysing recordings of the natural environment. They are subject to many effects including natural noise (such as wind and rain), man-made noise (such as cars and aeroplanes) and various forms of electrostatic interference. In addition, many animal species exhibit significant call variation and their call spectra vary depending on proximity to the microphone. Our research has identified the need to provide ecologists with flexibility when analysing acoustic data. To this end, we provide the following options for processing and analysing acoustic data:

- Manual analysis
- Automated call recognition
- Human-in-the-loop analysis

A. Manual Analysis

The Acoustic Workbench provides users with core functionality to playback recordings, identify vocalisations and annotate spectrograms with identification tags. This approach is primarily used for single species identification or detailed single species behavioural studies. Manual analysis by skilled users provides a highly accurate and comprehensive audit of acoustic data, however processing large volumes of data can be time consuming. Manual analysis may also be necessary in acoustically complex environments (e.g. avian dawn chorus), where automated tools fail to discriminate between simultaneous vocalisations.

Given the volume of data associated with long term acoustic sensing, the time required to manually analyse recordings can be prohibitive. Additionally, manual analysis typically requires highly trained users who can discriminate between vocalisations and identify many numbers of species. To help overcome some of these limitations, the Acoustic Workbench provides a number of additional tools to assist manual analysis:

- *Online collaboration*: enables users to scale manual analysis by allowing multiple users to collaborate in identifying and annotating large volumes of acoustic data. The workbench also incorporates a feedback and confidence rating system which provides the ability to rate the accuracy of collaborating users. Collaboration can also be used to focus the attention of ‘expert’ user’s on complex or difficult-to-identify calls.

- *Online species identification library*: assists users in call identification. To reduce the time taken to correctly identify vocalisations, users can compare a call in a spectrogram with spectrograms of previously identified exemplars. To reduce the number of exemplar spectrograms to compare, the library can be filtered on features such a frequency band, call duration and geographic proximity to the current recording.

- *Removal of silence and noise*: removes sections of recordings with long periods of silence or periods with continuous noise pollution (e.g. caused by wind or rain). Automated removal of these sections of a recording reduces the volume of acoustic data to analyse, and focuses manual effort on those parts of a recording mostly amenable to analysis.

- *Rapid spectrogram scanning*: allows a user to visualise a recording in a fraction of the time that it takes to listen to it. Many vocalisations have a characteristic spectral appearance that the human eye can recognise more easily than available automated techniques. Rapid spectrogram scanning allows user to scan quickly through an entire recording to search for a specific species or vocalisation.

TABLE I. RECOGNISER RESULTS FROM EXPERIMENTS USING FOUR RECOGNITION TECHNIQUES.

Call structure	**Recognition technique**	**Call type**	**Recordings (Files in Datasets)**	**# Files with Calls**	**Recall**	**Precision**	**Accuracy**
Single syllable	MFCC features + HMM	Currawong	29 x 4-minutes	7	28.6%	100%	75.9%
Oscillating single syllables in time domain	Detection of temporal oscillations within a characteristic frequency band of the STFT.	Cane Toad	337 x 2-minutes	55	92.5%	98.0%	98.5%
		Asian House Gecko	270 x 2-minutes	77	90.9%	89.7%	94.4%
		Male Koala (bellows)	115 x 4-minutes	12	75.0%	75.0%	94.8%
Static pattern in time and frequency	Detection of a characteristic pattern of acoustic events in the STFT. (AED + EPR)	Ground Parrot (one call type)	405 x 1-minute	23	87.0%	87.0%	98.5%
Complex single/ multiple-line patterns	Detect whistle followed by whip using Syntactic Pattern Recognition.	Whipbirds	38 x 2-mintues	14	100%	66.7%	81.6%

B. Automated Call Recognition

Perhaps due to the importance of birds as indicator species of environmental health, there is already a considerable body of work published on the automatic detection of bird vocalisations [20-28]. A common approach has been to adopt the well-developed tools of Automated Speech Recognition (ASR), which extract Mel-Frequency Cepstral Coefficients (MFCCs) as features and use Hidden Markov Models (HMMs) to model the vocalisations.

Unfortunately it is not so easy to translate ASR to the analysis of environmental recordings because there are far fewer constraints in the latter task. Two issues are noise and variability. ASR tasks are typically restricted to environments where noise is tightly constrained, for example over the telephone. By contrast, environmental acoustics can contain a wide variety of non-biological noises having a great range of intensities and a variety of animal sounds which are affected by the physical environment (vegetation, geography etc.). Furthermore, the sources can be located any distance from the microphone. Secondly, despite its difficulty, ASR applied to the English language requires the recognition of about 50 phonemes (or 150 tri-phones). By contrast, bird calls offer endless variety; variety in call structure between species, variety between populations of the one species and variety within and between individuals of the one population. Many species have multiple calls and many are mimics. To give some indication of the difficulty of bird call recognition, a state-of-the-art commercial system using an ASR approach that has been under development for more than a decade, achieves, on unseen test vocalisations of 54 species, an average accuracy of 65% to 75% [29].

In our experience, ASR techniques have not been effective for most animal calls (our work extends beyond bird recognition to include insects, reptiles and koalas). Two reasons would appear to be 1) the inappropriateness of cepstral coefficients as features to describe bird whistles and 2) the difficulty of having a suitable HMM noise model to cover the wide variety of situations that occur in an uncontrolled recording. Note that MFCC features were developed for ASR under conditions where noise and recording conditions were tightly controlled.

Our approach has been to identify the invariant features of calls of interest and to build recognisers for those features. Not all noise types and all species occur at all locations so it is possible to achieve useful recognition results without building a universal-classifier to recognise everything.

While some animal and bird calls have complex structures [27], species recognition does not necessarily require recognition of an entire call. For example it is not necessary to model the complex structure of 30 second male Koala (*Phascolarctos cinereus*) bellow. Instead the oscillatory characteristic of its exhales provides a suitable feature on which to train a recogniser. Likewise the Bush Stone Curlew (*Burhinus grallarius*) has a multi-syllable call structure with harmonics, but recognition can be limited to detection of a single characteristic formant. Even highly variable bird calls such as that of the Golden Whistler (*Pachycephala pectoralis*) may be confined to a particular frequency band and have characteristic frequency modulated whistles. Many multi-syllable calls consist of the same repeated syllable (e.g. the cane toad (*Bufo marinus*)) or different syllables varying in pitch (e.g. the ground parrot (*Pezoporus wallicus*)), duration or both (e.g. the whistle and whip of the Eastern Whipbird (*Psophodes olivaceus*)).

While all these call types exhibit some form of variability, nevertheless each has an invariant feature to which a recogniser can be tuned. Representative examples of recognition techniques we have implemented include:

- MFCC features + HMMs: We have found this technique to be suitable only for high quality single-syllable calls. We used the HMM Tool Kit [30] and applied it to the recognition of Pied Currawong (*Strepera graculina*) calls.
- Oscillation Detection (OD): We used a Discrete Cosine Transform to find repeating or oscillating elements of calls within a user specified bandwidth. This method is highly sensitive and does not require prior noise removal. For more details see [31].
- Event Pattern Recognition (EPR): This technique models a call as a 2D distribution of acoustic events in the spectrogram. Step 1: Acoustic Event Detection (AED). Extract acoustic events from the spectrogram.

Each call syllable should be isolated as a single event. Step 2: Detect a 2D pattern of events whose distribution matches a template. Note that the content of the syllables themselves is *not* modelled. The advantage of this method is that it is resistant to background noise and other acoustic events. For more details see [32].

- Syntactic Pattern Recognition (SPR): this technique models a call as a symbol sequence, each symbol selected from a finite alphabet representing ‘primitive’ elements of the composite pattern. In our case the primitives are short straight-line segments at different angles in the spectrogram. Step 1: Isolate Spectral Peak Tracks (SPTs) which appear as ridges in the spectrogram. Step 2: Describe the spectral tracks as piece-wise straight line segments. We apply this technique to Eastern Whipbird calls that can be modelled as a series of horizontal line segments (the whistle) followed by a series of near-vertical line segments (the whip).

To test these methods we used data sets selected by an ecologist based on judgements as to what selection of recordings at different times of the day would provide interesting information about the locality. An ecologist tagged all calls of interest, even those at the limits of audibility and not expected to be detected by automated means. Our objective was to devise experimental conditions that would reflect how an ecologist would use the acoustic workbench. Results are displayed in Table I. We use the following definitions of recall and precision:

recall = TP/(TP+FN)

precision = TP/(TP+FP)

Accuracy is defined as the total number of correctly classified 1-4-minute file segments in the test set. We adopted the convention that where a recogniser detected a true positive (TP) in a single 1-4 minute file yet made an error in the same file (either a false positive—FP—or false negative—FN) we labelled that file correctly classified. On the other hand we observed many instances where multiple TPs were obtained in one recording but offset by a single error in another file. The most common errors were FN due to a distant call or call lost in noise. We chose this form of presenting accuracy because it is more efficient for ecologists using the Acoustic Workbench to work with audio segments of 1-4 minute rather than manipulate hours of recording. Furthermore birds tend to call in clusters and reporting on a file basis reduces the length of a report.

It is notable that the use of MFCCs and HMMs was the least successful technique tested (Table I). Although the accuracy figures presented should only be regarded as general indications of performance in a real operational environment, they nevertheless demonstrate that useful accuracy rates can be achieved for automated recognition when appropriate algorithms are selected for specific vocalisations.

C. Human-in-the-Loop Analysis

Human-in-the-loop analysis provides a hybrid approach which addresses the respective strengths and weaknesses of the manual and automated techniques. Manual analysis utilises the sophisticated recognition capabilities of an expert user, but cannot be efficiently scaled to process the volumes of data collected in long-term sensor deployments. Automated techniques are effective for identifying targeted species in large volumes of data, but they require a high degree of skill to develop and are still not able to cope with the variability that animal calls present.

Combining manual and automated approaches provides users with the ability to interactively and systematically process large volumes of acoustic data in a semi-automated fashion. Human-in-the loop analysis recognises that: a) many species (particularly avian species) have a broad range of vocalisations and these vocalisations may have significant regional variation; b) environmental factors such as wind, rain, vegetation and topography can attenuate, muffle and distort vocalisations considerably and c) human analysis capabilities are currently far superior to that of automated computational analysis tools. Accurately identifying species with diverse vocalisations in diverse conditions requires a broad selection of characteristic vocalisations under different conditions. This problem is analogous to developing speech recognition tools to identify any number of words, in any number of languages for any number of accents, in almost any physical environment. The human-in-the-loop technique provides users with the ability to:

- Identify and associate many different vocalisations with a single species.
- Automate repetitive scanning and annotation tasks.
- Leverage expert user time by searching an entire recording or set of recordings with a number of identifying vocalisations.
- Identify and locate potential vocalisations which have not been identified i.e. identify novelty.
- Develop a comprehensive, geographical-specific library of vocalisations to apply to other recordings

To illustrate this technique, the following is an example of a typical human-in-the-loop scenario.

A user is tasked with producing a species list and associated call frequency data for avian species detected in a seven day (168 hour) continuous acoustic recording. The user is also tasked with building up a library of representative calls of species of interest in the recording. This library could be used later to assist call identification in other recordings at the same location.

The recording is uploaded to the environmental workbench, divided into four-minute segments for playback and processed segment-wise to remove background noise.

The analysis process begins by performing a manual scan of the first minutes of the recording to identify calls of interest. These are manually tagged as the identified species, processed and placed in the call library. At present we use a binary matrix to represent the shape of calls in a spectrogram. These steps are represented in Fig. 2 as the arrows from ‘Start’ to ‘New Tags’ to ‘Library of Calls’.

The automated part of the human-in-the-loop process (the top section of Fig. 2) is to scan the entire recording with the templates in the Call Library. The recall/precision trade-off is controlled with a sensitivity parameter. At present we use a nearest-neighbour (NN) recogniser but in principle a number of recognition algorithms could be used. This automated step returns a list of ‘hits’ some number of which will be false positive errors. The recogniser will also have missed some true calls (false negatives).

The user now identifies and corrects errors (see Error Correction box in Fig. 2) and adds new examples of calls including those wrongly identified in the previous scan. The expanded library is now used as the basis for a second scan of the entire recording.

The above process is iterated until all vocalisations of interest have been annotated. Note that iterative identification and annotation of vocalisations builds up a library that not only covers the species range but also the variation within species for that location. Since calls in the library are annotated with their location, filtering for geographic proximity reduces the number of vocalisations to be compared.

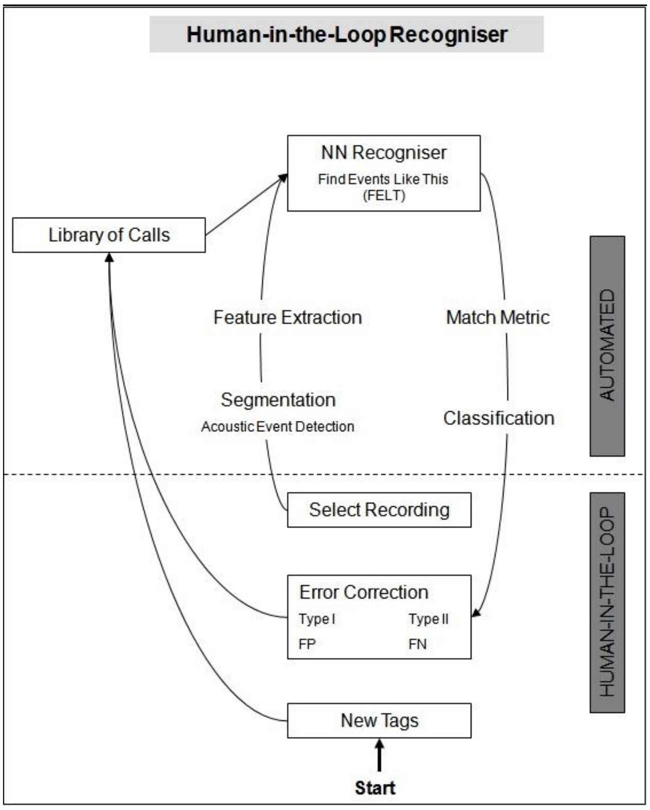

Figure 2. Semi-automated analysis (human-in-the-loop).

To give some idea of the performance of the nearest-neighbour recogniser (which also requires any similarity measure to exceed a threshold for positive identification) we used it to detect Bush Stone Curlew calls in a two hour recording. We used a single template describing just one of the several syllable types characteristic of a Bush Stone Curlew call. Dividing the recording into 4-minute segments, the single syllable recogniser achieved a recall rate of 63%, precision of 100% and accuracy of 76%. The addition of more syllables to the call library would increase recognition performance correspondingly.

IV. Discussion and Future Work

Acoustic sensors are set to play an important role in protecting biodiversity as we face increasing environmental challenges. Sensors provide scientists with the capability to collect data over large spatial and temporal scales, far exceeding what would be traditionally possible using manual methods. With this ability however comes the problem of analysing large volumes of data. This research has developed a ‘toolbox’ approach to the analysis of acoustic data, while providing scientists with an online environment to store, access and collaborate on data collected from acoustic sensors. Ultimately, these tools are aimed at providing scientists with the ability to detect and identify species in recordings. This information can be analysed over time to observe fluctuations in species richness, detect the presence of rare or invasive species, and to monitor the effects of climate change on the environment.

The automated recognition of animal calls has not yet reached a level of reliability that allows ecologists to use the methods without careful verification of results. Any application which offers analysis tools to ecologists must necessarily offer graded levels of utility from fully manual to fully automated. In particular it makes sense to offer semi-automated tools which allow an adjustable degree of user interaction with the data.

To this end, the automated recognisers in our Acoustic Environmental Workbench have a number of features that adapt them to the real world of semi-automated classification as opposed to the optimised world of a specialised machine learning laboratory. In particular:

1) We have constructed generic classifiers that respond to a particular feature which is common to many animal calls. The most obvious example in our work is the Oscillation Detector. Another feature of our generic recognisers is that they have parameters whose tuning is relatively intuitive. The only exception to this rule is the use of HMMs in HTK. These classifiers require IT expertise to construct. Reporting the accuracy of call classifiers based on carefully prepared data sets is not an accurate reflection of the typical ecologist’s requirements.

2) Except for our HMM classifiers, we have prepared generic classifiers that can be ‘trained’ with very few (even just one) instance. This is necessary because many bird species of interest are cryptic. As more calls are identified, the classifier can be improved in a boot-strap manner.

3) We have constructed classifiers that can be used in *both* a multi-class context (e.g. as a nearest-neighbour classifier) or as

stand-alone binary classifiers. The latter option is necessary because in many situations an ecologist is interested in a particular species and has no need of a classifier that recognisers multiple species. The difficulty to be solved in order to achieve this outcome is to normalise classification scores independently over a broad range of call types.

The identification of animal calls in arbitrary recordings of the environment remains a difficult task. We believe that it is more difficult than ASR, which is only just becoming a reliable technology after three decades and huge investment. From an economic standpoint alone, it is most unlikely that automated recognition of animal vocalisations will be achieved in the near future, certainly not having sufficient accuracy to replace human identification. Consequently human-in-the-loop will be required for analysis of environmental acoustic data for the foreseeable future. Our workbench recognises this reality, however, we anticipate that we will continue to improve on the accuracy of both semi-automated and fully-automated identification of species and these features will be added to the on-line acoustic workbench as they become available.

Acknowledgments

The Microsoft QUT eResearch Centre is funded by the Queensland State Government under a Smart State Innovation Fund (National and International Research Alliances Program), Microsoft Research and QUT. This research was conducted with the support of the QUT Institute of Sustainable Resources and the QUT Samford Ecological Research Facility.

References

[1] E. Nagy, *et al.*, "New Eyes on the World: Advanced Sensors for Ecology," *Bioscience,* vol. 59, p. 385, 2009.

[2] R. Mason, *et al.*, "Towards an Acoustic Environmental Observatory," in *eScience, 2008. eScience '08. IEEE Fourth International Conference on,* 2008, pp. 135-142.

[3] D. Cayan, *et al.*, "The wireless watershed at the Santa Margarita Ecological Reserve," *Southwest Hydrology,* vol. 2, pp. 18-19, 2003.

[4] S. L. Collins, *et al.*, "New opportunities in ecological sensing using wireless sensor networks," *Frontiers in Ecology and the Environment*, vol. 4, pp. 402-407, 2006.

[5] J. K. Hart and K. Martinez, "Environmental Sensor Networks: A revolution in the earth system science?," *Earth-Science Reviews,* vol. 78, pp. 177-191, 2006.

[6] A. Celis-Murillo, *et al.*, "Using soundscape recordings to estimate bird species abundance, richness, and composition," *Journal of Field Ornithology,* vol. 80, pp. 64-78, 2009.

[7] J. Haselmayer and J. S. Quinn, "A Comparison of Point Counts and Sound Recording as Bird Survey Methods in Amazonian Southeast Peru," *The Condor,* vol. 102, pp. 887-893, 2000.

[8] T. Penman, *et al.*, "A cost-benefit analysis of automated call recorders," *Applied Herpetology,* vol. 2, pp. 389-400, 2005.

[9] M. Thompson, *et al.*, "Heard but not seen: an acoustic survey of the African forest elephant population at Kakum Conservation Area, Ghana," *African Journal of Ecology,* vol. 48, pp. 224-231, 2009.

[10] K. Riede, "Monitoring Biodiversity: Analysis of Amazonian Rainforest Sounds," *Ambio,* vol. 22, pp. 546-548, 1993.

[11] A. Ali, *et al.*, "An empirical study of collaborative acoustic source localization," *Journal of Signal Processing Systems,* vol. 57, pp. 415-436, 2009.

[12] L. E. Freitag and P. L. Tyack, "Passive acoustic localization of the Atlantic bottlenose dolphin using whistles and echolocation clicks," *The Journal of the Acoustical Society of America,* vol. 93, pp. 2197-2205, 1993.

[13] J. Sueur, *et al.*, "Rapid Acoustic Survey for Biodiversity Appraisal," *PLoS ONE,* vol. 3, p. e4065, 2008.

[14] S. H. Gage, *et al.*, "Assessment of ecosystem biodiversity by acoustic diversity indices," *The Journal of the Acoustical Society of America,* vol. 109, pp. 2430-2430, 2001.

[15] J. Porter, *et al.*, "Wireless Sensor Networks for Ecology," *BioScience,* vol. 55, pp. 561-572, 2005.

[16] R. Butler, *et al.*, "Cyberinfrastructure for the analysis of ecological acoustic sensor data: a use case study in grid deployment," *Cluster Computing,* vol. 10, pp. 301-310, 2007.

[17] S. E. Moore, *et al.*, "Listening for Large Whales in the Offshore Waters of Alaska," *BioScience,* vol. 56, pp. 49-55, 2006.

[18] D. Mellinger, *et al.*, "Fixed Passive Acoustic Observation Methods for Cetaceans," *Oceanography,* vol. 20, p. 36, 2007.

[19] E. Fegraus, *et al.*, "Maximizing the value of ecological data with structured metadata: An introduction to ecological metadata language (EML) and principles for metadata creation," *Bulletin of the Ecological Society of America,* vol. 86, pp. 158-168, 2005.

[20] M. A. Acevedo, *et al.*, "Automated classification of bird and amphibian calls using machine learning: A comparison of methods," *Ecological Informatics,* vol. 4, pp. 206-214, 2009.

[21] S. T. Brandes, "Automated sound recording and analysis techniques for bird surveys and conservation," *Bird Conservation International,* vol. 18, pp. S163-S173, 2008.

[22] J. Cai, *et al.*, "Sensor network for the monitoring of ecosystem: Bird species recognition," in *Third International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Melbourne, 2007, pp. 293–298.

[23] Z. Chen and R. Maher, "Semi-automatic classification of bird vocalizations using spectral peak tracks," *The Journal of the Acoustical Society of America,* vol. 120, p. 2974, 2006.

[24] C. Juang and T. Chen, "Birdsong recognition using prediction-based recurrent neural fuzzy networks," *Neurocomputing,* vol. 71, pp. 121-130, 2007.

[25] C. Kwan, *et al.*, "Bird classification algorithms: theory and experimental results," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04).* 2004, pp. V-289-92 vol.5.

[26] A. L. McIlraith and H. C. Card, "Birdsong recognition using backpropagation and multivariate statistics," *IEEE Transactions on Signal Processing,* vol. 45, pp. 2740-2748, 1997.

[27] P. Somervuo, *et al.*, "Parametric Representations of Bird Sounds for Automatic Species Recognition," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 14, pp. 2252-2263, 2006.

[28] S. Anderson, *et al.*, "Template-based automatic recognition of birdsong syllables from continuous recordings," *Journal of the Acoustical Society of America,* vol. 100, pp. 1209-1219, 1996.

[29] I. Agranat, "Automatically Identifying Animal Species from their Vocalizations," presented at the Fifth International Conference on Bio-Acoustics, Holywell Park, 2009.

[30] S. Young, *et al.*, *The HTK book (for HTK version 3.4)*. Cambridge: Cambridge University Engineering Department, 2006.

[31] B. Planitz and M. Towsey, "Technical Report: A Toolbox for Animal Call Recognition," Queensland University of Technology, Brisbane 2010.

[32] M. Towsey and B. Planitz, "Technical Report: Acoustic analysis of the natural environment," Queensland University of Technology, Brisbane 2010.