ROBUST IN-FIELD AUTOMATED BIOACOUSTIC IDENTIFICATION OF SPECIES FOR RAPID BIODIVERSITY STUDIES

D. Chesmore Departrment of Electronics, University of York, Heslington, York, YO10 5DD, UK

1 INTRODUCTION

Recognition of birds, insects and mammals by means of their calls is widely used for bioacoustic species inventory and biodiversity studies, particularly for insects, birds and mammals [1, 2, 3]. Such surveys are time consuming and often require many years of training. They also commonly take place over short periods and at infrequent intervals (primarily because of the time required) leading to difficulties in data interpretation due to undersampling. Also, there is an increasing taxonomic impediment which means that there are too few trained taxonomists in a world where biodiversity is changing rapidly due to habitat loss [4, 5]. It is now feasible to develop computer-based systems to achieve automated bioacoustic species inventory with a number of significant advantages over manual methods including continuous long term unattended operation, automatic species logging and behavioural studies. This paper describes the development and testing of a field deployable species identification system which is aimed at addressing this problem in part by providing robust automated species identification for use by professional taxonomists, parataxonomists and ecologists.

2 AUTOMATED BIOACOUSTIC TAXON IDENTIFICATION

Automated identification of taxa can be achieved in a number of ways – image processing, acoustics, radar/sonar and other techniques. This paper only discusses acoustic applications which can potentially involve insects, birds, amphibia and mammals. Recent research into automated identification of insects has focussed on quarantine wood boring Coleoptera such as *Anoplophora glabripennis, A. chinensis* and *Hylotrupes bajulus* [6] and the development of non-invasive tools for determining the presence of Stag Beetle (*Lucanus cervus*) larvae in tree stumps [7]. These systems use low-cost vibration sensors to detect bites and stridulation from larvae with identification achieved using Time Domain Signal Coding (TDSC) for feature extraction. Other acoustic applications for insect identification include Orthoptera (grasshoppers, crickets and bushcrickets) [8, 9, 10, 11, 12, 13], cicadas [14], mosquitoes [15] and leafhoppers [16]. Systems also exist for the bioacoustic identification of migrating birds [17], birds in general [18, 19] and identifying individual elements of birdsong [20, 21, 22]. Relatively little work has been carried out on amphibian acoustic identification. One Australian project developed a field-deployable system to identify the presence of Cane Toads (*Bufo marinus*) in Australia [23]. To date, there have been relatively few applications involving automated identification based on bioacoustics. Most work on mammals has been on aquatic species such as dolphins and whales [24] and bats, recognising species from their echolocation calls [25, 26, 27].

One prevalent problem with bioacoustic signal identification is caused by the presence of interfering signals such as other taxa (same or different taxonomic group) and man-made sounds. As a consequence, the system described here is capable of not only recognising a number of target taxa but also other sounds including aircraft, vehicles and birds. This leads naturally to consideration of a generalised soundscape element identifier. The system described here is an initial attempt at recognising general sound categories divided into the following:

- biophony – all natural sounds of biological origin (excluding human sounds);
- geophony – all natural sounds of geological origin, e.g. earthquakes, water, precipitation and other meteorological phenomena.
- anthropophony – all sounds produced by man-made activities including speech, music, vehicles, alarms, etc.

Specifically, the system has been designed to recognise 13 sounds in the above three categories as indicated in Table 1.

Label	Sound Source	Sound Type
Ov	*Omocestus viridulus* (Linnaeus) (Orthoptera: Acrididae)	Biophonic – grasshopper
Mm	*Mymeleotettix maculatus* (Thunberg) (Orthoptera: Acrididae)	Biophonic – grasshopper
Cp	*Chorthippus parallelus* (Zetterstedt) (Orthoptera: Acrididae)	Biophonic – grasshopper
Ca	*Chorthippus albomarginatus* (De Geer) (Orthoptera: Acrididae)	Biophonic – grasshopper
Fl	Blowfly (Diptera, unknown species)	Biophonic – fly
B1	Unknown bird alarm call, species 1	Biophonic – bird
B2	Unknown bird alarm call, species 2	Biophonic – bird
B3	Unknown bird alarm call, species 3	Biophonic – bird
B4	Chiffchaff *Phylloscopus collybita*) (Vieill.) (Sylviidae)	Biophonic – bird
C1	Saloon-style vehicle on metalled surface	Anthropophonic
C2	Saloon-style vehicle on unmetalled surface	Anthropophonic
Pl	Single engine light aircraft	Anthropophonic
Bg	Background sounds (mainly wind)	Geophonic

Table 1. List of sounds the system has been trained to identify

3 SYSTEM DESIGN

Figure 1 shows a block diagram of the system which is divided into the following function blocks:

a) **Signal input**. This is a .wav encoded file obtained from a digital recording, 16 bits at a sampling rate of 44.1kHz. Recordings were made on various devices including a Sonifex Courier digital recorder and Marantz PM660 using low cost microphones.
b) **TDS coder**. This performs TDSC as described in Section 3.1 and generates an A-matrix as the feature set for classification.
c) **Sound classifier**. A multilayer perceptron (MLP) artificial neural network (ANN) which is trained to recognise a number of sounds.
d) **General sound classifier**. This groups sounds according to higher classes, in this case biophony (two sub-groups – insects and birds), geophony and anthropophony.

3.1 Feature Extraction using Time Domain Signal Coding

TDSC is based on time encoded speech (TES) developed by King and Gosling in 1978 [28] as a means of transmitting speech at very low data rates. It is based on encoding the shape of a waveform (often as the number of minima) between successive zero-crossings (termed an epoch) as indicated in Figure 2. TDSC extends the basic approach by incorporating amplitude scaling to overcome the loss of amplitude information in the encoding process. Each epoch is described by a doublet: the duration (D) in number of samples and the shape (S) as the number of positive minima or negative maxima. The number of possible combinations of (D, S) can be very large and TDSC employs a non-linear mapping to create a single codeword for each epoch.

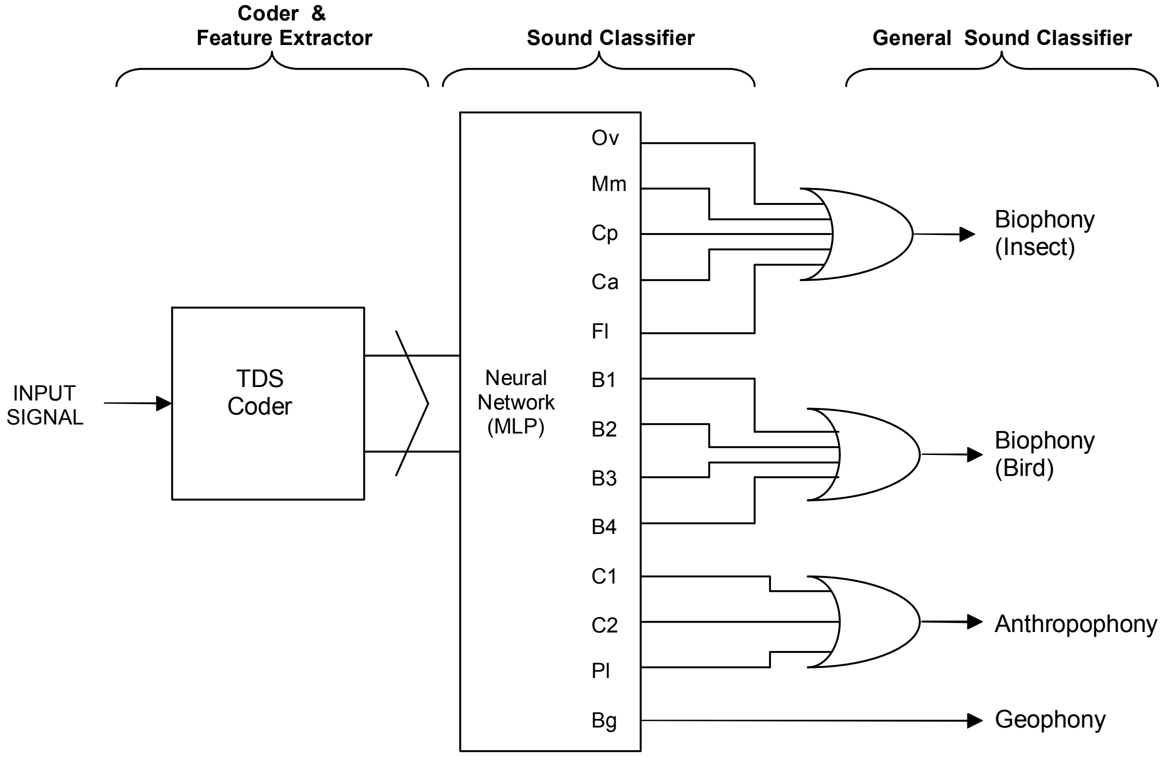


Figure 1. Block diagram of the identification system showing classification into individual sounds and general categories. Symbols Ov, etc are described Table 1.

The histogram of the frequency of occurrence of codewords over a given time interval is termed the S-matrix and a second matrix, the A-matrix, describes the number of occurrences of codeword i followed by codeword j as given by:

$$a_{ij} = \frac{1}{(N-1)} \sum_{n=2}^{N} x_{ij}(n) \qquad (1)$$

where a_{ij} = element (i,j) of matrix **A**
$x_{ij}(n)$ = 1 if t(n) = i and t(n-L) = j (0 otherwise)
L = lag (L=1 in this case)
and t(n) = n^{th} codeword

The A-matrix is a fixed size histogram (currently 28x28 codewords) with time-invariant dimensions representing the conditional probability of finding pairs of codewords and is the feature set employed here for classification purposes using a MLP neural network. It has been shown that the A-matrix provides a good feature for classification in a variety of acoustic applications ranging from fault detection in gearboxes [29] to heart defect classification [30] and insect identification [6, 8, 9]. The system is trained on mulitple examples of sounds, e.g. ten echemes for each species, and then tested on previously unseen signals.

3.2 Sound Classification

Classification is carried out using a MLP neural network with 28^2=784 inputs (one for each entry in the A-matrix) and 13 outputs, one for each sound to be classified (Table 1). The number of neurons in the hidden layer is variable and all tests described here use 40 neurons which is considered most

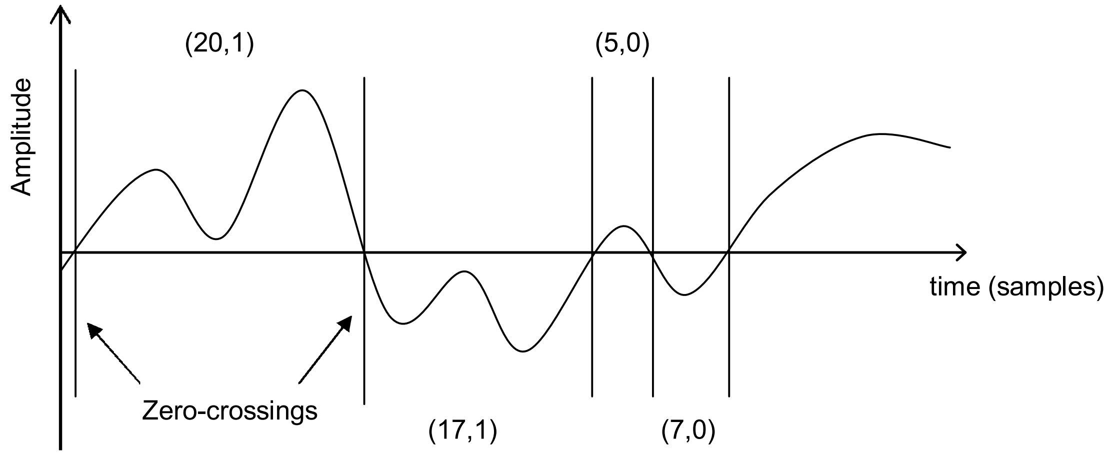


Figure 2. Principle of Time Domain Signal Coding

suitable (by trial and error). The MLP has been trained with exemplars of each sound type and then tested with new signals. The “winning” neuron is detemined as follows:

$$i_{max} : Q(i_{max}) = \underset{i=1..N}{MAX}\{Q(i)\} \tag{2}$$

where $Q(i) = i^{th}$ neuron
$Q(i_{MAX})$ = the neuron with the largest output
i_{MAX} = the number of the neuron with the largest output

The sound is therefore identified as sound $I=i_{MAX}$.

A number of tests have been carried out, concentrating on the identification of the four grasshopper species. Two specific tests were carried out, one for single echemes (equivalent to syllables) and one for the whole song. The trained system was then tested on previously unseen signals. Table 2 gives results for single echemes. In the first test, codewords were accumulated over the period of a typical echeme (2s) and an A-matrix created for each echeme. The lowest identification accuracy is 58.8% (Mm) and the highest 97.1% (Ov); this include all 13 sounds. However, despite the low result for Mm it is evident that there is no misidentification which is vital as far as a taxonomist, and therefore acceptability, is concerned.

	Ov	**Mm**	**Cp**	**Ca**	**Other**
Ov (34)	**97.1%**				2.9%
Mm (17)	5.9%	**58.8%**	11.8%	17.6%	5.9%
Cp (19)	10.5%	10.5%	**73.7%**	5.3%	
Ca (16)	6.3%		6.3%	**87.4%**	

Table 2. Identification accuracy for four grasshopper species based on single echemes. Numbers in brackets are the number of sounds included in the tests. Other denotes any of the remaining nine sounds.

The second test shown in Figure 3 is for a whole song. In the this test, the same approach has been taken but the A-matrix is created from the whole song, typically 5-10s. Also, a threshold is applied to the winning output neuron which has the range 0.0 – 1.0, applied by:

$$I = i_{MAX} \text{ if } i_{MAX} : Q(i_{MAX}) = \underset{i=1..N}{MAX}\{Q(i)\} \geq T_{REJ} \text{ else } I=0 \qquad (3)$$

where $0.0 < T_{REJ} \leq 1.0$

A low neuron value indicates a poor identification so the addition of a rejection threshold will remove these, resulting in a “don’t know” state; this is preferable to an incorrect identification. Figure 3 shows that a higher identification accuracy occurs for a higher threshold with Ov (Common Green grasshopper) achieving 100% and the lowest being Cp (Meadow grasshopper) at 80% for a threshold of 0.9.

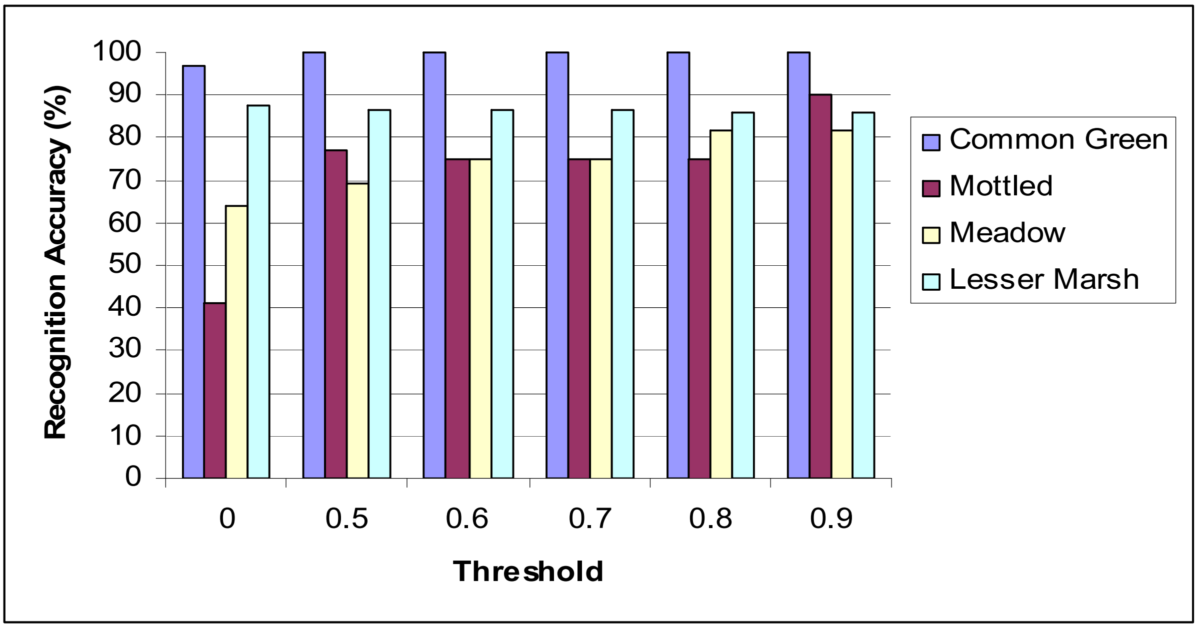


Figure 3. Recognition accuracy for four grasshopper species based on whole song.

3.3 Higher Level Classification

Classification to a higher level is achieved by grouping the sounds into categories as follows:

Biophony–insect = $(Ov \cup Mm \cup Cp \cup Ca \cup Fl)$

Biophony–bird = $(B1 \cup B2 \cup B3 \cup B4)$

Biophony = biophony-insect $\cup$ Biophony-bird

Geophony = (Bg)

Anthropophony = $(C1 \cup C2 \cup Pl)$

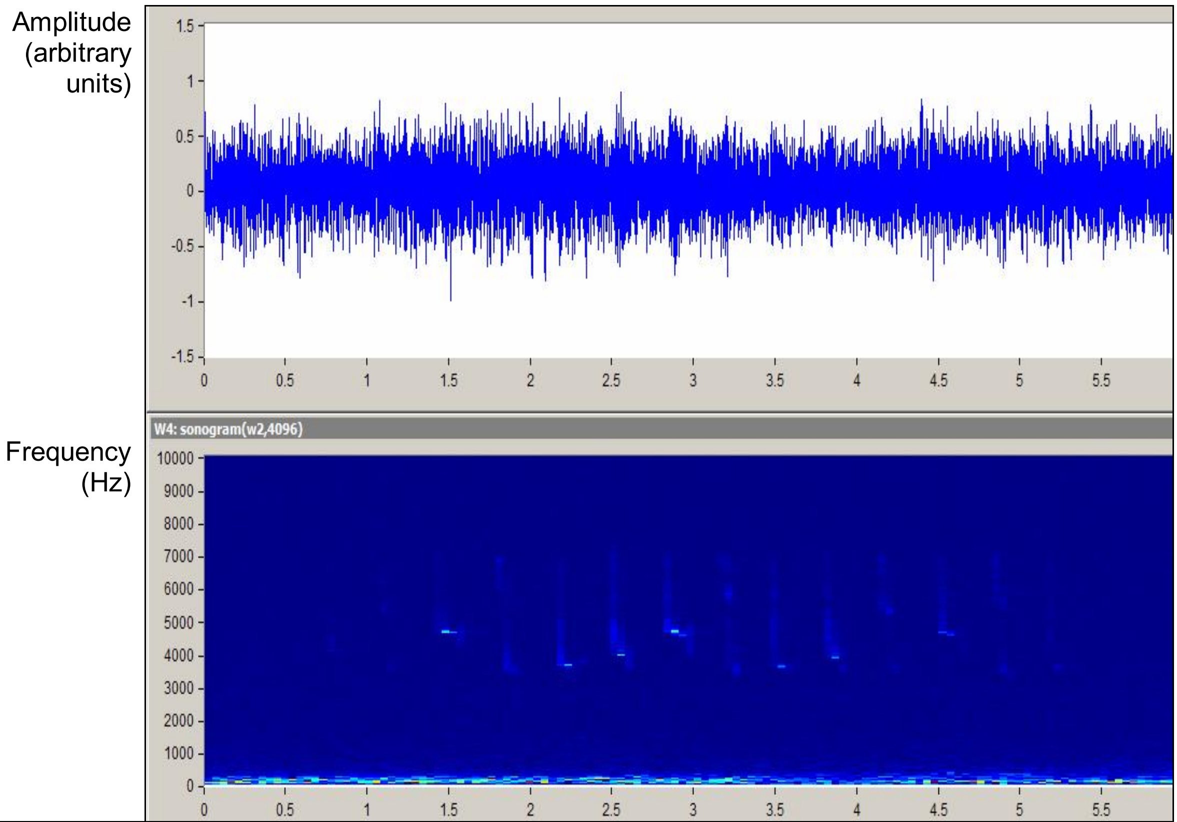
In Figure 1, this grouping is represented by OR gates, with the winning neuron output set to 1 and all others set to 0. Figure 4 illustrates the identification of sounds on a 1 second sample. The top graph is the time domain waveform, below is a spectrogram showing the characteristic call of a Chiffchaff and a low frequency component due to an aircraft. The values show the neuron outputs (0-1) indicating that the bird and aircraft are correctly recognised in blocks 1 to 4. It is evident that the system selects the dominant sound, for example, the aircraft dominates in block 6 even though the bird is still singing albeit quieter.

4 CONCLUSIONS

The paper has described a method for automatically identifying sounds offline, currently within a recording. Use of TDSC as a signal coder and feature generator is computationally simple and lends itself to implementation on a microcontroller-based system. It is therefore possible to create field deployable species identification systems of various forms including hand-held, data-logging and radio networked to provide large scale coverage.

The number of sounds to be classified depends on application. For example, it may be necessary to recognise up to 30 species of Orthoptera or, at the other extreme, optimally recognise only one sound and rejecting all others. The latter approach is well suited to the detection and identification of pest species for area protection as illustrated in Figure 5 where a vulnerable area is protected by a “ring-fence” of sensors. Depending on the distance between sensors, the links may be cable or radio, the latter enabling large areas to be protected.

At present, the system does not take into account unknown sounds; these will therefore be classified as one of the thirteen known sounds. However, unless the unknown sound is close to any of the sounds the system is trained with, the neural outputs will all have low values which will fall below the threshold and therefore be rejected.



Block (s)		1	2	3	4	5	6
Chiff-chaff	B		1	1	1		
Aircraft	A	0.45				1	0.16
Background	G	0.43					0.18

Figure 4. Recognition of sounds on a one second block length. Top graph is the time domain waveform; bottom graph is the spectrogram showing the characteristic call of the Chiffchaff. The aircraft sound is visible below 1kHz. B = biophony, A = anthropophony, G = geophony

REFERENCES

1. S.R. Baillie, ‘Monitoring terrestrial bird populations’, in ‘Monitoring for Conservation and Ecology’, Goldsmith, F.B. Chapman and Hall, 112-132. (1995).
2. FP. Fischer, U. Schulz, H. Schubert, P. Knapp and M. Schmoger, ‘Quantitative assessment of grassland quality: acoustic determination of population sizes of orthopteran indicator species’, Ecol. Appl., 7(3) 909-920. (1997).
3. T. Gardiner, J. Hill and D. Chesmore, ‘Review of the methods frequently used to estimate the abundance of Orthoptera in grassland ecosystems’, J. Insect Cons., 9(3), 151-174. (2005).
4. K. Gaston and M.A. O'Neill, ‘Automated species identification -- why not?’, Phil. Trans. Royal Soc. B*,* 359(1444), 655-667. (2004).
5. K. Riede, ‘Monitoring biodiversity: analysis of Amazonian rainforest sounds’. Ambio 22, 546-548. (1993).
6. I.J. Farr E.D. and Chesmore, E.D. ‘Acoustic Detection and recognition of wood-boring insects’, Royal Entomological Society National Meeting: Entomology 2005, University of Sussex. (September 2005).
7. I.J. Farr, D. Chesmore, D. Harvey, C. Hawes and A. Gange, ‘Bioacoustic detection and recognition of Stag Beetle (*Lucanus cervus*) larvae underground using vibration sensors’, Royal Entomological Society National Meeting: Entomology 2005, University of Sussex. (September 2005).
8. E.D. Chesmore and C. Nellenbach, Acoustic Methods for the Automated Detection and Identification of Insects’, Acta Horticulturae, 562, 223-231. (2001).
9. E.D. Chesmore and E. Ohya, ‘Automated identification of field-recorded songs of four British grasshoppers using bioacoustic signal recognition’, Bull. Ent. Res.*,* 94(4), 319-330 (2004).
10. F. Schwenker, C. Dietrich, H.A. Kestler, K. Riede and G. Palm, ‘Radial basis function neural networks and temporal fusion for the classification of bioacoustic time series’, Neurocomputing*,* 51, 265-275. (2003).
11. C. Dietrich, G. Palm and F. Schwenker, ‘Decision templates for the classification of bioacoustic time series’, Information Fusion*,* 4, 101-109. (2003).
12. C. Dietrich, G. Palm, K. Riede and F. Schwenker, ‘Classification of Bioacoustic Time Series based on the combination of global and local decision’, Int. J. Pattern Rec., 37, 2293-2305. (2004).
13. E.D. Chesmore, ‘Automated bioacoustic identification of species’, Anais da Academia Brasileira de Cienci*as*, 76(2), 435-440. (2004).
14. E. Ohya, Identification of *Tibicen* cicada species by a principal components analysis of their songs, Anais da Academia Brasileira de Ciencias, 76(2), 441-444. (2004).
15. R.H. Campbell, S.K. Martin, I. Schneider and W.R. Michelson, ‘Analysis of mosquito wing beat sound’, 132nd Meeting of the Acoustical Society of America, Honolulu, 1996.
16. C.D. Dietrich and C.D. Pooley, ‘Automated identification of leafhoppers (Homoptera: Cicadellidae: *Draeculacephala* Ball)’, Annals Ent. Soc. Am., 87(4), 412-423. (1994).
17. H. Mills, ‘Automatic detection and classification of nocturnal migrant bird calls’, J. Acoust. Soc. Am. 97, 3370-3371. (1995).
18. E.D. Chesmore, ‘Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals’, J. Appl. Acoust.*,* 62, 1359-1374. (2001).
19. A.L. McIraith and H.C. Card, ‘Birdsong recognition with DSP and neural networks’, IEEE WESCANEX’95 Proceedings, 409-414. (1995).
20. S.E. Anderson, A.S. Dave and D. Margoliash, ‘Template-based automatic recognition of birdsong syllables from continuous recordings’, J. Acoust. Soc. Am.100, 1209-1217. (1996).
21. J.A. Kogan and D. Margoliash, ‘Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models, a comparative study’, J. Acoust. Soc. Am., 103(4), 2185-2196. (1998).
22. A.M.R. Terry and P.K. McGregor, ‘Census and monitoring based on individually identifiable vocalizations, the role of neural networks’, Animal Cons., 5, 103-111. (2002).

23. A. Taylor, G. Grigg, G. Watson and H. McCallum, 'Monitoring frog communities, an application of machine learning', Eighth Innovative Applications of Artificial Intelligence Conference, AAAI Press. (1996).
24. S.O. Murray, E. Mercado and H.L. Roitblat, 'Characterizing the graded structure of false killer whale (*Pseudorca crassidens*) vocalizations', J. Acoust. Soc. Am., 104(3), 1679-1688. (1998).
25. N. Vaughan, G. Jones and S. Harris, 'Identification of British bat species by multivariate analysis of echolocation call parameters', Bioacoustics 7, 189-207. (1996).
26. S. Parsons, 'Identification of New Zealand bats in flight from analysis of echolocation calls by artificial neural networks', J. Zool (London) 253, 447-456. (2001).
27. S. Parsons and G. Jones, 'Acoustic identification of 12 species of echolocating bat by discriminant function analysis and artificial neural networks'. J. Exp. Biol. 203, 2641-2656. (2000).
28. R.A. King and W. Gosling, 'Time-encoded speech', Electronics Letters 15, 1456-1457. (1978).
29. W.G. Lucking, M. Darnell and E.D. Chesmore, 'Acoustical condition monitoring of a mechanical gearbox using artificial neural networks', IEEE Conference on Neural Networks, 28 June-2 July 1994, part of the IEEE World Congress on Computational Intelligence, Orlando, Florida (1994).
30. M.D. Swarbrick and E.D. Chesmore, 'Automatic classification of heart sounds and murmurs using time domain signal coding (TDSC), artificial neural networks and expert systems', Conf. on Recent Advances in Soft Computing 98, DeMontfort University, 37-46. (1998).

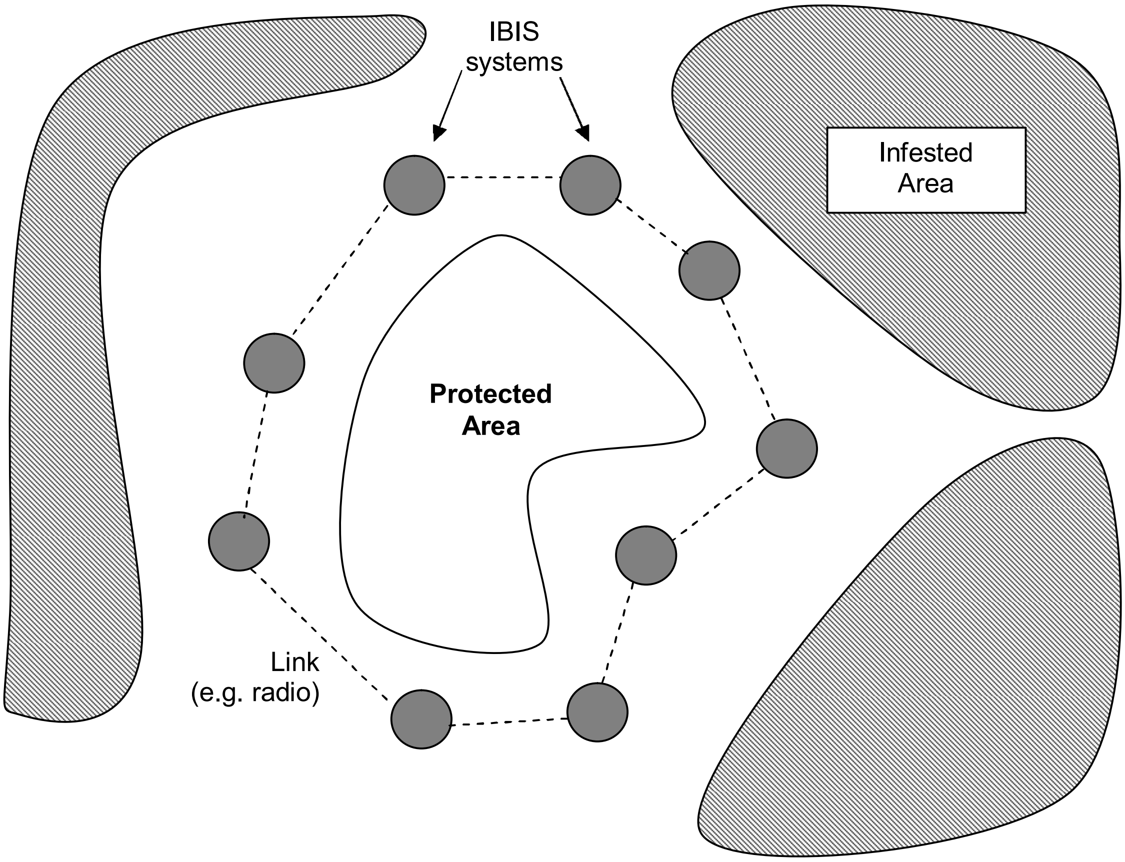


Figure 5. Example of ring-fencing protection of vulnerable areas.