Automated identification of field-recorded songs of four British grasshoppers using bioacoustic signal recognition

Article *in* Bulletin of Entomological Research · September 2004

Impact Factor: 1.91 · DOI: 10.1079/BER2004306 · Source: PubMed

2 authors:

David Chesmore
The University of York

Eiji Ohya
Forestry and Forest Products Research I…

Bulletin of Entomological Research (2004) **94**, 319–330 DOI: 10.1079/BER2004306

Automated identification of field-recorded songs of four British grasshoppers using bioacoustic signal recognition

E.D. Chesmore[1]* and E. Ohya[2]

[1]Department of Electronics, University of York, Heslington, York, YO10 5DD, UK: [2]Biodiversity Research Group, Tohoku Research Center, Forestry and Forest Products Research Institute, Shimokuriyagawa aza Nabeyashiki 92–25, Morioka 020–0123, Japan

Abstract

Recognition of Orthoptera species by means of their song is widely used in field work but requires expertise. It is now possible to develop computer-based systems to achieve the same task with a number of advantages including continuous long term unattended operation and automatic species logging. The system described here achieves automated discrimination between different species by utilizing a novel time domain signal coding technique and an artificial neural network. The system has previously been shown to recognize 25 species of British Orthoptera with 99% accuracy for good quality sounds. This paper tests the system on field recordings of four species of grasshopper in northern England in 2002 and shows that it is capable of not only correctly recognizing the target species under a range of acoustic conditions but also of recognizing other sounds such as birds and man-made sounds. Recognition accuracies for the four species of typically 70–100% are obtained for field recordings with varying sound intensities and background signals.

Introduction

Many animals produce sound either deliberately for communications (non-incidental) or as a by-product of their activity, such as eating, moving or flying (incidental sounds). Recognition of species from non-incidental sounds has been employed for many years for bird census, for identifying individuals and locating animals. Such surveys are carried out 'manually' and are slow, time consuming and rely strongly on the surveyor's expert knowledge of the group under investigation. Surveys also take place generally over short periods and at infrequent intervals mainly because of the time required, leading to difficulties in interpreting data. Due to rapid advances in computing and electronics it is now possible to develop automated recognition systems to provide long term continuous unattended monitoring in inhospitable regions. Such systems could be designed for hand held use and applications range from rapid biodiversity assessment especially in acoustically rich habitats (Riede, 1993) to electronic identification guides, acoustic autecology and the detection and recognition of pest species. Automated species identification using bioacoustics in entomology is still in its infancy and is more mature in other fields; examples include birds (McIlraith & Card, 1995; Anderson *et al.*, 1996), individual bird recognition and population counting (Terry & McGregor, 2002), nocturnal migrating birds (Mills, 1995), frogs and other amphibia (Taylor *et al.*, 1996), cetaceans (Murray *et al.*, 1998), deer (Reby *et al.*, 1997), elephants (Clemins & Johnson, 2002) and bats (Vaughan, *et al.*, 1996; Parsons & Jones, 2000; Parsons, 2001). Automated bioacoustic identification of insect species is still limited, examples include grasshoppers

*Fax: 01904 432335
E-mail: edc1@ohm.york.ac.uk

(Chesmore, 2001; Chesmore & Nellenbach, 2001; Chesmore *et al.*, 1998), crickets (Schwenker *et al.*, 2003), mosquitoes (Campbell *et al.*, 1996) and cicadas (Ohya & Chesmore, 2003).

In many applications it is simply sufficient to be able to detect the presence of animals, with species identification either unnecessary or a bonus. One of the major application areas for this is in pest detection where the presence of pests is primarily indicated by the generation of incidental sounds such as movement or feeding. Indeed, this is where most entomological acoustic research to date has been undertaken. The development of early warning and rapid detection systems would be of considerable commercial benefit and acoustic detection of many pest species is a viable alternative to trapping and for some species is the only method capable of detecting larvae. For example, in the USA, acoustics is used for detecting beetle larvae in rice grains (Shuman *et al.*, 1993, 1997). Other researchers have used similar techniques for monitoring *Rhizopertha dominica* Fabricius (Coleoptera: Bostrichidae) in wheat kernels (Hagstrum *et al.*, 1990). It is also possible to detect the presence of subterranean insect pests and stem borers (Mankin & Weaver, 2000; Mankin *et al.*, 2000), termites (Matsuoka *et al.*, 1996) and larvae feeding inside cotton bolls (Hickling *et al.*, 2000). More recently, attention has been focused on the detection of the Asian longhorn beetle *Anoplophora glabripennis* (Motschulsky) (Coleoptera: Cerambycidae) in live trees and solid wood packing materials (MacLeod *et al.*, 2002). It is a native of China and Korea where it is a major pest of hardwood trees such as elm, maple, willow and aspen (Haack *et al.*, 1997; MacLeod *et al.*, 2002). Early detection is essential to stop infestation and work is underway to develop systems capable of detecting early instar larvae.

The remainder of the paper describes the development of a novel bioacoustic signal recognition system and its application to the recognition of British Orthoptera. The technique employed is computationally simple and is known as time domain signal coding (TDSC) which, when coupled with an artificial neural network (ANN) classifier provides a powerful vehicle for bioacoustic signal analysis and recognition. It has previously been successfully tested on 25 species of British Orthoptera with 99% recognition accuracy (Chesmore *et al.*, 1998; Chesmore, 2000, 2001; Chesmore & Nellenbach, 2001) and ten species of Japanese bird with 100% accuracy (Chesmore, 1999, 2001). However, these results were for high signal to noise ratio (SNR) signals. This paper describes results of field trials where the SNR is more variable and sounds are corrupted by interference from other natural and man-made noise sources.

Materials and methods

Sound recordings and test data

Recordings were made between June and September 2002 at a variety of sites and habitats in North and East Yorkshire, England (table 1). The north of England does not have many species of Orthoptera and a total of six species were identified during the period, two of which were not used in the tests; these were *Tetrix undulata* (Sowerby) (Orthoptera: Tetrigidae) which does not use acoustic communication and *Conocephalus dorsalis* (Latreille) (Orthoptera: Tettigoniidae). *Conocephalus dorsalis* was not used in the study as only one specimen was found at a single locality. The four remaining species *Omocestus viridulus* (Linnaeus) (Orthoptera: Acrididae), *Chorthippus parallelus* (Zetterstedt) (Orthoptera: Acrididae), *Chorthippus albomarginatus* (Dr Geer) (Orthoptera: Acrididae) and *Mymeleotettix maculatus* (Thunberg) (Orthoptera: Acrididae) were used in the study. Sounds were recorded on a Sony MZ-R90 portable minidisc recorder with a Sony ECM-MS907 condenser microphone (frequency response 100 Hz-15 kHz) and transferred to a PC (Dell Inspiron 8100) via a Soundblaster sound card. The sounds were sampled at 44.1 kHz and stored as 16-bit signed mono .wav format files using the Avisoft-SASLab Pro software package. One point that must be emphasized here is that many bioacoustic researchers do not consider the minidisc to be appropriate for bioacoustics because the data compression algorithm is specific to speech and music, compression being achieved by utilizing knowledge of the human auditory system. The time and frequency domain signals of directly digitized Orthoptera sounds were compared with those from a minidisc and there were no significant differences. The recordings are also not being used for detailed bioacoustic analysis and the authors therefore consider the minidisc to be appropriate for this application. Minidisc recorders have a number of features useful for field recording – physically small, consume little power and compression enables longer recording times than digital audio tape (DAT) to be achieved.

Signal analysis and recognition

Figure 1 shows a schematic diagram of the species identification system which consists of the following functional blocks: (i) signal encoding using time domain signal coding to generate a code stream C(n); (ii) generation of A-matrices from the code stream; and (iii) identification of signal using a signal classifier, in this case a 3-layer multilayer perceptron artificial neural network (ANN).

The system operates in two phases – training phase and operational phase. In the training phase, high quality examples of sounds that are to be identified (known as exemplars) are used to train the ANN so that the correct ANN output is activated. For example, training for four acridid species requires a representative range of echemes (first order assemblage of syllables) or song samples from each species to be selected as exemplars, conversion of these into A-matrices and the network trained. Training occurs by repeated presentation of the sounds and modification of the weights within the network in such a way as to reduce the overall error between the current outputs and desired outputs. Training continues until the overall error is below a given threshold. In the example given, the number of ANN outputs might be four, each output representing a species. Once trained, the system is ready to use and unknown sounds can be classified. Each of the four outputs will give a value between 0.0 (zero match) and 1.0 (perfect match), the unknown sound being recognized as the output with the highest value.

Time domain signal coding

Time domain signal coding (TDSC) is based on a method known as time encoded speech (TES) which was developed in the 1970s (King & Gosling, 1978) as a purely time domain

Table 1. Location of recording sites in Yorkshire,UK in 2002.

Site	Grid reference	Owner	Habitat	Species recorded
Wheeldale Moor, North York Moors	SE 801 992	FC	Upland moorland	*Omocestus viridulus*
Cropton Forest, North York Moors	SE 826 952	FC	Pine plantations	*O. viridulus* *Tetrix undulata*
Wharram Quarry	SE 858 653	YWT	Disused chalk quarry	*O. viridulus* *Mymeleotettix maculatus*
Thixendale, East Yorkshire	SE 840 610	ERYCC	Chalk grassland	*O. viridulus*
Faxfleet, East Yorkshire	SE 864 241	ERYCC	Reed bed (River Humber)	*O. viridulus* *Chorthippus albomarginatus*
Allerthorpe Common	SE 762 474	YWT	Lowland wetland heath	*O. viridulus* *M. maculatus* *T. undulata*
Skipwith Common		EN	Lowland wetland heath	*O. viridulus* *M. maculatus* *T. undulata*
Spurn Point	TA 417 151	YWT	Salt marsh	*O. viridulus* *C. albomarginatus* *Conocephalus dorsalis*
Walmgate Stray, York	SE 617 508	Public land	Scrubland	*O. viridulus* *M. maculatus*

EN, English Nature; ERYCC, East Riding of Yorkshire County Council; FC, Forestry Commission; YWT, Yorkshire Wildlife Trust.

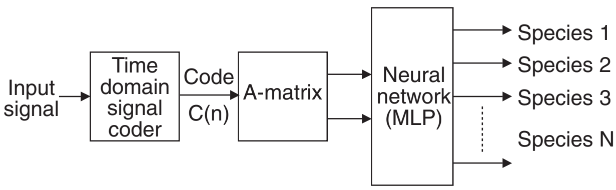

Fig. 1. Schematic diagram of identification system. The input signal is coded to give (D,S) pairs which are converted into a code per epoch C(n). Codes and lagged codes are accumulated for the duration of the complete sound and the resulting A-matrix forms the input to an artificial neural network. The result is a species identification per sound.

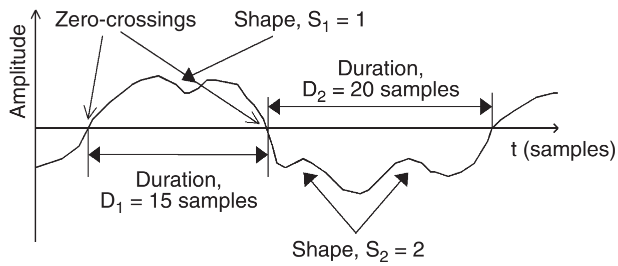

Fig. 2. Example of waveform epochs (D_1, S_1) and (D_2, S_2). Duration is taken as the number of samples between successive zero-crossings (termed an epoch) and shape as the number of minima in each epoch.

approach to the compression of speech for digital transmission over low capacity radio channels. Time encoded speech utilizes the fact that any bandlimited signal can be characterized by its real and complex zero locations; Time encoded speech only uses real zeros for simplicity. The basic concept of time encoded speech has been extended to include matrix normalization, matrix scaling and automated codebook generation (Swarbrick, 2001); the term time domain signal coding is now used to encompass these additions. Time domain signal coding has been used for a variety of acoustic applications including condition monitoring of machinery (Lucking *et al.*, 1994) and heart sound analysis and defect identification (Swarbrick & Chesmore, 1998; Swarbrick, 2001). Time domain signal coding is not limited to acoustic signals as it is possible to analyse any bandlimited signal.

Figure 2 shows the basic principle of time domain signal coding where the shape of the waveform between two successive zero-crossings is characterized; each waveform interval is termed an epoch and forms the basic unit for signal description. Each epoch is described in terms of its duration (D) in samples at the sampling rate (e.g. 44.1 kHz) and shape (S) usually taken as the number of positive minima (or negative maxima) as indicated in fig. 2. It is also possible to use signal energy and frequency-scaled signal energy as a measure of shape (Swarbrick, 2001). Time domain signal coding provides information on the fundamental frequency (D^{-1}) with the number of minima being related to the harmonic content of the signal. The number of possible D-S combinations is termed the natural alphabet which is unique to the overall structure of a given signal. Figure 3 shows D-S distributions for *C. parallelus* and *C. albomarginatus* on a logarithmic scale, indicating that the majority of durations are less than 20 samples (453 µs at 44.1 kHz sampling rate) and shape less than 10 for both species. This is also true for the other two species under consideration. The number of different D-S pairs can in fact be large and it is usual to non-linearly map them onto a smaller set known as a codebook, creating a single code for a range of D-S values. This mapping is also dependent on sound structure and will not necessarily be the same for different sound groups. The codebook must therefore be generated for each sound group by manual examination of the distribution of D-S pairs. A typical codebook may consist of 20–30 codewords. On examination of other Orthoptera species, the general D-S distributions follow the same pattern as in fig. 3, and so a single codebook of size 28 is suitable for all British Orthoptera. The output of the coding scheme is a stream of codewords which can be further analysed by two different methods: (i) a histogram of the frequency of occurrence of the codes in a given number of epochs – the S-matrix; and (ii) a two-dimensional histogram of the frequency of occurrence of pairs of codes – the A-matrix.

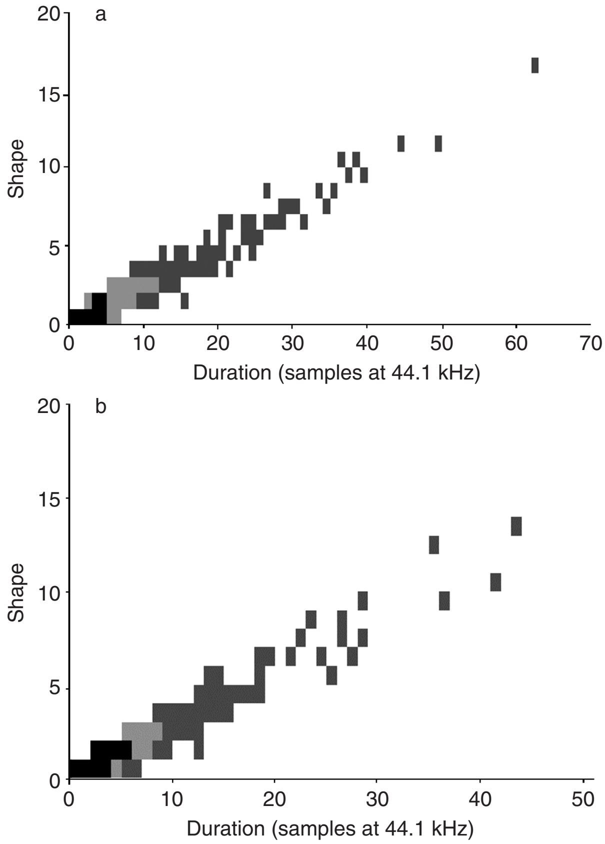

Fig. 3. Duration-shape distribution for (a) *Chorthippus parallelus* and (b) *C. albomarginatus*. The graph shows the number of occurrences of shape-duration values as a shaded block. The darker the shade, the more likely the occurrence. The intensity plot is logarithmic. Sample rate is 44.1 kHz.

Each entry in the S-matrix is given by:

$$s(i) = \sum_{j=1}^{N} x(j) \quad 1 \le i \le M$$

where s(i) = element (i) of matrix S; N = number of epochs in the signal; M = number of codes in codebook; j^{th} codeword in the signal; x(j) = 1 if x (j) = j (0 otherwise).

The A-matrix describes the number of occurrences of codeword i followed by codeword j as given by:

$$a_{ij} = \frac{1}{(N-1)} \sum_{n=2}^{N} x_{ij}(n) \quad 1 \le i \le M \quad 1 \le j \le M$$

where a_{ij} = element (i,j) of matrix A; L = lag; M = number of codes in codebook; $x_{ij}(n)$ = 1 if t(n) = i and t(n-L) = j (0 otherwise); and t(n) = n^{th} codeword

The A-matrix is a fixed size histogram with time-invariant dimensions representing the conditional probability of finding pairs of codewords and is the feature set used here for subsequent classification purposes using an artificial neural network. Figures 5 to 8 show S-matrices and A-matrices for the four acridid species under investigation.

Signal identification using backpropagation neural networks

Artificial neural networks are now widely used in many classification and identification problems as they can be trained, are good at handling fuzzy and disparate data and are able to perform non-linear discrimination (see, for example Schalkoff, 1992; Looney, 1997; Bishop, 2000). There are many forms of ANN which can be divided into supervised (requires training) and unsupervised classification (no training). Much of the research carried out in ANN applications to date uses multilayer perceptrons (MLP) using backpropagation for training. An algorithm compares the computed result with the expected results and the systematically modifies the weightings of individual neurons within the network. This application uses a multilayer perceptron as indicated in fig. 4; each neuron in the input layer is connected to the A-matrix, requiring 28^2 = 784 inputs. In general for an N codeword system, the size of the A-matrix is N^2. The number of outputs depends on the number of sounds to be identified. Here, 13 outputs represent the four target species and a variety of other sounds which represented the most common sound types recorded at the various sites. The number of neurons in the hidden layer can be varied and it is only possible to find the correct number through trial and error. Too few will result in the network not converging and it will never train; too many will simply increase the training time due to large numbers of calculations. In this application, 40 neurons give good convergence and short training time. Representative samples of high quality sound 2 s in duration were used to generate exemplars, A-matrices for each species and the ANN to train using a backpropagation algorithm (Looney, 1997).

In this application, the neuron outputs are scaled between 0.0 and 1.0 and the output with the highest value is selected as follows:

$$i_{MAX} : Q(i_{MAX}) = \max_{i=1..N} \{Q(i)\}$$

where Q(i) = ith neuron; $Q(i_{MAX})$ = the neuron with the largest output; and i_{MAX} = the number of the neuron with the largest output.

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

Results

The time domain waveforms, S-matrices and A-matrices for each of the four species of Orthoptera under investigation are shown in figures 5–8. It is evident that the A-matrices look similar whilst the S-matrices show differences in the proportion of codes in each signal. Despite the similarities between A-matrices, they are sufficiently different to be used as features for species recognition.

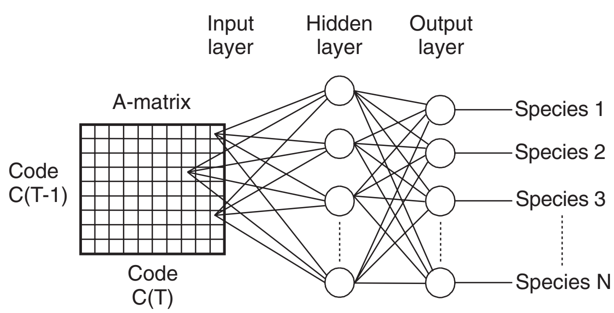

Fig. 4. Multilayer perceptron neural network architecture. This is a three layer network which is trained on A-matrices from each species. The A-matrix is described in the text.

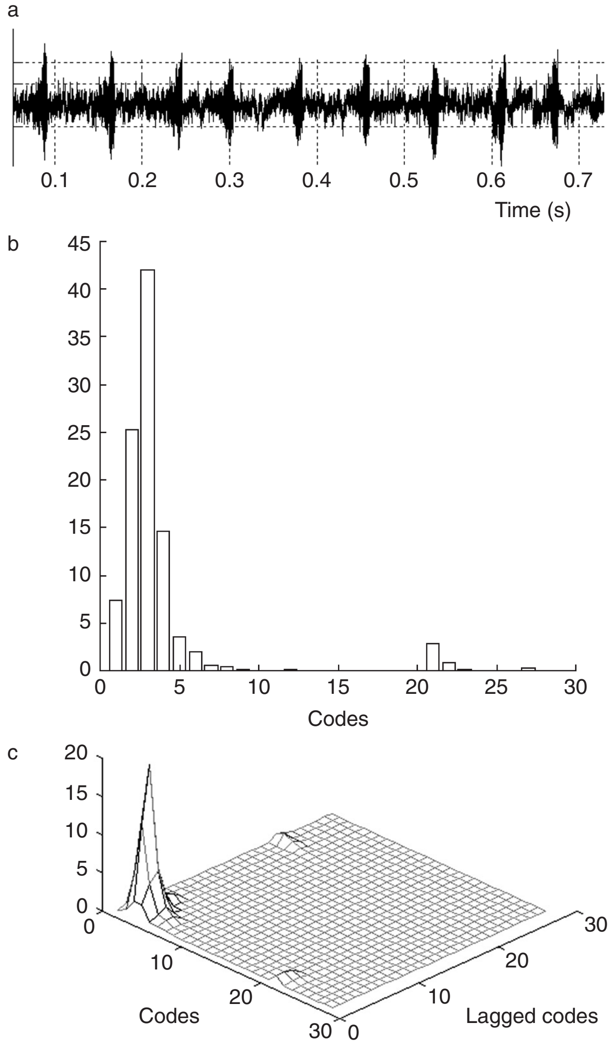

Fig. 5. *Omocestus viridulus* time waveform (a), S-matrix (b) and A-matrix (c). The vertical scale on the S- and A-matrices is percentage of code occurrence.

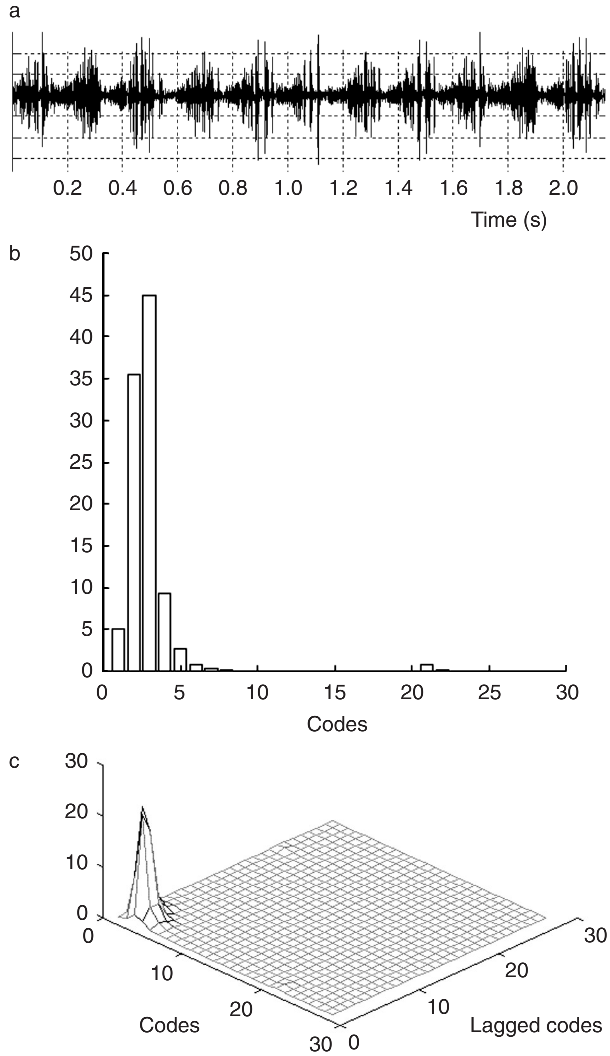

Fig. 6. *Chorthippus parallelus* time waveform (a), S-matrix (b) and A-matrix (c). The vertical scale on the S- and A-matrices is percentage of code occurrence.

Before training of the system was carried out, each recording was listened to, to determine the number of different sound types present. These were divided into three categories: (i) animal (including insects and birds) sounds; (ii) natural sounds (e.g. wind) and (iii) man-made sounds (e.g. vehicles and aircraft). Representative sounds for each category were selected, stored as separate .wav files and used to train the ANN. The following sound sources were used in training: (i) four grasshopper species; (ii) one blowfly sound (wing beats of unknown species); (iii) four bird sounds (three different alarm calls of undetermined origin and chiffchaff call *Phylloscopus collybita*) (Vieill.) (Passeriformes: Sylviidae); (iv) two vehicle (car) sounds (metalled road and dirt road); (v) one single engine light aircraft sound; and (vi) one background sound (sound when no other sources present – includes wind noise).

Figures 9 and 10 show time domain waveforms, S-matrices and A-matrices for two other sounds – a light engine aircraft and a blowfly flying past the microphone. These have been included to illustrate the differences between Orthoptera and non-Orthoptera sounds. These differences are due mainly to the low frequency components of aircraft sounds and fly wing beats.

Typically 10 × 2 s blocks of each sound type were selected from each source and used as the training set. The number of neurons in the hidden layer was selected at 40 as a compromise between slow training (fewer neurons) and over specification. The system was tested by considering the following possibilities: (i) single echeme recognition; (ii) whole song recognition; and (iii) generalized recognition of sounds from files in two second increments. This approach does not rely on extracting any grasshopper specific information (e.g. beginning and end of song) but simply allocates an identification for each 2 s sample.

Rejection threshold selection

It is evident upon examination of the data in tables 2 to 5 that in some cases all neuron outputs are low (see for

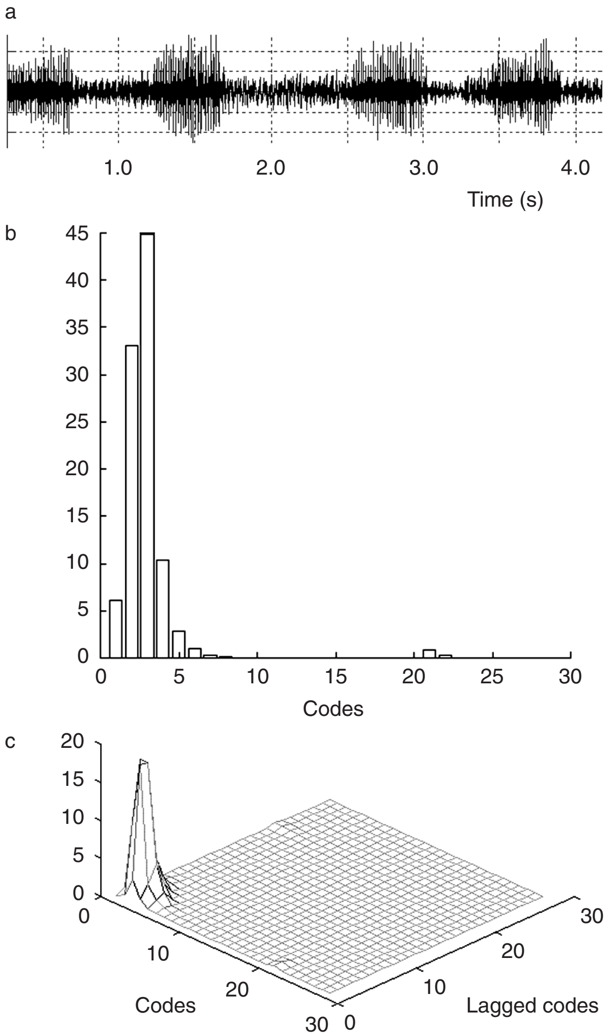

Fig. 7. *Chorthippus albomarginatus* time (a), S-matrix (b) and A-matrix (c). The vertical scale on the S- and A-matrices is percentage of code occurrence.

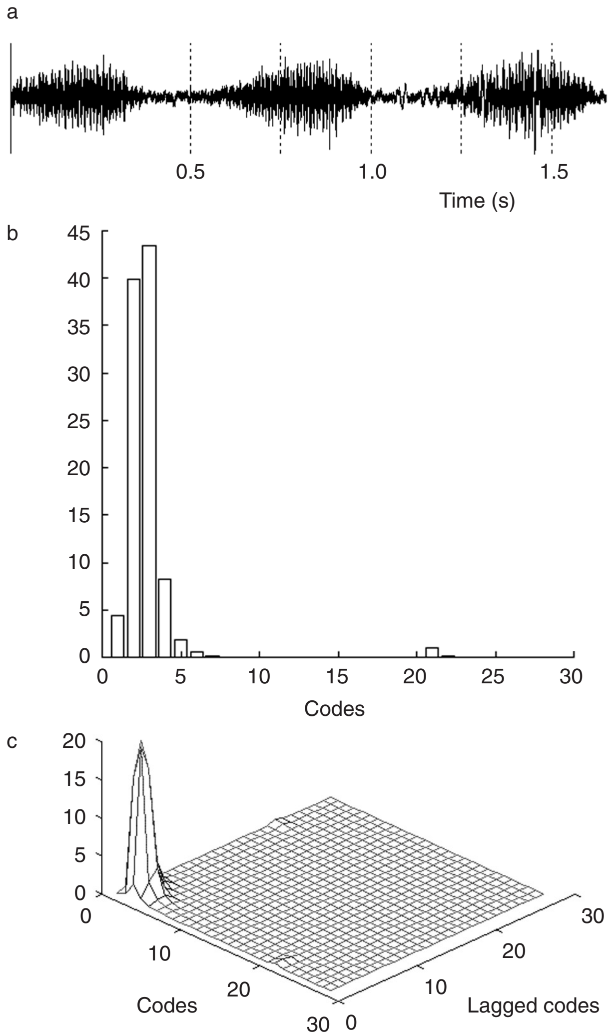

Fig. 8. *Mymeleotettix maculatus* time waveform (a), S-matrix (b) and A-matrix (c). The vertical scale on the S- and A-matrices is percentage of code occurrence.

example CP08 in table 2 or CA09 in table 4) indicating that the sound has not been recognized with any degree of certainty. It is possible to reject these low values by adding a rejection threshold such that all values less than the threshold are rejected. This has the advantage of reducing incorrect identifications due to low signal level or interference. For a threshold value T_{REJ} ($0.0 \leq T_{REJ} \leq 1.0$) the identified sound I is given by:

$$I = i_{MAX} : Q\left(i_{MAX}\right) = \max_{i=1}\left\{Q(i)\right\} \geq T_{REJ} \quad \text{else} \quad I = 0$$

If I = 0 then the sound has not been identified to a high enough accuracy and is therefore rejected.

Single echeme recognition

The results for echeme recognition of the four grasshopper species are presented in tables 2 to 5 and table 6 gives a summary of the recognition accuracies for varying rejection thresholds. Generation of test files consisted of manual extraction of approximately 2 s of sound covering 1 echeme (1 second for *M. maculatus*). The system then identified each sound. Entries in the tables give all non-zero outputs, for example in table 2, CP05 is correctly recognized as *C. parallelus* as the output is 0.999. CP08 is incorrectly recognized as *C. albomarginatus* with an output of 0.57, the next highest output being *C. parallelus* with a value of 0.549. Table 2 shows all 13 possible sound categories whereas the remaining tables only show sound categories with non-zero outputs.

Chorthippus parallelus

There are nine correct identifications, two misidentifications as *O. viridulus*, two as *M. maculatus* and one as *C. albomarginatus* (table 2). The overall recognition accuracy is therefore 64% which is considered to be low. Recognition accuracy improves if a threshold is applied as indicated in table 8; a threshold of 0.9 gives an overall recognition accuracy of 81.8%.

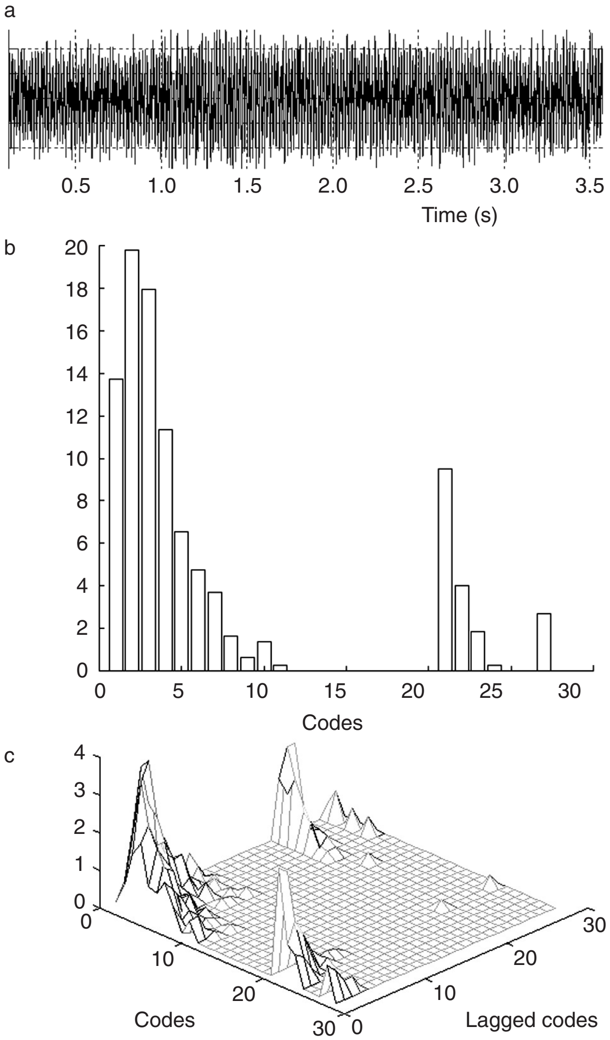

Fig. 9. Time domain waveform (a), S-matrix (b) and A-matrix (c) for a single engine light aircraft. The vertical scale on the S- and A-matrices is percentage of code occurrence.

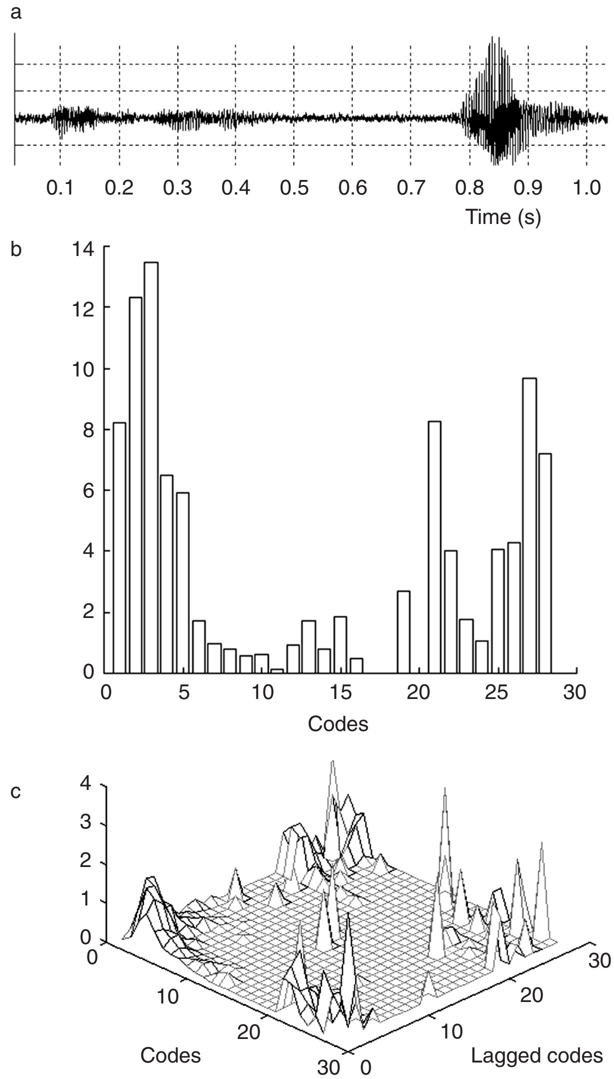

Fig. 10. Time domain waveform (a), S-matrix (b) and A-matrix (c) for a blowfly passing the microphone. The vertical scale on the S- and A-matrices is percentage of code occurrence.

Omocestus viridulus

Omocestus viridulus produces long relatively loud songs comprising many repeated echemes. Here, in table 4 there is only one misidentification in 34 echemes (as B11) giving 97% accuracy. When a threshold is applied (any value between 0.5 and 0.9) the accuracy rises to 100% as shown in table 6.

Chorthippus albomarginatus

An accuracy of 87.5% (14 out of 16 echemes) is obtained for this species (table 4). However, both misidentified echemes (CA13 and CA14) had high levels of interference from a singing bird and movement of the microphone respectively. Applying a threshold decreases recognition accuracy since several correctly identified echemes have low values and are rejected (table 6).

Myrmeleotettix maculatus

This species produces echemes that are approximately 1 s in duration. Misidentification is higher in this species on a single echeme basis with an overall accuracy of 41.2% with no threshold as indicated in table 5. Application of a threshold greatly improves the accuracy with a maximum of 90% for a 0.9 threshold (table 6).

Whole song recognition

Tests on the recognition of complete songs were carried out for two species only: *C. albomarginatus* (table 7) and *C. parallelus* (table 8). The ANN was not retrained for these tests but was the same as used for single echemes. Considering *C. albomarginatus* first, only one misidentification was obtained (CAF06) which, upon investigation, had interference from a very loud light aircraft The overall recognition accuracy was 87.5% which remained the same for different thresholds as indicated in table 9. *Chorthippus parallelus* song was recognized 100% of the time irrespective of threshold level (table 9).

It is evident that recognition of the complete song gives

Table 2. *Chorthippus parallelus* results of 2-s time samples.

Sound sample	Ov	Mm	Cp	Ca	FLY	CR1	CR2	B11	B21	BD1	CHF	BGR	PL
CP01		**0.997**					0.018						
CP02		0.129	**0.976**										
CP03			**0.997**										
CP04		**0.999**											
CP05			**0.999**										
CP06			**0.999**										
CP07	**0.73**		0.013	0.02									
CP08	0.02		0.549	**0.57**									
CP09	**0.155**		0.06	0.116									
CP10			**0.98**										
CP11		0.046	**0.997**										
CP12			**0.999**										
CP13			**0.999**										
CP14			**0.999**										

The complete matrix is shown here. Blank entries indicate the output is less than 0.001. Maximum outputs are shown in bold; these are the identified sounds. B11, B21, BD1, alarm calls of three bird species; BR, background sound; Ca, *Chorthippus albomarginatus*; CHF = sound of chiff-chaff warbler; Cp, *C. parallelus;* CR1, CR2, different sounds of car; FLY, blowfly sound; Mm, *Mymeleotettix maculatus*; Ov, *Omocestus viridulus*; PL, single engine aircraft.

Table 3. *Omocestus viridulus* results, single echeme samples.

Sound sample	Ov	Mm	Cp	Ca	B11
OV01	**0.999**				
OV02	**0.999**				
OV03	**0.999**				
OV04	**0.97**		0.059		
OV05	**0.999**				
OV06A	**0.65**				0.096
OV06B	0.005				**0.292**
OV07A	**0.153**				0.047
OV07B	**0.919**				0.004
OV08	**0.999**				
OV09	**0.999**				
OV10	**0.999**				
OV11A	**0.999**				
OV11B	**0.763**				0.015
OV13	**0.999**				
OV13B	**0.999**				
OV15A	**0.999**				0.2
OV15B	**0.999**				0.2
OV16A	**0.999**				
OV16B	**0.999**				0.2
OV17	**0.999**				
OV18	**0.999**				
OV20	**0.999**				

This table only shows outputs that are significant; all others (see table 2) are less than 0.001. Each sample was divided into two halves A and B giving a total of 34 sounds. Only sounds with results other than 0.999 are shown in two halves. Maximum outputs are shown in bold; these are the identified sounds. See table 2 for key to abbreviations.

higher reliability than individual echemes. This does, however, rely on the detection of end points, i.e. the start and end of the song which may be difficult if the level of interference is high. It is possible to make use of the fact that songs contain multiple echemes and to accumulate echeme recognition over the song. This approach has not yet been tested.

Generalized recognition for 2-s intervals

Song and echeme recognition rely on the reliable detection of the end-points of an echeme or song which, as

Table 4. *Chorthippus albomarginatus* results, single echeme samples. See table 2 for key to abbreviations.

Sound sample	Ov	Mm	Cp	Ca
CA01		0.001		**0.999**
CA02				**0.999**
CA03		0.06		**0.753**
CA04				**0.999**
CA05				**0.999**
CA06				**0.999**
CA07				**0.999**
CA08			0.001	**0.92**
CA09*	0.038		0.011	**0.365**
CA10			0.001	**0.981**
CA11	0.007			**0.97**
CA12				**0.999**
CA13*	**0.999**			
CA14*			**0.952**	0.019
CA15				**0.999**
CA16				**0.999**

* Very noisy signal due to movement of microphone and bird singing.

This table only shows outputs that are significant; all others (see table 2) are less than 0.001. Maximum outputs are shown in bold; these are the identified sounds.

stated above, may be difficult under interference. It is possible to ignore song structure and simply recognize sounds at a suitable interval. The third series of tests reported here were carried out by recognizing sounds during each 2 s. Results for these tests are given in figs 11 and 12 for *O. viridulus* and *C. parallelus*, respectively. Identification was obtained for each 2-s interval as described earlier. Each figure shows the time domain waveform and identification below together with the neuron output level. Results in fig. 11 show highly reliable recognition of three short songs of *O. viridulus* together with a light aircraft and a bird alarm call (unknown species). Figure 12 gives similar results for *C. parallelus* with a light aircraft in the background. Both tests indicate that reliable identification is possible without carrying out additional processing to find the beginning of a song or echeme. This approach has considerable potential for creating temporal sound maps where individual sounds can be identified together with their intensity on a continuous basis. Selection of the best interval will depend on the application; here 2 seconds appears to be most appropriate.

Discussion

The system described here uses a computationally simple but powerful approach to the analysis and automatic recognition of bioacoustic signals. Results indicate that it is possible to automatically identify grasshopper species in field recordings with the additional ability to identify background sounds including birds and man-made sounds such as aircraft. At present, the system has been trained to discriminate between 13 sound types including four grasshopper species, the small number of Orthoptera species being due to the lack of species in the north of England.

Basic recognition accuracy of echemes varies with species and ranges from 41.2% (*M. maculatus*) to 97% (*O. viridulus*). Application of a rejection threshold removes low confidence results and increases the overall recognition accuracy for these two species to a minimum of 76.9% and 100% respectively. Recognition of complete songs has been tested for two species with a minimum accuracy of 87.5%.

It should be noted that the sample data sets are not large and more data needs to be analysed before reliable statistical results can be obtained. These results represent early tests for the system and there are many possible future areas of investigation including: (i) investigating more accurate waveform descriptors, including tree grammars and inclusion of energy level; (ii) increasing the number of species to include all British singing Orthoptera as well as the number of 'interfering' sounds; (iii) expansion of the dataset to include song types, such as courtship and aggression; (iv) development of systems for other insect groups, such as cicadas; and (v) investigating other ANN architectures.

One potential problem with the current system is that of multiple simultaneous calls, both intraspecific and interspecific. This is a difficult problem to overcome since the sounds received are additive and cannot be separated

Table 5. *Mymeleotettix maculatus* results, single echeme samples.

Sound sample	Ov	Mm	Cp	Ca	CR2
MM01A		**0.521**	0.004	0.065	
MM01B		**0.999**			
MM02A		**0.999**			
MM02B		**0.999**			
MM03A					0.153
MM03B	**0.972**				
MM04A		**0.999**			
MM04B	0.02		0.137	**0.41**	
MM05A			0.11	**0.275**	
MM05B			**0.95**		
MM06A			**0.999**		
MM06B		0.145	0.003	**0.198**	
MM07A		**0.823**	0.017	0.009	
MM08B		**0.89**	0.017	0.004	
MM09		**0.999**			
MM10		**0.999**			
MM11		**0.999**			

This table only shows outputs that are significant; all others (see table 2) are less than 0.001. Maximum outputs are shown in bold; these are the identified sounds.. See table 2 for key to abbreviations.

Table 7. *Chorthippus albomarginatus* recognition accuracy for complete song.

	Ov	Mm	Cp	Ca
CAF01				**0.999**
CAF02				**0.999**
CAF03				**0.999**
CAF04	0.002		0.03	**0.92**
CAF05	0.001		0.15	**0.555**
CAF06*	**0.999**			
CAF07				**0.999**
CAF08				**0.999**

* Interference from very loud light aircraft.

This table only shows outputs that are significant; all others (see table 2) are less than 0.001. Maximum outputs are shown in bold; these are the identified sounds. See table 2 for key to abbreviations.

Table 8. *Chorthippus parallelus* recognition accuracy for complete song.

	Ov	Mm	Cp	Ca
CPF01		0.267	**0.935**	
CPF02			**0.999**	
CPF03			**0.999**	
CPF04			**0.948**	0.024
CPF05			**0.72**	0.3

This table only shows outputs that are significant; all others (see table 2) are less than 0.001. Maximum outputs are shown in bold; these are the identified sounds. See table 2 for key to abbreviations.

Table 6. Identification accuracy for different threshold levels, short interval samples (1–2 s).

Threshold		Ov	Mm	Cp	Ca
0.5	Rejected	2	4	1	1
	Accuracy	100% (32/32)	76.9% (10/13)	69.2% (9/13)	86.7% (13/15)
0.6	Rejected	2	5	2	1
	Accuracy	100% (32/32)	75% (9/12)	75% (9/12)	86.7% (13/15)
0.7	Rejected	3	5	2	1
	Accuracy	100% (31/31)	75% (9/12)	75% (9/12)	86.7% (13/15)
0.8	Rejected	4	5	3	2
	Accuracy	100% (30/30)	75% (9/12)	81.8% (9/11)	85.7% (12/14)
0.9	Rejected	4	7	3	2
	Accuracy	100% (30/30)	90% (9/10)	81.8% (9/11)	85.7% (12/14)
None	Accuracy	97% (33/34)	41.2% (7/17)	64.3% (9/14)	87.5% (14/16)
Sample size		34	17	14	16

Assumes all outputs less than threshold are rejected. See table 2 for key to abbreviations.

Table 9. Identification accuracy of *Chorthippus parallelus* and *Chorthippus albomarginatus* for different threshold levels, complete song.

Threshold		Cp	Ca
0.5	Rejected	0	0
	Accuracy	100% (5/5)	87.5% (7/8)
0.6	Rejected	0	1
	Accuracy	100% (5/5)	85.7% (6/7)
0.7	Rejected	0	1
	Accuracy	100% (5/5)	85.7% (6/7)
0.8	Rejected	1	1
	Accuracy	100% (4/4)	85.7% (6/7)
0.9	Rejected	1	1
	Accuracy	100% (4/4)	85.7% (6/7)
None		100% (5/5)	87.5% (7/8)
Sample size		5	8

Assumes all outputs less than threshold are rejected. See table 2 for key to abbreviations.

unless they occur in different frequency bands. If this is the case, then appropriate filtering of the signal can separate the sounds. Multiple intraspecific calls are more problematic. The authors are working on an alternative approach involving the recognition of sounds over very short time scales (50–200 ms) using a different form of ANN (a self-organizing feature map – SOFM) similar to that developed for human heart sound analysis (Swarbrick & Chesmore, 1998; Swarbrick, 2001). The SOFM approach groups similar S-matrices accumulated over 20–40 epochs (typically 20–100 ms) to identify sounds on a short time scale. This technique relies on occasions when fewer (preferably one) species or individuals are calling and accumulating recognition data over long time periods to ensure as many species as possible are detected. Preliminary tests on six species of Japanese cricket and bushcricket show promise but more work needs to be carried out.

Applications of the system are varied and include automated biodiversity assessment, habitat assessment

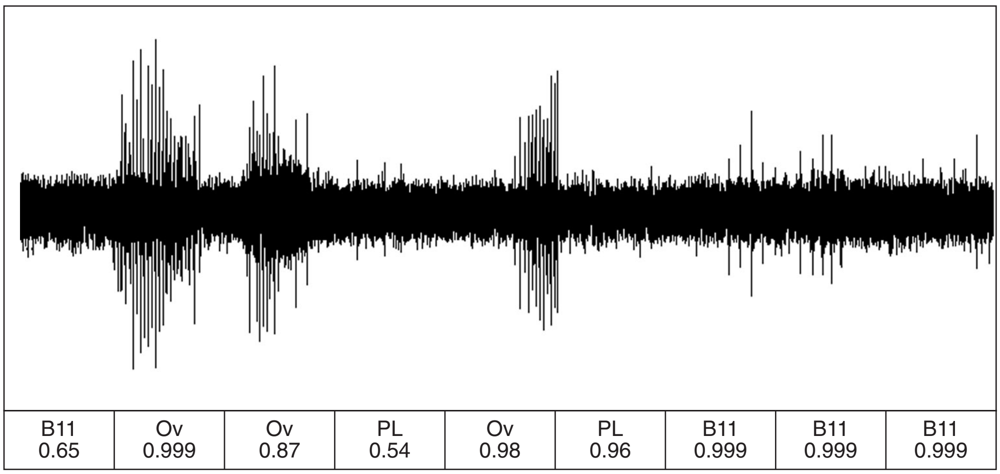

Fig. 11. Classified sounds from an 18-s sequence on a 2-s interval recorded at Allerthorpe Common on 15 July 2002. The system correctly recognizes three short songs by *Omocestus viridulus* (Ov), a light aircraft (PL) and a bird alarm call (B11).

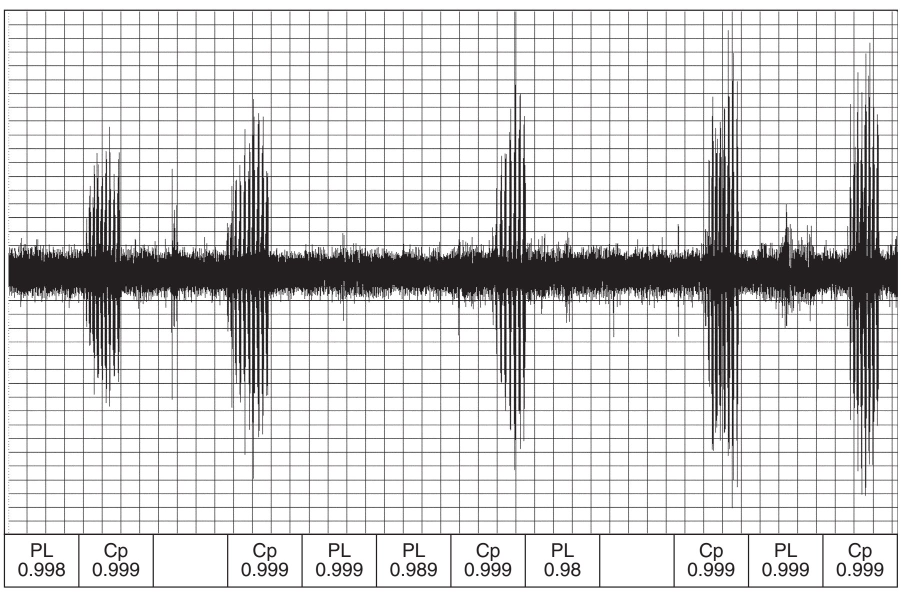

Fig. 12. Classified sounds from a 24-s sequence on a 2-s interval recorded at Allerthorpe Common on 15 July 2002. The system correctly recognizes five songs of *Chorthippus parallelus* (Cp) and an aircraft (PL). The two blank intervals are below the threshold of 0.9.

using Orthoptera as bioindicators, hand held identification aids and continuous monitoring of calls for autecology.

Acknowledgements

The authors would like to acknowledge the funding of Dr Ohya's visit to the UK as a collaborator, via a fellowship under the OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agriculture Systems. Permission for access to various sites was given by English Nature, the Forestry Commission, East Riding County Council of Yorkshire and the Yorkshire Wildlife Trust.

References

Anderson, S.E., Dave, A.S. & Margoliash, D. (1996) Template-based automatic recognition of birdsong syllables from continuous recordings. *Journal of the Acoustical Society of America* **100**, 1209–1217.

Bishop, C.M. (2000) *Neural networks for pattern recognition*. Oxford, Oxford University Press.

Campbell, R.H., Martin, S.K., Schneider, I. & Michelson, W.R. (1996) Analysis of mosquito wing beat sound. *132nd Meeting of the Acoustical Society of America*, Honolulu.

Chesmore, E.D. (1999) Technology transfer: applications of electronic technology in ecology and entomology for species identification. *Natural History Research* **5**, 111–126.

Chesmore, E.D. (2000) Methodologies for automating the identification of species. pp. 3–12 in *Proceedings of 1st BioNET International Working Group on Automated Taxonomy, July 1997*.

Chesmore, E.D. (2001) Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals. *Applied Acoustics* **62,** 1359–1374.

Chesmore, E.D. & Nellenbach, C. (2001) Acoustic methods for the automated detection and identification of insects. *Acta Horticulturae* **562**, 223–231.

Chesmore, E.D., Swarbrick, M.D. & Femminella, O.P. (1998) Automated analysis of insect sounds using TESPAR and expert systems – a new method for species identification. pp. 273–287 *in* Bridge, P., Jeffries, P., Morse, D.R. & Scott, P.R. (*Eds*) *Information technology, plant pathology and biodiversity*. Wallingford, Oxon, CAB International.

Clemins, P.J. & Johnson, M.T. (2002) Automatic type classification and speaker verification of African elephant vocalizations. *Animal Behaviour 2002*, Vrije Universiteit, Amsterdam.

Haack, R.A., Law, K.R., Mastro, V.C., Ossenbruggen, H.S. & Raimo, B.J. (1997) New York's battle with the Asian long-horned beetle. *Journal of Forestry* **95**, 11–15.

Hagstrum, D.W., Vick, K.W. & Webb, J.C. (1990) Acoustic monitoring of *Rhizopertha dominica* (Coleoptera: Bostrichidae) populations in stored wheat. *Journal of Economic Entomology* **83**, 625–628.

Hickling, R., Lee, P., Velea, D., Denneby, T.J. & Patin, A.I. (2000) Acoustic system for rapidly detecting and monitoring pink bollworm in cotton bolls. pp. 1–10 in *Proceedings of the Beltwide Cotton Conference*, San Antonio.

King, R.A. & Gosling, W. (1978) Time-encoded speech. *Electronics Letters* **15**, 1456–1457.

Looney, C.G. (1997) *Pattern recognition using neural networks*. Oxford, Oxford University Press.

Lucking, W.G., Darnell, M. & Chesmore, E.D. (1994) Acoustical condition monitoring of a mechanical gearbox using artificial neural networks. pp. 120–124 in *IEEE Conference on Neural Networks*, Orlando, Florida.

MacLeod, A., Evans, H.F. & Baker, R.H.A. (2002) An analysis of pest risk from an Asian longhorn beetle (*Anoplophora glabripennis*) to hardwood trees in the European community. *Crop Protection* **21**, 635–645.

Mankin, R.W. & Weaver, D.K. (2000) Acoustic detection and monitoring of *Cephus cinctus* sawfly larvae in wheat stems. *International Congress of Entomology*, Brazil.

Mankin, R.W., Brandhurst-Hubbard, J., Flanders, K.L., Zhang, M., Crocker, R.L., Lapointe, S.L., McCoy, C.W., Fisher, J.R. & Weaver, D.K. (2000) Eavesdropping on insects hidden in soil and interior structures of plants. *Journal of Economic Entomology* **93**, 1173–1182.

Matsuoka, H., Fujii, Y., Okumura, S., Imamura, Y. & Yoshimura, T. (1996) Relationship between the type of feeding behaviour of termites and the acoustic emission (AE) generation. *Wood Research* **83**, 1–7.

McIlraith, A.L. & Card, H.C. (1995) Birdsong recognition with DSP and neural networks. pp. 409–414 in *IEEE WESCANEX'95 Proceedings*, Winnipeg, Canada.

Mills, H. (1995) Automatic detection and classification of nocturnal migrant bird calls. *Journal of the Acoustical Society of America* **97**, 3370–3371.

Murray, S.O., Mercado, E. & Roitblat, H.L. (1998) The neural network classification of false killer whale (*Pseudorca crassidens*) vocalizations. *Journal of the Acoustical Society of America* **104**, 3626–3633.

Ohya, E. & Chesmore, E.D. (2003) Automated identification of grasshoppers by their songs. *Annual Meeting of the Japanese Society of Applied Entomology and Zoology*, Iwate University, Morioka, Japan.

Parsons, S. (2001) Identification of New Zealand bats in flight from analysis of echolocation calls by artificial neural networks. *Journal of Zoology* **253**, 447–456.

Parsons, S. & Jones, G. (2000) Acoustic identification of 12 species of echolocating bat by discriminant function analysis and artificial neural networks. *Journal of Experimental Biology* **203**, 2641 – 2656.

Reby, D., Lek, S., Dimopoulos, I., Joachim, J., Lauga, J. & Aulagnier, S. (1997) Artificial neural networks as a classification method in the behavioural sciences. *Behavioural Processes* **40**, 35–43.

Riede, K. (1993) Monitoring biodiversity: analysis of Amazonian rainforest sounds. *Ambio* **22**, 546–548.

Schalkoff, R. (1992) *Pattern recognition: statistical, structural and neural approaches*. New York, John Wiley & Sons.

Schwenker, F., Dietrich, C., Kestler, H.A., Riede, K. & Palm, G. (2003) Radial basis function neural networks and temporal fusion for the classification of bioacoustic time series. *Neurocomputing* **51**, 265–275.

Shuman, D., Coffelt, J.A., Vick, K.W. & Mankin, R.W. (1993) Quantitative acoustical detection of larvae feeding inside kernels of grain. *Journal of Economic Entomology* **86**, 933–938.

Shuman, D., Weaver, D.K. & Mankin, R.W. (1997) Quantifying larval infestation with an acoustical sensor array and cluster analysis of cross-correlation outputs. *Applied Acoustics* **50**, 279–296.

Swarbrick, M.D. (2001) *Acoustic diagnosis of heart defects using time-domain signal processing and artificial neural networks*. PhD thesis, University of Hull.

Swarbrick, M.D. & Chesmore, E.D. (1998) Automatic classification of heart sounds and murmurs using time domain signal coding (TDSC), artificial neural networks and expert systems. pp. 37–46 in *Proceedings, Conference on Recent Advances in Soft Computing 98*, DeMontfort University.

Taylor, A., Grigg, G., Watson, G. & McCallum, H. (1996) Monitoring frog communities, an application of machine learning. *Eighth Innovative Applications of Artificial Intelligence Conference*, 1996, AAAI Press.

Terry, A.M.R. & McGregor, P.K. (2002) Census and monitoring based on individually identifiable vocalizations: the role of neural networks. *Animal Conservation* **5**, 103–111.

Vaughan, N., Jones, G. & Harris, S. (1996) Identification of British bat species by multivariate analysis of echolocation call parameters. *Bioacoustics* **7,** 189–207.

(Accepted 23 February 2004)