Nat. Hist. Res., Vol. 5 No. 2: 111-126, March 1999

Technology Transfer: Applications of Electronic Technology in Ecology and Entomology for Species Identification

David Chesmore

School of Engineering, University of Hull
Hull, HU6 7RX, England

Abstract Recent advances in electronics and computing technology are leading to new applications in biology, ecology and conservation; one particular research area, computer-aided taxonomy, is only becoming a reality because of these advances. Computer-aided taxonomy encompasses automatic species identification, computer-based identification keys and taxonomic methods such as cladistics. Whilst there is a reasonable research effort being put into the development of computer identification keys for economically important groups, there is relatively little research on automated identification of species. Applications involving automated identification are diverse and include insect counting and sorting, pest monitoring and biodiversity assessment. Accordingly, this paper concentrates on automated identification research and describes recent and ongoing work at Hull University and elsewhere on two topics-bioacoustic identification of insects and birds, and image processing applications for discrimination of quarantine species of fungi and insects. The paper is based on the Open Lecture given at the Natural History Museum and Institute, Chiba, Japan on March 21, 1998.

Key words: computer-aided taxonomy, automated species identification and recognition, bioacoustics, image processing, computer applications, entomology, ecology.

In recent years, the growth in the electronic and computing industries has been remarkable, mainly due to the development of high speed, low cost microprocessors and memory. The advent of computers operating at 300 MHz or higher, with hundred's of megabytes of memory have led to new applications unheard of only a few years ago. In biology and ecology, there are now many possibilities for the development of research tools, commercial instrumentation and software packages for biological analysis. It is also becoming increasingly apparent that these new applications can only be properly designed if biologists have a modest knowledge of engineering and vice versa. Indeed, it is the author's belief that the only way forward is to increase multidisciplinary teaching and research in universities and colleges. It is from the multidisciplinary background of the members of the Control and Intelligent Systems Engineering Research Group, a merger of the Control Research Group and the Environmental Electronics Research Group, at Hull University that the research projects described in this paper have arisen. The paper describes a number of projects that fall under the heading of computer-aided taxonomy (CAT), a new term which was first defined at the inaugural meeting of the BioNET International Group for Computer-aided Taxonomy (BIGCAT) in Cardiff, Wales in July 1997 (Chesmore, 1998b). CAT is defined as the application of any computer or computer technique for taxonomic purposes.

CAT can be divided into 3 groups: automated identification systems, identification keys and software techniques for taxonomy such as cladistics. It has great potential for aiding species identification, especially during biodiversity studies where there may be many species. The quality of such surveys is dependent on the accuracy of the identification process which may be difficult and time-consuming. Edwards and Morse (1995) give a detailed account of the development of computer-aided identification mainly in key-based systems. It is beyond the scope of this paper to discuss key systems which are generally regarded to be well developed (see for example, Dallwitz *et al.*, 1998 and Pankhurst, 1991). Instead, the paper concentrates on more recent work involving automated and semi-automated species identification. More detailed discussions of CAT can be found in (Chesmore, 1997a, b; Chesmore and Morse, 1997).

Automated Species Identification

Before considering automated identification applications it is important to make a distinction between "identification" as described here and the term used for classification which is in wide use in taxonomy. In the applications described in this paper, "identification" is used in the context of associating an unknown signal (acoustic signals, images, etc.) with one in a pre-defined set of classes (species or groups). The identification system is trained with signals from each class prior to use. There is an ongoing debate as to whether "identification", "recognition" or "classification" should be used; it is the author's experience from meetings such as the Systematics Association Biennial Conference (Chesmore and Morse, 1997) and Fifth Workshop of the ESF Network in Systematic Biology on New Directions in Systematics (Chesmore, 1997b) that "identification" is the preferred term in Europe and that "classification" is to be restricted to taxonomy even though it is used commonly in engineering applications.

The concept of automated identification of species has received relatively little attention until recently; Table 1 shows some examples of current and past research areas. The techniques employed generally fall into 2 broad categories—acoustics and image processing, with a few miscellaneous methods involving flow cytometry for algae and phytoplankton (Balfoort *et al.*, 1992; Boddy *et al.*, 1994), and radar for aerial insect migration (Smith *et al.*, 1993).

Species identification by electronic means

Table 1. Examples of species identification systems.

Species/Group	Sensor(s)	Classification Method(s)
Fish species	Active acoustics (sonar)	PDF, cluster analysis, ANN
Orthoptera	Passive acoustics	TDSC+ANN
Amphibia	Passive acoustics	FFT+machine learning
Birds	Passive acoustics	FFT, LPC, TDSC+ANN
Mosquito	Passive acoustics	Wingbeat frequency
Flying insects	Radar	Scatter
	Infra-red Doppler	Wing beat frequency
Lepidoptera	Monochrome & colour image	Colour discrimination
Phytoplankton	Flow cytometry	ANN
	Monochrome image	Linear discrimination, ANN
Hymenoptera	Monochrome image	Wing venation; PCA
Leaf-miners	Monochrome & colour image	ANN
Fungal spores	Monochrome image	ANN, shape discrimination
Plants, weed species	Monochrome & colour image	FFT, colour discrimination
Nanofossils	Scanning electron microscope	General image processing
Pollen	Scanning electron microscope	Texture analysis

Notes: ANN, artificial neural network; PDF, probability density function; TDSC, time domain signal coding; FFT, fast Fourier transform; LPC, linear predictive coding.

can be considered to be an application of general pattern recognition in which an unknown (specimen) is placed into one of a number of possible classes depending on features extracted from measurements on the species. Pattern recognition has many applications ranging from handwriting recognition to speech analysis and identification of faults in machinery (condition monitoring). Automated species identification is very similar to many of these applications. Two main levels of automation can be identified—partially and fully automated as described below:

a) Fully Automated. Complete identification without user interaction; this requires highly reliable identification with a very low (ideally zero) probability of misclassification.

b) Semi-automated. This category is perhaps more realistic than a) as it allows prior sorting into higher taxonomic categories and presents the user with data for further manual identification if required. It is a relaxation of fully automated identification and is therefore more likely to be feasible in the short term.

It is anticipated that semi-automated systems will be the most viable as they allow the user to perform or verify the final identification. This is considered to be more acceptable in the short term as there is a tendency for humans to mistrust computers or consider them as a threat which may result in a possible impediment to CAT. It is therefore expected that semi-automated systems will play an important role in validation of techniques and in obtaining general acceptance of automation. In addition, such systems must not be seen as replacements for trained taxonomists but as identification aids.

Applications of automated identification systems are diverse and include insect counting, biodiversity assessment, ecological monitoring and detection/identification of pests. Each area is discussed in more detail below:

a) Insect Counting. Little research has been carried out in this potentially important area. Gonzales (1986) developed a pilot image processing system for identifying insects from agricultural surveys in an effort to speed up the often time consuming sorting process. It has been suggested that systems of this nature could aid considerably in sorting from large catches even if the sorting process only identifies to order or genus. Such pre-sorting could reduce the identification time by an order of magnitude. This also links to biodiversity assessment and pest monitoring (see below).

b) Biodiversity Assessment. Systematics Agenda 2000 (1994) has stated that biodiversity inventories must be carried out for as many habitats as possible. However, increasing destruction of many habitats has led to the necessity for developing more rapid biodiversity assessment in an attempt to identify habitats and areas rich in biodiversity. Riede (1993) suggested that since many rainforest species produce sounds, it may be possible to use acoustic analysis for monitoring fauna. Riede used Orthoptera for more rapid biodiversity estimation in a tropical lowland forest in Ecuador and Oba (1994, 1995) used bird song as a measure of the "natural sound diversity" in Japan. Until very recently, it has only been possible to identify species manually from recordings which is both costly and time consuming; the development of automated identification systems will speed up the process and lead to continuous real-time monitoring.

c) Ecological Monitoring. Potential applications for ecological monitoring are diverse and include recording the occurrence of call types and correlating with environmental conditions, long term continuous monitoring and determination of bird species for species-specific bird strike avoidance (bird scarers) in airports. It is also theoretically possible to identify and monitor individuals in populations of some taxa (e.g. birds).

d) Pest Monitoring. Many animal pests, particularly insects, can be detected by their sound production. In the USA, Shuman *et al.* (1993) used acoustics for detecting beetle larvae in rice grains. Hagstrum *et al.* (1990) used similar techniques for monitoring of *Rhizopertha dominica* (Coleoptera: Bostrichidae) in wheat kernels. It is theoretically possible to detect and, more importantly, identify many different insect and animal pests in a variety of agricultural and horticultural environments although very little work has been done to date.

Identification Using Bioacoustics

Many insect, bird and animal species produce sounds either deliberately for communications or as a byproduct of movement (flying, eating, walking, etc.). In many cases, such sounds can be used to detect the presence of animals or, if sufficiently distinctive, identify species. The majority of acoustic applications are passive, i.e. they rely on calls and sounds produced by animals. The majority of research to date has been on the identification of birds in general (McBraith and Card, 1995; Anderson *et al.*, 1996), nocturnal migrant birds (Mills, 1995), frogs and other amphibia (Taylor *et al.*, 1996) and insects (Chesmore *et al.*, 1998; Chesmore and Nellenbach, 1997). Applications involving detection rather than identification are less complex and have mainly been for stored grain insect pests (Shuman *et al.*, 1993; Shuman *et al.*, 1997; Weaver *et al.*, 1997; Anderson *et al.*, 1996).

An alternative to passive listening is active sonar which relies on scatter from transmitted sound, often ultrasound. Underwater sonar applications are currently restricted to identification of fish species in shoals (Simmonds *et al.*, 1996; Scalabrin *et al.*, 1996) and zooplankton classification (Martin *et al.*, 1996). The latter research project does not identify to species but determines approximate groupings according to acoustic spectral reflection characteristics.

Development of the Intelligent Bioacoustic Identification System (IBIS)

IBIS is a multipurpose testbed for automated bioacoustic identification applications. It is based on a technique known as time encoded speech (TES) which was developed in the 1970's by King (King and Gosling, 1978) as a purely time domain approach to the compression of speech for digital transmission. It has subsequently been used in a number of applications including acoustic

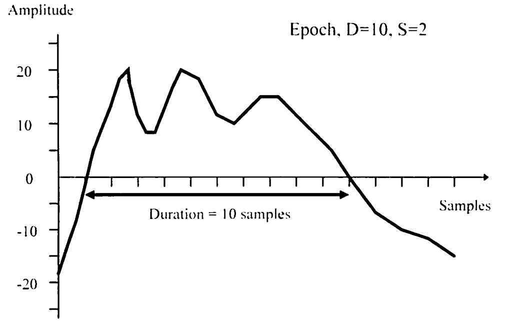

Fig. 1. Definition of a TDSC epoch.

condition monitoring of machinery (Lucking *et al.*, 1994) and heart sound analysis and defect identification (Swarbrick and Chesmore,1998). TES is now termed time domain signal coding (TDSC) by the author (Chesmore, 1998c) since the term is more general than the original application. TDSC characterises any bandlimited signal by its shape between successive real zeros (termed an epoch); generally, this shape is taken between actual zero-crossings. Each epoch is described in terms of its duration in samples (D) and shape (S) usually taken as the number of minima or signal energy as indicated in Fig. 1 which shows a 10 sample epoch with 2 minima (D=10, S=2). The number of possible D–S combinations (symbols) is termed the natural alphabet which can often be nonlinearly mapped onto a smaller symbol set to give signal compression. In the original speech application, the coded symbols were transmitted and used to regenerate the speech signal at the receiver thus providing digital speech transmission at substantially reduced data rates.

TDSC can be described as the concatenation of a signal's D–S symbols, i.e. it produces a sequential stream of symbols and one analysis method is to examine the occurrence of pairs of symbols over time to give a histogram, A, which describes the number or proportion of symbols i and j occurring in succession, i.e. the number of times i is followed by j by a lag L. A 2-dimensional histogram, the A-matrix, can be formed, expressed mathematically as:

$$a_{ij}=\frac{1}{(N-L)}\sum_{n=L+1}^{m=N} x_{ij}(n)$$

where L = lag

$x_{ij}(n)=1$ if $t(n)=i$ and $t(n-L)=j$ (0 otherwise)

$t(n)=n^{\text{th}}$ TES symbol

This fixed size histogram with time-invariant dimensions is the feature set used for classification purposes. The entry at position (i, j) represents the number (percentage) of occurrences of the TDSC symbol pair i and j where j is delayed relative to the first (in epochs). In this application, a lag of 1 epoch is used; multiple lags may also be employed giving rise to multi-dimensional matrices. The A-matrix is independent of any gain factors if the input signal has no dc component and is therefore insensitive to relative energies of different segments of the signal.

Each sound (call, syllable, etc.) is transformed into a single A-matrix which forms the basic feature for pattern classification using an artificial neural network (ANN). ANNs 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. There are many forms of ANN which can be divided into supervised (requires training) and unsupervised classification (no training). The majority of classification methods used in IBIS are based on multilayer perceptrons (MLP) using backpropagation for training. More recently, self-organising features maps have been investigated with some success.

All software has been developed using Matlab, a generic mathematical programming language optimised for matrix manipulation. Matlab has been chosen for two reasons—it is simple to use and a wide range of toolboxes are available, in this case the neural networks toolbox. Matlab runs on any 486 computer (or better) with at least 4 Mbytes of RAM and a hard disc, and is capable of running on portable PCs. The sounds are sampled at 44 kHz, 16-bits per sample using a program called Goldwave V3.03 via the line input of a Soundblaster card.

Bioacoustic Identification Examples

To date, 2 different test groups have been selected—25 species British Orthoptera and

Table 2. List of British Orthoptera species for IBIS tests.

ID Code	Latin Name	English Name
OR01	*Meconema thalassinum*	Oak Bush-cricket
OR02	*Tettigonia viridissima*	Great Green Bush-cricket
OR03	*Decticus verrucivorus*	Wart-biter
OR04	*Pholidoptera griseoaptera*	Dark Bush-cricket
OR05	*Platycleis albopunctata*	Grey Bush-cricket
OR06	*Metrioptera brachyptera*	Bog Bush-cricket
OR07	*Metrioptera roeselii*	Roesel's Bush-cricket
OR08	*Conocephalus discolor*	Long-winged Cone-head
OR09	*Conocephalus dorsalis*	Short-winged Cone-head
OR10	*Acheta domesticus*	House-cricket
OR11	*Gryllus campestris*	Field-cricket
OR12	*Nemobius sylvestris*	Wood-cricket
OR13	*Gryllotalpa gryllotalpa*	Mole-cricket
OR14	*Stethophyma grossum*	Large Marsh Grasshopper
OR15	*Stenobothrus lineatus*	Stripe-winged Grasshopper
OR16	*Stenobothrus stigmaticus*	Lesser Mottled Grasshopper
OR17	*Omocestus rufipes*	Woodland Grasshopper
OR18	*Omocestus viridulus*	Common Green Grasshopper
OR19	*Chorthippus brunneus*	Field Grasshopper
OR20	*Chorthippus vagans*	Heath Grasshopper
OR21	*Chorthippus parallelus*	Meadow Grasshopper
OR22	*Chorthippus albomarginatus*	Lesser Marsh Grasshopper
OR23	*Euchorthippus pulvinatus*	Jersey Grasshopper
OR24	*Gomophocerippus rufus*	Rufous Grasshopper
OR25	*Myrmeleotettix maculatus*	Mottled Grasshopper

10 species of bird occurring in Japan. Results for each test group will be considered separately.

1. Results for British Orthoptera

Table 2 lists the 25 species used in 1 set of tests. The sounds were derived from a

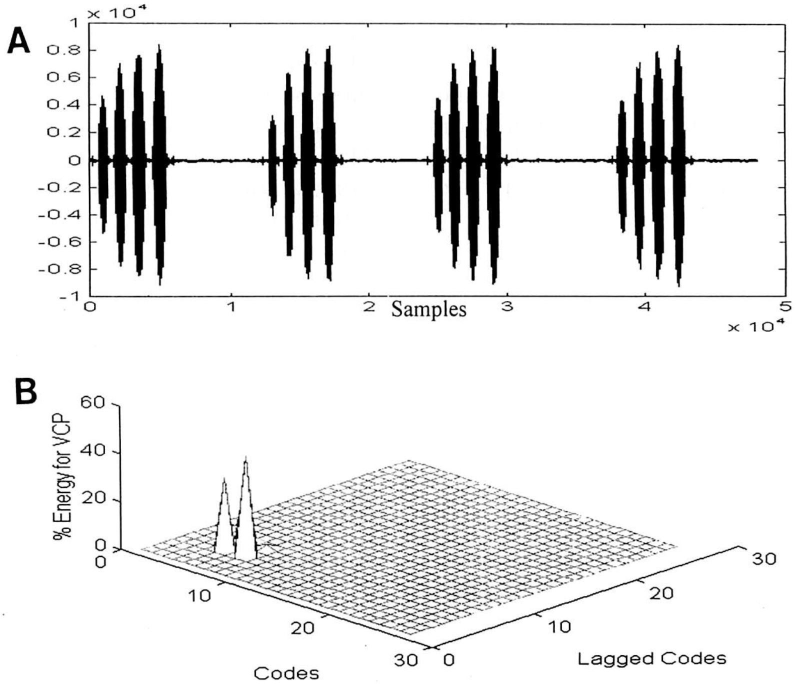

Fig. 2. *Gryllus campestris* (OR11): A, Time domain waveform; B, Average scaled A-matrix.

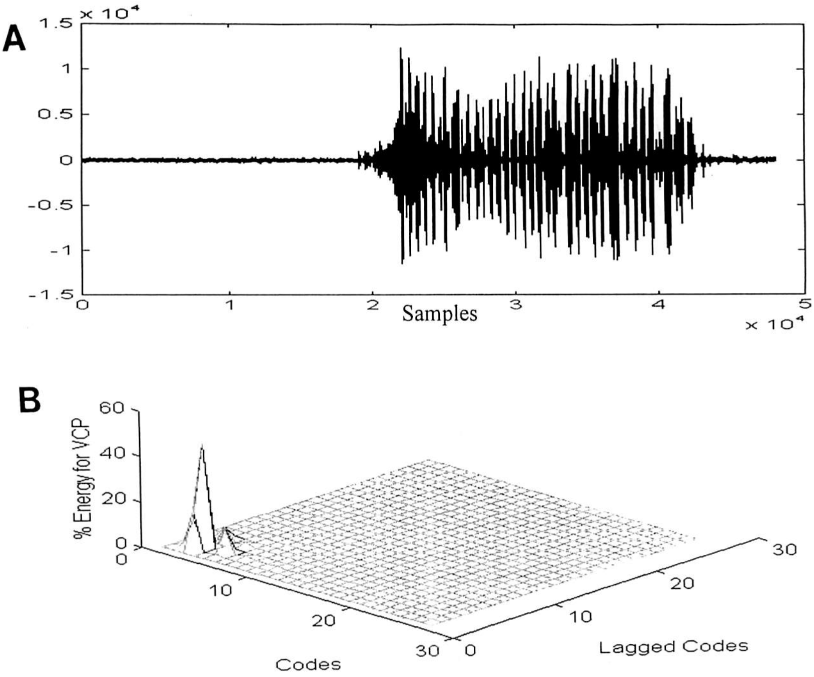

Fig. 3. *Chorthippus albomarginatus* (OR22): A, Time domain waveform; B, Average scaled A-matrix.

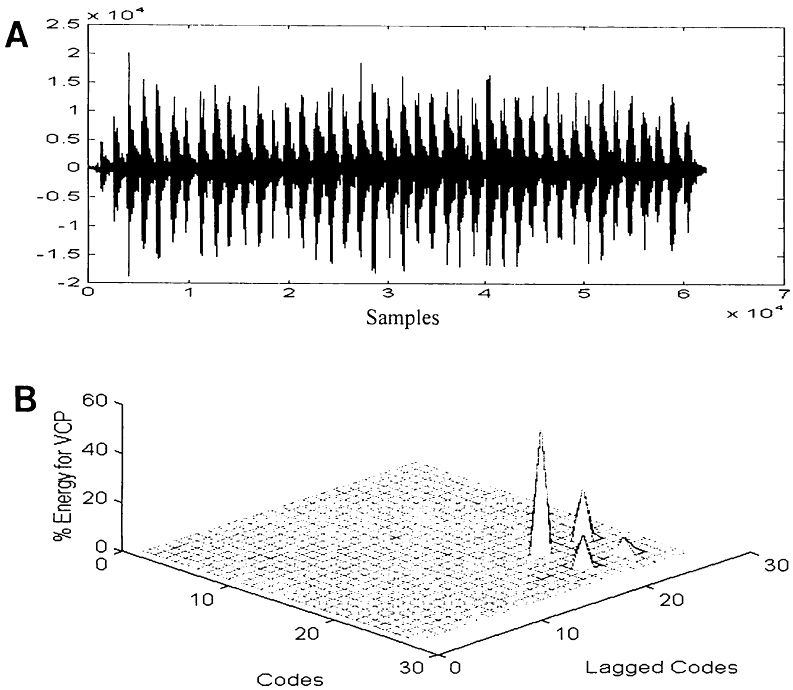

Fig. 4. *Meconema thalassinum* (OR01): A, Time domain waveform; B, Average scaled A-matrix.

widely available audio cassette (Burton and Ragge, 1987) available as an accompaniment to The Grasshoppers and Allied Insects of Great Britain and Ireland (Marshall and Haes, 1988) and digitised as previously described. TDSC symbols for up to 2 seconds of sound for each species. The subsequent A-matrices were used to train a single layer Perceptron consisting of 784 inputs (1 for each location in the A-matrix) and 25 output neurons, 1 for each species. Figs. 2, 3 and 4 show time plots and A-matrices for 3 species—*Gryllus campestris* (OR11), *Chorthippus albomarginatus* (OR22) and *Meconema thalassinum* (OR01). The single layer Perceptron was developed by Rosenblatt in the 1960's (Rosenblatt, 1962) and is a competitive learning network that performs n-dimensional discriminant analysis. The results suggest that TES is a good pre-processor providing wide separation of sounds which would have similar spectra. This is evident when the example A-matrices in Fig. 2B, 3B and 4B are examined. The differences between species are obvious, especially for *M. thalassinum* which has predominantly low frequencies compared with the other 2 species.

Once trained, the system was tested with new sounds; Table 3 shows identification results with various levels of added noise to simulate response to poorer conditions. Each entry is an average of 1000 normally distributed random A-matrices (zero mean, unity variance) added to the A-matrices which simulates Gaussian white noise over the whole frequency spectrum. It is evident from Table 3 that identification is very high (99–100%) under low noise conditions with the exception of *Metrioptera brachyptera* (OR06) and that mis-identification is zero until 30% noise is added. The latter is a fundamental requirement for an automated system.

However, it is important to note that some species exhibit a very rapid decline in identification accuracy (OR01—*Meconema thalassinum* and OR13—*Gryllotalpa gryllotalpa*). Both species have characteristically low dominant frequencies and this may contribute to confusion. However, it is known that the former species produces substrate-based sounds and would not be detected in a bioacoustic survey. Further work on determining the reasons needs to be carried out.

Table 3. Orthoptera species identification accuracy with added noise. For ID code, see Table 2.

ID Code	Noise Level (%)							
	1	3	5	10	20	30	40	50
OR01	99.9	83.5	54.2	24.3	8.7	5.3	4.7	4.3
OR02	100	100	100	100	96.7	89.3	82.3	76.8
OR03	100	99.6	94.8	76.8	63.8	60.6	58.6	53.6
OR04	100	100	100	99.9	94.1	86.8	79.3	72.0
OR05	100	100	99.4	90.1	71.1	66.9	59.0	57.9
OR06	85.6	65.5	58.8	54.3	54.3	52.0	52.0	47.9
NR07	100	100	99.9	94.2	65.0	45.1	29.1	21.6
OR08	100	100	100	100	99.1	95.1	87.4	79.0
OR09	100	100	100	100	100	99.7	98.4	95.5
OR10	100	100	100	100	99.4	94.8	84.5	76.0
OR11	100	100	100	100	100	100	100	100
OR12	100	100	100	100	99.1	93.8	84.1	73.8
OR13	100	84.9	63.7	34.2	16.5	9.3	7.6	6.0
OR14	100	100	100	100	93.4	85.7	76.1	73.6
OR15	100	100	99.7	89.3	74.9	61.5	58.9	52.2
OR16	100	99.9	98.0	84.9	68.0	63.8	59.2	59.4
OR17	100	86.4	74.4	63.4	58.2	52.8	50.2	45.8
OR18	100	98.1	86.7	59.7	47.0	38.1	35.4	32.8
OR19	100	100	99.9	95.6	80.7	71.5	62.8	62.9
OR20	100	100	99.9	91.3	77.1	67.7	63.7	60.5
OR21	100	100	100	100	99.8	97.5	92.7	81.8
OR22	99.8	86.1	74.1	62.3	54.5	56.4	53.5	54.5
OR23	100	100	100	99.7	94.1	84.6	78.5	71.9
OR24	100	99.6	92.0	65.8	48.9	39.6	36.7	30.6
OR25	100	100	100	100	100	99.8	98.9	96.6
Mean positive ident. (%)	99.4	96.1	91.8	83.4	74.6	68.7	63.7	59.5

Table 4. Species of Japanese bird used in IBIS tests.

ID Code	Latin Name	English Name
JB01	*Acrocephalus arundinaceus*	Great Reed Warbler
JB02	*Cuculus canorus*	Common Cuckoo
JB03	*Cettia diphone*	Bush Warbler
JB04	*Cuculus poliocephalus*	Little Cuckoo
JB05	*Emberiza variabilis*	Gray Bunting
JB06	*Ficedula narcissina*	Narcissus Flycatcher
JB07	*Megalurus pryeri*	Japanese Marsh Warbler
JB08	*Parus major*	Great Tit
JB09	*Phylloscopus tenellipes*	Pale-legged Willow Warbler
JB10	*Turdus chrysolaus*	Brown Thrush

Variations in classifier using MLPs and expert system identification have been assessed, all with reasonable results (Chesmore *et al.*, 1998; Chesmore, 1997c, 1998c).

2. Results for Birds in Japan

During the author's visit to Japan in March 1998, IBIS was tested on 10 species of bird (Table 4) which occur in woodland and grassland in Japan. The same approach was employed but the classifier was a MLP with 784 inputs (as before), 40 neurons in the hidden layer and 10 outputs (1 for each species). Sounds for training were obtained from CD (Kabaya and Matsuda, 1996a, b, c) and results were very encouraging. Table 5 gives the

Table 5. Preliminary results for bird identification matrix. For ID code, see Table 4.

	JB01	JB02	JB03	JB04	JB05	JB06	JB07	JB08	JB09	JB10
JB01	**1**	0	0	0	0	0	0	0	0	0
JB02	0	**1**	0	0	0	0	0	0	0	0
JB03	0	0	**1**	0	0	0	0	0	0	0
JB04	0	0	0	**1**	0	0	0	0	0	0
JB05	0	0	0	0	**1**	0	0	0	0	0
JB06	0	0	0	0	0	**1**	0	0	0	0
JB07	0	0	0	0	0	0	**1**	0	0	0
JB08	0	0	0	0	0	0	0	**1**	0	0
JB09	0	0	0	0	0	0	0	0	**1**	0
JB10	0	0	0	0	0	0	0	0	0	**1**

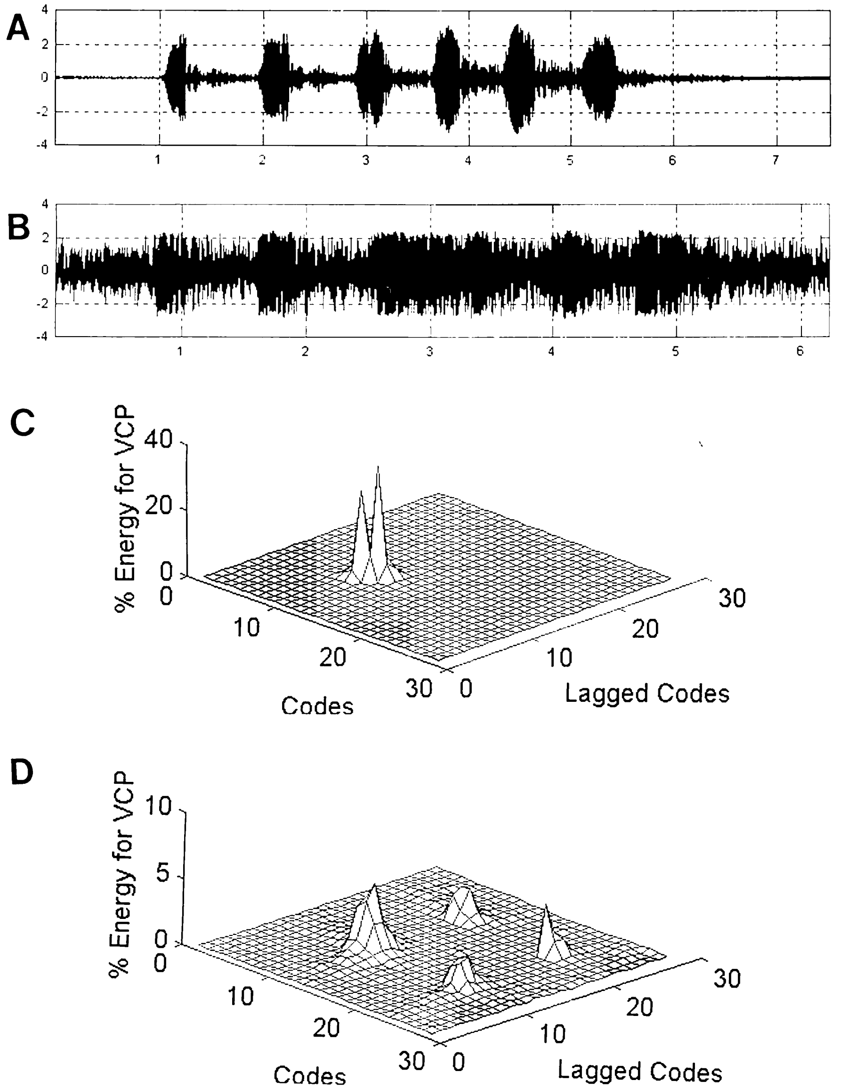

Fig. 5. *Cuculus poliocephalus*: A, Time domain waveform of the sound for training; B, Time domain waveform under natural conditions (woodland); C, A-matrix for the sound for training; D, A-matrix under natural conditions (woodland).

results in the form of a confusion matrix. It is evident that the recognition accuracy for good quality sounds is 100% with no mis-identification. The system was tested further by extracting a single call of *Cuculus poliocephalus* from a woodland recording by Dr Oba from the Natural History Museum and Institute, Chiba (CBM Acc. No. 043-0086) and presenting this to the system. Fig. 5A shows the time signal under good conditions, 5B the time signal under natural conditions, 5C the A-matrix for the bird under good conditions and 5D the A-matrix for the bird under natural conditions. Under these conditions the species was recognised with 86% accuracy.

Again, the research is still in its infancy and much work needs to be done. One particular area of research which is considered essential is to preserve the time structure of the sounds since bird (and higher animals) song is complex in both the time and frequency domains. Work at Hull is currently investigating the separation of individual syllables and applying syntactic pattern recognition techniques.

Image Processing Applications

Image processing is the second major sensor system for species identification. It is possible to use monochrome or colour images captured digitally, scanned from photographs or from scanning electron microscopes. It is evident from Table 1 that the range of applications is wider than for bioacoustics as image processing is more generic. For the purposes of this paper, it is convenient to divide the applications into species identification or discrimination and morphological analysis of individuals within populations.

Species identification research can be further sub-divided into categories describing the final potential application such as biodiversity assessment, discrimination between closely similar quarantine species, agricultural applications, paleobiology and paleobotany. Mention has already been made about image-based insect counting which aims to speed up sorting; other applications include identification of braconid hymenoptera (Weeks *et al.*, 1997; Weeks and Gaston, 1997) and solitary bees by their wing venation, leafhopper species (Dietrich and Pooley, 1994), blue-green algae (Thiel, 1994) and cyanobacteria. Discrimination of closely similar species for quarantine purposes will be described in detail later in the paper. Agricultural applications have concentrated on real-time identification of weed species for herbicide placement, part of the concepts of integrated crop management (ICM) and integrated pest management (IPM). Recent research has used fractals and Fourier analysis (Critten, 1996) and colour chromaticity (Shulin and Runtz, 1995). A related application using spectral reflectance to discriminate between plants and soil is described in Vrindts and De Baerdemaeker (1997). It should be noted that the latter project does not use image processing but near IR diffuse spectral reflection. Paleobotany research has been on identification of pollen from sediment cores in lakes to recreate environmental histories (Langford *et al.*, 1990); no other research projects have been found to date.

Morphological analysis (measurements) using image processing is well established in many fields such as biomedicine (e.g., X-ray images) and engineering (e.g. robotics, fault diagnosis) but less well so in biology and even less in entomology. The main entomological applications are for Lepidoptera to measure genetic influences on wing patterns, for example Windig *et al.* (1994) used an image processing system for quantifying seasonal polyphenism in species of *Bicyclus* butterflies.

Image Processing Applications at Hull

The image processing projects at Hull can be divided into two categories—morphological analysis of Lepidoptera and automatic discrimination of closely similar species. All the research has been carried out using custom software written in Visual C for Windows or Matlab. Matlab has the advantage of simplicity in prototyping algorithms but is interpreted and hence slow whereas C requires a detailed knowledge of programming but is very fast. Matlab also has a number of toolboxes such as image processing and artificial neural networks which speed up the development process. Images are derived

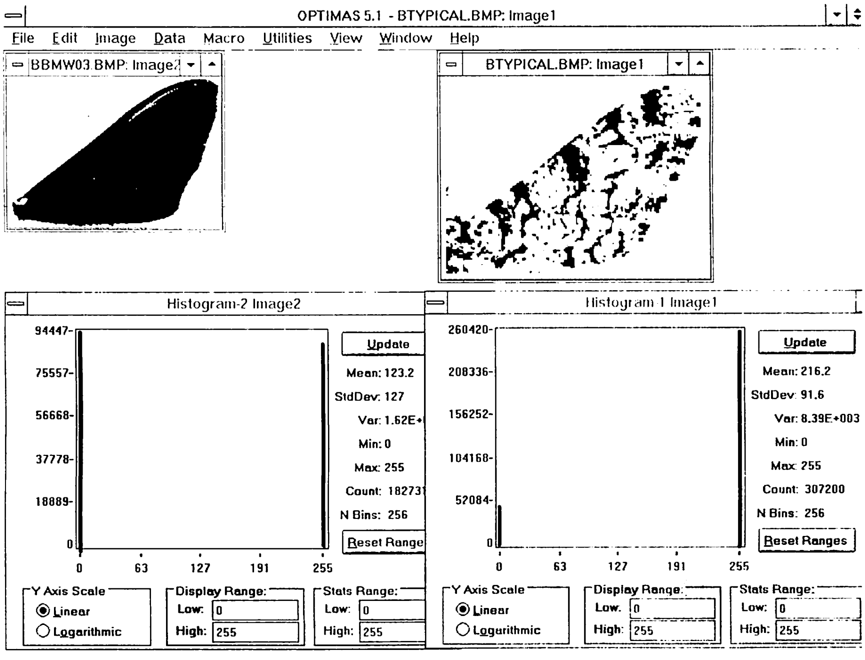

Fig. 6. Image processed *Biston betularia* forewings.

from digital cameras attached to microscopes or in the case of Lepidoptera directly from a digital camera; live specimens can be photographed without harm using an electroimmobilisation unit developed at Hull (Chesmore and Monkman, 1994).

Morphological Analysis

Morphological analysis of Lepidoptera has been attempted for a number of different applications including analysis of variation in British butterflies, quantitative analysis of melanism in *Biston betularia* for industrial melanism studies and analysis of the colour morphs of *Noctua pronuba*. One future application will be determination of fluctuating asymmetry in *Melanargia galathea*, a species of butterfly which occurs in sporadic populations in the Yorkshire Wolds not far from Hull. The first 2 applications will be discussed in more detail.

Quantitative Analysis of Melanism. Melanism is exhibited in a number of moth species in several countries in the World. Britain appears to have the highest number of species, the most notable being the Peppered moth, *Biston betularia* which has 2 common forms—f. *typica* which has a white background and many black spots, and f. *carbonaria* which is entirely black. F. *carbonaria* was thought to exist in very small proportions prior to the industrial revolution in Britain but became the predominant morph very rapidly after the onset of the revolution. In some heavy industrial localities such as Manchester, the proportion of f. *carbonaria* became 100% and remained so until the 1960's when a clean air act was introduced, reducing smoke pollution dramatically. Since then, the proportion of f. *carbonaria* has reduced in most localities as lichens have re-established. However, the traditional explanation of selective predation due to the light form being visible on dark, smoke covered tree trunks and hence being eaten, has some problems. In America, the same increases and decreases of a very closely related species have followed the same trend as in Britain but without the loss of any lichens. Also, *B. betularia* has an intermediate form, f. *insularia*, which is thought to be continuously variable between the 2 extremes. In addition, in 1 study in Helsinki using 2 closely related moth species (*Oligia latruncula* and *O. strigilis*) has further confused the situation in that *O. strigilis* has increased in melanic pro-

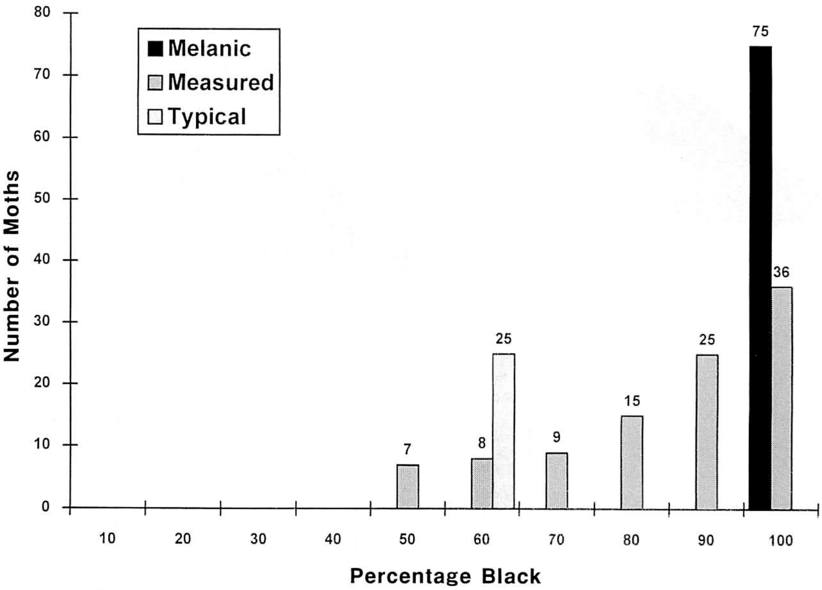

Fig. 7. Example of continuous variation in *Biston betularia.*

portions while *O. latruncula* has decreased over the same period. Image processing has been used to measure the proportions of black and white in a forewing of *B. betularia* in order to quantify the amount of melanism (Chesmore and Yorke, 1997a, b; Chesmore, 1998a). This is achieved by detecting the wing edge, using thresholding to discriminate between the white and black pixels and then counting the number of black pixels. The result is normalised against wing area and expressed as a percentage. Fig. 6 shows 2 wings, 1 of each form, processed in this manner, together with estimates of "blackness". Fig. 7 shows how it is possible to give more detail using this technique instead of classifying "normal" or "melanic" (includes f. *insularia*). Research is ongoing and it is hoped that more extensive trials will take place in 1999.

Analysis of *Noctua pronuba* Forms. This moth is extremely common in Britain and can be a pest in gardens. It exhibits up to 7 named colour forms and is sexually dimorphic. Because of this and it's abundance, it was chosen as a test subject for image analysis with an emphasis on colour discrimination of forms. In Yorkshire only 4 forms are commonly found; these are f. *ochrea* (female), f. *rufa* (female), f. *innuba* (male) and f. *ochrea-brunnea* (male), with the last being the commonest recorded at the trapping site (using a low power ultraviolet light trap). Images of detached forewings were obtained using a JVC colour CCD camera and digitised with a frame grabber installed in a Pentium computer. Each image was digitised to 272×320 pixels, 24-bit colour. Each wing was digitised with a blue background which provided a strong contrasting colour to the wings and could be removed by simple thresholding of the blue image. In 1 study 58 features were extracted from each wing image, including mean, median, mode, standard deviation, Kurtosis, energy and entropy of the green (G), red (R) and blue (B) channels, R–G, B–G, R–B covariance and correlation, moments and edge densities (Chesmore *et al.*, 1996). This number was reduced to 16 features describing each wing by considering the ranges of variability of the parameters and rejecting those with very small variability. The methods included principal components analysis, cluster analysis and unsupervised neural networks. In cluster analysis and neural networks, 5 clusters were selected to correspond to the 4 forms and a fifth as a "catch-all". Different colour forms can be distinguished successfully, f. *ochrea* being almost always classified perfectly. It should also be noted that some entomologists consider several of the forms as continuously variable and the named forms are extremes; this is borne out to some degree in these results since in most cases, f. *ochreabrunnea* is spread across several groups.

Automatic Discrimination of Closely Similar Species

This research is more recent than previously described work and there are fewer results. The two projects under way are both related to the early detection of quarantine species which are considered to be too dangerous (in terms of economy) to be allowed into Britain. The work is in association with Central Science Laboratory (CSL) located near York and one of its roles is to examine consignments of plants for import to assess the presence of pests. It is important to be able to detect the presence of pests as early and accurately as possible and this is carried out manually by experts. The taxa selected for the projects are *Liriomyza* spp (Diptera: Agromyzidae), a leafminer and *Colletotricrichum* spp, a fungus causing strawberry black spot.

Detection of *Liriomyza* spp. *Liriomyza trifolii* is an established pest in Britain on various horticultural crops such as tomato and chrysanthemum whereas *L. huidobrensis* is a quarantine pest and cannot be allowed into the country. It is therefore important to provide rapid and accurate identification, allowing for more specific targetting of pesticides and other control and eradication measures, minimising environmental damage, preventing further infestation and saving money. Identification is important but time consuming, often requiring identification of the larvae or waiting until they become adults. Image processing of the leaf mines may provide a solution. The ongoing project (Pether, 1998) is investigating whether it is possible to discriminate mines from each species; this is complicated by the fact that both species are polyphagous and the mine character is dependent on hostplant species. Fig. 8A shows a typical leaf mine and Fig. 8B shows how it is possible to discriminate between the leaf and mine using simple edge detection.

Discrimination between *C. acutatum* and *C. gloeosporioides*. *C. acutatum* is an EC listed quarantine organism and is separated from *C. gloeosporioides* in part by examination of conidia which are morphologically similar. One of the key characters used is the shape of the spore's apex—described as acute for *C. acutatum* and obtuse for *C. gloeosporioides*. The aim of the project was to ascertain if image analysis could quantify this difference and reliably separate these species. Images were obtained from CSL and were derived

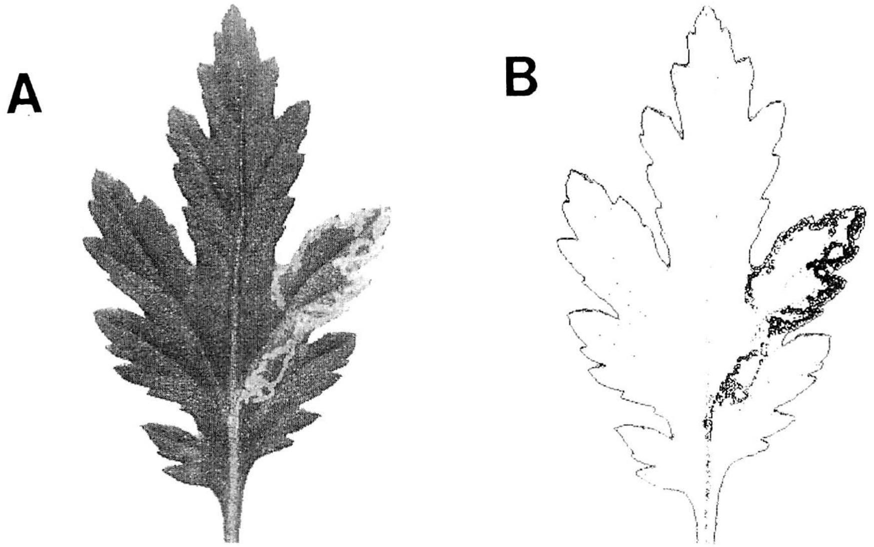

Fig. 8. *Chrysanthemum*: A, Leaf with leaf mine; B, Image processed leaf showing mine.

from microscope slides at 400 times magnification, digitised using a three chip colour camera (JVC-TK 1270E) and captured with a Snapper 24 frame grabber using computer software from Optimas (Seattle, USA). Initially, analysis was performed using Optimas (version 5.2) at CSL. Length (principal axis), breadth (minor axis) and area of conidia were measured and the data exported to Microsoft Excel where the area bounding the spore (length × breadth) and the area/bounding area were calculated. Images were then analysed using Matlab at Hull. A theoretical ellipse was generated from the principal and minor axes and then compared with the actual spore contour (outline). Secondly, the principal and minor axes were used to separate each spore into 4 segments and individual areas calculated. These were placed in a vector and multiplied by the Hadamard matrix to calculate the Hadamard-Walsh spectrum of the spore shape (related to symmetry properties). Results showed that there was a statistically significant difference between *C. acutatum* and *C. gloeosporioides* for the area/bounding area for conidial line drawings. Although, a similar trend was obtained for conidia from cultures it was not statistically significant. Comparison of spores with a theoretical ellipse failed to separate species reliably. The Hadamard function, however, showed greater promise. The vertical component for *C. acutatum* ranged from 1–2 whilst for *C. gloeosporioides* it was between 0–1, showing reasonable discrimination (Lane *et al.*, 1998).

Conclusions

The paper has given a brief outline of computer-aided taxonomy, concentrating on automated species identification. It is evident that there are many possible applications for automatic identification, perhaps the most important being the development of more rapid biodiversity assessment methods.

References

Anderson, S. E., A. S. Dave and D. Margoliash. 1996. Template-based automatic recognition of birdsong syllables from continuous recordings. J. Acoust. Soc. Am. 100 no. 2 (1): 1209–1217.

Balfoort, H. W., J. Snoek, J. R. M. Smits, L. W. Breedveld, J. W. Hofstraat and J. Ringelberg. 1992. Automatic identification of algae: neural network analysis of flow cytometric data. J. Plankton Res. 14: 575–589.

Boddy, L., C. W. Morris, M. F. Wilkins, G. A. Tarran and P. H. Burkill. 1994. Neural network analysis of flow cytometric data for 40 marine phytoplankton species. Cytometry 15: 283–293.

Burton, J. F. and D. R. Ragge. 1987. Sound Guide to the Grasshoppers and Allied Insects of Great Britain and Ireland. Harley Books, Great Britain.

Chesmore, E. D. 1997a. Methodologies for automating the identification of species. First Bionet International Working Group on Automated Taxonomy, University of Wales, Cardiff.

Chesmore, E. D. 1997b. Future directions in computer-assisted taxonomy. Fifth Workshop of the ESF Network in Systematic Biology on New Directions in Systematics, Crete.

Chesmore, E. D. 1997c. Neural networks and expert systems for automated insect identification. British Ecological Society Ecological Computing Group Workshop on Machine learning methods for ecological applications, Manchester Metropolitan University.

Chesmore, E. D. 1998a. New ways to measure melanism. Meeting of the Institute of Biology, Royal Entomological Society and Lancashire and Cheshire Entomological Society on A Fresh Look at Melanic Insects, Liverpool Museum.

Chesmore, E. D. 1998b. BioNET INTERNATIONAL Group for Computer-assisted Taxonomy (BIGCAT). BioNET INTERNATIONAL NEWS 2: 8.

Chesmore, E. D. 1998c. Automated identification of insects by sound: application to detecting and identifying species using acoustics in the field. Entomology '98, University of Exeter.

Chesmore, E. D., A. G. Hall and S. Flandin. 1996. Colour analysis of Lepidoptera for species identification. Conference on Computer-aided Species Identification, University of Kent.

Chesmore, E. D. and Monkman, G. 1994. Automated analysis of variation in Lepidoptera. The Entomologist 113 (3/4): 171–182.

Chesmore, E. D. and D. R. Morse. 1997. Automating the identification of species—a viable proposition? First Biennial International Conference of the Systematics Association, University of Oxford.

Chesmore, E. D. and C. Nellenbach. 1997. Acoustic Methods for the Automated Detection and Identification of Insects. Third International Symposium on Sensors in Horticulture, Tiberias, Israel.

Chesmore, E. D., M. D. Swarbrick and O. P. Femminella. 1998. Automated analysis of insect sounds using TESPAR and expert systems—a new method for species identification. *In* Information Technology, Plant Pathology and Biodiversity, pp. 273–287. CAB International, Wallingford, UK.

Chesmore, E. D. and L. Yorke. 1997a. New approaches for the quantification of melanism in Lepidoptera. Royal Entomological Society Northern Meeting, Liverpool Hope University.

Chesmore, E. D. and L. Yorke. 1997b. Image processing applications for Lepidoptera. Royal Entomological Society Northern Region Meeting, Newcastle.

Critten, D. L. 1996. Fourier based techniques for the identification of plants and weeds. J. Agric. Enging. Res. 64: 149–154.

Dallwitz, M. J., T. A. Paine and E. J. Zurcher. 1998. Interactive keys. *In* Information Technology, Plant Pathology and Biodiversity, pp. 201–212. CAB International, Wallingford, UK.

Dietrich, C. D. and C. D. Pooley. 1994. Automated identification of leafhoppers (Homoptera: Cicadellidae: *Draeculacephala* Ball). Annals of the Entomological Society of America 87(4): 412–420.

Edwards, M. and D. R. Morse. 1995. The potential for computer-aided identification in biodiversity research. TREE 10(4): 153–158.

Gonzales, N. N. 1986. Insect Identification using Template Matching: a Pilot Study. Master's Thesis, New Mexico State University.

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

Kabaya, T. and M. Matsuda. 1996a. The Songs and Calls of 333 Birds in Japan. Vol. 1, CD No. 3. 200 pp. Shogakukan, Tokyo.

Kabaya, T. and M. Matsuda. 1996b. The Songs and Calls of 333 Birds in Japan. Vol. 2, CD No. 5. 200 pp. Shogakukan, Tokyo.

Kabaya, T. and M. Matsuda. 1996c. The Songs and Calls of 333 BIrds in Japan. Vol. 2, CD No. 6. 200 pp. Shogakukan, Tokyo.

King, R. A. and W. Gosling. 1978. Time-encoded speech. Elec. Lett. 15: 1456–1457.

Lane, C. R., H. Jeuland, A. G. Hall and E. D. Chesmore. 1998. Image analysis as an aid to discrimination between *Colletotrichum acutatum* and *C. gloeosporioides*, Seventh International Conference of Plant Pathology (ICPP98), Edinburgh.

Langford, M., G. E. Taylor and J. R. Flenley. 1990. Computerised identification of pollen grains by texture analysis. Rev. Palaeobot. Palynol 64(1/4): 197–203.

Lucking, W. G., M. Darnell and E. D. Chesmore. 1994. Acoustical condition monitoring of a mechanical gearbox using artificial neural networks. IEEE Conference on Neural Networks, Orlando, Florida.

Marshall, J. A. and E. C. M. Haes. 1988. Grasshoppers and Allied Insects of Great Britain and Ireland. Harley Books, Great Britain.

Martin, L. V., T. K. Stanton, P. H. Wiebe and J. F. Lynch. 1996. Acoustic classification of zooplankton. ICES Journal of Marine Science 53: 217–224.

McBraith, A. L. and H. C. Card. 1995. Birdsong recognition with DSP and neural networks. IEEE WESCANEX '95 Proceedings, pp. 409–414.

Mills, H. 1995. Automatic detection and classification of nocturnal migrant bird calls. J. Acoust. Soc. of Am. 97(5): 3370–3371.

Oba, T. 1994. Sampling methods for the study of the natural sound environment in Japan: consideration of the sample time unit. Nat. Hist. Res. 3(1): 27–32.

Oba, T. 1995. What is the natural sound diversity? A consideration for the local natural amenity. Nat. Hist. Res. 3(2): 173–185.

Pankhurst, R. J. 1991. Practical Taxonomic Computing. 202 pp. Cambridge University Press, Cambridge.

Pether, A. M., 1998. Computer-aided Taxonomy of Leaf-miners. PhD Transfer Report, University of Hull.

Riede, K. 1993. Monitoring biodiversity: analysis of Amazonian rainforest sounds. Ambio 22(8): 546–548.

Roesnblatt, F. 1962. Principles of Neurodynamics. Spartan, New York.

Scalabrin, C., N. Diner, A. Weill, A. Hillion and M. Mouchot. 1996. Narrowband acoustic identification of monospecific fish shoals. ICES J. Marine Sci. 53: 181–188.

Shulin, D. and K. Runtz. 1995. Image processing methods for identifying species of plants. IEEE WESCANEX '95 Proceedings, pp. 403–408.

Shuman, D., J. A. Coffelt, K. W. Vick and R. W. Mankin. 1993. Quantitative acoustical detection of larvae feeding inside kernels of grain. J. Economic Entomol. 86(3): 933–938.

Shuman, D., D. K. Weaver and R. W. Mankin. 1997. Quantifying larval infestation with an acoustical sensor array and cluster analysis of cross-correlation outputs. J. Appl. Acoust. 50(4): 279–

296.

Simmonds, E. J., F. Armstrong and P. J. Copland. 1996. Species identification using wideband backscatter with neural network and discriminant analysis. ICES J. Marine Sci. 53: 189–195.

Smith, A. D., J. R. Riley and R. D. Gregory. 1993. A method for routine monitoring of the aerial migration of insects by using a vertical-looking radar. Phil. Trans. R. Soc. Lond. B 340: 393–404.

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

Systematics Agenda 2000. 1994. Systematics Agenda 2000: Charting the Biosphere. Technical Report, Systematics 2000.

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

Thiel, S. U. 1994. The Use of Image Processing Techniques for the Automated Detection of Blue-Green Algae. PhD Thesis, University of Glamorgan.

Vrindts, E. and J. De Baerdemaeker. 1997. Optical discrimination of crop, weed and soil for on-line weed detecton. Bios Scientific Publishers 1: 537–544.

Weaver, D. K., D. Shuman and R. W. Mankin. 1997. Optimising and assessing the performance of an algorithm that cross-correlates acquired acoustics emissions from internally feeding larvae to count infested wheat kernels in grain samples. J. Appl. Acoust. 50(4): 297–308.

Weeks, P. J. D. and K. J. Gaston. 1997. Image analysis, neural networks and the taxonomic impediment to biodiversity. Biodiversity and Conservation 6: 263–274.

Weeks, P. J. D., I. D. Gauld, K. J. Gaston and M. A. O'Neill. 1997. Automating the identification of insects: a new solution to an old problem. Bull. Entomol. Res. 87: 203–211.

Windig, J. J., P. M. Brakefield, N. Reitsma and J. G. M. Wilson. 1994. Seasonal polyphenism in the wild: survey of wing patterns in five species of *Bicyclus* butterflies in Malawi. Ecol. Entomol.19: 285–298.

(Accepted 9 March 1999)

技術移転：電子工学技術を生態学と昆虫学における種識別へ適用する

David Chesmore

School of Engineering, University of Hull
Hull, HU6 7RX, England

電子工学やコンピューター技術の最近の発達は，生物学，生態学および環境保全に新しい適用の道を開いている．なかでもこうした発展ゆえに際立った研究領域として，コンピュータ依存型分類学が現実化しつつある．コンピュータ依存型分類学は自動種識別，コンピュータを基礎とする検索表，および分岐学のような分類法を包含する．一方で，経済的に重要な分類群のコンピュータを基礎とする検索表の発展には相応な研究努力が投入されているが，種の自動識別には相対的に僅かな研究しか行われていない．自動識別が必要とされる応用は多様で，昆虫の個体数のカウント，種類の選別，有害昆虫のモニタリングおよび生物学的多様性評価を含む．したがって，本稿は自動識別研究に着目し，ハル大学他で近年および目下進行中の課題である，昆虫と鳥類の生物音響学的識別と検疫菌類や昆虫類の区別への画像解析の応用について記述する．本稿は，1998年3月21日に千葉県立中央博物館で行われた公開講演会に基づくものである．