ELSEVIER

Discriminating signal from noise: Recognition of a movement-based animal display by artificial neural networks

Richard A. Peters [a,*], Colin J. Davis [b]

[a] *Animal Behaviour Lab & Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia*
[b] *Macquarie Centre for Cognitive Science, Macquarie University, Sydney, NSW 2109 Australia*

Abstract

In this study, we investigated the feasibility of applying neural networks to understanding movement-based visual signals. Networks based on three different models were constructed, varying in their input format and network architecture: a Static Input model, a Dynamic Input model and a Feedback model. The task for all networks was to distinguish a lizard (*Amphibolurus muricatus*) tail-flick from background plant movement. Networks based on all models were able to distinguish the two types of visual motion, and generalised successfully to unseen exemplars. We used curves defined by the receiver-operating characteristic (ROC) to select a single network from each model to be used in regression analyses of network response and several motion variables. Collectively, the models predicted that tail-flick efficacy would be enhanced by faster speeds, greater acceleration and longer durations.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Lizards; Animal communication; Movement; Neural network; Classification

Research on animal communication systems has considered the information content of signals and the factors that influenced their design. Constraints on signal design affect signal generation, transmission through the environment, and reception (Endler, 1992). Of recent interest is the role that receivers (e.g., social companions, opponents or potential mates) have played in signal evolution (Ryan and Rand, 1990; Endler, 1991; Guilford and Dawkins, 1991; Pagel, 1993; Endler and Basolo, 1998). Signals must be designed to stimulate the sense organs of intended receivers. Some signals have evolved to exploit these pre-existing biases (Ryan and Rand, 1993; Basolo, 1990), while others are likely to be a product of the interaction between receiver sensory systems and environmental conditions, or 'sensory drive' (Endler, 1992). A detailed understanding of signal structure has helped to disentangle the evolutionary factors responsible for their design such as for acoustic signals (summarised in Hopp et al., 1998) and static visual signals (e.g., colour patterns: Endler, 1990). It also allows for investigations of the processes involved in signal detection, and the mechanisms that ensure signals are discriminated from the background.

The quantitative analysis of signal structure is often combined with empirical investigations whereby predictions based on structure can be tested. Designing experiments of this kind becomes increasingly challenging as the number and complexity of parameters that can be manipulated increases. Typically, however, researchers make intuitive judgements about the parameters that are likely to be important to receivers, based on their own knowledge of the system. An alternative approach is to use models such as artificial neural networks (henceforth called neural networks) that act in the place of a receiver to identify a subset of variables worthy of closer attention. Neural networks draw upon the general properties of nervous systems, such that sensory information is distributed across a network of neuron-like 'units', and response to incoming information is determined by the strength of connections between them. A major advantage of this approach over more traditional computational models is their ability to learn. Neural networks have been used for classifying animal signals, and have already made important contributions to understanding evolutionary processes, sensory biases and constraints on signal design (Enquist and Arak, 1993; Johnstone, 1994; Bateson and Horn, 1994; Phelps and Ryan,

* Corresponding author. Present address: Centre for Visual Sciences, Research School of Biological Sciences, Australian National University, Canberra, ACT 2601, Australia. Tel.: +61 2 6125 5066; fax: +61 2 6125 3808.
  E-mail address: richard.peters@anu.edu.au (R.A. Peters).

1998; Deecke et al., 1999; Borst and Haag, 2002; Enquist and Ghirlanda, 2005).

There is increasing evidence that neural networks offer a useful alternative for classifying animal signals. Traditional methods for classification, such as discriminant function analysis, cluster analysis, or multi-dimensional scaling, can be relatively inefficient. In contrast, neural networks perform well when the input is noisy and there is no *a priori* basis for (within species) classification. Reby et al. (1997) trained a neural network to discriminate between the vocalisations of four male fallow deer (*Dama dama*), achieving a discrimination accuracy of 88%. Potter et al. (1994) used a similar network to discriminate marine mammal vocalisations (bearded seal, *Erignathus barbatus*; beluga whales, *Delphinapterus leucuas*) from noise (including wave noise and a variety of sounds produced by ice). Increased sample sizes and an alternative approach to defining input resulted in an accuracy of 98.5%. Neural network classification of marine mammal vocalisations have recently been used to provide an index of acoustic similarity suitable for comparisons across studies, species, and time (Deecke et al., 1999).

Neural networks have also been used to explore how sensory and perceptual processes might influence the evolution of biological signals. Enquist and Arak (1993) modelled one way in which biases inherent in signal recognition systems could influence signal design, and Johnstone (1994) used a similar network and training procedure to illustrate preferences for low levels of fluctuating asymmetry and ornament elaboration. Similar strategies were adopted to model the evolution of exploitation and honesty (Krakauer and Johnstone, 1995), courtship rituals (Wachtmeister and Enquist, 2000), and mimicry (Holmgren and Enquist, 1999). In a different approach, Phelps and Ryan (1998) used genetic algorithms, whereby network 'fitness' reflects the effect of selection, evolving networks to recognise calls of the Túngara frog (*Physalaemus pustulosus*) against noise with the same amplitude envelope. Remarkably, responses of female frogs in playback experiments showed the same biases as the networks. Work of this kind illustrates how modelling can elucidate the relationship between signal structure and receiver responses (see also Ghirlanda and Enquist, 1998; Phelps and Ryan, 2000; Phelps et al., 2001). Used in this way, neural networks help to focus research strategy before time-consuming fieldwork or captive experiments are undertaken.

To our knowledge, neural networks have not been applied to classify movement-based animal signals. Recent developments in the quantitative analysis of visual motion (Zeil and Zanker, 1997; Peters et al., 2002) now make it possible to extend their use to this class of signals. Our goal in the present study was to investigate whether neural networks can be trained to recognise two different classes of visual motion, and to examine which motion parameters are likely to mediate network responses; importantly, our goal was to generate testable predictions rather than to recreate the processing capabilities of the brain. We constructed several types of networks, however, the task for each was to discriminate between the motion characteristics of a movement-based animal signal and irrelevant background motion. Lizards use movement-based displays in territorial disputes between rival males that function to provide

information about the signaller, such as size and condition. Displays by the Australian jacky lizard (*Amphibolurus muricatus*) are comprised of five distinct motor patterns delivered in an obligatory sequence and typically begin with a series of tail flicks (Peters and Ord, 2003). Therefore the two classes of visual motion presented to the networks comprised the introductory tail-flicks of jacky lizards as the class requiring detection, and the movement of wind-blown plants as the irrelevant background motion against which tail-flicks are typically seen. Building upon analyses of signal structure (Peters and Evans, 2003a), the modelling work presented herein helped design playback experiments testing the efficacy of different tail-flicking strategies in this system (Peters and Evans, 2003b). By including neural network modelling in our research program we identified both expected (speed and acceleration effects: e.g., Fleishman, 1992) and unexpected (a duration effect) mediators of recognition. We have since obtained strong experimental evidence that the duration effect is principally responsible for signal efficacy (Peters and Evans, 2003b).

1. Materials and methods

We constructed three different neural network models that varied in their input format and network architecture. In the *Static Input model*, the input was collapsed over time. The response of a population of artificial sensory units, each tuned to a particular velocity and summed across the time domain (see Peters et al., 2002), was used as the input to a feed-forward network with backward propagation of error to modify connection weights (e.g., Haykin, 1999). In subsequent models, we introduced time and considered both signal and noise as temporal sequences. The *Dynamic Input model* considered sensory unit responses over time and similarly comprised a feed-forward architecture. However, this model was not error-correcting, rather, weights were updated using an associative learning rule (Grossberg, 1976) in a fashion conceptually similar to Hebbian learning. The *Feedback model* used a recursive architecture specifically designed to capture temporal variation (Elman, 1990; Phelps and Ryan, 1998). In contrast to feed-forward networks, an additional layer stores activity so that the network 'remembers' its previous internal state. The inputs to the Feedback model were estimates for speed over time.

The training strategy for all three models was the same irrespective of architecture. Tail-flick and wind-blown plant sequences were randomly allocated to non-overlapping training and test sets. The training set was used to train each network to distinguish the signal from noise. Network performance was assessed in terms of its ability to generalise to the unseen exemplars in the test set. The Static Input and Feedback models were implemented using the Neural Network Toolbox (v.3) for Matlab (v.5) on a Macintosh computer. The Dynamic Input model was custom written in C++ and implemented on a PC running Windows 2000.

The complete data set used for training and testing each of the models comprised 181 tail-flick sequences and 64 sequences of wind-blown plants (*Acacia longifolia*, *Grevillea linearfolia*,

Kunzea ambigua, *Pteridium esculentum*, *Xanthorrhea arborea*) recorded at the field site where these lizards were captured and representative of the plant species that form the background to their displays. Recordings of wind-blown plants were undertaken in summer during light to typical wind conditions over the range from 0.7 to 2.0 m s^{-1} (see Peters and Evans, 2003a for details). The tail-flick sequences included variations in the angle and distance between the lizard and the camera used to record the display. Successful training thus required that the model develop some degree of translation invariance. Tail-flicks were recorded in response to the introduction of a conspecific male placed in front of a resident's enclosure (see Peters and Ord, 2003 for details). All sequences were quantified using the approach described by Peters et al. (2002). Briefly, digital video footage is converted to image sequences of intensity (greyscale) values at 25 frames per second (PAL standard). A computational motion analysis algorithm is used to calculate velocity estimates by tracking changes in image intensity. The algorithm assumes a locally constant image structure and calculates the velocity field from temporal and spatial derivatives of filtered versions of image intensity. Output from this procedure comprises estimates of velocity for each frame of the image sequence. As velocity is determined separately for all spatial locations (i.e., all intensity values) a given frame will therefore comprise many estimates. In the first two models, the velocity field is summarised by a family of 48 sensory units that have a preferred direction and speed of movement. Velocity information is therefore reduced to the response levels of these sensory units. Input to the third model comprises a single estimate of vector magnitude (or 'speed') per frame.

1.1. Static Input model

1.1.1. Network input

The velocity field of a subset of tail-flick and plant sequences was first quantified using the procedure described above. The resulting velocity estimates were then summarised using a set of 48 artificial sensory units defined by Gaussian functions with maximum values tuned to a particular direction and speed of movement (16 directions × 3 speeds; as per Peters et al., 2002; Peters and Evans, 2003a,b, the sensory units are used to summarise velocity rather than reflect actual units in the lizard brain). All estimates from a given frame are presented to each of the sensory units. For a given sensory unit, if an estimate exactly matches its preferred direction and speed, a value of 1 is added to the response level of this unit; if the estimate is outside the range for a given sensory unit, in the opposite direction for example, nothing is added to the response level of this unit. This is done for each frame of the sequence, so that sensory unit response levels represent the combined response to all frames of the image sequence. Sensory unit responses were then length normalised to restrict input to the range [0,1]:

$$I_i = R_i((R_1{}^2 + R_2{}^2 + R_3{}^2 + \cdots + R_{48}{}^2)^{-0.5}) \quad (1)$$

The average normalised response of sensory units to tail-flicks and wind-blown plants in the training set is presented in Fig. 1;
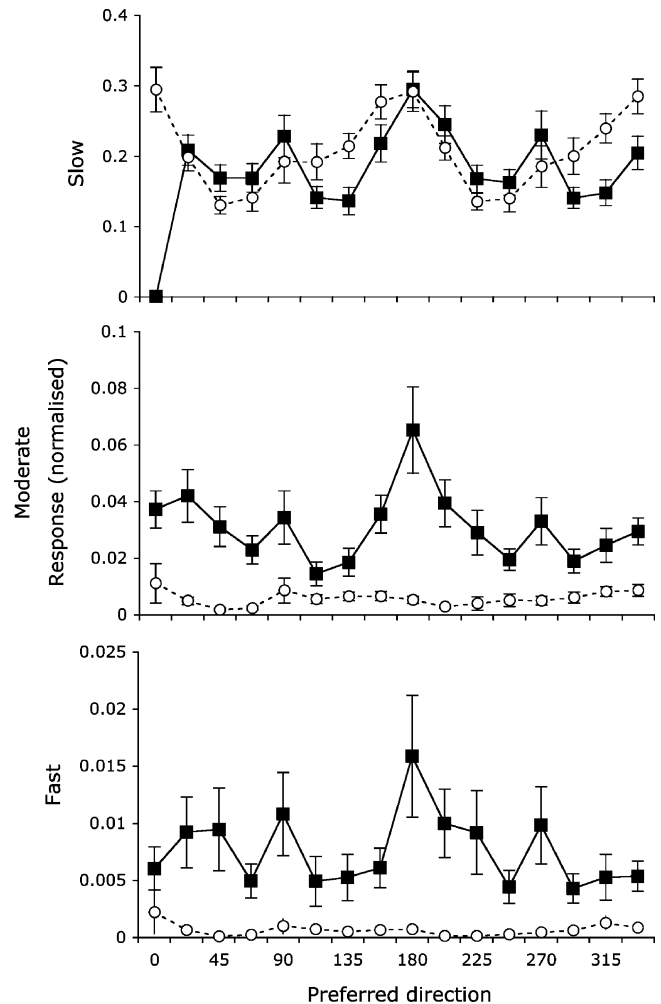

Fig. 1. Mean (±S.D.) normalised sensory unit response to the tail-flick (■) and plant (○) sequences presented to the Static Input model. Preferred direction is plotted on the *x*-axis, and separate plots are presented for the slow (top), moderate (middle) and fast (bottom) speed preferences.

this defined the task for the network. Sequences were randomly allocated to either a training (35 tail-flick and 33 plant sequences) or test set (50 tail-flick and 31 plant sequences).

1.1.2. Network architecture

Preliminary modelling determined an appropriate architecture (Peters, 2003). The model comprised the input, two hidden layers and an output layer (Fig. 2a). The 48 inputs were separated according to their relative speed preference resulting in three input layers with 16 elements each (i.e., 16 preferred directions). Each input layer was connected to a hidden layer of four units via fixed connections that were not updated during training (see Fig. 2a). These units were connected to all units in a second hidden layer, where the number of units varied (*n* = 2, 7, 13 or 18). These hidden layer units were, in turn, all connected to a single unit in an output layer. Connections between the first and second hidden layers, and from the second hidden layer to the output layer were subject to learning. The activity level of hidden units in the second hidden layer and the output layer unit were defined by a sigmoid transfer function, limiting unit output

in the range [0,1]:

$$Y = (1 + \exp^{-a})^{-1} \tag{2}$$

where *a* is the sum of weighted inputs plus the unit's bias.

1.1.3. Network training

We supervised training by giving the network a target output for all training vectors: 0.8 for tail-flicks and 0.2 for plant

sequences. As the response from the output unit is constrained in the range [0,1], a target lower than the maximum allows for the detection of 'super-normal' tail-flick sequences that exceed those upon which the networks were trained. The gradient descent back-propagation algorithm (Werbos, 1974; Rumelhart et al., 1986) was used to train the networks. Briefly, network error (i.e., the difference between the target and the output produced by the network) is a function of the weights connecting network units and so adjustments should be made so as to minimise this error function. Weight changes are made in an iterative manner whereby the error is passed backwards through the network so that the relative contribution of a given weight to the overall error can be determined and small changes to the weight can be performed. New weights are selected in the direction of the steepest gradient (or slope) of the error function, until it reaches a minimum. We varied both the number of units in the second hidden layer (2, 7, 13 and 18) and the duration of training, measured in *epochs* (i.e., the number of times the input set is presented to the network: $10^3 \times 1$, 5, 10, 20, 50 and 100 epochs). The factorial combination of hidden layer unit size (four levels) and training duration (six levels) resulted in a total of 24 networks.

1.1.4. Analysis of network performance

The performance of each of the 24 networks was examined first by presenting the training set to the trained networks.
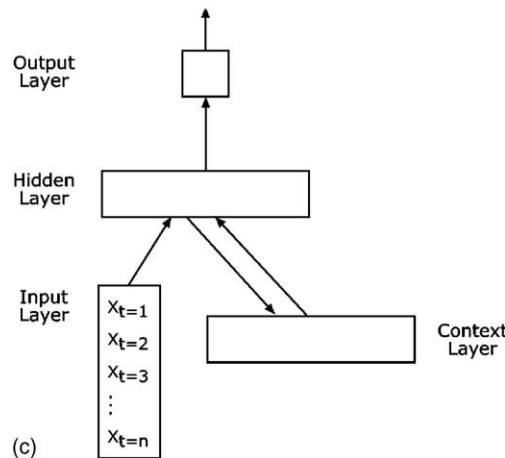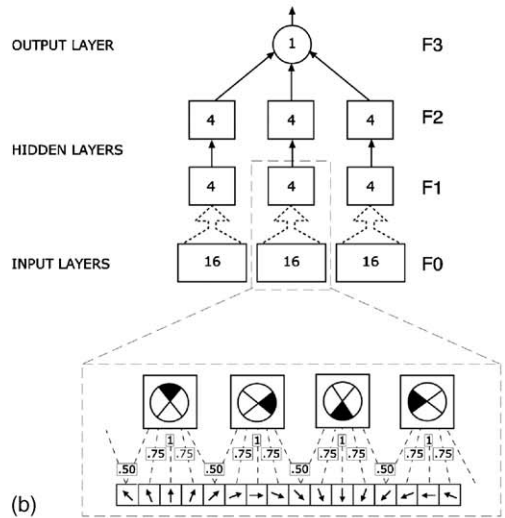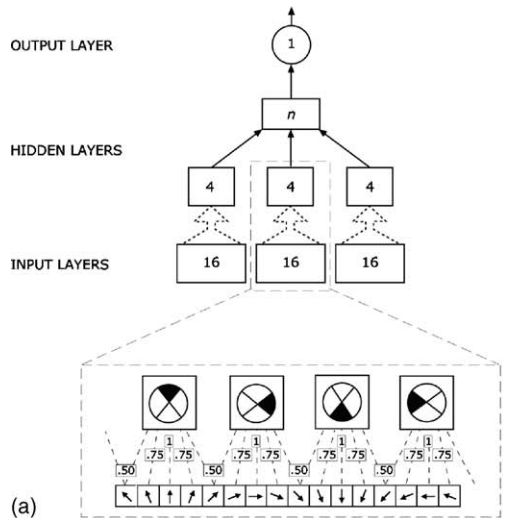
Fig. 2. Schematic representation of the neural network models used to discriminate the tail-flick signal from plant movement. In all models, dashed lines represent fixed weights not modified during training. (a) The *Static Input model* was a feed-forward network with two hidden layers and an output layer. There were 48 inputs to the network representing different velocities summed over time. Inputs were separated according to their relative speed preference resulting in three input layers of 16 elements each (i.e., 16 preferred directions). Each input layer was connected to a hidden layer of four units via fixed connections that were not updated during training. This pre-processing reduced direction preferences from 16 to 4 units. These units were connected fully to a second hidden layer, where the number of units varied ($n = 2, 7, 13$ or 18). These hidden layer units were, in turn, all connected to a single unit in an output layer. Connections between the first and second hidden layers, and from the second hidden layer to the output layer were subject to learning. The activity level of hidden units in the second hidden layer and the output layer unit were defined by a sigmoid transfer function; no transfer function was used in the first hidden layer. Training was achieved using the gradient descent back-propagation algorithm. (b) The *Dynamic Input model* also comprised a feed-forward network with two hidden layers and an output layer. The input layer (F0) comprising 48 sensory unit responses by *N*-frames, separated into three speed classes, were collapsed into four general directions of movement and length normalised in the first hidden layer (F1). The output from F1 was passed to the second hidden layer (F2) on a 1-to-1 basis after being passed through a filter that prevented low levels of activity (less than the threshold *f*) from moving to the second hidden layer. The four units per speed class in the second hidden layer were linked to the output layer unit (F3) via weighted connections, which were the only connections modified by training. The activity of F2 and F3 units is described fully in the text. (c) The *Feedback model* was a recurrent network using estimates for speed over time as the input. The input layer is connected to the hidden layer, which is then connected to the output layer. However, in order to represent time and to provide the network with a form of memory, the hidden layer units are also connected to a context layer, which stores the activity levels of each unit in the hidden layer. This is achieved using connection weights of 1, which are not modified during training. The context layer is connected back to the hidden layer; these weights are updated during training. Hidden layer units therefore integrated current input (input layer) as well as activity from the previous time step (context layer). The sigmoid transfer function defined the response probabilities of hidden units and the output unit.

This was followed by presentation of the test set to assess the networks' ability to generalise to unseen exemplars. In both cases, the response of the network was determined by the output from the single unit in the output layer. Network responses to tail-flick and plant sequences were compared individually using *t*-tests, separately for the training and test sets. We also assessed performance by converting network output to a binary response indicating whether a tail-flick was present in the input sequence. Network output was transformed from a continuous score in the range [0,1] to a score of 1 if it exceeded a given threshold; otherwise it was set to 0. Reducing network responses in this way enabled us to perform signal detection analysis on the networks' performance (Swets, 1961). We then shifted the threshold for detection between 0 and 1 (increments of 0.1) whereby the change in correct decisions and false alarms defined a receiver operating characteristic (ROC) curve for each network.

To conclude this approach to modelling recognition, we considered which, if any, characteristics of the test set predicted network responses. For this purpose, we chose the network that showed the greatest sensitivity to tail-flicks, as defined by the ROC curve (i.e., the network that maximised the hit rate, while minimising the number of false alarms) and performed stepwise regression of network output against sequence duration, as well as average and maximum measures for speed and acceleration. All calculations were made according to procedures outlined in Peters et al. (2002).

1.2. Dynamic Input model

1.2.1. Network input

The second model differed from the first in that it considered changes in sensory unit response over time. Calculation of sensory unit responses to the velocity field were identical to that described for the Static Input model with one important difference: responses were stored separately for each frame of the sequence rather than summed over time. The input now comprised 48 sensory units (16 directions × 3 speeds) by *N*-frames; the number (*N*) of frames varied between sequences. An illustration of the type of input presented to the network is shown in Fig. 3. We used a different training strategy than for the Static Input model. First, as the temporal properties of the signal were now preserved, and because jacky lizard tail-flicks vary considerably in terms of signal duration (Peters and Ord, 2003), a subset of the tail-flick sequences was selected (range: 9–231 frames; mean: 71 frames; median: 58 frames). Plant sequences were then chosen to match as close as possible the distribution of tail-flick sequence length (range: 15–235 frames; mean: 81 frames; median: 72 frames). Second, as described below, the process by which the network learns is different to that presented in the previous model. Connections between units are strengthened when both are 'firing', and importantly, are weakened when the output unit does not 'fire' (i.e., when the network is told that the input sequence features wind-blown plants). Consequently, the network learns that certain units are not associated with tail-flicks. We therefore wished to examine how much experience with wind-blown plants is needed before the network effectively learns to ignore certain unit activity. In order to implement
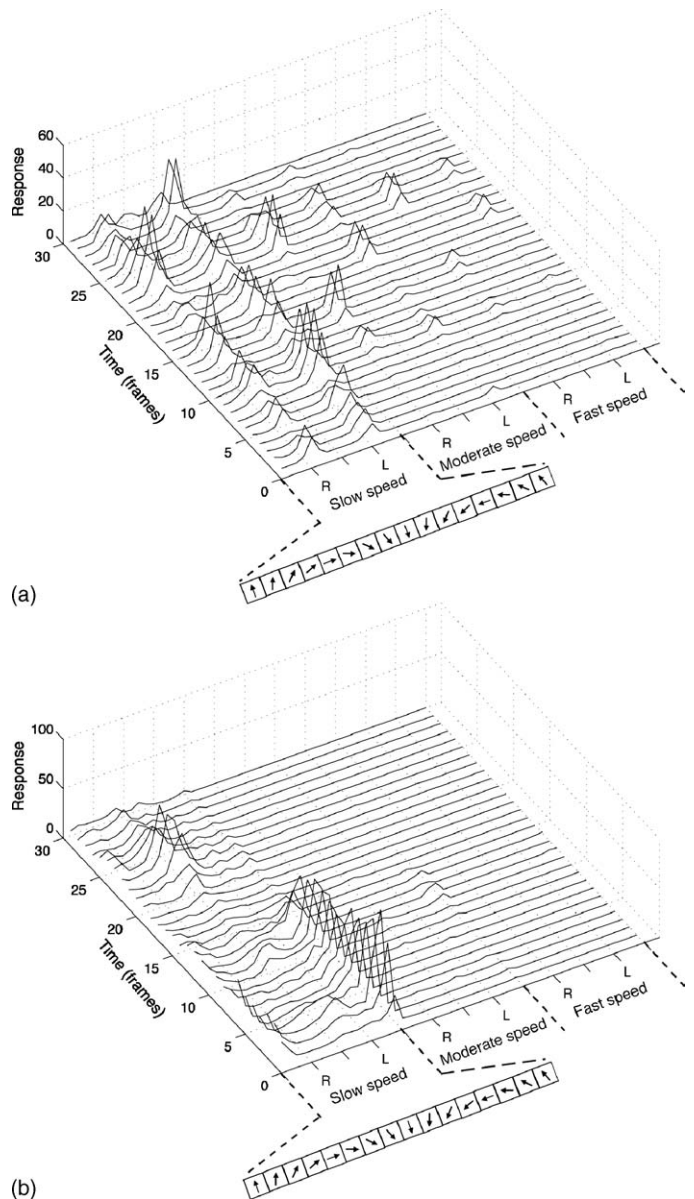

(a)

(b)

Fig. 3. Sensory unit response over time for representative (a) tail-flick and (b) plant sequences presented to the Dynamic Input model. Units are presented from upward movement then to the right (R), down and left (L).

this, we draw on the fact that tail-flicks are rare relative to the motion of wind-blown plants so we adjusted the frequency of plant sequences to reflect the difference in probability of the two classes of visual motion events. This was achieved by increasing the number of plant sequences for a given duration to generate input data sets that varied in the number of tail-flicks relative to plant sequences (signal-to-noise ratios of 1:10, 1:20 and 1:100).

1.2.2. Network architecture

The architecture of the Dynamic Input model is depicted in Fig. 2b; it comprises an input layer, two hidden layers and an output layer. The model processes each speed class from the input layer separately. Within each speed class, the raw sensory unit responses in the input layer (F0) were first collapsed into four general directions of movement and length normalised. The

output from the first hidden layer (F1) was passed to the second hidden layer (F2) on a 1-to-1 basis after being passed through a filter that prevented low levels of activity (less than the threshold f) from moving to the second hidden layer. The four units per speed class in the second hidden layer were linked to the output layer unit (F3) via weighted connections, which were the only connections modified by training.

The inputs to the network were the raw sensory unit responses over time (Fig. 3). The first hidden layer (F1) performed pre-processing as described above. The activity of F2 units was modelled over time according to the following differential equation:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -Ax_i + ms_i \qquad (3)$$

whereby s_i is the activity of the ith unit in the F1 layer, x_i the activity of the ith unit in the F2 layer, A the rate of decay of unit activity and m is a multiplier (10) that scales the response of F2 units to activity in F1 units.

The activity of the F3 output unit (denoted y) was updated through a process of interactive activation. The first step was to calculate the net input to the unit:

$$I = \left(\sum\nolimits_{i=1:N}(x_i z_i)\right) + 100T \qquad (4)$$

where z_i is the weight from the ith unit in F2 to F3 and T defined the input during training as a tail-flick or plant sequence (1 and 0, respectively); T was set to 0 during network testing, regardless of input type. The change in activity for the output unit was defined by:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \begin{cases} -By + I(1-y), & \text{if } (I \geq 0) \\ -By + Iy, & \text{if } (I < 0) \end{cases} \qquad (5)$$

where B is the rate of decay of the output unit. This ensured that F3 unit activity decreased when plant sequences were presented, but increased when a tail-flick was presented; activity decayed toward zero in the absence of input.

The connection weights (z_i) between units in the second hidden layer (F2) and the output unit (F3) were modified during training. Weight changes were determined by the differential equation:

$$\frac{\mathrm{d}z_i}{\mathrm{d}t} = \lambda(1 - z_i)x_j(y - z_i) \qquad (6)$$

In this equation, learning in z_i is stabilised (i.e., $\mathrm{d}z_i/\mathrm{d}t = 0$) when $(y - z_i) = 0$. That is, the weight z_i tracks the value of y: the weight increases when the output unit fires and decreases (toward 0) when the output unit is not firing. The inclusion of x_i in Eq. (6) implies that learning in the weight from the ith F2 unit to the output unit can only take place when the ith F2 unit is itself active. The term λ is a parameter that controls the rate of learning, and the term $(1 - z_i)$ ensures that weights do not exceed an upper bound of 1. Thus, the simultaneous activation of an F2 unit and the F3 unit results in the strengthening of the connection between these units (i.e., the network learns that activity in this F2 unit can signal the presence of a conspecific), as in Hebbian learning. However, unlike standard Hebbian learning, activation of an F2

unit in the absence of F3 unit activity results in the weakening of the connection between these units (i.e., the network learns that activity in this F2 unit does not signal the presence of a conspecific). There is no change in the weight when the F2 unit is inactive (i.e., when $x_i = 0$). Learning rules with such properties have been referred to as non-Hebbian or gated steepest descent (Grossberg, 1976).

The set of differential Eqs. (3), (5) and (6) were solved numerically using Euler's method with a step size of 0.01. A process of empirical adjustment of parameters determined that setting all parameters (f, A, B and λ) at 0.05 resulted in satisfactory learning.

1.2.3. Network training and performance assessment

We did not vary the size of hidden layers in this model or the training duration, and as such created only one version of this model. During training, each input sequence was presented once in a random order. The network modified connection weights between the 12 units in F2 (4 per speed class) and the output unit in F3. Discrimination performance was examined as in the Static Input model.

1.3. The Feedback model

1.3.1. Network input

Input to the Feedback model consisted of a single estimate of speed per frame for 100 frames (4 s). Recall that for any given frame of the image sequence there will be many estimates of velocity, according to the amount of movement detected in the scene by the motion analysis algorithm (see Peters et al., 2002). In the previous two models we retained both components of velocity (i.e., direction and speed), however, here we retain only vector magnitude, or speed. This value is determined for each estimate using Pythagoras' Theorem. A single estimate for speed for each frame was then calculated by averaging across these values; an illustration of the type of input used is presented in Fig. 4. From the subset of tail-flick sequences completed within 100 frames, we randomly selected 30 to be used for training; 30 sequences of wind-blown plant movement were also randomly selected. The test set comprised 55 tail-flick sequences and 34 plant sequences not presented during training. Due to software constraints, it was necessary to standardise the duration of input sequences. Consequently, we zero-padded the tail-flick sequences that were less than 100 frames in duration so that the movement started at different times within each sequence but all finished at frame 100.

1.3.2. Network architecture

The Feedback model was a recurrent network (Elman, 1990) consisting of four layers: the input layer, two hidden layers (one of which is a context layer), and an output layer (see Fig. 2c). The input layer is connected to the hidden layer, which is then connected to the output layer. However, in order to represent time and to provide the network with a form of memory, the hidden layer units are also connected to a context layer, which stores the activity levels of each unit in the hidden layer. This is achieved using connection weights of 1, which are not modified
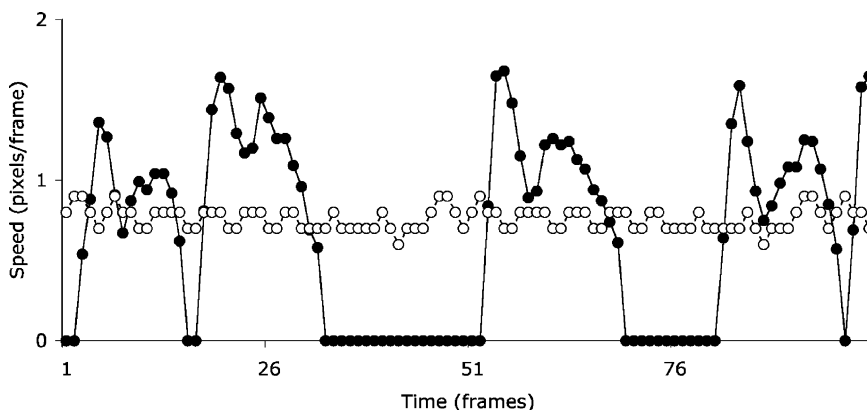
Fig. 4. Representative speed-time profiles for tail-flicks (●) and wind-blown plant movement (○) presented to the Feedback model.

during training. The context layer is connected back to the hidden layer; these weights are updated during training. Hidden layer units therefore integrated current input (input layer) as well as activity from the previous time step (context layer). The network produces an output at each time-step (equal to the duration of the input), however the final output value was considered to be the networks' response to the input sequence. Response properties of all hidden units and the output unit were defined by the sigmoid transfer function (2).

1.3.3. Network training

The input sequences were presented randomly during training. The network was trained using the gradient descent back-propagation algorithm, using the sum of squares error term. Weights were only updated after all input sequences had been presented. We trained the network to a response of 1 for tail-flicks and 0 for plant sequences. A total of 15 networks were constructed varying in hidden layer size (1, 2 and 5 units) and training duration (25, 50, 100, 300 and 500 epochs).

1.3.4. Analysis of network performance

The performance of each of the 15 networks was examined by presenting the training set to the trained network. This was followed by presentation of the test set to the network. The response of the network was determined by the final output from the single unit in the output layer. *t*-Tests were used to compare responses to the two types of visual motion for a given network, separately for the training and test sets. ROC curves were generated to identify the best performing network, and stepwise regressions of its response against predictor variables were conducted in the same manner as for the other models.

2. Results

2.1. Static Input model

Mean network response ($\pm$S.D.) to the training set by training duration and size of the second hidden layer is presented in Fig. 5a. As can be seen, the networks were able to discriminate between tail-flick and plant movement sequences ($t = 9.3–31.6$;

d.f. $= 1$; $p < 0.0001$ for each). The response of the networks to the previously unseen test set is presented in Fig. 5b. While performance was reduced relative to the training set, the networks were still capable of discriminating reliably between the tail-flick and plant movement sequences ($t = 4.8–14.6$; d.f. $= 1$; $p < 0.0001$ for each). The network responses to both the training and test sets suggest moderate improvement with longer training durations, while the effect of increasing the number of hidden layer units was most notable when increased from 2 to 7; further increases in the number of hidden layer units did not improve performance.

Shifting the criterion for tail-flick recognition between 0 and 1 assessed the signal detection performance of each network. The change in correct decisions and false alarms is defined by the ROC curve for each network. Representative ROC curves from the 24 networks are presented in Fig. 6. Each plot depicts the best (closed circles) and worst (open circles) performing networks at each hidden layer size, and showed that all networks performed better than chance (diagonal line). Furthermore, the distance of the ROC curve from chance level performance is positively related to its ability to discriminate signals from noise, and consequently, the best performing network can be regarded as the one that maximised this separation. In the present analysis, the best network featured seven hidden layer units and a training duration of 50,000 epochs (Fig. 6b closed circles) and was subsequently used in regression analysis. Of the variables that were used to summarise the input sequences of the test set, only average and maximum speed were found to be significant predictors of network response ($F(2,78) = 93.391$, $p < 0.0001$, adjusted $R^2 = 0.698$; Table 1).

2.2. Dynamic Input model

The mean (+S.D.) response of the Dynamic Input model to sequences after training is presented in Fig. 7, and the relationship between network response and sequence duration is presented in Fig. 8. The network was unable to discriminate tail-flicks from the background noise when these were common during training (1:10 signal-to-noise ratio: $t = 0.54$; d.f. $= 92$; $p = 0.589$), and the response was highly correlated
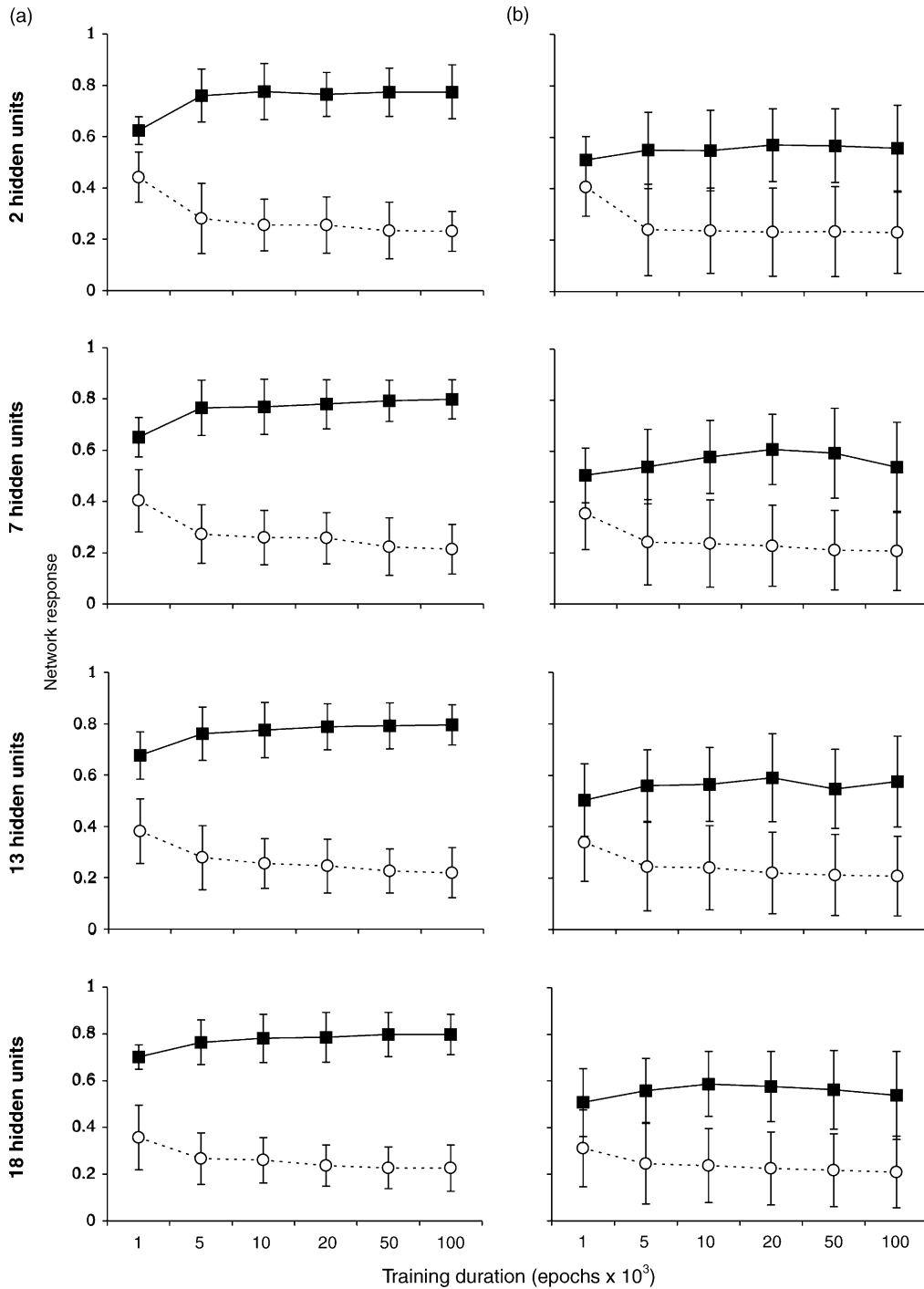
Fig. 5. Mean (±S.D.) response of Static Input model networks to the (a) training and (b) test sets by training duration for tail-flicks (■) and wind-blown plant movement (○). Separate plots are presented for different numbers of units in the second hidden layer (2, 7, 13 and 18). All networks were successful at discriminating the two classes of visual motion in both training and test sets (*t*-tests, $p < 0.05$).

with sequence duration (tail-flick: $R^2 = 0.82$, $p < 0.001$; plants: $R^2 = 0.91$, $p < 0.001$; Fig. 8a). Discrimination performance started to improve when the signal-to-noise ratio was reduced to 1:20 $t = 2.23$; d.f. $= 92$; $p = 0.028$); however response was still highly correlated with sequence duration (tail-flick: $R^2 = 0.72$, $p < 0.001$; plants: $R^2 = 0.95$, $p < 0.001$; Fig. 8b). Further reduction in the signal-to-noise ratio (1:100) during training substantially improved discrimination (Fig. 7; $t = 5.02$; d.f. $= 92$;

$p < 0.0001$). This is apparent primarily in the very small response of the network to plant sequences, which was now unrelated to sequence duration ($R^2 = 0.08$; Fig. 8c). The response to tail-flicks remained sensitive to sequence duration ($R^2 = 0.71$, $p < 0.001$; Fig. 8c). A stepwise regression indicated that sequence duration and maximum speed explained approximately half of the total variance in network response ($F(2,78) = 41.59$, $p = 0.000$, adjusted $R^2 = 0.504$; Table 1).
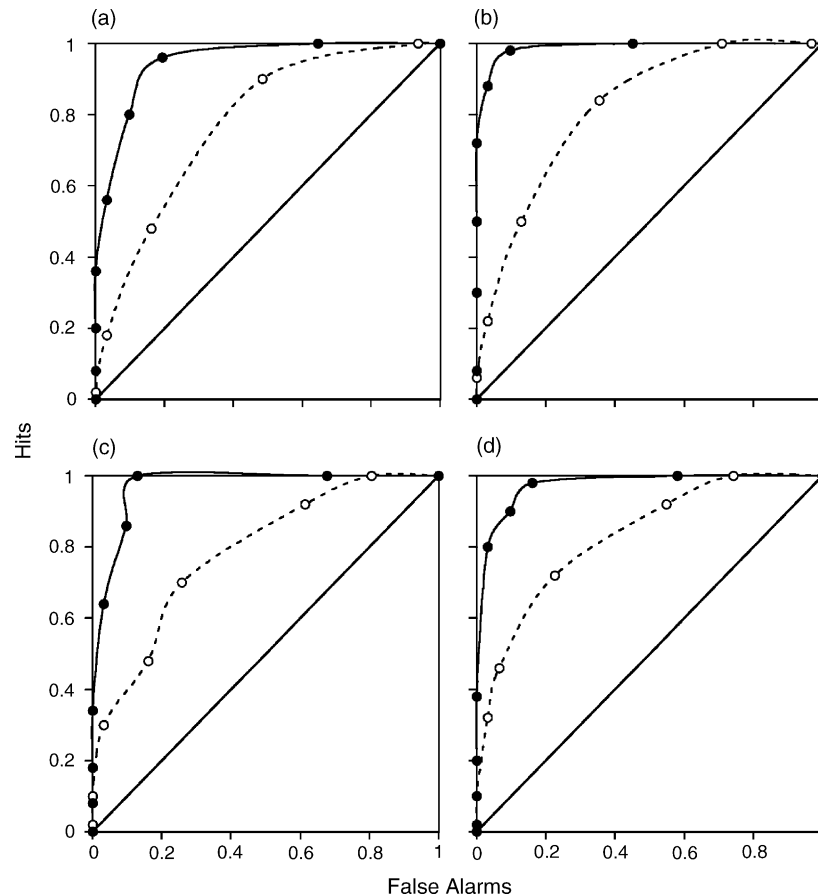
Fig. 6. Receiver operating characteristic (ROC) curves for the best (●) and worst (○) performing networks of the Static Input model for hidden layer sizes of (a) 2, (b) 7, (c) 13 and (d) 18 units. The worst performing network for each hidden layer size was trained for 1×10^3 epochs. The best performing networks for (a, c and d) were trained for 10×10^3 epochs, while (b) was trained for 50×10^3 epochs. The diagonal straight line represents chance performance.

2.3. The Feedback model

Mean network response (±S.D.) to the training set by training duration and hidden layer size is presented in Fig. 9a. There was a clear effect of hidden layer size as well as training duration. Networks with either 1 or 2 units in the hidden layer and a training
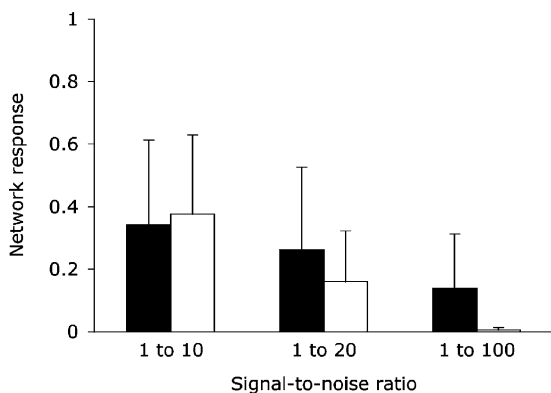

Fig. 7. Mean (+S.D.) network response to tail-flicks (■) and wind-blown plant movement (□) by the signal-to-noise ratio during training for the Dynamic Input model. (*) A significant difference between network response to the tail-flicks and plant movement ($p < 0.05$).

duration of less than 500 failed to discriminate input sequences ($t = 0.588$–1.272; d.f. = 58; $p > 0.10$); the longer training time of 500 epochs resulted in better performance ($t = 5.563$; d.f. = 58; $p < 0.001$ and $t = 17.70$; d.f. = 58; $p < 0.0001$ for 1 and 2 hidden units, respectively). Discrimination improved with five hidden units, and when coupled with longer training durations (300 or 500 epochs) performance was very high (i.e., very low variance in network response) ($t = 0.632$; d.f. = 58; $p > 0.1$ following training for 25 epochs, while $t = 7.4$–620.0; d.f. = 58; $p < 0.0001$ for durations of 50, 100, 300 and 500 epochs). The response of the networks to the previously unseen test set mirrored that of the training set (Fig. 9b); networks with five hidden layer units and longer training performed best. Networks with one or two hidden layer units and less than 500 epochs during training failed to discriminate the input patterns ($t = 0.658$–1.333; d.f. = 87; $p > 0.10$). Longer training improved performance for one ($t = 6.024$; d.f. = 87; $p < 0.001$) and two ($t = 15.939$; d.f. = 87; $p < 0.001$) hidden units. The network with a short (25 epochs) training duration and five hidden units was unable to discriminate the input patterns ($t = 0.668$; d.f. = 87; $p > 0.10$), while longer training durations (50, 100, 300 and 500 epochs) resulted in significant discrimination performance ($t = 8.737$–197.241; d.f. = 87; $p < 0.0001$). ROC curves for networks with five hidden layer units are presented in Fig. 10. Networks with long train-

Table 1
Results of stepwise regression of network response and motion parameters for each of the three neural network models

Variables	Un-standardised coefficients		Standardised coefficients		
	B	S.E.	Beta	t	p
Static Input model					
Constant	0.023	0.037		−0.614	0.541
Average speed	0.216	0.049	0.573	4.372	0.000
Maximum speed	0.104	0.047	0.290	2.209	0.030
Dynamic Input model					
Constant	−0.233	0.037		−6.315	0.000
Sequence duration	0.001	0.000	0.402	5.063	0.000
Maximum speed	0.143	0.021	0.550	6.929	0.000
Feedback model					
Constant	0.014	0.055		0.262	0.793
Average speed	0.569	0.044	0.784	12.961	0.000
Maximum speed	−0.193	0.027	−0.415	−7.197	0.000
Maximum acceleration	0.199	0.041	0.318	4.901	0.000

Excluded variables	Standardised coefficients		
	Beta In	t	p
Static Input model			
Sequence duration	−0.041	−0.439	0.662
Average acceleration	0.018	0.152	0.880
Maximum acceleration	0.024	0.272	0.786
Dynamic Input model			
Average speed	0.056	0.367	0.714
Average acceleration	−0.032	−0.253	0.801
Maximum acceleration	−0.237	−1.815	0.073
Feedback model			
Average acceleration (sequence duration not entered)	−0.031	−0.446	0.656

ing times (300 and 500 epochs) resulted in near perfect signal detection performance, while the network with the shortest training period (25 epochs) actually performed worse than chance (open circles). Regression analysis of network (five hidden units, trained for 300 epochs) response against predictor variables revealed that average and maximum speed, as well as maximum acceleration were significant predictors of network response ($F(3,194) = 130.152$, $p < 0.0001$, adjusted $R^2 = 0.663$; Table 1).

3. Discussion

Neural network methodology has proven a useful tool for understanding signal design and evolution (e.g., Enquist and Arak, 1993; Johnstone, 1994; Ryan et al., 2001). To date, this approach has been mostly applied to artificial input (Enquist and Arak, 1993) and bioacoustics (Phelps and Ryan, 1998; Deecke et al., 1999). The present paper extends the application of neural networks to the discrimination of a movement-based animal signal from background plant movement, using both static and time-varying input patterns. We show that networks of varying design can be trained to distinguish input derived from two different classes of visual motion. It is important to emphasise that our goal was to examine the feasibility of using neural networks as a statistical tool for understanding behaviour, rather than to investigate the neural mechanisms involved in discriminating types of visual motion.

Three models were presented that varied in the input structure, network architecture and the manner in which they 'learned'. The Static Input model was designed to capture the structure of visual motion input in terms of direction and speed (i.e., velocity), while ignoring temporal information by collapsing across the time domain (Fig. 1). In contrast, the other two models incorporated temporal information. The Dynamic Input model considered estimates for velocity over time (Fig. 3), while the Feedback model was presented with sequential estimates for velocity magnitude (i.e., speed; Fig. 3). The first two models comprised a feed-forward architecture with two hidden layers and an output layer (Fig. 2a and b), whereas a recurrent network was used for the Feedback model (Fig. 2c), which is designed to capture temporal variation by means of a context layer that 'remembered' its previous state. All models were supervised during training, meaning that correct outputs were provided to the network. In the case of the Static Input and Feedback models, connection weights were adjusted to better match its output to the target, using the back-propagation learning algorithm. In contrast, in the Dynamic Input model, error was not propagated backward through the network. Weights were adjusted during training when the connected units (from different layers) were both 'firing', according to an associative (gated steepest descent) learning rule.

Several versions of the Static Input and Feedback models were created with systematic variation in hidden layer size and training duration (see Section 1). Signal detection analysis was
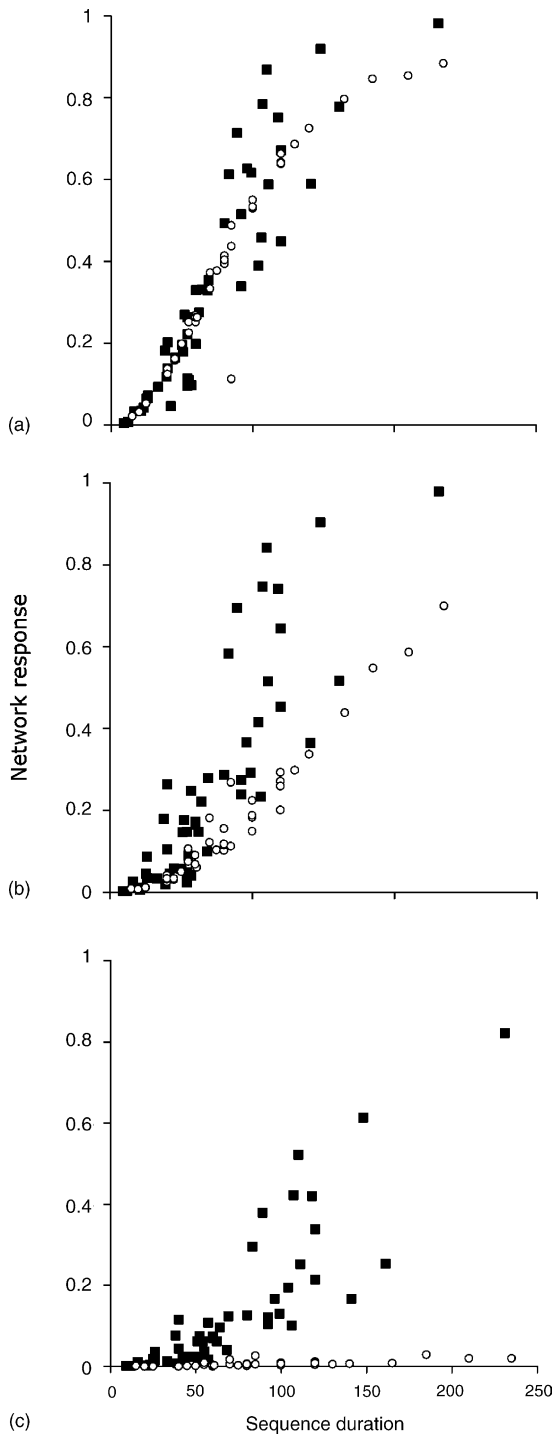
Fig. 8. Response of Dynamic Input model networks to tail-flicks (■) and wind-blown plant movement (○) as a function of sequence duration. Plots are for different signal-to-noise ratios of (a) 1:10, (b) 1:20 and (c) 1:100.

then used to determine the relative performance of each network. Comparisons between ROC curves (Figs. 6 and 10) allowed for selection of a single optimal network for each model. All models correctly classified the two different classes of visual motion used during training, and generalised successfully to previously unseen input (Figs. 5, 7 and 9). Regression analyses were then used to examine the relationship between network response and average and maximum values for speed and acceleration, as well

as sequence duration (Table 1). Maximum speed was shown to be a significant predictor of network response for all three models, while average speed was significant for the Static Input and Feedback models, and maximum acceleration for the Feedback model only. Sequence duration only predicted network response for the Dynamic Input model. It is not surprising that sequence duration did not affect the other two models, as the nature of the input to these models would likely have masked any effect of this variable. The sensory unit input to the Static Input model was calculated frame-by-frame and then summed across time. As a consequence, the same input profile could have been derived from a short sequence with lots of activity, or a long sequence with intermittent activity. Similarly, the input to the Feedback model was standardised to 100 frames (4 s), by zero padding of sequences that were shorter.

These results suggest that the Static Input and Feedback models used structural differences in visual motion (i.e., velocity) as the basis for learning the discrimination between signal and noise during training. Furthermore, discrimination could be performed based on calculation of velocity magnitude (speed) alone. This is consistent with behavioural findings in other lizards (*Anolis auratus*; reviewed in Fleishman, 1992). Fleishman (1986) showed that response probability to artificial lures was greatest when the stimulus had high velocity and acceleration values; a characteristic of the initial portion of their visual display (Fleishman, 1988). In circumstances where the goal is to classify motor patterns and other types of visual motion, it is likely that relatively simple networks, like those of the Static Input and Feedback models, will be suitable. However, conspicuousness is not sufficient to fully explain signal design (Peters and Evans, 2003a). The jacky lizard tail-flick is the first of five distinct motor patterns, which are performed in response to a conspecific intruder and expressed in an obligatory sequence (Peters and Ord, 2003). We have shown previously that four out of the five motor patterns have structural characteristics distinct from background noise (Peters and Evans, 2003a), so structure alone cannot account for the tail-flick being the introductory component. Predictions from the Dynamic Input model suggested that duration is an important factor determining signal efficacy. This prediction was tested in a recent playback study investigating the relative importance of several motion variables (Peters and Evans, 2003b). Results demonstrated that tail-flick efficacy does indeed depend upon duration, suggesting that short displays may be ineffective for attracting attention, regardless of intensity. This close match between the response characteristics of the Dynamic Input model and the behaviour of these lizards suggests that this architecture might be the most appropriate for predicting the effect of signal variation on real receivers.

Neural networks can hence be used in combination with quantitative analyses of signal structure and experimental analyses of receiver responses. By identifying patterns of structural variation in visual motion input, these models help to isolate the subset of variables that are likely to be most important, so that these can be manipulated experimentally. This approach compares favourably with the alternative of making unguided or intuitive choices, which are perhaps less likely to correspond
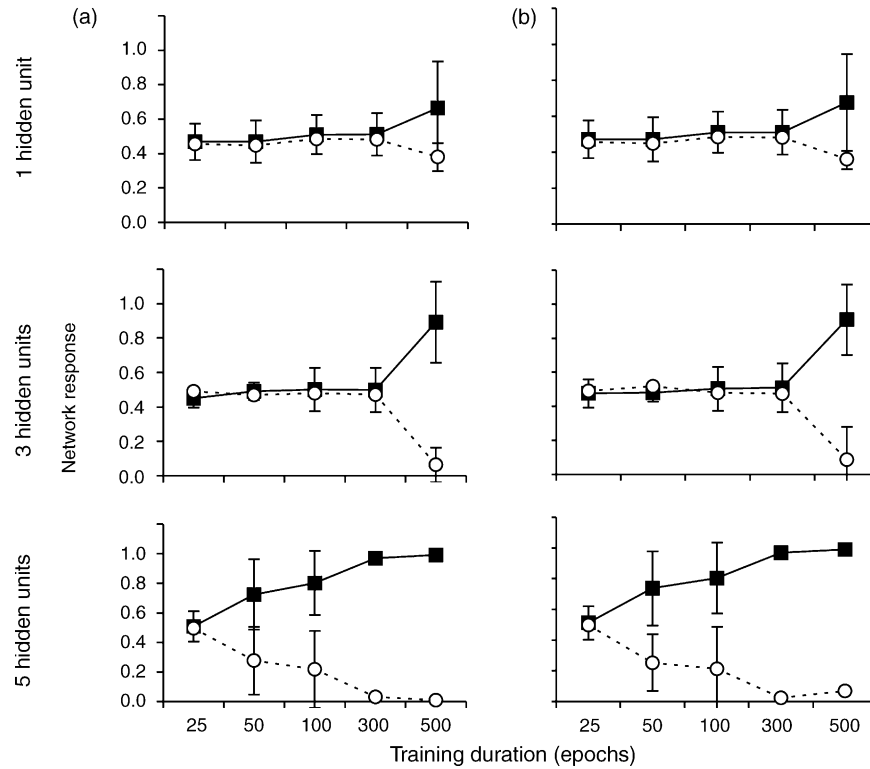
Fig. 9. Mean (±S.D.) response of Feedback model networks to the (a) training and (b) test sets by training duration (epochs) for tail-flicks (■) and wind-blown plant movement (○). Separate plots are presented for different numbers of hidden units (1, 3 and 5). (*) A significant difference between network response to the tail-flicks and plant movement ($p < 0.05$).

with the parameters that are important to the lizards. In our current research, we utilize 3D digital animations to explore signal recognition in the jacky lizard (Peters and Evans, 2003b), which allows us to construct any number of motor patterns in a mathematically-precise manner. Trained networks will be used

to explore recognition of a large number of synthetic sequences, so that we can select a subset of these for presentation to live lizards using digital video playback. The general approach we are advocating here is not limited to movement-based signals; a similar strategy has already proven successful in predicting response biases in the acoustic domain (Phelps and Ryan, 1998).

In conclusion, the present analysis has shown that the study of movement-based animal signals has the potential to benefit from the application of neural network methodology. Networks from each of the three models were highly successful at discriminating the signal from noise. However, we expect that neural networks will be particularly useful for identifying more subtle differences in structure than those we have presented, such as in comparisons between individuals performing the same motor pattern. Furthermore, the performance of the Dynamic Input model is encouraging because it predicted behavioural findings that would not be expected based on analyses of structure alone. Further development of this approach to studying signal design will help to direct future research into the evolution of this class of animal signals.
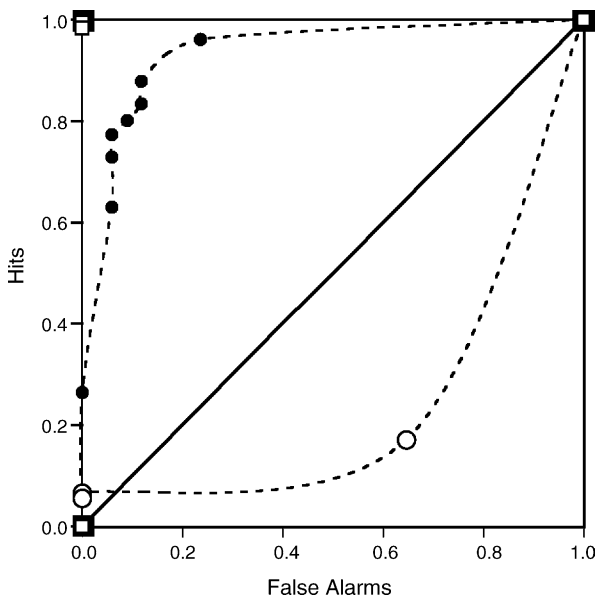

Fig. 10. ROC curves for Feedback model networks with five hidden layer units and training durations of: 25 (○), 100 (●), 300 (□) and 500 (■) epochs (not shown: 50 epochs). The diagonal line reflects chance performance.

Acknowledgements

References

Basolo, A.L., 1990. Female preference predates the evolution of the sword in swordtail fish. Science 250, 208–210.

Bateson, P., Horn, A.G., 1994. Imprinting and recognition memory: a neural net model. Anim. Behav. 48, 695–715.

Borst, A., Haag, J., 2002. Neural networks in the cockpit of the fly. J. Comp. Physiol. A Sens. Neural Behav. Physiol. 188, 419–437.

Deecke, V.B., Ford, J.K.B., Spong, P., 1999. Quantifying complex patterns of bioacoustic variation: use of a neural network to compare killer whale (*Orcinus orca*) dialects. J. Acoust. Soc. Am. 105, 2499–2507.

Elman, J.L., 1990. Finding structure in time. Cogn. Sci. 14, 179–211.

Endler, J.A., 1990. On the measurement and classification of colour in studies of animal colour vision. Biol. J. Linnean Soc. 41, 315–352.

Endler, J.A., 1991. Variation in the appearance of guppy color patterns to guppies and their predators under different visual conditions. Vis. Res. 31, 587–608.

Endler, J.A., 1992. Signals, signal condition and the direction of evolution. Amer. Nat. 139, S125–S153.

Endler, J.A., Basolo, A.L., 1998. Sensory ecology, receiver biases and sexual selection. Trends Ecol. Evol. 13, 415–420.

Enquist, M., Arak, A., 1993. Selection by exaggerated male traits by female aesthetic senses. Nature 361, 448–449.

Enquist, M., Ghirlanda, S., 2005. Neural Networks & Animal Behavior. Princeton University Press, Princeton, NJ, p. 253.

Fleishman, L.J., 1986. Motion detection in the presence or absence of background motion in an *Anolis* lizard. J. Comp. Physiol. A Sens. Neural Behav. Physiol. 159, 711–720.

Fleishman, L.J., 1988. Sensory influences on physical design of a visual display. Anim. Behav. 36, 1420–1424.

Fleishman, L.J., 1992. The influence of the sensory system and the environment on motion patterns in the visual displays of anoline lizards and other vetebrates. Amer. Nat. 139 (Suppl.), S36–S61.

Ghirlanda, S., Enquist, M., 1998. Artificial neural networks as models of stimulus control. Anim. Behav. 56, 1383–1389.

Grossberg, S., 1976. Adaptive pattern classification and universal recoding. Part I. parallel development and coding of neural feature detectors. Biol. Cybernet. 23, 121–134.

Guilford, T., Dawkins, M.S., 1991. Receiver psychology and the evolution of animal signals. Anim. Behav. 42, 1–14.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. Prentice Hall, New Jersey.

Holmgren, N.M.A., Enquist, M., 1999. Dynamics of mimicry evolution. Biol. J. Linnean Soc. 66, 145–158.

Hopp, S.L., Owren, M.J., Evans, C.S., 1998. Animal Acoustic Communication: Sound Analysis and Research Methods. Springer–Verlag, Berlin.

Johnstone, R., 1994. Female preference for symmetrical males as a by-product of selection for mate recognition. Nature 372, 172–176.

Krakauer, D.C., Johnstone, R.A., 1995. The evolution of exploitation and honesty in animal communication: a model using artificial neural networks. Philos. Trans. R. Soc. Lond. B 348, 355–361.

Pagel, M., 1993. The design of animal signals. Nature 361, 18–20.

Peters, R.A., 2003. Design of movement-based animal signals: insights from an Australian lizard, Amphibolurus muricatus. Ph.D. Thesis. Macquarie University, Sydney, Australia.

Peters, R.A., Clifford, C.W.G., Evans, C.S., 2002. Measuring the structure of dynamic visual signals. Anim. Behav. 64, 131–146.

Peters, R.A., Evans, C.S., 2003a. Design of the Jacky dragon visual display: signal and noise characteristics in a complex moving environment. J. Comp. Physiol. A Sens. Neural Behav. Physiol. 189, 447–459.

Peters, R.A., Evans, C.S., 2003b. Introductory tail-flick of the Jacky dragon visual display: signal efficacy depends upon duration. J. Exp. Biol. 206, 4293–4307.

Peters, R.A., Ord, T.J., 2003. Display response of the Jacky dragon, *Amphibolurus muricatus* (Lacertilia: Agamidae), to intruders: a semi-Markovian process. Aust. Ecol. 28, 499–506.

Phelps, S.M., Ryan, M.J., 1998. Neural networks predict response biases of female tungara frogs. Proc. R. S. Lond. B 265, 279–285.

Phelps, S.M., Ryan, M.J., 2000. History influences signal recognition: neural network models of túngara frogs. Proc. R. Soc. Lond. B Biol. Sci. 267, 1633–1639.

Phelps, S.M., Ryan, M.J., Rand, A.S., 2001. Vestigial preference functions in neural networks and túngara frogs. Proc. Natl. Acad Sci. U.S.A. 98, 13161–13166.

Potter, J.R., Mellinger, D.K., Clark, C.W., 1994. Marine mammal call discrimination using artificial neural networks. J. Acoust. Soc. Am. 96, 1255–1262.

Reby, D., Lek, S., Dimopoulos, I., Joachim, J., Lauga, J., Aulagnier, S., 1997. Artificial neural networks as a classification method in the behavioural sciences. Behav. Process. 40, 35–43.

Rumelhart, D.E., Hinton, D.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature (London) 323, 533–536.

Ryan, M.J., Phelps, S.M., Rand, A.S., 2001. How evolutionary history shapes recognition mechanisms. Trends Cogn. Sci. 5, 143–148.

Ryan, M.J., Rand, A.S., 1990. The sensory bias of sexual selection for complex calls in the tungara frog *Physalaemus pustulosus* (sexual selection for sensory exploitation). Evolution 44, 305–314.

Ryan, M.J., Rand, A.S., 1993. Sexual selection and signal evolution: the ghosts of biases past. Philos. Trans. R. Soc. Lond. B 340, 187–195.

Swets, J.A., 1961. Is there a sensory threshold? Science 134, 168–177.

Wachtmeister, C.-A., Enquist, M., 2000. The evolution of courtship rituals in monogamous species. Behav. Ecol. 11, 405–410.

Werbos, 1974. Beyond regression: new tools for prediction and analysis in the behavioral sciences. Harvard University, Boston, MA.

Zeil, J., Zanker, J.M., 1997. A glimpse into crabworld. Vis. Res. 37, 3417–3426.