Bulletin of the Seismological Society of America, Vol. 85, No. 1, pp. 308–319, February 1995

# Artificial Neural Network-Based Seismic Detector

by Jin Wang and Ta-Liang Teng

**Abstract** An artificial neural network-based pattern classification system is applied to seismic event detection. We have designed two types of Artificial Neural Detector (AND) for real-time earthquake detection. Type A artificial neural detector (AND-A) uses the recursive STA/LTA time series as input data, and type B (AND-B) uses moving window spectrograms as input data to detect earthquake signals. The two AND's are trained under supervised learning by using a set of seismic recordings, and then the trained AND's are applied to another set of recordings for testing. Results show that the accuracy of the artificial neural network-based seismic detectors is better than that of the conventional algorithms solely based on the STA/LTA threshold. This is especially true for signals with either low signal-to-noise ratio or spikelike noises.

## Introduction

The availability of increasingly low-cost, high-speed computers and rapid data transmission systems has made real-time earthquake monitoring an easy task that will eventually lead to earthquake early warning. Therefore, real-time seismology has recently gained much attention from seismologists and public agencies (U.S. National Research Council, 1991). Real-time earthquake monitoring aims at (1) detecting seismic events of potential interests, (2) locating the source and estimating the size of detected events, (3) predicting the areas that could be affected by strong shaking, and (4) initiating an early warning to users. A number of algorithms providing real-time information have been developed by many seismologists, and research continues in improving the reliability of these algorithms and in reducing the overhead time. However, to timely provide reliable earthquake information, such as magnitude, location, and potential damage patterns in populated areas, is still an important field to be studied. In real-time seismology, some of the problems involve inversion, such as for the location, rupture process, and source mechanism; and others belong to signal discrimination and pattern recognition, such as earthquake detection, magnitude determination of an earthquake, and prediction of its intensity distribution. There is a practical need to seek a quick and reliable solution if the solution is to be useful for emergency response.

Recent developments of Artificial Neural Networks (ANN's) indicate that they may be useful in solving problems in signal discrimination and classification. A neural network can learn patterns from a sample data set and determine the class of new data based on previous knowledge. Compared with some statistical pattern recognition methods and expert systems, a neural network requires no explicit knowledge about probability distribution models, and can handle numerical data better. A few applications of ANN's to seismology have been carried out recently. Dowla *et al.* (1990) applied the Multi-Layered Perceptron (MLP) neural network to discriminate natural earthquakes versus underground nuclear explosions. In that study, spectral amplitudes of the picked phase window are used as a training data set for event discrimination. Dystart and Pulli (1990) applied the MLP neural network to classification of chemical explosions and earthquakes based on the spectral ratios *Pn*/*Sn*, *Pn*/*Lg*, and the mean cepstral variance of *Pn* and *Sn*. Wang (1992) and Wang and Mendel (1992) developed a new adaptive minimum prediction–error deconvolution procedure based on the Hopfield neural network. Their method is applied to the reflection seismology for reflectivity location detection, reflectivity magnitude estimation, and source wavelet extraction. Alexander *et al.* (1992) applied the frequency-slowness seismic image as the input of the neural network to identify earthquakes and explosions. Katz and Aki (1992) developed an approach to earthquake prediction using neural network techniques. The methodology was tested both in reversed-time and real-time modes.

We are applying the artificial neural network to real-time seismology. Two types of artificial neural detectors (AND-A and AND-B) are designed for real-time earthquake monitoring. The AND-A uses the recursive STA/LTA time series as the input feature. The AND-B uses the moving window spectrum as the input feature. Training and testing procedures are designed and implemented using seismic time series recorded in the field. Experimental results for both types of artificial neural network-based seismic detectors show that the accuracy of the neural detector is better and more intelligent than the conventional methods, which are based solely on the threshold parameters. These AND's are especially useful for events either with low signal-to-noise ratios or with spikelike noises.

## Artificial Neural Networks

Artificial neural networks (ANN's) are inspired by biological systems where large numbers of neurons, which individually function rather slowly and imperfectly, collectively perform tasks that even the largest computers have not been able to match. This field is among the most rapidly developing scientific inquiries today. The field is interdisciplinary in nature, and its potential applications are in such important areas as speech and image recognition, linear and nonlinear optimization, automatic control, and so on. As in a biological nervous system, the fundamental element of an ANN is the artificial neuron (also called a cell, unit, or node). The function of artificial neurons is identical to that of real neurons: they must integrate input from other neurons and communicate the integrated signal to a decision-making center. Computational models of a neural network try to emulate the physiology of real neurons. There are two principal functions for artificial neural networks. One is the input-output mapping or feature extraction. The other is pattern association or generalization. The mapping of input and output patterns is estimated or "learned" by the neural network with a representative sample of input and output patterns. The generalization of the neural network is an output pattern in response to an input pattern, based on the network memories that function like the human brain. In their current state, neural networks are probably most useful in problems related to pattern recognition (Antognetti and Milutinovic, 1991). Among the many different types of neural networks, we introduce a particular type called the Multi-Layered Perceptron (MLP). The MLP that is used in our application study has been proven to be most useful in engineering applications.

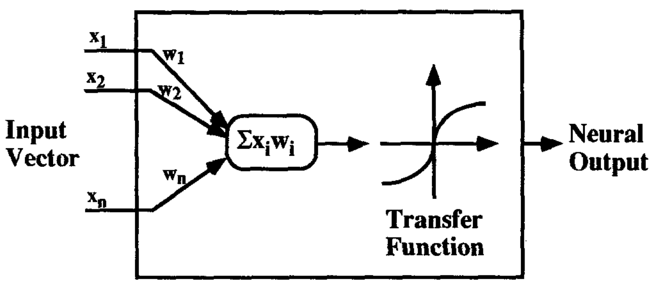


Figure 1. Architecture of a single artificial neuron consisting of $N$ input units. The input vector fans in from the left and fans out to the right. Interconnection of artificial neurons forms a complex architecture called an artificial neural network.

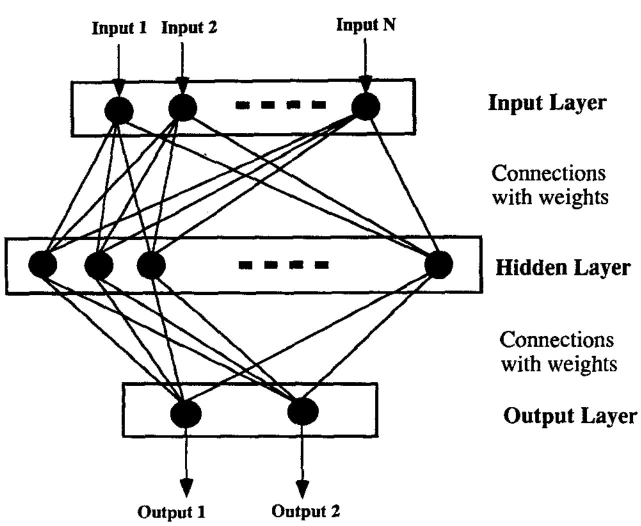


Figure 2. Structure of a multi-layer perceptron neural network. It is a feedforward and partially connected network. Neurons are only connected to the adjacent layer neurons.

The perceptron may be viewed as a neuron that computes activation with a sigmoidal activation function, which takes the form (Pao, 1989, Morgan and Scofield, 1991)

$$x_i = F(d_i) = \frac{1}{1 + \exp(-\beta d_i)}, \tag{1}$$

and

$$d_i = \sum_{j}^{N} w_{ij} x_j - \theta_i, \tag{2}$$

where $x_i$ is the activity output at neuron $i$, constant $\beta$ controls the slope of the semi-linear region, $d_i$ is the weighted sum of neuron $i$ for the input from the last lower layer with $N$ neurons, $x_j$ is the input element from neuron $j$, $w_{ij}$ is the weight between neurons $i$ and $j$, and $\theta_i$ is the intrinsic threshold that can be treated as an individual weight with a negative sign.

The structure of an artificial neuron is shown in Figure 1. The one-to-one correspondence between the artificial neuron and the biological neuron is commonly drawn: the inputs correspond to the dendrites of a biological neuron; the weights correspond to the synapses; the summation and the transfer function unit correspond to the cell body; and the output of transfer fanning out to other units corresponds to axons. Artificial neural networks can be constructed by linking the neurons through the weights. In this way, a complex architecture may be formed. A schematic illustration of a multi-layer feedforward network is shown in Figure 2. The input data are propagated or feedforwarded through neurons and their connections in the hidden layer to the output neurons. As Figure 2 shows, the feedforward architecture only permits unidirectional connection between adjacent layers.

The weights and thresholds of an ANN are obtained by learning through the training data set. The most widely used supervised learning algorithm for a multi-layer feedforward neural network is the backward error propagation algorithm (Rumelhart *et al.*, 1986). Define an objective function (mean squared error) in the output layer as

$$J_p = \frac{1}{2M} \sum_{k=1}^{M} [T_k(p) - O_k(p)]^2, \qquad (3)$$

and define the normalized sum of objective function over total training set as

$$J = \frac{1}{P} \sum_{p=1}^{P} J_p, \qquad (4)$$

where $M$ is the number of outputs of the neural network, $O_k(p)$ is the network output, and $T_k(p)$ is the desired (or target) output of the $k$th neuron for the pattern $p$. The back propagation learning algorithm uses a least-squares error minimization criterion to minimize $J$. This can be accomplished by adjusting the weights according to the negative gradient of the error with respect to the weights. The threshold $\theta_i$ and weights $w_{ij}$ are sequentially updated for each sample pattern $p$ from the output to the input layer. That is, the learning is supervised by the given input-output pattern set. The second order supervised learning algorithm is given by

$$w_{ij}(p + 1) = w_{ij}(p) + \Delta w_{ij}(p), \qquad (5)$$

$$\Delta w_{ij}(p) = \eta\, \delta_i(p) O_j(p) + \alpha \Delta w_{ij}(p - 1), \qquad (6)$$

in which $\eta$ is called the learning rate and $\alpha$ is the momentum rate that adjusts the direction of descent by using a fraction of the previous weight modification.

For equations (5) and (6), define

$$\delta_i(p) = F'(d_i)\left(\frac{\partial J_p}{\partial x_j}\right). \qquad (7)$$

After some mathematical derivation, we obtain the expression for neurons in the hidden layer(s),

$$\delta_i(p) = O_i(p)[1 - O_i(p)] \sum_k \delta_k(p) w_{ik}, \qquad (8)$$

where $O_i(p)$ is the output of the $j$th neuron for the $p$th pattern. And for neurons in the output layer,

$$\delta_i(p) = O_i(p)[1 - O_i(p)][T_i(p) - O_i(p)]. \qquad (9)$$

The learning process will end if $J$ has converged to a desired value. The trained neural network will be used to process unknown input data and classify them into different classes.

## Artificial Neural Detector (AND) for Earthquakes

To apply the artificial neural network to real-time seismology, we present an application of a neural network-based pattern classification system to the seismic event detection problem. It is very important for a seismic network to miss no real earthquakes, and at the same time to keep down the number of false triggers. A similar requirement is made of the operation of a single portable seismic station in the field. The most commonly used method in automatic event detection is the STA/LTA threshold classification method (Tottingham *et al.*, 1989). The conventional STA/LTA ratio triggering algorithm requires a careful setting of the triggering threshold parameters that must be adjusted to suit various field environments. The threshold setting seeks a compromise between two conflicting criteria, namely to simultaneously minimize both the false alarm rate and the possibility of missing real events. Even with care and experience, it is very hard to avoid excessive false triggers. For a stand-alone seismic station, it is not uncommon to have much more than 50% of its recordings being false triggered events. This adds the time-consuming job of subsequent data manipulation and processing. The proper choices of these trigger parameters are very difficult, yet important for a successful seismic network operation.

To enhance the STA/LTA event detection, we introduce a pattern recognition approach that consists of two basic steps: feature extraction (input selection) and classification. The two steps are closely related in a manner that a better feature extraction will result in an easier classification. The success of a neural network system for pattern classification depends primarily on the designer's physical understanding of the problem. Seismic event detection is a pattern recognition problem in the real-time sense. When data come in continuously, the computer of a seismic network needs to determine whether this signal represents an earthquake event and then handle it as such.

### Artificial Neural Detector in the Time Domain (AND-A)

The most commonly used seismic triggering algorithm is the STA/LTA ratio threshold method. The first step in this algorithm is to take the absolute value of the difference of the amplitude of the incoming time series ($\Delta x$). This filters out the long-term variation in the signal. Then the short- and long-term averages (STA and LTA, respectively) are calculated recursively in a moving window. The equations are given in the following sequence:

$$\Delta x(k) = |x(k) - x(k-1)|, \tag{10}$$

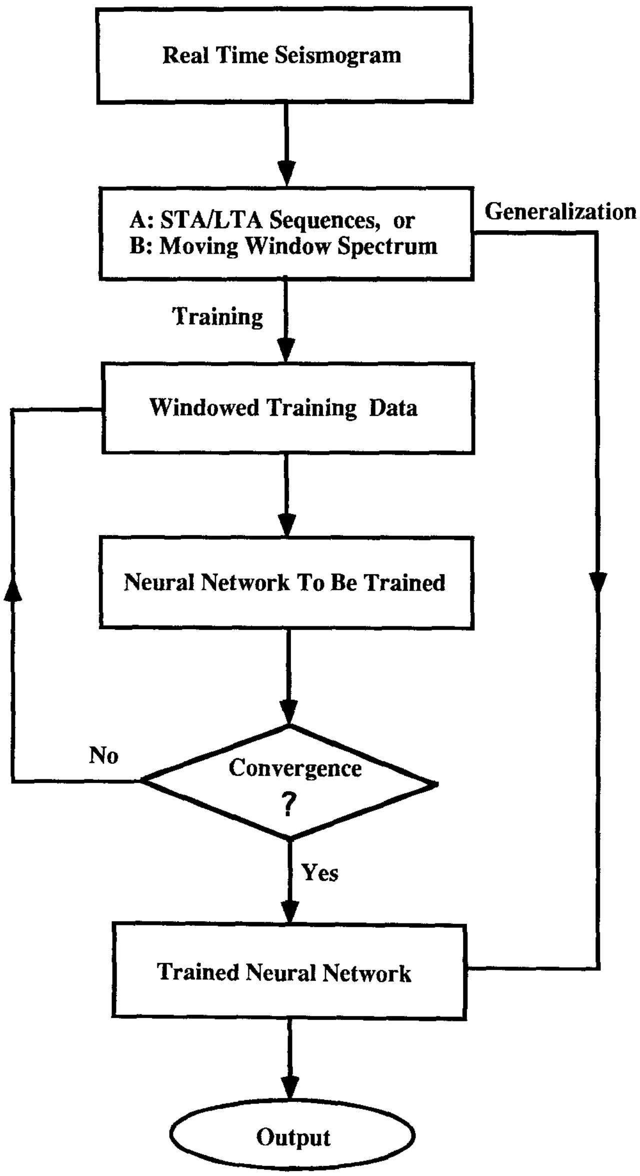


Figure 3. Block diagram of the artificial neural detectors in this study. There are two processes in the neural network: supervised learning and generalization. After an initial learning, the neural network can be used to classify incoming new data, called generalization. Seismograms are preprocessed to feed the AND's. The STA/LTA ratio is computed for the AND-A, and the moving window spectrum for the AND-B.

$$\mathrm{STA}(k) = \mathrm{STA}(k-1) + [\Delta x(k) - \mathrm{STA}(k-1)]/_{T_{st}}, \tag{11}$$

$$\mathrm{LTA}(k) = \mathrm{LTA}(k-1) + [\mathrm{STA}(k) - \mathrm{LTA}(k-1)]/T_{lt}, \tag{12}$$

where $x(k)$ is the incoming data, and $T_{st}$ and $T_{lt}$ are the short-term and long-term moving time window lengths, respectively. Then the system continuously computes $\alpha$, the ratio of the difference to the long-term average. It takes the form

$$\alpha(k) = \Delta x(k)/\mathrm{LTA}(k). \tag{13}$$

Then the system computes the ratio of the short-term average to the long-term average, $\beta$, which takes the form

$$\beta(k) = \mathrm{STA}(k)/\mathrm{LTA}(k). \tag{14}$$

Whenever the signal changes feature abruptly, the $\alpha$ value will be very high. If $\alpha$ exceeds a preset critical value $\alpha_c$, a trigger flag is set. To confirm the trigger, $\beta$ should exceed a preset threshold value $\beta_c$. If $\beta$ does not exceed a preset threshold level $\beta_c$ at least for $N_{tcc}$ (trigger confirmation count) samples, the trigger is discarded. This weeds out spikes and other types of transient signals. So there are three threshold parameters, $\alpha_c$, $\beta_c$, and $N_{tcc}$ that are strongly dependent on the seismic monitoring environment and signals to be detected. In practice, these parameters have to be determined for each new experimental site, and only then can they be preset for a monitoring network.

In this study, we enhance the performance of the STA/LTA detecting method by taking the entire waveform of STA/LTA sequences within a certain window as a detecting pattern instead of using a few threshold parameters. The AND-A is designed with the Multi-Layered Perceptron (MLP) discussed before. It consists of one input layer, one hidden layer, and one output layer. The input layer has 50 units that corresponds to a 1-sec moving windowed STA/LTA time series $\beta(i)$, $i = 1, \ldots, 50$, from one channel of seismometer output. Input vector $\mathbf{X}$ can be represented as $\mathbf{X} = [\beta(1), \beta(2), \beta(3), \ldots, \beta(50)]$. For seismic detection, there are two possible outputs, earthquake or nonearthquake onset. Thus, the output layer of the neural network detector is composed of two neurons. The first neuron is coded for an earthquake onset, if the output of the first neuron is 0.9, the corresponding input is an earthquake onset, and the second neuron is coded for a nonearthquake onset. The number of neurons in the input and output layers is usually determined by the features of the problem. The number of neurons in the hidden layer, however, is a parameter decided by the designer. Here, eight neurons are chosen to construct the hidden layer; the reason for choosing eight neurons will be discussed later. The structure of AND-A results in 51 * 8 + 9 * 2 = 426 connections (weights and thresholds) between the input and output. The block diagram in Figure 3 illustrates the whole training and automatic detecting process. For a given seismogram, the sequences of STA/LTA are computed. The neural detector is trained by a set of training data consisting of a number of input training vectors **X** extracted from windowed STA/LTA data for either earthquake or nonearthquake onset. After training, the trained neural network detector is used to determine incoming signals continuously.

In this experiment, the input data are extracted form real seismograms of aftershocks of 28 June 1992, Landers earthquake ($M$ = 7.5) recorded by the same Reftek instrument in the near-field site. The whole training data set for the AND-A, normalized by the maximum value, consists of 50 pattern samples, of which 25 patterns are earthquake onsets and the rest are not. The training patterns for earthquake onset and the patterns of nonearthquake onset are given in Figures 4a and 4b, respectively. The patterns of earthquake onset are defined by very similar features: 10 points of a flat line before the onset and 40 points with high normalized values. Each pattern looks approximately like a step function. The other patterns in Figure 4b with quite irregular variations are defined as nonearthquake onsets. The neural network was trained with the training set. The desired target is a vector that has two components corresponding to the class of pattern [0.9 0.1] for the earthquake onset and [0.1 0.9] for the nonearthquake onset.

Since the neural network is an adaptive system, the initial value of weights can be any random number, and they will be updated to final values after training. In the

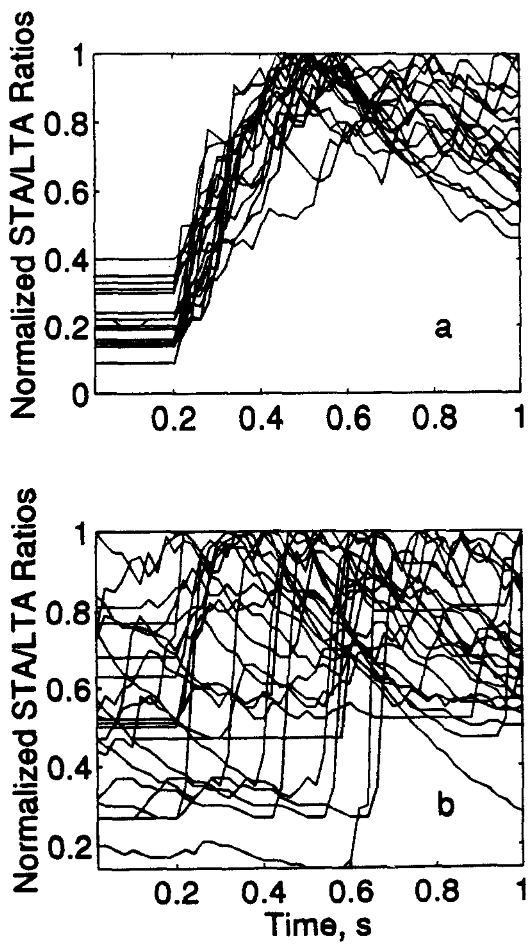


Figure 4. Training patterns of (a) earthquake onset, and (b) nonearthquake onset for AND-A.

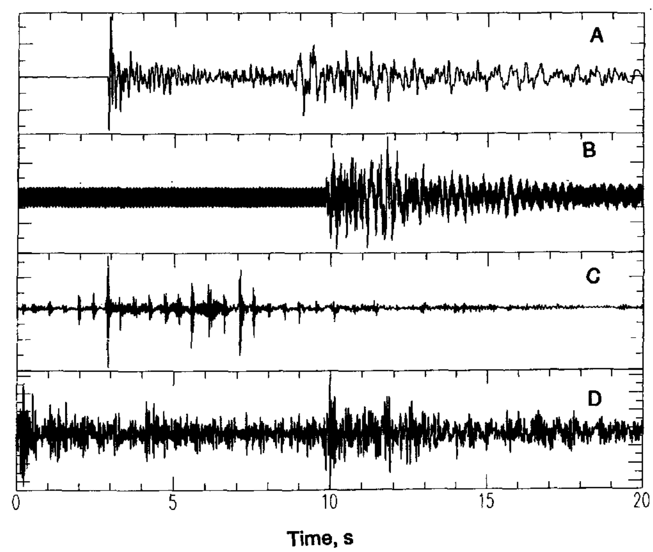


Figure 5. Four samples of seismograms used in this study: (a) earthquake recording with high STA/LTA ratio; (b) earthquake recording with low STA/LTA ratio; (c) nonearthquake recording with high STA/LTA ratio; and (d) nonearthquake recording with low STA/LTA ratio.

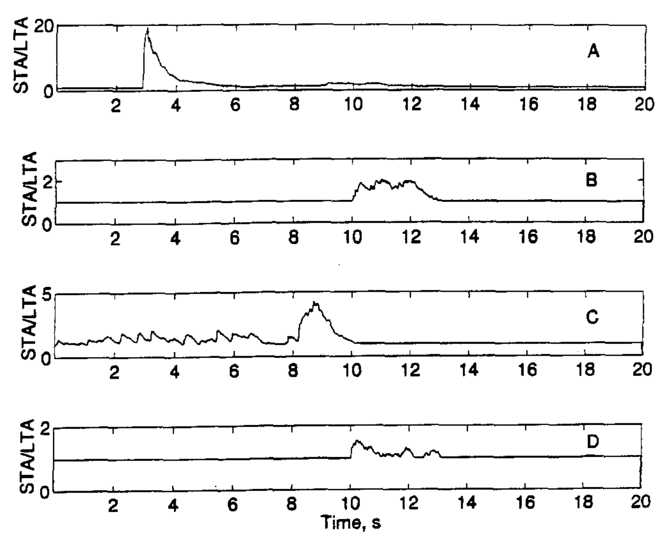


Figure 6. Corresponding STA/LTA ratios for the sample recordings in the Figure 5.

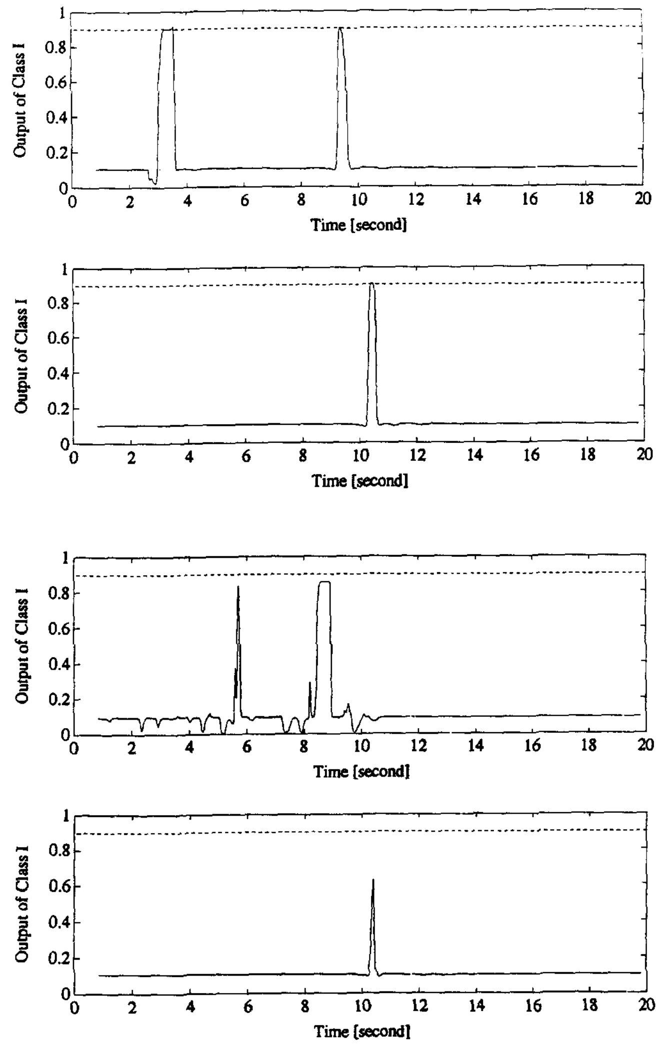


Figure 7. Corresponding output of the AND-A for the recordings in Figure 5. If the output value is equal to 0.9 or larger, the input signal is classified as an earthquake onset.

Table 1
Testing Results for the AND's and Comparison with STA/LTA Threshold Method

| Method | Threshold | Correct Identification | Correct Rate |
|---|---|---|---|
| STA/LTA | 2 | 48/60 | 80.0% |
| | 3 | 43/60 | 75.0% |
| | 4 | 37/60 | 61.7% |
| | 5 | 34/60 | 58.8% |
| AND-A | no | 55/60 | 91.6% |
| AND-B | no | 59/60 | 98.3% |

training process, the weights that connect the input and output neurons are set to small random values to initialize the neural network. Taking advantage of adaptive characteristic of the neural network, a random initialized neural network is first trained with a small subset of training data, say 10 patterns. Later on, the primary trained neural network is updated by learning from more training data. With this training strategy, the training process will be more efficient and can sometimes significantly decrease learning time.

After training with a set of 1-sec window STA/LTA patterns, the trained AND is used to discriminate 60 seismograms, each of which is 20-sec long. Since the AND-A detects the signal continually and recursively with 0.02-sec intervals, one 20-sec-long seismogram has 950 windows to be examined. The results of the test indicate that the neural network is indeed an excellent classification system.

Figures 5 through 7 show some examples of discrimination. Trace A in Figure 5 is a recording of a seismic event with high signal-to-noise ratio. Trace B is a seismic recording with a very noisy background. Trace C is a recording of a nonseismic event, and trace D is a recording of background noise. For each incoming signal, the STA and LTA in a moving window over the data are calculated recursively by equations (11) and (12). The short-term moving time window used here is 0.4 sec, and the long-term moving time window is 6 sec. Figure 6 shows four sequences of STA/LTA corresponding to traces A, B, C, and D, respectively. From Figure 6 we can see that trace A has a very high STA/LTA ratio of 19.6 at the onset, trace B has a low STA/LTA ratio of 2.03 at the onset, trace C has a peak of STA/LTA of 4.2, and trace D has a low peak value of 1.6. Applying the conventional threshold triggering method, these four cases will be difficult to differentiate. However, the artificial neural network can handle this kind of nonlinear boundary decision problem better.

Figure 7 gives the output results of neuron one for traces A, B, C, and D, shown in Figure 5. If output of neuron one in our AND-A equals the desired value 0.9 that was assigned in the training pattern, the incoming signal is regarded as an earthquake onset. The output of trace A has two pulses exceeding 0.9; the first one corresponds to *P*-wave onset and the second one corresponds to the *S*-wave arrival. As soon as the system is triggered by the first detected earthquake onset, signals within a preset time window will be recorded automatically, so the second pulse of *S* wave is ignored. The output of trace B has one pulse, which corresponds to the *P*-wave arrival. Although the STA/LTA ratios of trace B are very low, our AND-A can still detect the *P*-wave arrival. In contrast, the AND-A does not classify trace C as an earthquake despite the fact that the STA/LTA ratio of trace C is higher than that of trace B. Even though there are two output pulses for trace C, none exceeds the

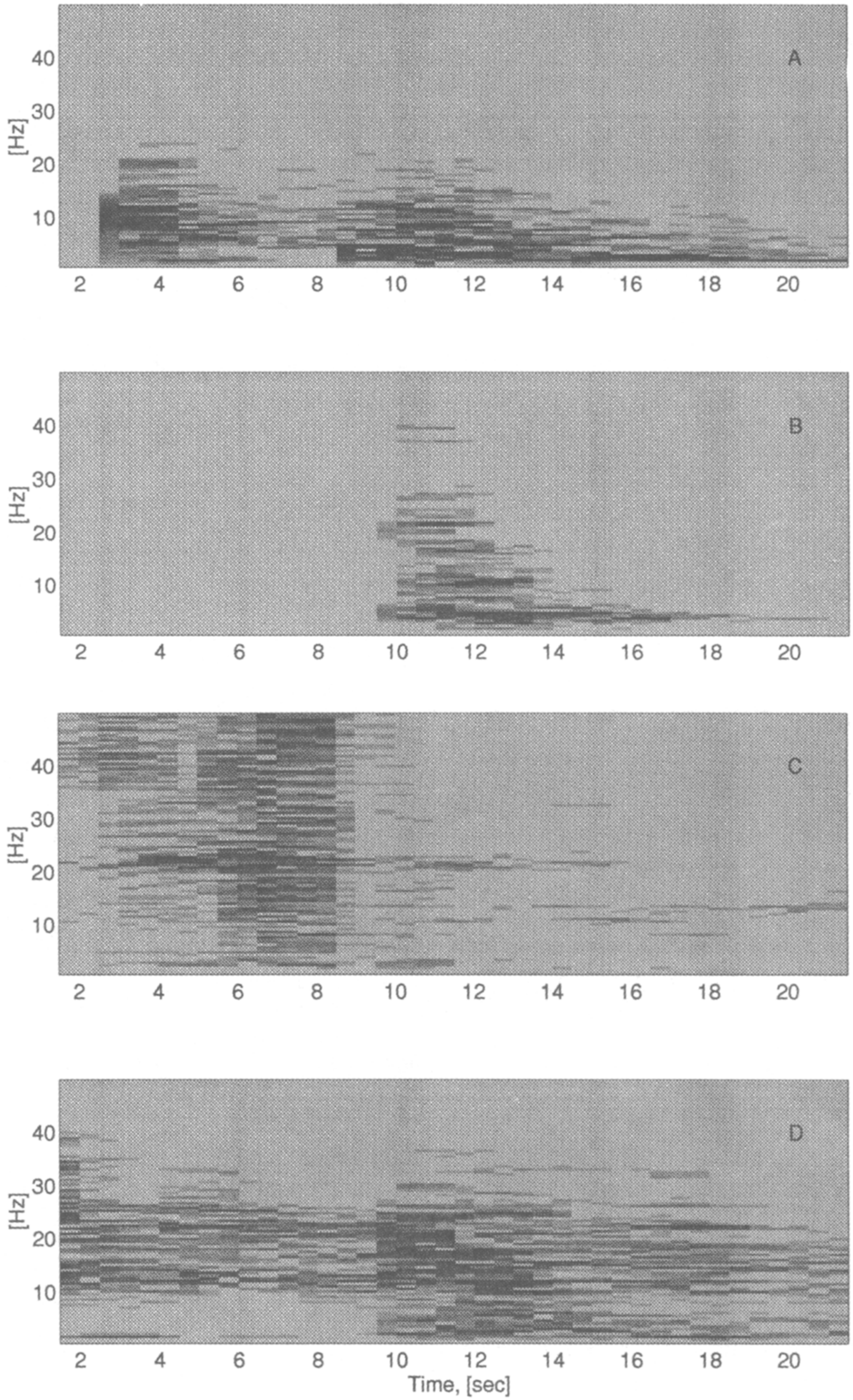


Figure 8. Corresponding spectrograms for the recordings in Figure 5. They show the moving window spectrum varied with time. The width of moving window is 2 sec and the moving step is 0.5 sec.

trigger value of 0.9. For trace D, the AND-A will not be triggered since the output is much lower then 0.9. It is interesting to see how the AND-A differentiates between traces B and D. The arrivals at 10 sec for B and D look very similar in the time traces (Fig. 5) and in the STA/LTA ratios (Fig. 6), but it is the trained neural network that tells the subtle differences in the composition of the waveforms: a more dispersive signal in B versus more monotonous fluctuations as a consequence of having a nearly "white" spectrum in D. Comparing Figures 5, 6, and 7, we find that the AND-A can discriminate events quite well, regardless of their STA/LTA ratios.

Table 1 gives the detection results obtained by the AND-A and a comparison with the conventional STA/LTA threshold method for the whole test data set. For the sake of simplicity, we only change the threshold value of $\beta$ and fix other thresholds. The conventional method as a different correct rate of detection with a different threshold value. The AND-A, however, with no preset threshold needed, has a correct rate as high as 91.6%. It is clear that the AND-A outperforms the threshold method, especially for the low signal-to-noise ratio data.

### Artificial Neural Detector in the Frequency-Time Domain (AND-B)

Although the AND-A outperforms the conventional STA/LTA threshold method, there are still some misclassifications for the given test data set. This implies that using the STA/LTA sequence in the time domain as input features has limitations even for our neural network-based seismic detector (AND-A). Since the spectral contents for earthquake and nonearthquake are intrinsically different, we develop another type of neural network-based seismic detector using the moving window spectrum of incoming signals as the input feature of the neural network and handle the detection problem in the frequency domain. We call this type of artificial neural detector the AND-B. To maintain real-time operation, the moving window spectra are computed continually for 2-sec (200 points) window lengths with 75% overlapping, i.e., the moving window interval is 0.5 sec. This process of moving window spectra becomes practical because of the modern fast CPU. Figure 8 gives the spectrograms of the same input traces A, B, C, and D in Figure 5. The spectrogram is a two-dimensional image that includes both temporal and frequency information. From Figure 8, we can see that the spectrograms of earthquakes have basically different characteristics from the spectrograms of nonearthquake signals.

As in the AND-A, the AND-B also adopts three layers of the MLP (Multi-Layered Perceptron) neural network with the supervised learning algorithm. The input layer consists of 100 units containing the normalized spectral distribution in a frequency band from 0.5 to 50 Hz. This frequency band is wide enough for most local earthquake recordings. The hidden layer consists of four neurons and the output layer has only one neuron. The reason for selecting four hidden neurons is based on the performance of the learning process, which will be discussed later. The earthquake triggering decision is represented with two values of one output neuron. The structure of the AND-B results in $101 * 4 + 5 * 1 = 409$ connections (weights and thresholds) between the input and output. The normalized spectra of the $P$-wave signals and nonearthquake signals are selected as the training data set. It consists of 60 patterns, of which 30 patterns are the spectra of seismic $P$ waves and the other 30 are the spectra of various noise sources.

The same testing data set used for AND-A is used here to test the AND-B. Since the AND-B detects the signal continually and recursively with 0.5-sec intervals, one 20-sec-long seismogram has 40 windows to be examined. The training samples of the AND-B are selected from the spectra of seismic $P$ waves and noises, so the output of the AND-B is not directly related to the onset of the earthquake. The system will be triggered when the first moving window gives rise to an AND-B output of 0.9, and then a fixed length of signal will be recorded as an event. The later AND-B output will be ignored until the system has finished the recording. Figure 9 gives the output results from the AND-B for the seismograms in

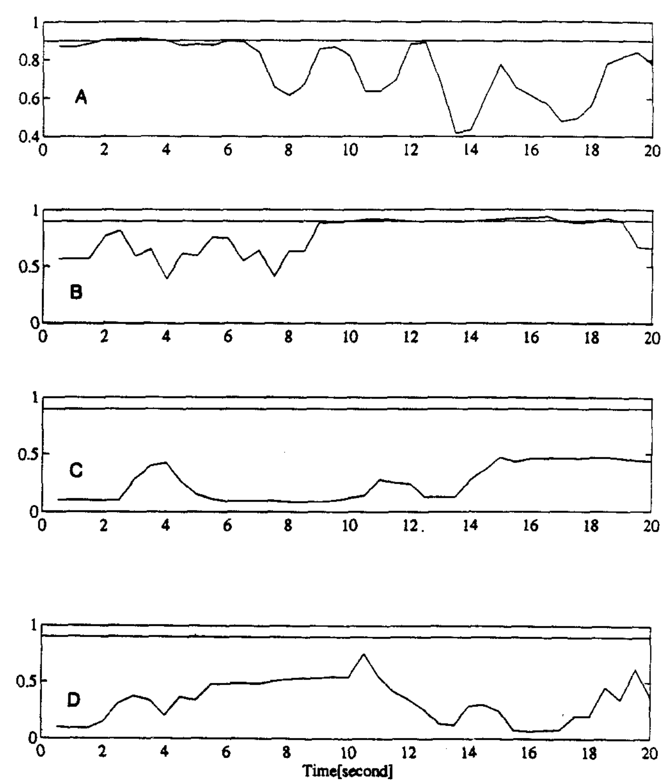


Figure 9. Corresponding output sequences of the AND-B for the recordings in the Figure 5. If the output value is equal to 0.9 or larger, the input signal is classified as the $P$ wave of an earthquake.

Figure 5. From Figure 9 we can see that trace A is detected at 2 sec, where the output of the AND-B reaches 0.9. After 7 sec, the output values become much lower. This means that the *P* wave of trace A has a 5-sec duration. The later parts belong to *S* waves that have different spectral features. Trace B is detected at 9 sec and the *P* waves last about another 9 sec. Traces C and D are classified as noises since *P* waves are not detected. For the same testing data set, the AND-B classifies these data very well except for one mistake out of 60 test seismograms, or a 98.3% correct rate. No trigger detection was produced by strong peaks of spikes or other nonearthquakes. A more interesting example is shown in Figure 10, in which the whole seismogram consists of two large nonearthquake pulses and one small local earthquake signal. By the conventional STA/LTA threshold method, the seismic detector will trigger at the first pulse. Even for the AND-A developed in last section, these features are discriminated correctly. For the AND-B, however, the output of neural network recognizes the correct event at 12 sec.

The uncertainty of trigger time of the AND-B is 0.5 sec, since a 0.5-sec window interval is used. The moving window itself cannot be set too long or too short. A longer window produces a larger error in arrival time and a shorter window will not offer a good spectral representation. Compared with the conventional STA/LTA detector, the AND-B is found to be more efficient in discriminating events and nonevents, either with low signal-to-noise ratios or with spikelike noises.

## Discussion and Conclusions

We use prerecorded aftershocks of the Landers earthquake of 1992 as our test data for the artificial neural detector, and found that the neural network-based detector can be used in real-time automatic seismic monitoring. The normalized recursive STA/LTA time series and the normalized moving window spectra in the frequency-time domain are selected as input of the artificial neural networks. Other features of seismograms can be developed in further enhancing this classification capability.

Learning is an important characteristic of artificial neural networks. The learning ability varies with the network architecture. Since the number of input units and output units is determined by the input pattern dimension and the classes of output, the number of the hidden neurons is the only free parameter that needs to be determined by the designer. The problem of size choice is under intensive study, with no conclusive answer available thus far for different tasks. No strict rules appear to exist for choosing the number of units in the hidden layer. According to the discussion of Zurada (1992), the

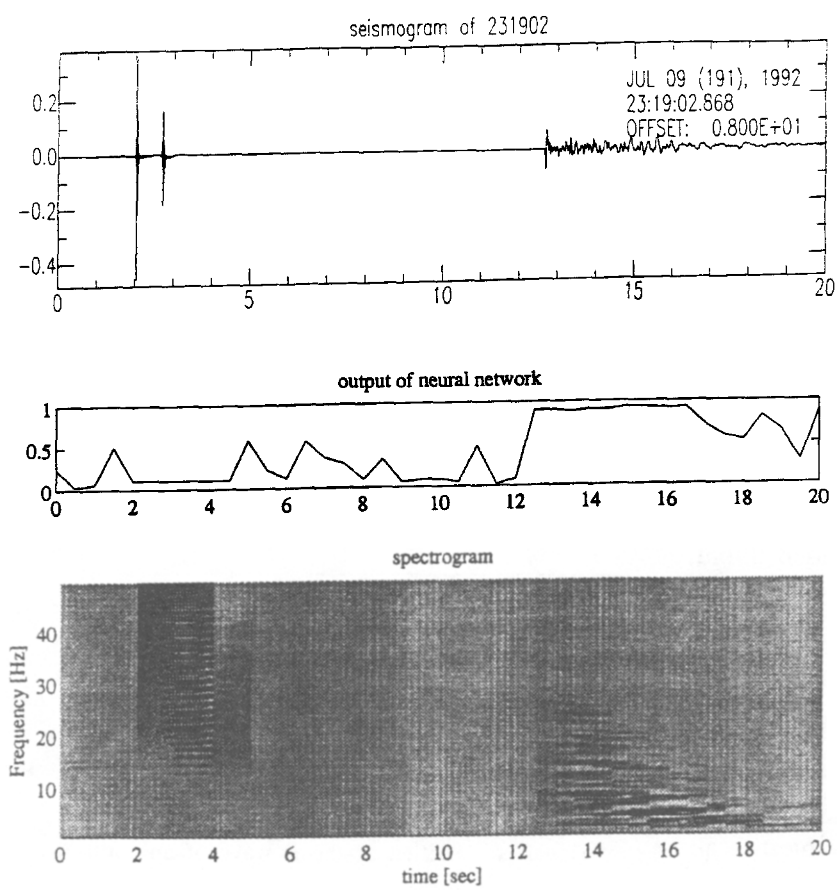


Figure 10. An example of application of the AND-B. Based on the spectrum feature, the AND-B can classify nonearthquake spikes and weak earthquake signal correctly.

number of training data should be much larger than the number or weights of the neural network.

In our approach, the number of units in the hidden layer is decided by the corresponding learning performance of the neural network during the training process. The performance of the neural network detector with different numbers of hidden neurons can be studied from the learning curves. The learning curves of the AND-A and AND-B with different numbers of hidden neurons, from one to ten, are plotted in Figure 11. From Figure 11a we can see that the training process of the AND-A converges most quickly for our detection problem with eight hidden neurons. So, eight neurons are implemented in the hidden layer for the AND-A. From Figure 11b we also find that the AND-B with four hidden neurons learns the training patterns most quickly and converges smoothly. Hence, the hidden layer of the AND-B in our problem consists of four neurons. In connection with the input and output dimensions, we note that AND-A has a total of $51 * 8 + 9 * 2 = 426$ weights, and the input layer has a total of 50 (training patterns) * 50 (points of each pattern) = 2500 data points. So, the total number of input data is larger than total number of weights. AND-B has a total of 6000 training data points, which is much larger than the total number of $101 * 4 + 5 * 1 = 406$ weights.

"Adaptivity" is another important property of artificial neural networks. The performance of the neural network can be improved by updating the weights by learning from new training data sets, rather than by

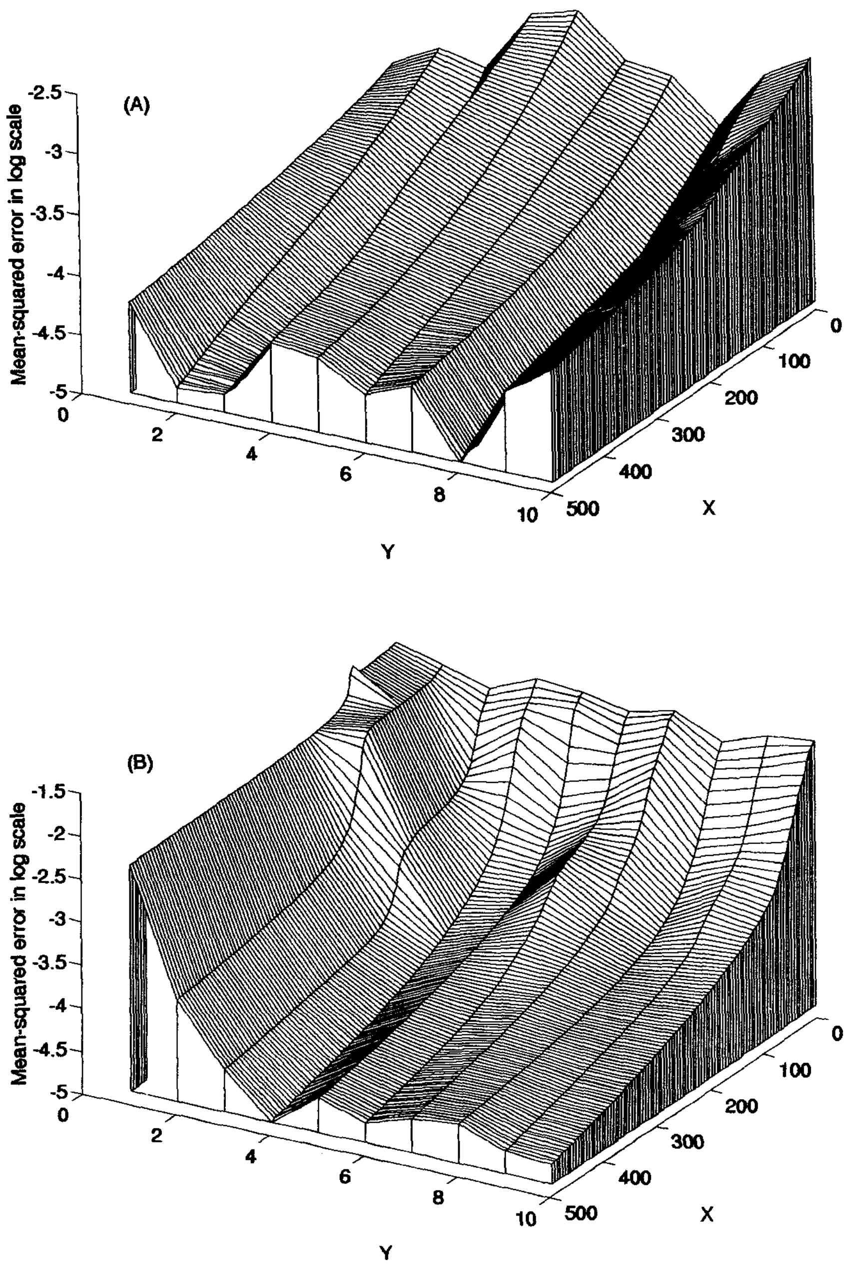


Figure 11. Learning surfaces with different hidden neuron numbers for (a) the AND-A, and (b) the AND-B. The $x$ axis represents the number of iterations and the $y$ axis represents the number of hidden neurons. AND-A with eight neurons gives the minimum error and the AND-B with four neurons gives the minimum error.

changing the algorithms for computation or the structure of neural network. The performance of ANN's depends crucially on what patterns have been learned. One must be certain that the training set consists of correctly identified patterns, otherwise false information may be incorporated into a poorly trained neural network. To show the adaptive property of the neural network, numerical experiments are conducted for the AND-B. Among the 60 total training patterns, a certain number of training samples, says 5, 10, 15, ..., are selected randomly. The neural network is then trained by these selected training sample subsets. The trained AND-B is applied to classify 60 patterns, and the objective function $J$ of equation (4) that measures average system error is computed. For a certain number of training patterns, the experiments are repeated three times with different training sample subsets by randomly choice. The average results and their standard deviations of the whole experiment are plotted in Figure 12. Clearly, from Figure 12, the larger the number of training events, the better the neural network performs. Starting with a primitive training data set, a successful detector would refine its performance, which may eventually approach the skill of an experienced human operator.

Taking advantage of the adaptive characteristics of neural networks, the learning strategy can begin with the AND trained by a small subset of training data for a randomly initialized neural network. The system improves itself with more training data when available by adaptive learning. This strategy can sometimes reduce the learning time significantly. A number of numerical experiments for the AND-A are conducted on two learning strategies, adaptive learning and initialized learning. The latter means the entire training data set is used in one training process. The results show in Figure 13 that the adaptive learning strategy can indeed save much learning time, especially for the case of a large training data set. The closed circles in Figure 13 represent the learning time for the initial network with different sizes of training data sets. The closed squares in Figure 13 represent the cumulative time for the adaptive learning process, in which the neural network is updated with five new training patterns each time. As shown in Figure 13, the cumulative time of the adaptive learning process is less than the learning time for initialized learning in general. There are no large differences between the cumulative time and the initial learning time for small training data sets. However, the adaptive learning process can save up to 40% of the learning time for a large training data set.

Although neural networks have exciting possibilities, they are of course not without their limitations. Conventional artificial intelligence approaches have in their favor the more transparent mechanisms that can often be expressed in terms of logic operations and rule-based representations, and that are meaningful to us in our everyday lives. On the other hand, artificial neural networks do not use structured knowledge with symbols used by humans to express the reasoning process. The weights of a neural network are just a set of numbers that in most cases has no obvious physical meaning to humans. Thus, a neural network is to some extent a "black box" solution to problems.

Through the application of neural networks to seismic detection problems, this study has reached the following conclusions.

1. The recently developed artificial neural network technique can be applied to routine seismic monitoring work, and can be potentially useful in real-time seismology. Two types of artificial neural network-based

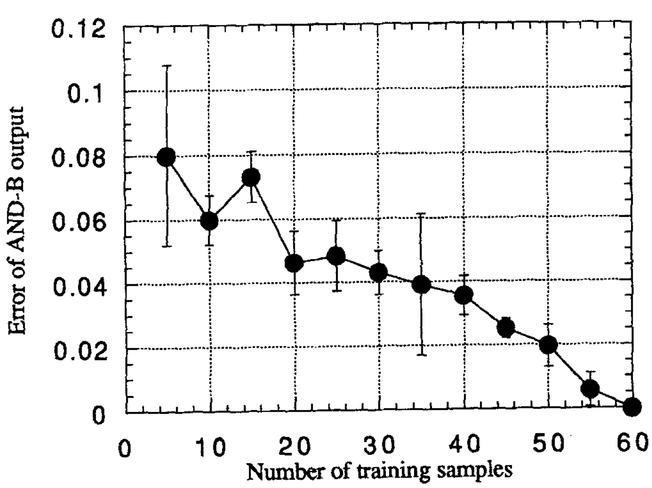


Figure 12. Average errors of the output by adaptive training of the AND-B. It shows that the errors decrease with more training data.

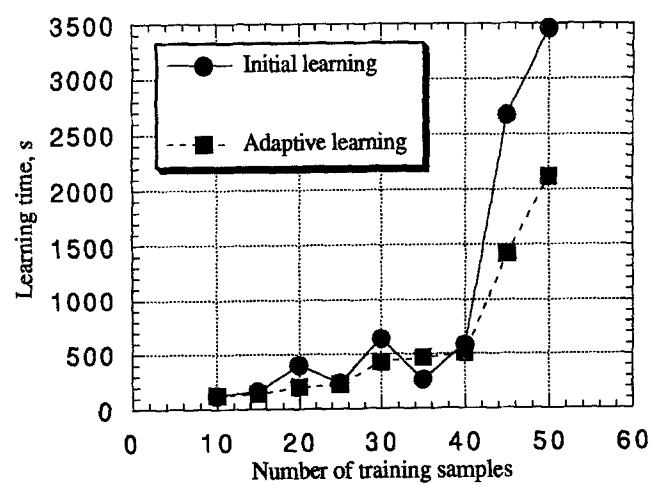


Figure 13. Comparison of learning times for different learning strategies. Closed circles represent the learning time for the initial neural network. Closed squares represent the cumulative time for adaptive learning.

seismic detectors (AND) are designed for seismic event detection. Trained and tested with real seismic data, both types of AND out-perform the conventional STA/LTA threshold classification method for events either with low signal-to-noise ratio or with spikelike noises.

2. The success of the artificial neural network for pattern classification depends primarily on the designer's physical insight into the problem. A good selection of input features by the designer will result in an easier classification for ANN's. The number of neurons in the input and output layers of ANN's is determined by the selected features of the problem. The number of neurons in the hidden layer can be determined by the learning performance of ANN's.
3. The artificial neural network approach is superior to the conventional threshold classification method, which bases its prediction on an individual parameter, while the neural network is capable of handling collectively the complex nonlinear problems involving many implicitly correlated parameters.
4. The neural network can be applied to other seismic classification, discrimination, and inversion problems and it will emerge as an important seismological tool in the future.

## Acknowledgments

The authors are grateful to K. Aki, S. Katz, Y. G. Li, D. Adams, L. X. Wang, and Y. P. Du for their kindly help and discussions. This research is supported by the USGS Contract Agreement 14-08-0001-A0620.

## References

Alexander, S. S., R. Hsu, and D. R. Baumgardt (1992). Identification of earthquakes and explosions, using Pattern record recognition techniques on frequency-slowness seismic image (abstract), *EOS* **73**, no. 43, 360.

Antognetti, P. and V. Milutinovic (1991). *Neural Networks: Concepts, Applications, and Implementations*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Dowla, F. U., S. R. Taylor, and R. W. Anderson (1990). Seismic discrimination with artificial neural networks: preliminary results with regional spectral data, *Bull. Seism. Soc. Am.* **80**, 1346–1373.

Dysart, P. S. and J. J. Pulli (1990). Regional seismic event classification at the NORESS array: seismological measurements and the use of trained neural networks, *Bull. Seism. Soc. Am.* **80**, 1910–1933.

Katz, S. and K. Aki (1992). Experiments with a neural net based earthquake alarm, *EOS* **73**, no. 43, 366.

Morgan, D. P. and C. L. Scofield (1991). *Neural Networks and Speech Processing*, Kluwer Academic Publishers, Boston, 43–44.

Pao, Y.-H. (1989). *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, Massachusetts, 269–290.

Rumelhart, D. E., J. L. McClelland, and the PDP Research Group (1986). Learning representations by back-propagating errors, *Nature* **332**, 533–536.

Tottingham, D. M., W. H. K. Lee, and J. A. Rogers (1989). *User Manual for MDETECT.IASPEI Software Library*, Vol. 1, Seismological Society of America, El Cerrito, 89–118.

U.S. National Research Council (1991). *Real-Time Earthquake Monitoring: Early Warning and Rapid Response*, National Academy Press, Washington, D.C.

Wang, L. X. (1992). A neural detector for seismic reflectivity sequences, *IEEE Trans. Neural Networks* **3**, 338–340.

Wang, L. X. and J. Mendel (1992). Adaptive minimum prediction-error deconvolution and source wavelet estimation using Hopfield neural networks, *Geophysics* **57**, 670–679.

Zurada, J. M. (1992). *Introduction to Artificial Neural Systems*, West Publishing Company, St. Paul, 216–218.

Department of Earth Sciences
University of Southern California
Los Angeles, California 90089-0740

Manuscript received 6 July 1993.