Geophys. J. Int. (1995) **120,** 758–774

Automatic picking of seismic arrivals in local earthquake data using an artificial neural network

Hengchang Dai[1,2] and Colin MacBeth[1]

[1] *British Geological Survey, Murchison House, West Mains Road, Edinburgh* EH9 3LA, *UK*
[2] *Department of Geology and Geophysics, University of Edinburgh, Kings Buildings, West Mains Road, Edinburgh* EH8 3JW, *UK*

Accepted 1994 September 23. Received 1994 September 23; in original form 1993 November 4

SUMMARY

A preliminary study is performed to test the ability of an artificial neural network (ANN) to detect and pick seismic arrivals from local earthquake data. This is achieved using three-component recordings by utilizing the vector modulus of these seismic records as the network input. A discriminant function, $F(t)$, determined from the output of the trained ANN, is then employed to define the arrival onset. 877 pre-triggered recordings from two stations in a local earthquake network are analysed by an ANN trained with only nine P waves and nine noise segments. The data have a range of magnitudes (M_L) from −0.3 to 1.0, and signal-to-noise ratios from 1 to 200. Comparing the results with manual picks, the ANN can accurately detect 93.9 per cent of the P waves and also 90.3 per cent of the S waves with a $F(t)$ threshold set at 0.6 (maximum is 1.0). These statistics do not include false alarms due to other non-seismic signals or unusable records due to excessive noise. In 17.2 per cent of the cases the ANN detected false alarms prior to the event. Determining the onset times by using the local maximum of $F(t)$, we find that 75.4 per cent of the P-wave estimates and 66.7 per cent of the S-wave estimates are within one sample increment (10 ms) of the reference data picked manually. Only 7.7 per cent of the P-wave estimates and 11.8 per cent of the S-wave estimates are inaccurate by more than five sample increments (50 ms). The majority of these records have distinct local P and S waves. The ANN also works for seismograms with low signal-to-noise ratios, where visual examination is difficult. The examples show the adaptive nature of the ANN, and that its ability to pick may be improved by adding or adjusting the training data. The ANN has potential as a tool to pick arrivals automatically. This algorithm has been adopted as a component in the early stages of our development of an automated subsystem to analyse local earthquake data. Further potential applications for the neural network include editing of poor traces (before present algorithm) and rejection of false alarms (after this present algorithm).

Key words: arrival time, artificial neural network, pattern recognition, picking.

1 INTRODUCTION

The primordial task of estimating arrival times for the primary (P) and secondary (S) waves found in recordings of an earthquake event still forms an important foundation for schemes employing automatic processing for event location, event identification, source mechanism analysis and spectral analysis. There is no shortage of techniques which profess to tackle this problem; however, they do tend to be data specific and are not generally available. The goal of global automation is far from achieved and such elementary seismogram interpretation still forms a bottleneck in the routine work of many observatories.

A great deal of effort, stretching back several decades, has been devoted to the automation of arrival picking, and many different varieties of algorithm exist. Only a few notable procedures are mentioned in this review for reasons of brevity. The method popularized by Allen (1978) uses a short-term and long-term average ratio to pick P waves. Numerous variants of this general scheme have been implemented. For example, Bear & Kradolfer (1987) used an envelope function for each signal trace in this algorithm

and then passed it through a non-linear amplifier. The resulting signal is then subjected to a statistical analysis to yield *P*-wave arrival times and a measure of reliability for the picking. Pisarenko, Kushnir & Savin (1987) developed an optimal *P*-wave picker using an asymptotic approximation of the likelihood function. Roberts, Christoffersson & Cassidy (1989) made an assessment of whether data are consistent with the arrival of a *P* wave or linearly polarized *S* wave using the auto- and cross-correlations of three-component data. Takanami & Kitagawa (1988, 1993) developed a method for *P* and *S* waves by fitting a locally stationary autoregressive model. Kracke (1993) developed a simple method based on the displacement vector of a seismic trace in a spherical coordinate system for *P* waves. Cichowicz (1993) developed an *S*-wave picker based on a filter which combines polarization and energy ratios. Joswig & Schulte-Theis (1993) used a master-event-correlation method to detect *P*-wave arrivals in weak local earthquake records. All of these are traditional programs: they quantify some attribute of the seismic trace and use this as the basis of the decision. These algorithms are not adaptive, working well under certain conditions, but quite often not producing good results. Analysts are still required to interactively check the quality of the result.

The tasks performed by the trained analyst in manually picking arrival times involve an intensive amount of pattern recognition. Experience provides a judicial balancing of wave characteristics such as amplitude, frequency and polarization from previous records at the same station to determine the most likely onset time. If questioned about a particular decision, however, the analyst may offer a few rules for guidance but can often give no obvious systematic reasoning because the decision has been partly subjective. This reasoning is based upon past experience. Consequently, different trained analysts give different answers, and the same analyst may choose a different interpretation after some time has elapsed. With the increase in the number of digital seismic networks being established world-wide, there is a pressing need to provide a more reliable and robust alternative, which is less time-consuming and more objective. The application of artificial intelligence methods to earthquake analysis is a relatively recent development which attempts to tackle these objectives. Various methods have been applied to the interpretation of seismic signals from a local seismic network such as knowledge-based systems according to the blackboard method used by Chiaruttini, Roberto & Saitta (1989), and later developed by Chiaruttini & Salemi (1993). Bache *et al.* (1990) developed an intelligent monitoring system based on this approach, data-base management systems and signal processing. As an alternative strategy Joswig (1990) developed a pattern recognition approach for *P* waves using a sonogram. Klumpen & Joswig (1993) used a pattern recognition technique to identify generic polarization patterns to estimate *P*-wave and *S*-wave onset times.

Artificial neural networks (ANNs), another group of techniques from the area of artificial intelligence, provide a natural alternative to this type of earthquake analysis as they have proven useful at handling complicated pattern recognition problems in other applications. ANNs have been used to solve a diversity of geological and geophysical problems. For example, Dystart & Pulli (1990) use them for the problem of automatic event classification. McCormack (1991) uses an ANN to combine synthetic spontaneous potential (SP) and resistivity logs to estimate lithology logs. Poulton, Sternberg & Glass (1992) use an ANN to estimate the offset, depth and conductivity-area product of a conductive target given an electromagnetic image of the target. Wang & Mendel (1992) use a Hopfield network to implement an adaptive minimum prediction-error deconvolution. The ANN also has been used in the first-break picking of surface seismic data (Murat & Rudman 1992; McCormack, Zaucha & Dushek 1993). The wide range of applications emphasizes the particular strength of this approach over traditional methods incorporating a fixed algorithm to solve a particular problem, as the ANNs utilize a learning scheme to develop an appropriate solution so that the network is flexible and adaptive to different data sets. In our application, the ANN may be likened to an analyst and is trained by presenting it with many different earthquake records. After training is accomplished, the ANN should be able to recognize new arrivals from a variety of new seismograms.

2 THE THEORY OF ARTIFICIAL NEURAL NETWORKS (ANNs)

The ANN is designed to simulate the neural connections of the human brain. It is made up of sets of nodes arranged in layers, consisting of an input layer, one or more intermediate hidden layers and an output layer. Fig. 1 schematically illustrates the architecture of such a neural network. Each node, the basic processing unit of the neural network, is usually drawn as a solid circle. Fig. 2 shows the components of this node, and what it represents. The outputs of the nodes in one layer are transmitted to nodes in another layer through links called 'weights'. These weights are real numbers, which are applied as simple multiplicative scalars, and effectively amplify or attenuate the signals. With

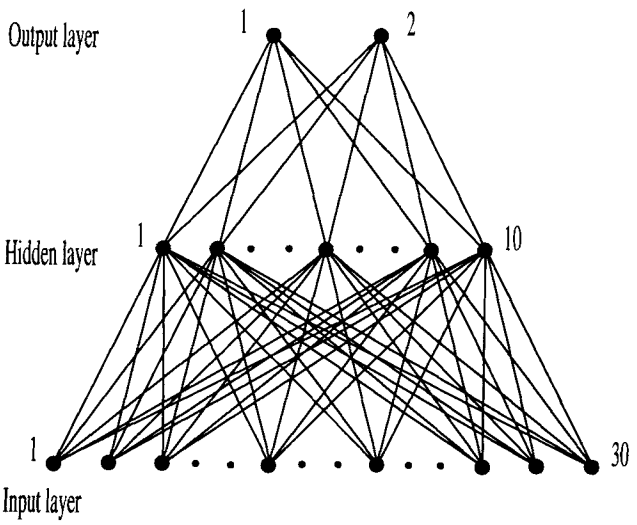

Figure 1. The neural network structure for picking arrivals on local earthquake records. Solid circles represent nodes, and straight lines represent weights. This is a three-layered neural network. The segment of vector modulus is input according to a time-series order. The output nodes give the results which are (1, 0) or (0, 1) for *P* wave or noise during training and are defined by the $F(t)$ during testing.

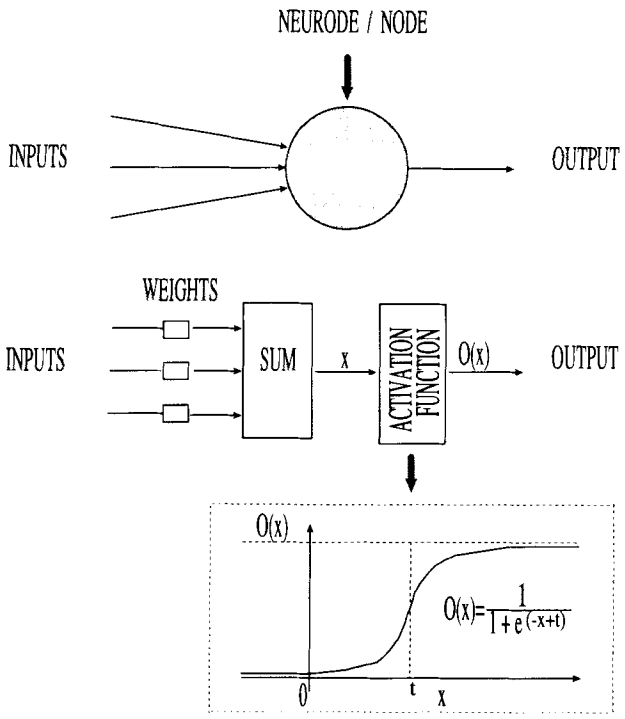

Figure 2. Schematic showing various constituents of a node in the neural network of Fig. 1. X represents summed input to the node, t the threshold, and $O(t)$ is the output. The node input is the sum of the weighted output of nodes in the previous layer and the node is then activated in accordance with the summed input using a pre-set function (usually a sigmoid function) and its threshold. All weights and thresholds are determined during the learning procedure.

the exception of the nodes in the input layer, the net input to each node is the sum of the weighted outputs of nodes in the previous layer. Each node is then activated in accordance with the summed input using a pre-set activation function (usually a sigmoidal function), and a threshold parameter for the function. In the input layer, the net inputs to each node are the components of the input pattern.

The ANN used in this present study is non-linear, multilayer, feed-forward and back-propagation of error (Rumelhart, Hinton & Williams 1986). This means that the activation function is non-linear, there are many layers in the network, and the signals feed through the network only in a forward direction. This is the most popular type of ANN in use today as it is well understood. It also incorporates a back-propagation learning algorithm, or *Delta Rule* which is usually used to train this type of neural network—a good mathematical summary is given by Pao (1988). This method attempts to find the most suitable solution (numerical values of weights and thresholds) for a global minimum in the mismatch between the desired output pattern and its actual value for all of the training examples. The degree of mismatch for each input–output pair is quantified by solving for unknown parameters between the hidden layer and output layer and then by propagating the mismatch backwards through the network to adjust the parameters between the input layer and hidden layer. In this learning procedure, the first pattern is presented as input to a randomly initialized network, and these weights and thresholds are then adjusted in all the links. Other patterns are then presented in succession, and the weights and thresholds adjusted from the previously determined values. This process continues until all patterns in the training set are exhausted (*an iteration*). The final solution is generally accepted to be independent of the order in which the example patterns are presented. A final check can be performed by looking at the *pattern error*, which is defined as the square of the mismatch between desired and actual output for each pattern, and the *system error*, which is defined as the average of all of these pattern errors, to determine whether the final network solution satisfies all the patterns presented to it within a certain threshold error. The set of weights and thresholds in the network are now specifically tailored to 'remember' each input and output pattern, and can consequently be used to recognize or generate new patterns given an unknown input. The network is now trained, and can be used in subsequent analyses.

3 DETECTION AND PICKING OF SEISMIC ARRIVALS USING AN ANN

3.1 Current objective

In order to present our current objective, it is necessary to distinguish between the following:

(1) *Arrival detection*—specification of an arrival time close to which an arrival may be bracketed within a pre-defined time window.

(2) *Arrival picking*—reliable and accurate estimation of the onset time of a definite seismic arrival. An arrival in our case is seismic motion defined by a wavelet for which the character resembles the training set wavelets.

(3) *Arrival identification*—this classifies individual arrivals into categories relating to their polarization (not necessarily linear) and to the nature of their propagation, for example *Pg*, *Sg* and *Lg* waves.

(4) *An event*—a transient seismic signal generated by a phenomenon such as an earthquake, quarry blast, sonic boom or underwater explosion, which is recorded as a time sequence. The event possesses a fine structure given by a definite hierarchy of arrivals, which are of importance in defining the event type and the Earth's structure.

(5) *A false alarm*—a spurious signal of non-seismic character or a disturbance sufficiently different in statistical character from an event that it cannot be readily utilized in defining the Earth's structure. Examples include electrical spikes, and continuous traffic noise.

(6) *An event window*—a time sequence whose endpoints bracket a seismic event of interest. This window is usually obtained through use of a triggered seismic network, and may contain many possible false alarms in addition to the main event. Determination of this feature is a robust and inherently stable operation. Arrival picking, which requires more resolution with a concomitant increase in sensitivity, is usually preceded by this stage.

(7) *Signal-to-noise ratio* (*SNR*)—the ratio between maximum vector amplitude of signal and quiescent period immediately before the arrival onset. For our purposes the noise level is evaluated within a window of 290 ms,

coinciding with the length of the input segment to the neural network.

In this paper we use the ANN for arrival detection and picking for *P* and *S* waves in local events, with the intention of constructing a hierarchical scheme of analysis in which (1) and (2) act as the basis for stage (3), and subsequently is used to define (4), (5) and (6). Hence, for the purposes of this work we do not consider the concept of an event window, and all examples are presented without consideration of false alarms. The distinction between a false alarm and an event is considered as a secondary objective in later work.

3.2 Input characteristics of data

The detection of the different arrivals is accomplished using the vector modulus of the three-component motion. This is useful as the recorded signal is strongly dependent on the source position and ray direction, which may otherwise give rise to a misleading interpretation. The instantaneous vector modulus $M(t)$, calculated at each individual three-component sample along the traces, separates the geometric dependency from our recorded vector motion whilst retaining the character of the seismogram for picking. $M(t)$ is then used as direct input for the ANN. It is believed that this attribute facilitates an easier identification, regardless of the polarization of the wave (Lomax & Michelini 1988). We do not use the polarization properties of individual arrivals as we believe it may not provide a satisfactory indicator due to such factors as phase changes during propagation, fine structure of the waveforms such as that due to shear-wave splitting, and directional dependency. The overriding concern is the uncertain applicability of such a parametric model of the wavefield in a heterogeneous crust (Der, Baumgardt & Shumway 1993).

$M(t)$ is presented to the network in segments which are selected from a sliding window which passes across the entire three-component seismogram. Each segment is individually normalized so that it is not dependent upon the magnitude and the distance of an earthquake, as otherwise it may bias the estimates with large changes. This reduces the number of training examples, which would otherwise have to cover a range of magnitudes and distances required, with a consequent increase in the training time and a larger network structure. This means that the network is forced to sense the relative amplitude and frequency content of the signals, and uses this information to detect the onset. For a high signal-to-noise ratio (SNR), the onset is characterized by a distinct change in the amplitude of the seismic activity. However, if the SNR is low, the major discriminating factor is a frequency change due to the different spectra of the background signal and earthquake signal. The network is similar to a sophisticated wavelet transform.

The data we use are local earthquake events recorded at stations DP and AY on the TDP-H1 seismic network (Crampin, Evans & Ucer 1985; Lovell 1989) between 1984 April and 1984 December. Several hundred local earthquakes are recorded on three-component seismometers at a 10 ms sampling interval. These recordings are not continuous and are triggered by a digital system (Evans *et al.* 1987). All are local, with depths from 2 km to 14 km and epicentral distances less than 30 km; most are closer to station DP than to station AY. For these local events, we identified predominant *Pg* and *Sg* waves in the seismogram records. Most events have magnitudes (M_L) between −0.3 and 1.0, and possess a wide distribution of SNRs which are shown in Figs 3 and 4 for the complete data set. All SNRs lie between 1 and 200, with station DP being of higher fidelity than station AY.

For the first test of the performance of the ANN we will apply it to a small subset of these data. The data set consists of 210 high-quality events recorded at stations DP (120 events) and AY (90 events). In the second test, we will use this trained neural network to process the complete data set of 877 recordings at stations DP and AY on the TDP-H1 network, with a mixture of good and bad data.

3.3 Neural network structure

A sliding window length is fixed at 290 ms (30 samples), and is chosen to include several complete cycles of the waves, giving 30 input nodes. There are two nodes in the output layer to flag the result: the output is (1, 0) for an arrival; and (0, 1) for pure noise. The number of hidden nodes depends on various factors such as input nodes, output nodes, system error, pattern error and training samples. There is no fixed generic relationship between the number and these factors for this type of network. However, we do know that in network learning, generalization is increased and memory is reduced by limiting the number of hidden nodes (Dowla, Taylor & Anderson 1990). Too few hidden nodes will lead to a long learning process or no convergence. In this case we finally chose 10 hidden nodes after a process of trial and error with different training runs. Although this solution is considered optimal for the current application, further architecture optimization could undoubtably be achieved by a more exhaustive search procedure on a more powerful computer.

3.4 Training procedure

A small number of recordings are used to train the network, and the remainder are used to test the performance of the trained network. The performance of a trained neural network depends on the training data sets. If we use incorrect or inconsistent data to train the neural network, we cannot expect it to give a correct answer for new data. For training, the $M(t)$ segments include either background signal or the *P* wave with some early background signal. The *P*-wave training segments are chosen to include waves with different characters. We do not include *S* waves in the training data set because $M(t)$ for these displays appears to exhibit a similar character. The function of this primary picker is to flag as many changes of $M(t)$ as possible and discard those which are neither *P* nor *S* arrivals. The segments are arranged so that the predicted onset time of every signal lies at the tenth sample, for which the network output flags (1, 0). This behaviour is imprinted on every training example. Fig. 5 shows the nine *P*-wave and nine background signal training segments used in the study. In the first experiment, we use seven training segments of *P*

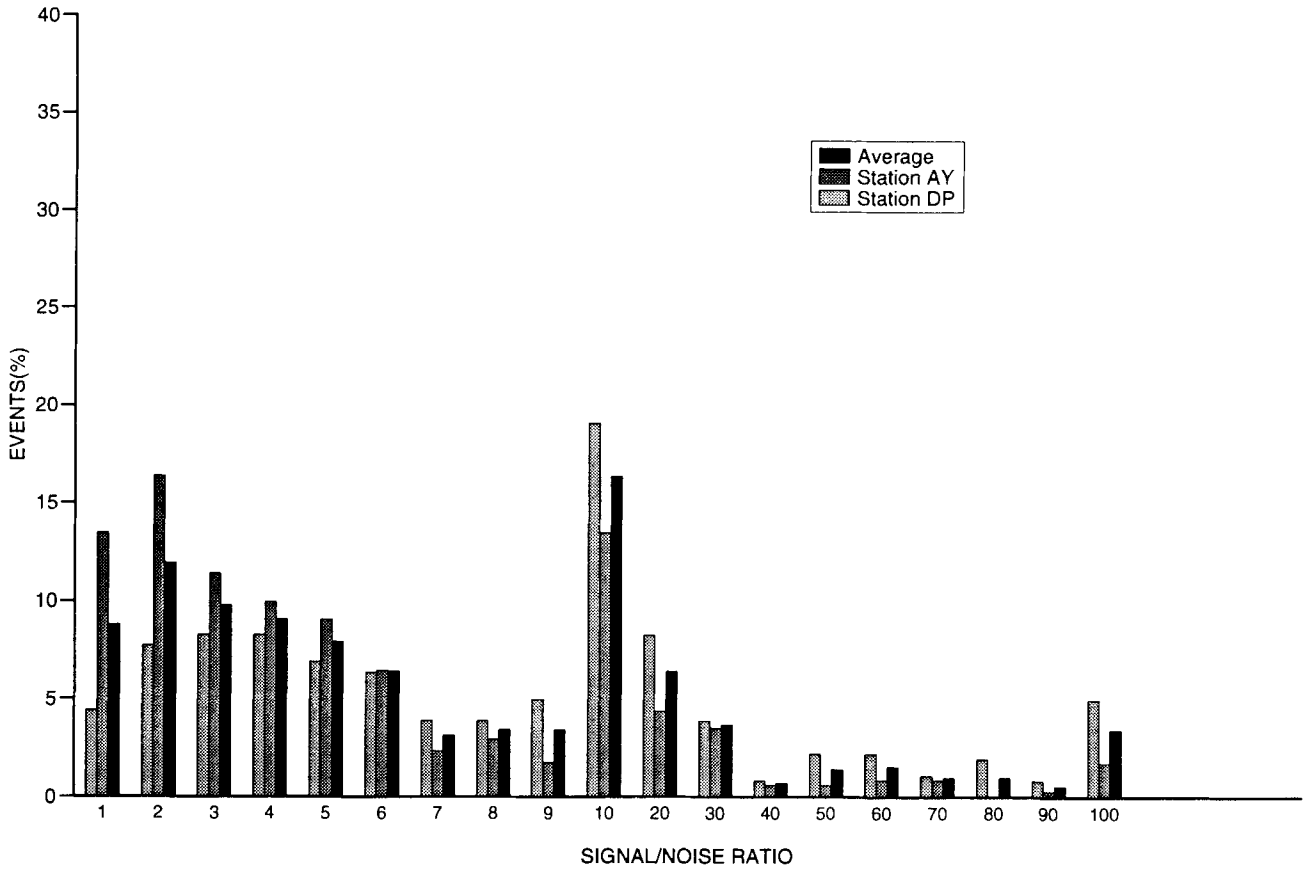

Figure 3. The distribution of signal-to-noise ratio (SNR) of P waves manually picked in the complete data set which includes 877 recordings. SNR is defined as the ratio of the $M(t)$ maxima before and after the onset time.

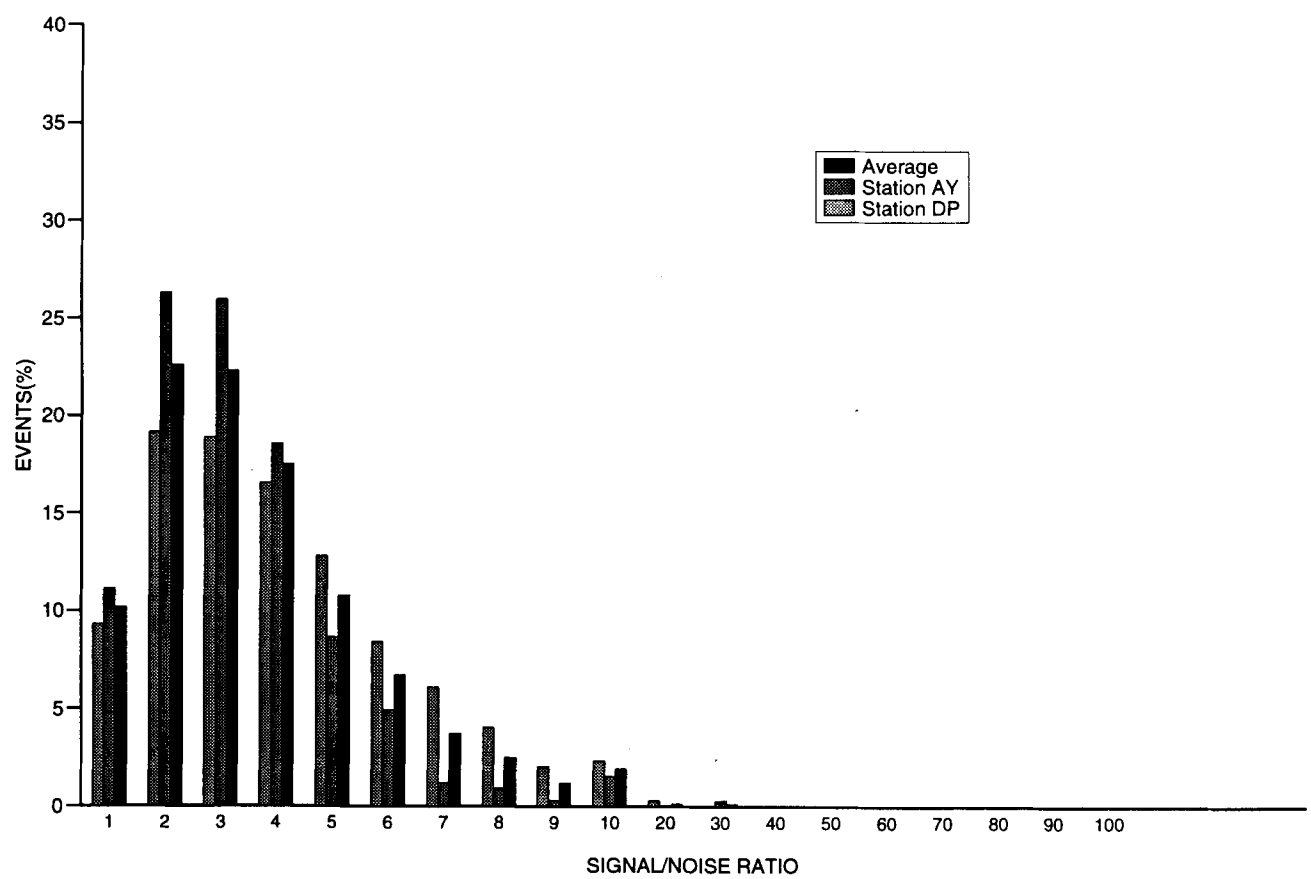

Figure 4. As Fig. 3 but for the S waves manually picked in the complete data set which includes 877 recordings.

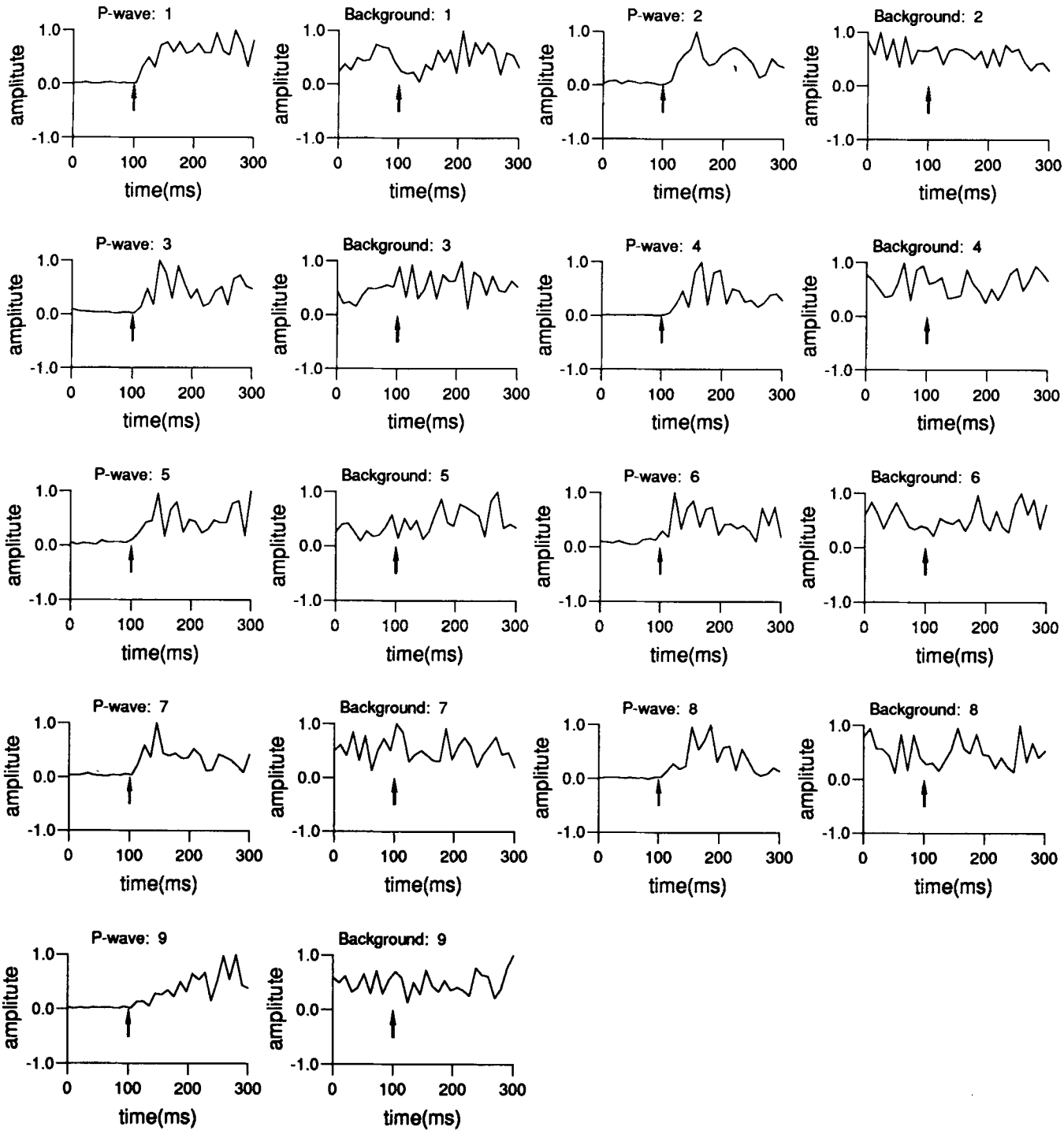

Figure 5. Nine *P*-wave segments and nine noise segments used for training the network. Noise segments are extracted prior to the *P*-wave arrival in the same seismogram. Arrows on *P*-wave segments indicate arrival times used to train the network, all are at the tenth sample. These segments are individually normalized before being input into the neural network.

arrivals and noise (Fig. 5). The training procedure takes 498 iterations (less than one minute of CPU time on a VAX4000). The system error reached is 2.5×10^{-5}, with a pattern error of 10^{-4}. After training, the neural network is ready to pick the waves.

3.5 Arrival detection

To detect or pick an arrival we first calculate the observed $M(t)$, take each windowed segment of this, and then feed it into the trained neural network. We shift the window by one

sample at a time and feed each segment into the network, storing the output. The procedure is repeated until the end of the seismogram is reached. In general, the output $(o_1(t), o_2(t))$ lies between the ideal for a signal or for a noise (for example, (0.8, 0.2) or (0.4, 0.6)). To provide a single indication of the onset, we use a function $F(t)$ which highlights the difference between the actual output and ideal noise:

$$F(t) = \tfrac{1}{2}[(o_1(t))^2 + (1 - o_2(t))^2]. \quad (1)$$

Figure 6 shows an example of $F(t)$ for one of the data segments in our chosen data set. The peaks in $F(t)$ correspond to abrupt changes in $M(t)$, with a small value implying a smooth change. These in turn are dependent on changes of the amplitude and frequency through the weightings in the network. In this curve, there are two large peaks corresponding to P and S waves. The positions of their maxima occur exactly at the manually chosen onset times of the P and S waves. We find that for most cases an arrival corresponds to a sharp change in $F(t)$, so that a threshold may be sufficient to detect the arrival. With a threshold of 0.6 it is possible to detect 200 P-wave arrivals (95.2 per cent) and 184 S-wave arrivals (87.6 per cent) from 210 sets of three-component data. If we decrease this

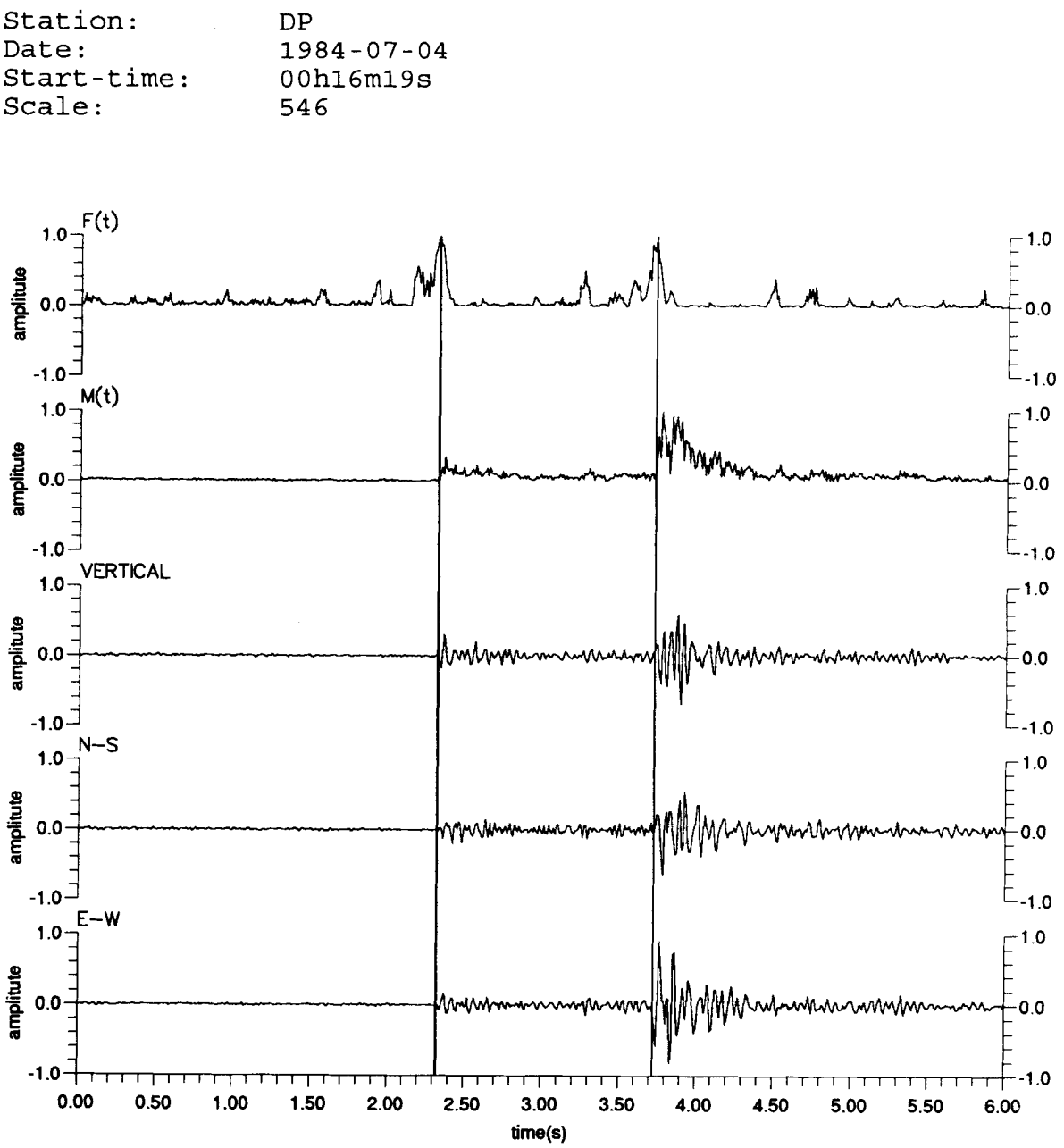

Figure 6. Three-component vector modulus $M(t)$ and $F(t)$ function computed from output of the neural network. The vertical lines are automatically picked by the network, and indicate exact picks without error. The maxima correspond to the P and S waves, although there are some small maxima in $F(t)$ prior to the main arrivals which show small changes in $M(t)$. There is a small precursor before each main peak in $F(t)$ which disappears in better trained networks.

threshold to 0.5, 205 *P*-wave arrivals (97.6 per cent) and 200 *S*-wave arrivals (95.2 per cent) are detected. Given that only seven *P*-waves from station DP were used to train this network, this result is extremely encouraging. It is interesting to note other peaks in this function, which indicate other wave arrivals, spikes or noise bursts. Fig. 7 shows such an example including some spikes in the seismogram. Other techniques are necessary to disregard these signals. If the peaks correspond to spikes which are typically one or two sample points of anomalously large amplitude relative to the background noise, they may be flagged using a conventional algorithm. It is also possible to discriminate false arrivals by examining the peak of $F(t)$, the arrival being rejected if only one or two points are greater than the threshold and if the $M(t)$ maximum is less than a pre-specified amount. Otherwise, we confine such discrimination to a secondary stage of our analysis scheme.

3.6 Picking of onset time

As the segments of $M(t)$ are fed into the trained neural network, $F(t)$ reaches a maximum when the arrival time is at the tenth point (see Section 3.4). Either side of this maximum, $M(t)$ is shifted and the network output $F(t)$ decreases as shown in Figs 6 and 7. This implies that the onset time may be estimated by searching for a local

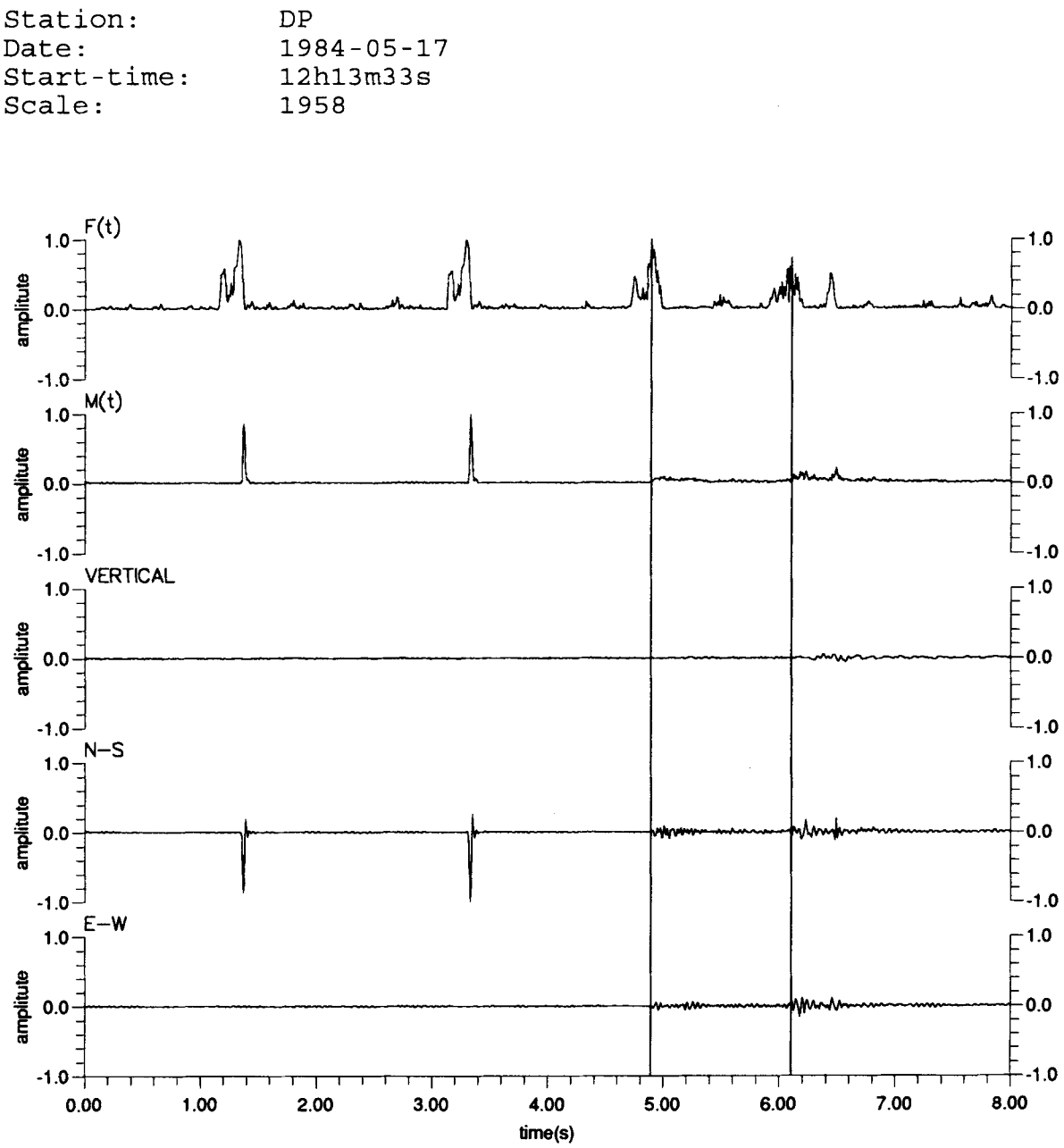

Figure 7. Three-component vector modulus $M(t)$ and $F(t)$ function computed from output of the neural network. Data have interfering electrical spikes, and $F(t)$ has some peaks with maxima greater than 0.6 corresponding to them. These spikes may be disregarded by a post-processing routine. *P* and *S* waves are automatically detected by this method.

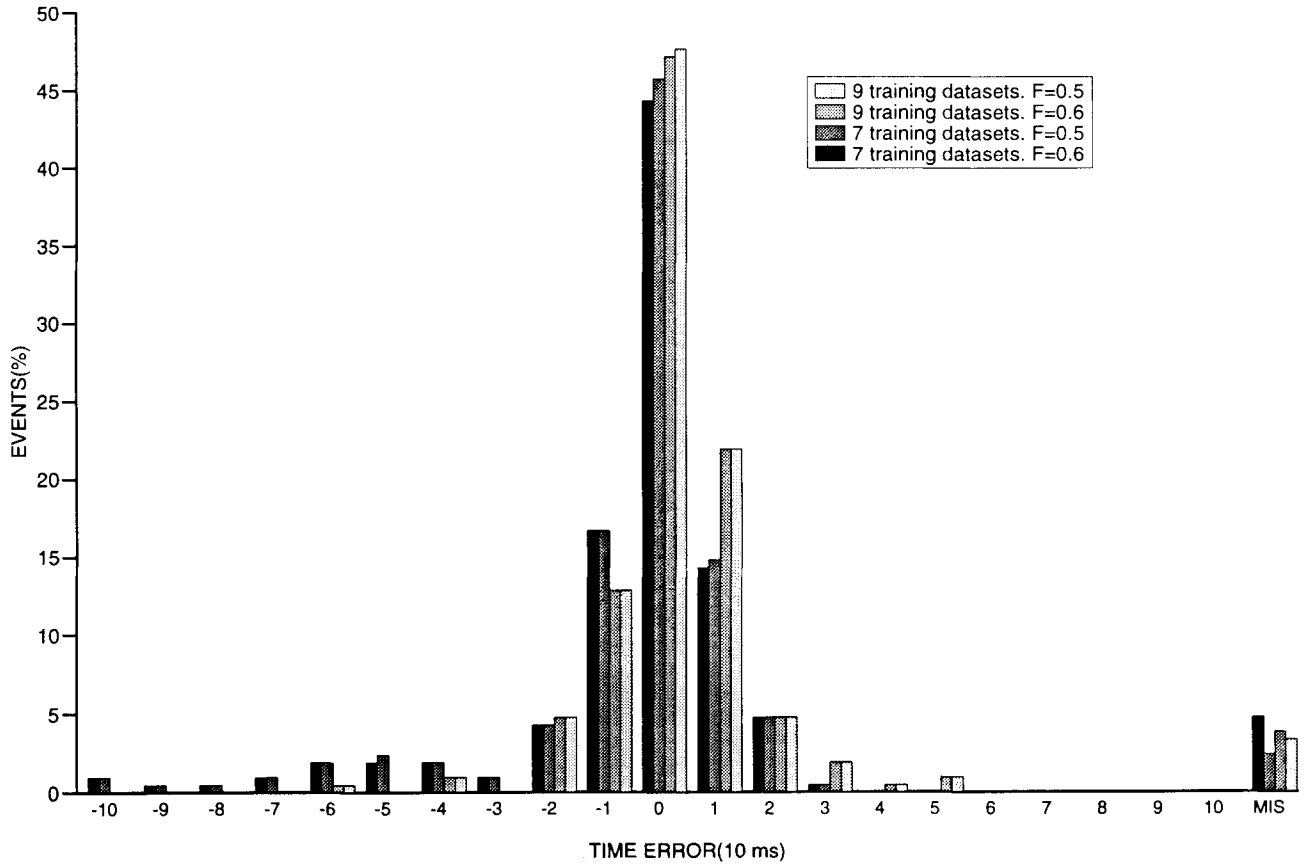

Figure 8. Statistics for *P*-wave picks on a data set of 210 local earthquake records. A negative value indicates a later pick than that given by the visual analysis. MIS refers to all those unpicked arrivals, defined as picks with errors larger than 10 sample increments (100 ms).

maximum after $F(t)$ exceeds the threshold. The $F(t)$ maxima exactly indicate the *P*- and *S*-arrival onset times in Figs 6 and 7. Note that each maximum has a small precursor which connects with the main peak, together with a base length equal to the input segment. According to the training procedure, small values of input nodes before the tenth point increase the value of $F(t)$. As the arrival onset comes into the input window, the relative values of input nodes before the onset decrease and the $F(t)$ should increase. For a well trained network, the $F(t)$ should gradually increase until the onset reaches the tenth point and then the $F(t)$ should gradually decrease until the onset moves out of the input window. However, for a less well trained network, as the onset moves forward in the window, $F(t)$ may decrease at some points forming a precursor. We find that as more training data are included, this precursor and the large peak merge into one wide peak (Fig. 14). Figs 8 and 9 show the statistical results of using this method for a network trained with the seven training sets of *P* arrivals and noise (Fig. 5). Here, we can estimate at least 75.2 per cent of the *P* waves and 50.0 per cent of the *S* waves having onset times of less than or equal to one sample increment using a threshold of 0.6, or 77.1 per cent of the *P* waves and 53.8 per cent of the *S* waves with a reduced threshold of 0.5. The onset time is relatively insensitive to the threshold, confirming that it is determined by the local maximum.

Let us now consider the possibility of improving the *P*-wave picks missed by the neural network. The arrivals not picked have no clear first motion, and the change of $M(t)$ is not visually obvious, with a small corresponding maximum in $F(t)$. A more suitable strategy is to retrain the neural network by including this type of data. This approach may be outweighed if too many training data are used, as this increases the training time, and there is the possibility of having to accommodate more subtle variations using a larger neural network structure. To tackle this, we include two extra *P*-wave segments (segments 8 and 9 in Fig. 5) which have different shapes from the other *P*-wave segments and two corresponding noise segments (Fig. 5). The training procedure takes 712 iterations, with a system error of 2×10^{-5} and a maximum pattern error of 10^{-4}. Figs 8 and 9 compare the results for this new trained network, with the previously trained network. The retrained neural network has an improved performance over the previous one. It detects 202 (96.1 per cent) *P* waves and 207 (98.6 per cent) *S* waves using the threshold for $F(t)$ of 0.6. The estimated onset times are also more accurate, with at least 172 (81.9 per cent) *P* waves and 162 (77.1 per cent) *S* waves (for $F(t) > 0.6$) or 173 (82.3 per cent) *P* waves and 163 (77.6 per cent) *S* waves (for $F(t) > 0.5$) having onset times $\leq$ one sample increment. Only one *P* wave and three *S* waves have onset times with errors $\geq$ five sample increments, and only

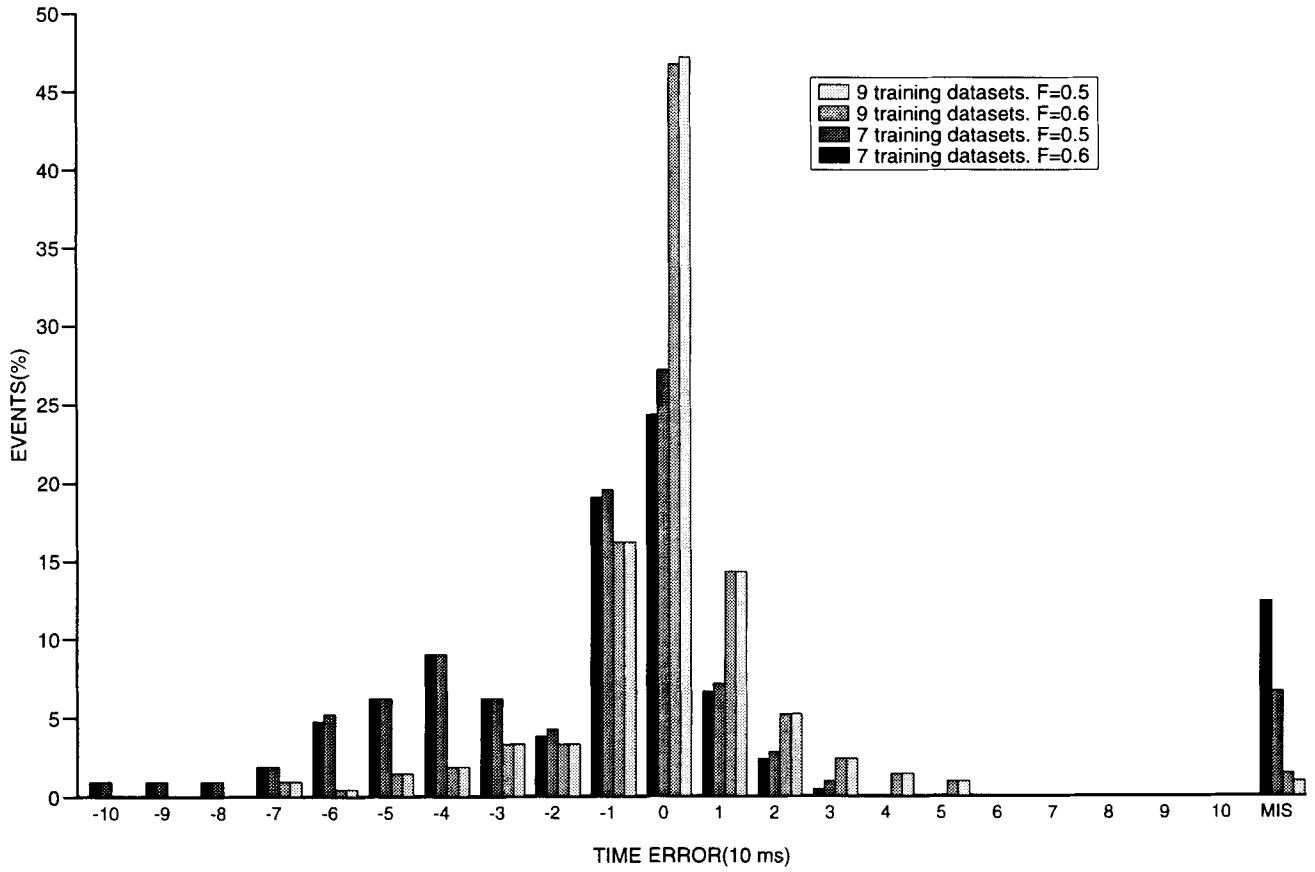

Figure 9. Notation similar to Fig. 8 but for *S*-wave picks.

eight *P* waves and three *S* waves are missed completely. It is interesting to note that only two extra *P* waves are detected but 23 more *S* waves, with a commensurate increase in the picking ability. The *S*-wave picking ability is improved as the additional *P*-wave segments actually resemble many of the *S* waves. Again, decreasing the $F(t)$ threshold to 0.5 does not significantly improve the picking, with only one more *P* wave and one more *S* wave picked. This comparison substantiates the well-known adaptive behaviour of neural networks, that improvement can always be achieved by judicious choice of the training data sets.

3.7 Sensitivity to segment length

The time taken during analysis depends on the neural network structure. The structure can be decreased by decreasing the nodes in the input layer and in the hidden layer. A reduction in the input nodes was tested by reducing from 30 to 20 nodes, keeping the hidden nodes and output nodes the same. We use the nine pairs of *P*-wave and noise segments to train this network. In this case, the training procedure is slower, and it took 2018 iterations to reach a satisfactory convergence point. The results are now worse than before, although the number of unpicked *P* waves remains the same (eight sets of records), with the number of unpicked *S* waves increasing from 3 to 14 events. The onset estimation is worse, with only 57.6 per cent of the *P* waves and 55.2 per cent of the *S* waves having estimates with errors $\leq$ one sample increment. There is also a larger number of spurious picks. Figs 10 and 11 summarize the comparison for this case and the original. The input nodes are also increased to 40 for comparison, with the network now taking 362 iterations to converge. Now, only 67.8 per cent of *P* waves and 63.8 per cent of *S* waves are picked with an error $\leq$ one sample increment (Figs 10 and 11). There is a larger number of unpicked arrivals.

We should point out that the input segment length depends on the characteristics of the signals. It is suggested that this segment should include several complete cycles of a wavelet. It appears that reducing or increasing the number of input nodes dramatically affects the performance and there appears to be an optimum number of input nodes for our particular configuration. This reflects the general observation that network architecture must be specifically tailored to individual applications. This represents the 'Achilles heel' of ANN applications, and further optimization is required to adapt to particular event types.

4 APPLICATION TO COMPLETE DATA SET WITH A RANGE OF SEISMOGRAM QUALITY

4.1 Data character and adaptation of ANN processing

Here, we test our neural network further by incorporating additional recordings from stations DP and AY, which now form the complete data set of 1754 three-component sets

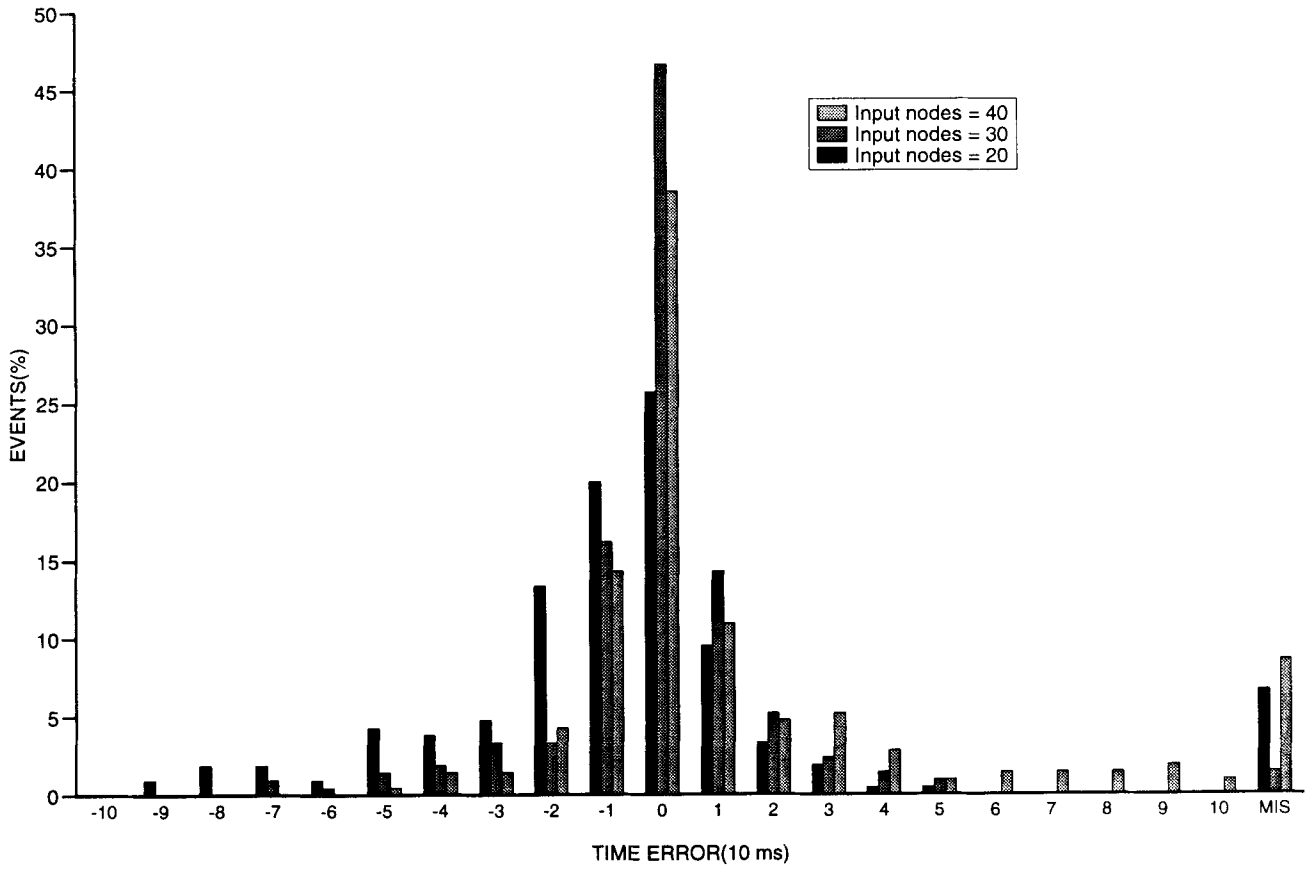

Figure 10. Notation similar to Fig. 8. Comparison of *P*-wave picks for three neural networks with 20, 30 and 40 input nodes.

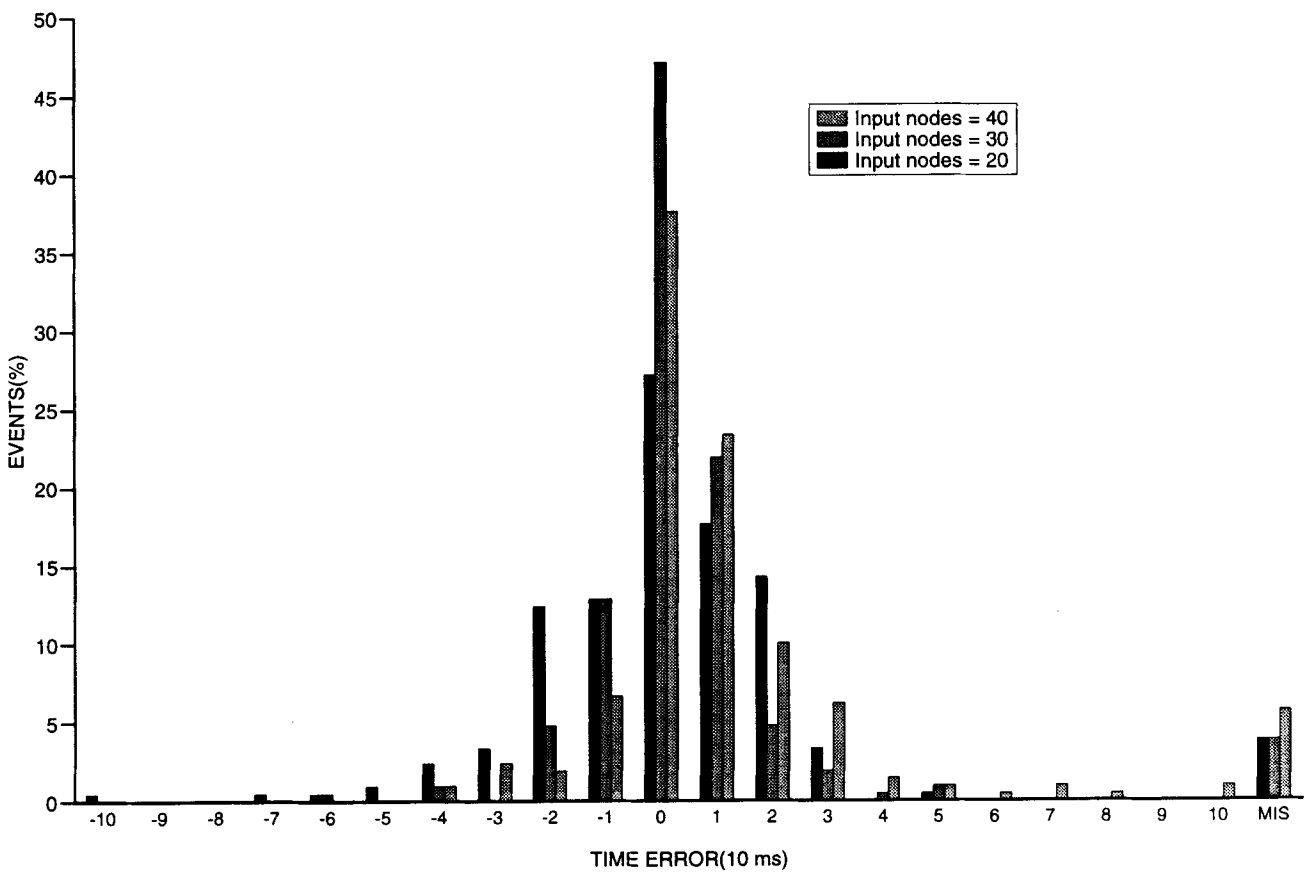

Figure 11. Notation similar to Fig. 8. Comparison of *S*-wave picks for three neural networks with 20, 30 and 40 input nodes.

(877 from DP and 877 from AY), from which only the high-quality data were selected in the previous section. In many of these cases the seismometers did not function properly, and either one or two components were inactive or possessed high-amplitude noise so that some of the three-component sets were incomplete. For the recordings in which one or two components do not record seismic signals, the ANN described above can still use $M(t)$ and analyse them with significant error. The ANN used to accomplish this task has a similar design to the one described above, but uses different threshold parameters for arrival detection. There are also 159 recordings from station DP and 60 recordings from station AY with excessive noise preceding the event or ringing throughout the record. They produce false alarms, and necessitate an additional component to the system to act as a quantity-control procedure. This may be achieved by the use of neural networks, and such a trace editing system is not uncommon; for example, McCormack, Zaucha & Dushek (1993) designed an ANN to detect noisy and dead traces in raw surface seismic field data. Here we select the recordings manually, leaving the design of this procedure to a later investigation. Consequently, we must be aware that our statistics will appear more successful than if this procedure had been applied to all the data irrespective of quality. We find that 325 recordings from station DP and 198 recordings from station AY are unusable, leaving 373 and 504 recordings respectively for further processing. These data are reduced further as manual picks for comparison with the ANN results are only possible for 360 *P* waves and 341 *S* waves at station DP and 342 *P* waves and 320 *S* waves at station AY.

The ANN used in this more extensive test is trained by nine pairs of *P* waves and noise segments, and again has 30 input nodes. With an $F(t)$ threshold of 0.6, the ANN can detect 348 *P* waves (96.7 per cent) and 317 *S* waves (93.0 per cent) from Station DP, and 311 *P* waves (90.9 per cent) and 280 *S* waves (87.5 per cent) from station AY. Most of the failures arise at the low SNRs of between 1 and 3. The method appears to pick all phases with an SNR > 3. Figs 12 and 13 show the estimation results. For station AY, 234 (68.6 per cent) *P* waves and 207 (64.7 per cent) *S* waves have onset times within one sample increment (10 ms) of the expected manual values. For station DP, 295 (81.9 per cent) *P* waves and 234 (68.4 per cent) *S* waves have onset times within the same tolerance. In addition, only 7.7 per cent of *P* waves and 11.8 per cent of *S* waves have onset times with errors greater than five sample increments or are missed entirely. Once picked, the SNR does not affect the accuracy of the estimate. In the situation where not all *S* waves picks are manually possible, a valid comparison with the ANN results cannot be made and they are excluded from our statistics. This produces a small bias in our results as it only affects 10 seismogram sets.

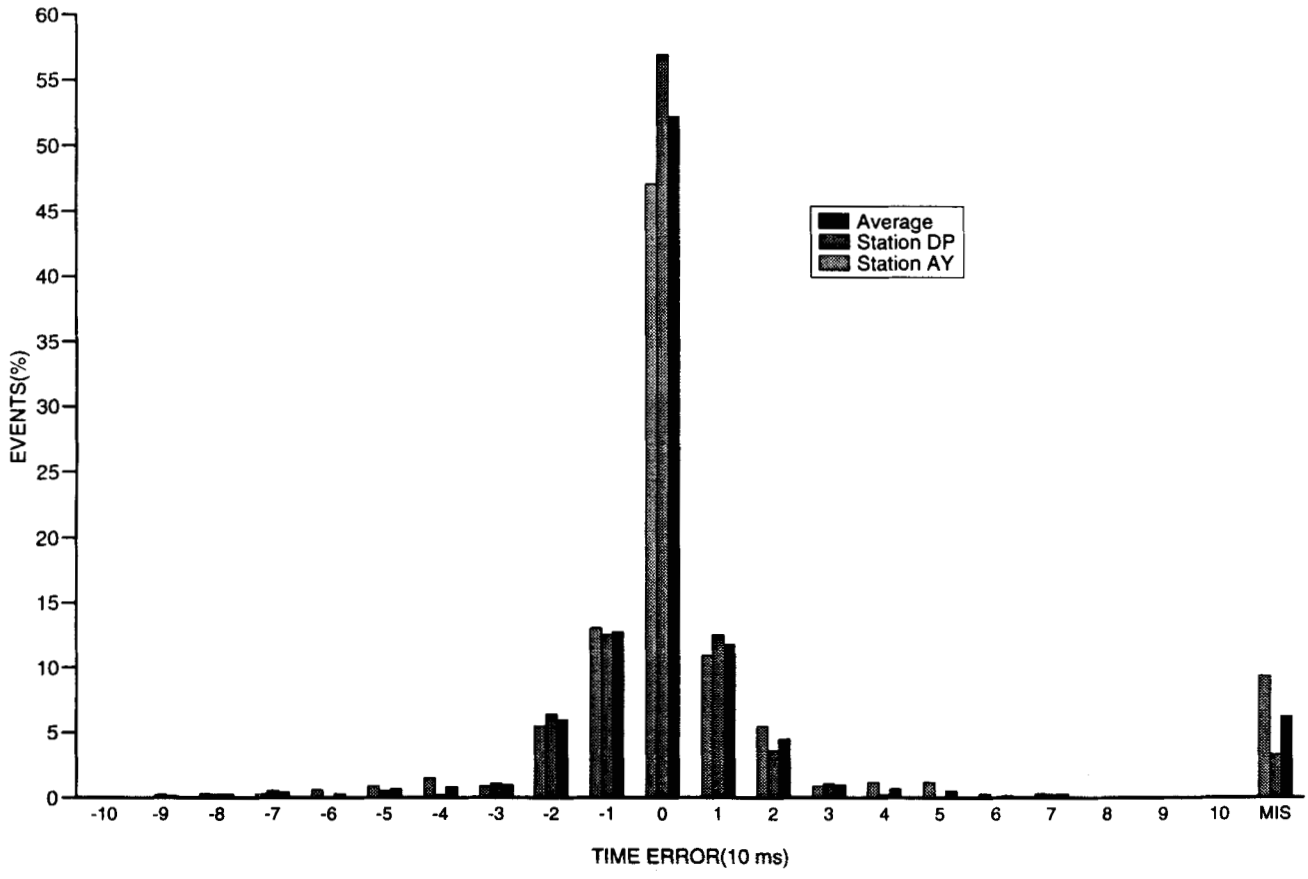

Figure 12. Statistics for *P*-wave picks on the complete data set of 877 local event recordings, with notation similar to Fig. 8. The success of the ANN relative to manual reference picks is quoted as a percentage.

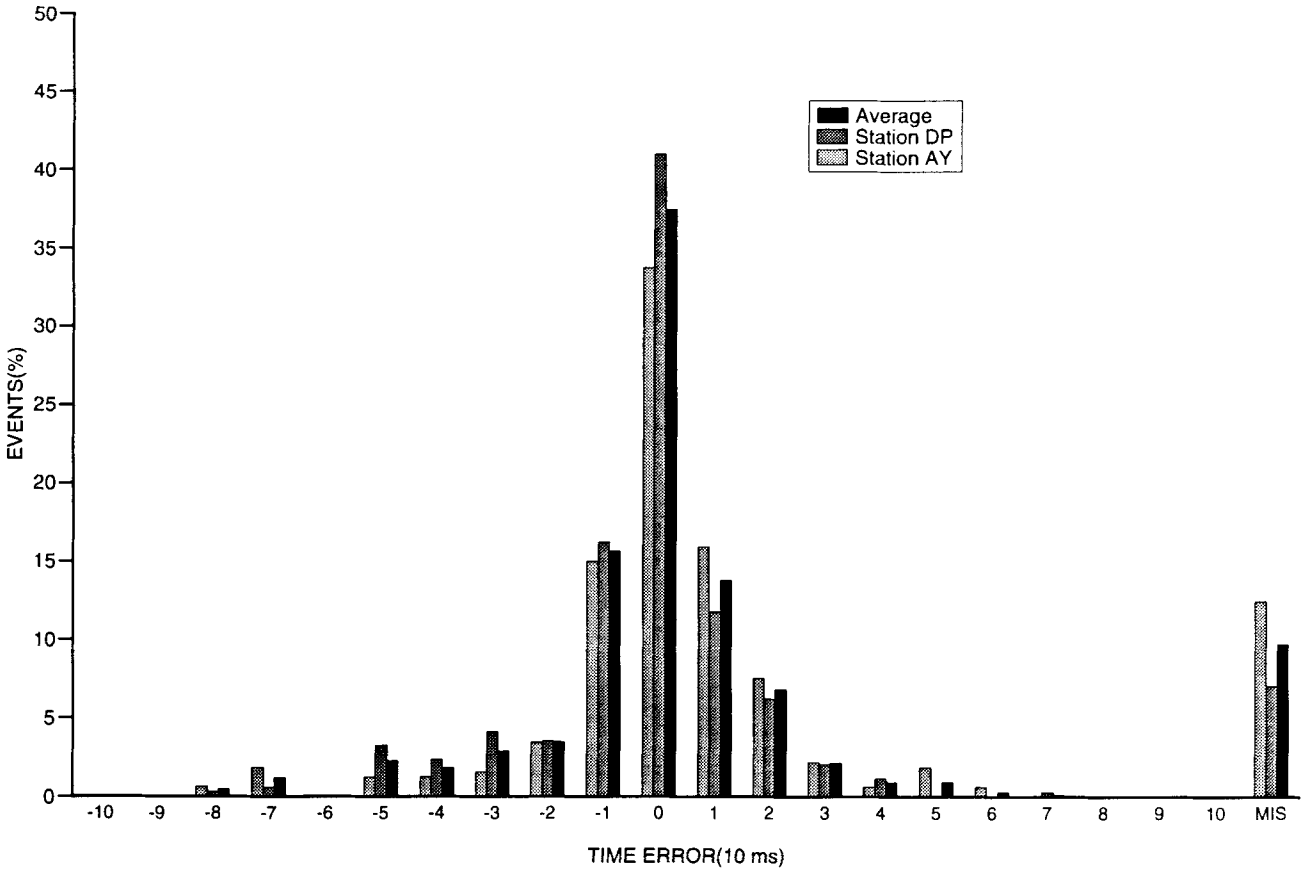

Figure 13. As in Fig. 12 but for *S* waves.

4.2 Sensitivity to signal-to-noise ratio

One additional benefit in using a trained neural network is its ability to pick waves in low SNR conditions with only a high SNR training data set. Fig. 14 shows an example in which $F(t)$ has two peaks, indicating a high SNR seismogram set. The peaks are rectangular in shape, with a width of roughly 30 sample points. Fig. 15 shows an example in low SNR, for which the *P*-wave peak appears sharper. Although the three-component particle motion does not display a significant difference between the signal and the noise for this case, $M(t)$ does indicate the change in the nature of the signal, and this translates to the narrow peak in $F(t)$ which now indicates the onset. This reveals the possibility of interpreting the shape of these peaks to extend the ability of the trained neural network beyond the boundaries of the training set. Fig. 16 shows a low SNR example in which the *S* wave (SNR = 1.8) can still be automatically picked by the ANN.

5 DISCUSSION AND CONCLUSIONS

5.1 ANN performance

A multilayered neural network is used as a tool to pick *P* and *S* waves from local earthquake data. The neural network input is the vector modulus of each set of three-component records. The results are encouraging, and demonstrate that a neural network trained using a small subset of the data (only nine *P* waves and commensurate noise segments in this case) can detect 93.9 per cent of the *P* waves and 90.3 per cent of the *S* waves. Using this to further pick the onset times, we find a success rate of 75.8 per cent (with $F > 0.6$) for the *P* waves and 66.7 per cent (with $F > 0.6$) for the *S* waves with an error of one sample increment (10 ms). Although false arrivals and spikes can be discarded by using pre-processing steps, here we chose to include these in the selection, to be discriminated at a later stage. However, a pre-processing stage is required to discard excessively noisy and unusable recordings. They are not included in the statistics which we quote. For the data set in Section 4, 17.4 per cent of the detections include false alarms. Inspection of the seismograms revealed that most of the false arrivals are similar to the *P* or *S* wavelets, and in fact it would not be possible to distinguish them visually if only one segment were available. Additional information is required for this task.

These results, combined with the advantage of not requiring programs to construct special variables and parameters with complicated mathematics, suggest that the ANN is a natural choice for such applications. All that is necessary is to select suitable example arrivals in a training set. The method is adaptive, and training sets can be altered to enhance particular features of different data sets. Adding new training data sets and retraining the network is easy and quick, and can improve the performance of the network. Although the training time can be long, especially as the network architecture becomes involved, once trained the

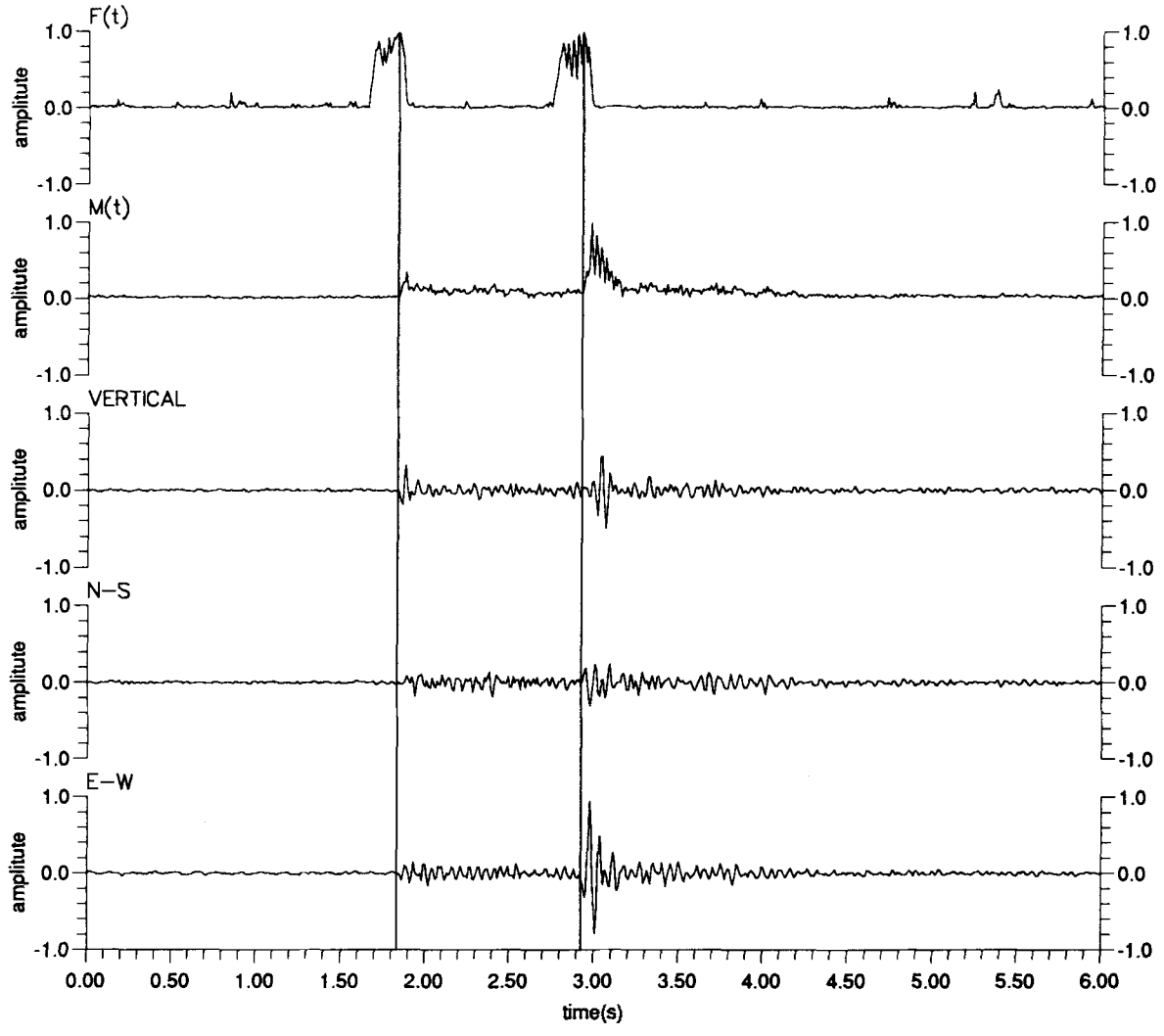

Figure 14. Example of picking *P* and *S* waves in a high signal-to-noise ratio. Three-component vector modulus $M(t)$ and $F(t)$ function computed from output of the neural network trained with nine pairs of *P* waves and noises. Compared with Figs 6 and 7, there are no precursors before the main peaks in $F(t)$ corresponding to the *P* and *S* arrivals. The peaks have a base length equal to the ANN input segment (290 ms).

network is sufficiently quick to operate in most real-time applications. However, the network cannot be viewed as all encompassing, as the performance still depends upon the training set and its ability to predict cannot lie too far outside its experience. The exact boundaries of this behaviour have not yet been completely explored. Another limitation is in finding an optimal architecture for a particular application. This is not yet fully understood for multilayered back-propagation networks, although there are certain other network designs where this is possible (Falman & Lebiere 1990; Kusuma & Brown 1992). The current network design is limited to three-component recordings, but this method has now also been adapted for single-component recordings.

5.2 Comparison with other picking algorithms

As discussed earlier, there are many picking algorithms already in use on many seismic networks world-wide. Table 1 gives a comparison of the performance of our technique for local earthquake data and the performance of a few selected techniques in common use. Because articles tend to

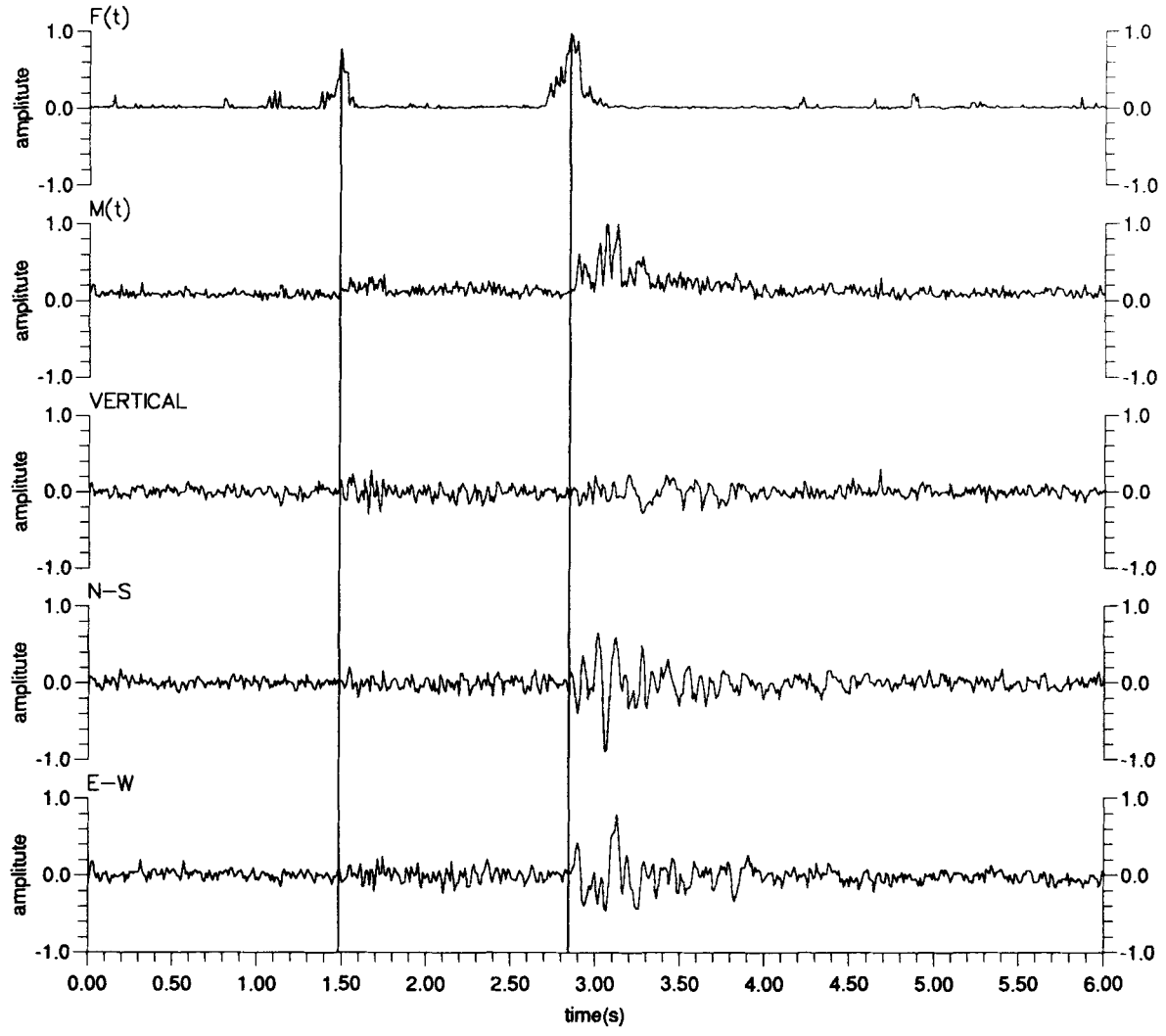

Figure 15. Example of picking P wave in a low signal-to-noise ratio. Three-component vector modulus $M(t)$ and $F(t)$ function computed from output of the neural network trained with nine pairs of P waves and noises. Although the three-component particle motion does not display a significant difference between the P wave and the noise, $M(t)$ does indicate the change in the nature of signal, and this translates to a narrow peak in $F(t)$ which indicates the P onset.

describe principles and show a few examples these cannot be directly or wholly compared with our result which is applied to a specific data set of local events, so that this table may not be truly representative of the optimal forms of each technique. As false alarms were not fully treated in our algorithm, we do not suggest without further tests and development that our method is better. However, it does appear that the small estimation error for both P-wave and S-wave analysis is potentially encouraging for future work. We believe that an additional strength of the neural network is that it can deal with raw data once it has been trained appropriately. This contrasts with many other techniques, which rely upon pre-processing steps to generate control parameters. The network presented here is relatively quick to train and has been shown to be adaptive to various types of waves. To demonstrate this feature fully, and avoid data-specific results, further tests are to be performed on the network, to extend application to other three-component data sets and single-component data sets. We suggest that it may be possible to process regional or

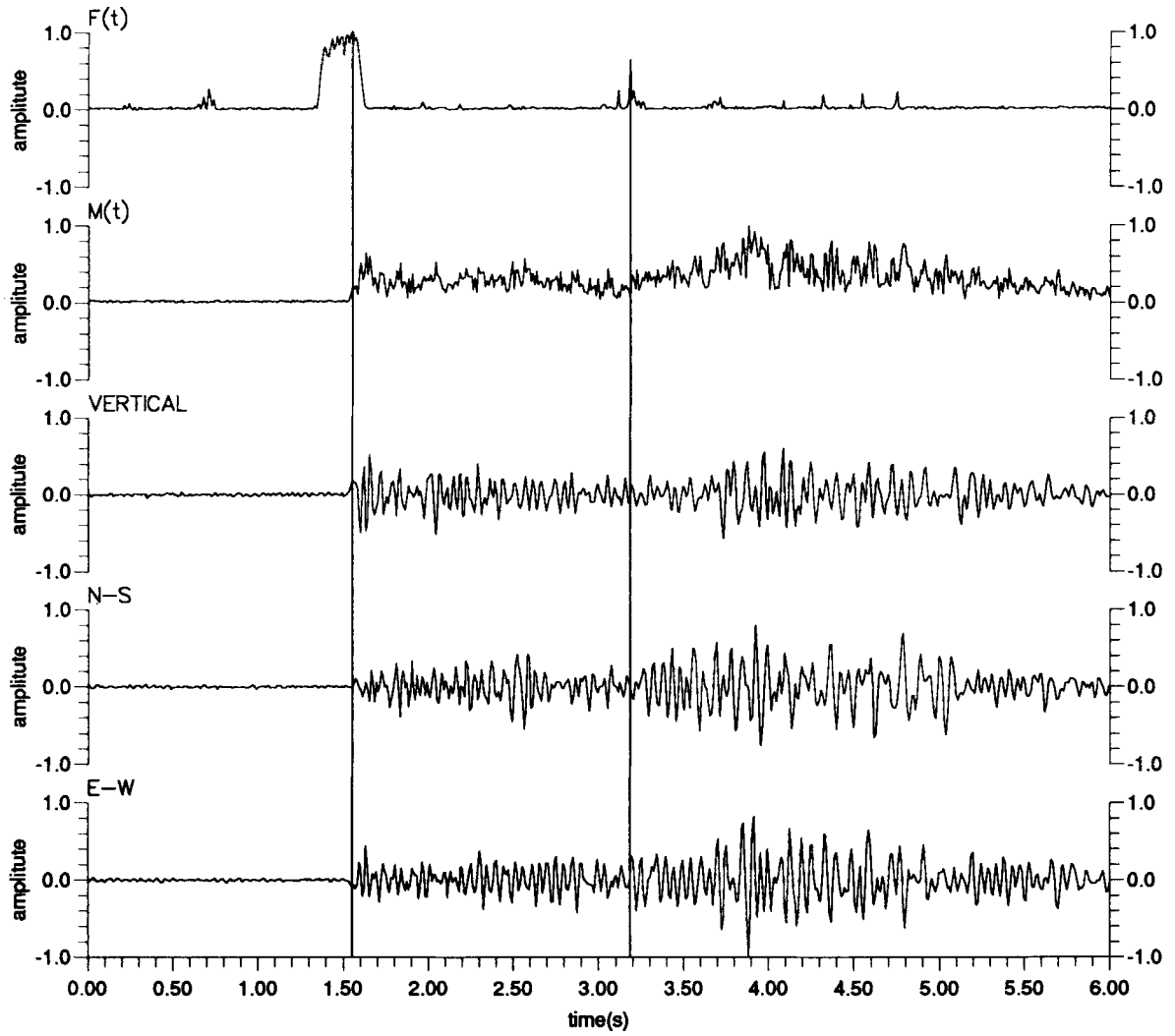

Figure 16. Example of picking *S* wave in a low signal-to-noise ratio. Three-component vector modulus $M(t)$ and $F(t)$ function computed from output of the neural network trained with nine pairs of *P* waves and noises. Due to *P* wave coda it is difficult to indicate the *S* arrival where the SNR is only 1.8. The peak in $F(t)$ is sharper and its maximum is only 0.64.

teleseismic observations also using this approach. In these cases, however, the network architecture may need to be adjusted to suit the behaviour of these data.

5.3 Future work

This work forms part of an ongoing programme of research to develop a fully automatic subsystem for earthquake analysis, for which picking the seismic arrivals is the key procedure. In this analysis we make a concerted effort to develop an approach which avoids any preconceived notions regarding the polarization properties of the individual arrivals in the earthquake record. This is important as we recognize that a parametric model may not be generally applicable to local and teleseismic events in a heterogeneous crust where the arrivals may be composed of several distinct signal processes (Der *et al.* 1993). Consequently, in future schemes we intend to direct our current approach towards the task of arrival identification using generic polarization characteristics, so that more complex wave trains such as *Sn* and *Lg* may be identified and analysed. Ultimately we hope to integrate other ANN units into a processing flow for record editing and event classification and mechanism determination.

Table 1. A summary comparison of selected picking methods.

Author	Input Data	Method	Wave Type	Picking Result	Time Error
Allen (1978)	single trace	STA/LTA	P	60-80%	≤0.05 sec
Bear & Kradolfer (1987)	single trace	modified STA/LTA	P	Local: 65.9% Region:79.5% Tele:90%	≤1 sample ≤3 sample
Joswig & Schulte-Theis (1993)	single trace	Master-event correlation	P	80 % for weak events	≤1 sample
Klumpen & Joswig (1993)	3-component recording	generic polarizatio n	P & S	67% for P & S	≤50ms
Kracke (1993)	modulus of 3-component recording	LTA/ threshold	P	96.5%	not mentioned
Dai & MacBeth (1994, this paper)	modulus of 3-component recording	neural network	P & S	94% for P 90% for S 75% for P 67% for S	≤100ms ≤100ms ≤10ms ≤10ms

ACKNOWLEDGMENTS

This research was sponsored by the Global Seismology Research Group (GSRG) at the British Geological Survey (BGS), Edinburgh, and is published with the approval of the Director of the BGS. We thank Chris Browitt, Programmes Director of BGS, David Booth and John Lovell for supplying the earthquake data. We also thank Yun Liu for assisting with the manual picking. Thanks are also extended to the staff and students of GSRG for their support and encouragement with this work.

REFERENCES

Allen, R.V., 1978. Automatic earthquake recognition and timing from single trace, *Bull. seism. Soc. Am.*, **68,** 1521–1532.

Bache, T.C., Bratt, S.R., Wang, J., Fung, R.M., Kobryn, C. & Given, J.W., 1990. The intelligent monitoring system, *Bull. seism. Soc. Am.*, **80,** 1833–1851.

Bear, M. & Kradolfer, U., 1987. An automatic phase picker for local and teleseismic event, *Bull. seism. Soc. Am.*, **77,** 1437–1445.

Chiaruttini, C. & Salemi, G., 1993. Artificial intelligence techniques in the analysis of digital seismograms, *Comput. Geosci.*, **19,** 149–156.

Chiaruttini, C., Roberto, V. & Saitta, F., 1989. Artificial intelligence techniques in seismic signal interpretation, *Geophys. J. Int.*, **98,** 223–232.

Cichowicz, A., 1993. An automatic S-phase picker, *Bull. seism. Soc. Am.*, **83,** 180–189.

Crampin, S., Evans, R. & Ucer, S.B., 1985. Analysis of records of local earthquakes: the Turkish Dilatancy Projects (TDP1 and TDP2), *Geophys. J. R. astr. Soc.*, **83,** 1–16.

Der, Z.A., Baumgardt, D.A. & Shumway, R.H., 1993. The nature of particle motion in regional seismograms and its utilization for phase identification, *Geophys. J. Int.*, **115,** 1012–1024.

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

Dystart, P.S. & Pulli, J.J., 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.

Evans, S.A., Beamish, D., Crampin, S. & Ucer, S.B., 1987. The Turkish Dilatancy Project (TDP3): multidisciplinary studies of a potential earthquake source region, *Geophys. J. R. astr. Soc.*, **91,** 265–286.

Fahlman, S.E. & Lebiere, C., 1990. The cascade-correlation learning architecture, in *Advances in Neural Information Processing Systems 2*, ed. Touretzky, D.S., Morgan Kauffmann.

Joswig, M., 1990. Pattern recognition for earthquake detection, *Bull. seism. Soc. Am.*, **80,** 170–186.

Joswig, M. & Schulte-Theis, H., 1993. Master-event correlations of weak local earthquake by dynamic waveform match, *Geophys. J. Int.*, **113,** 562–574.

Klumpen, E. & Joswig, M., 1993. Automated reevaluation of local earthquake data by application of generic polarization patterns for P- and S-onsets, *Comput. Geosci.*, **19,** 223–231.

Kracke, D., 1993. A three-component event detector based on waveform analysis, *Comput. Geosci.*, **19,** 117–122.

Kusuma, T. & Brown, M., 1992. Cascade-correlation learning architecture for first-break picking and automated trace editing, in *Proceedings of SEG 62nd Annual International Meeting*, pp. 10–13, Society of Exploration Geophysics.

Lomax, A.J. & Michelini, A., 1988. The use of spherical coordinates in the interpretation of seismograms, *Geophys. J.*, **93,** 405–412.

Lovell, J., 1989. Source parameters of microearthquake swarm in Turkey, *MPhil thesis*, University of Edinburgh, Edinburgh.

McCormack, M.D., 1991. Neural computing in geophysics, *Geophysics: The Leading Edge of Exploration*, **10,** 11–15, January 1991.

McCormack, M.D., Zaucha, D.E. & Dushek, D.W., 1993. First-break refraction event picking and seismic data trace editing using neural network, *Geophysics*, **58,** 67–78.

Murat, M. & Rudman, A., 1992. Automated first arrival picking: A neural network approach, *Geophys. Prospect.*, **40,** 587–604.

Pao, Y.H., 1988. *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, New York.

Pisarenko, V.F., Kushnir, A.F. & Savin, I.V., 1987. Statistical adaptive algorithms for estimation of onset moments of seismic phases, *Phys. Earth Planet. Inter.*, **47,** 4–10.

Poulton, M.M., Sternberg, B.K. & Glass, C.E., 1992. Location of subsurface target in geophysical data using neural networks, *Geophysics*, **57,** 1534–1544.

Roberts, R.G., Christoffersson, A. & Cassidy, F., 1989. Real-time event detection, phase identification and source location estimation using single station three-component seismic data, *Geophys. J.*, **97,** 471–480.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J., 1986. Learning internal representation by backpropagating errors, *Nature*, **332,** 533–536.

Takanami, T. & Kitagawa, G., 1988. A new efficient procedure for the estimation of onset times of seismic waves, *J. Phys. Earth*, **36,** 267–290.

Takanami, T. & Kitagawa, G., 1993. Multivariate time-series model to estimate the arrival times of S-waves, *Comput. Geosci.*, **19,** 295–301.

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