

**Application of Back-propagation Neural Networks to Identification of Seismic Arrival Types**

Hengchang Dai<sup>1,2</sup> and Colin MacBeth<sup>1</sup>

<sup>1</sup> *British Geological Survey, Murchison House, West Mains Road, Edinburgh EH9 3LA Scotland, UK.*

<sup>2</sup> *Department of Geology and Geophysics, University of Edinburgh, Edinburgh EH9 3JW, Scotland, UK.*

**Intended for publication in:** *Physics of the Earth and Planetary Interiors*

**Page heading:** *Arrival identification using BPNN*

**Address for correspondence**

Dr. Hengchang Dai (before 1 March)

Dr. Colin MacBeth (after 1March)

British Geological Survey

Murchison House

West Mains Road

Edinburgh EH9 3LA

Scotland, U.K.

## Application of Back-propagation Neural Networks to Identification of Seismic Arrival Types

Hengchang Dai<sup>1,2</sup> and Colin MacBeth<sup>1</sup>

<sup>1</sup> *British Geological Survey, Murchison House, West Mains Road, Edinburgh EH9 3LA Scotland, UK.*

<sup>2</sup> *Department of Geology and Geophysics, University of Edinburgh, Edinburgh EH9 3JW, Scotland, UK.*

### Abstract

A back-propagation neural network approach is developed to identify *P*- and *S*-arrivals from three-component recordings of local earthquake data. The BPNN is trained by selecting trace segments of *P*- and *S*-waves and noise bursts converted into an attribute space based on the degree of polarization (DOP). After training, the network can automatically identify the type of arrival on earthquake recordings. Compared with manual analysis, a BPNN trained with nine groups of DOP segments can correctly identify 82.3% of the *P*-arrivals and 62.6% of the *S*-arrivals from one seismic station, and when trained with five groups from a training dataset selected from another seismic station, it can correctly identify 76.6% of the *P*-arrivals and 60.5% of *S*-arrivals. This approach is adaptive and needs only the onset time of arrivals as input, although its performance cannot be improved by simply adding more training dataset due to the complexity of DOP patterns. Our experience suggests that other information or another network may be necessary to improve its performance.

## 1. Introduction

The most important procedure for analysing local earthquake events is the estimation of the arrival times of the primary ( $P$ ) and secondary ( $S$ ) waves, as these measurements form the basis of subsequent analysis schemes employing processing for event location, event identification, source mechanism analysis and spectral analysis. These tasks are often performed by a trained analyst who manually picks arrival times according to his individual experience, involving an intensive amount of pattern recognition. With the increase in the number of digital seismic networks being established worldwide, there is a pressing need to provide an automatic alternative, which is more reliable, robust, objective and less time-consuming.

The estimation of arrival onset times involves reliable and accurate estimation in addition to the identification of individual arrivals from their polarization, amplitude and propagation characteristics. A great deal of effort, stretching back several decades, has been devoted to the automation of arrival picking [see Dai and MacBeth (1995) for an extensive list]. However, for the purposes of automation, identifying arrival types is more difficult than picking their onset times. In some cases, identification based on horizontal velocity from a  $f-k$  filter can provide a major simplification of the interpretation task (Mykkeltveit and Bungum, 1984; Bache *et al.*, 1990; Kvaerna and Ringdal, 1992). Der, Baumgardt and Shumway (1993) have also investigated the feasibility of adaptive and automatic recognition of regional arrivals by a wavefield extrapolation scheme for data from a mini-array. However, for single station data, only a few methods can be used to pick special types of arrivals. Roberts, Christoffersson and Cassidy (1989), based on the auto- and cross-correlations of the three orthogonal components within a short time window, detect the arrival of a  $P$ -wave or a linearly polarized  $S$ -wave. Cichowicz (1993) developed an  $S$ -phase picker which depends on a well-defined pulse for the first-arrival  $P$ -wave. Tong (1995) developed an arrival separator based on certain "features" extracted from intelligent segmentation of the seismograms. Tong and Kennett (1995) developed an approach to identify late arrivals by analysing the energy content of seismic traces. There is no general method to identify the  $P$ - and  $S$ -arrival simultaneously, and automation is still an unresolved issue. The intensive amount of pattern recognition

involved suggests that a natural choice should be artificial neural networks. In this paper, we use the back-propagation neural network (BPNN) as it has been employed successfully in the past (Rumelhart, Hinton, and Williams, 1986; Pao, 1989; Haykin, 1994; Fausett, 1994; Dai and MacBeth, 1995).

## 2. Degree of polarization

One of the obvious features of distinguishing  $P$ - and  $S$ -arrivals is their polarization directions: the polarization direction of a  $P$ -arrival is parallel to its propagation direction, and the polarization direction of an  $S$ -arrival is perpendicular to its propagation direction in an isotropic medium. It appears a simple task to identify the  $P$ - and  $S$ -arrivals with the 3-C recordings now available. However, it is not practical to directly use the polarization direction of an arrival as it is related to the event location which is not available prior to analysis. Another feature of distinguishing the  $P$ - and  $S$ -arrivals is their polarization state. In general, it is observed that the direct  $P$ -arrival is predominantly linearly polarized whilst the following arrivals, such as  $S$ -arrivals, have considerably more complicated polarization patterns involving phase shifts (Basham and Ellis, 1969; Roberts and Christoffersson, 1990). The polarization state can be measured by using the *degree of polarization*,  $F(t)$ , (appendix A) which is independent of the source-receiver direction. For a linearly polarized wave,  $F(t)$  is unity, and for a completely unpolarized or circularly polarized wave it is zero (Cichowicz, Green and Brink, 1988). Cichowicz (1993) pointed out that both  $P$ - and  $S$ -wave arrivals exhibit a high degree of linear polarization, but the  $P$ -wave coda manifests a generally elliptical polarization with a significantly lower value of  $F(t)$ . For real data, although the  $S$ -arrival is usually associated with a far larger value of  $F(t)$  than that of the  $P$ -wave coda, it does not reach a value of 1.0. The variation of this quantity along the seismogram forms the pattern which may indicate the wave type. Figure 1 shows a typical example in which  $F(t)$  is of high amplitude for a  $P$ -arrival, a median level value for a  $S$ -wave, and then a low level pattern for a noise burst.

We have investigated the  $F(t)$  patterns for data from two seismograph stations DP and AY from the TDP3 network (Lovell, 1989, Dai and MacBeth, 1995). Usually, most seismograms conform to our expectations from Figure 1, with the  $P$ -arrivals having well-defined linear polarization patterns.

Although most  $P$ -arrivals differ from  $S$ -arrivals and noise bursts, some noise bursts were found similar to the seismic arrivals. Comparing the data from stations DP and AY, the  $F(t)$  patterns are different even for the same arrivals from the same earthquake source (Figure 2).

It is important to note that this particular definition of  $F(t)$  does not consider the signal amplitude. To consider the amplitude information, we define a modified function:

$$MF(t) = F(t) / \overline{M}(t) \quad (1)$$

where  $\overline{M}(t)$  is a smoothed relative function of the modulus  $M(t)$  of the 3-C recording within a window. This value is also independent of the source-receive direction. The normalization factor is calculated using the window between the onset point and the following ten points. Note that  $MF(t)$  and  $F(t)$  have slightly different patterns (Figure 1). The  $MF(t)$  patterns are complex and identifying them requires an intensive amount of pattern recognition for which the BPNN is suitable. The  $MF(t)$  is presented to the BPNN in a segment selected from a window in which the arrival's onset-time is at the centre.

### 3. Approach

Figure 3 shows a flow chart for identifying arrival types using a BPNN. This approach is used as the second stage following the arrival picking. In this approach, unlike the BPNN picking approach which uses a BPNN as a filter to deal with an entire seismic trace (Dai and MacBeth, 1995), only arrival segments are input into the BPNN. An arrival segment is selected by its onset time which is obtained by using the other procedure. This  $MF(t)$  segment is then fed into a BPNN for training and testing.

This arrival segment is selected by an onset time which has been picked previously and it is then positioned at the centre of the segment. Due to picking errors, the onset time may not be exactly at the centre of the segment and it is found that this may affect the BPNN output. In order to avoid this effect, an adjustment of the onset time is necessary to ensure that the performance of the trained BPNN is not affected. For each  $MF(t)$  segment, the first local maximum after the onset-time is set at the centre of the segment. The  $MF(t)$  segment has 60 samples (590ms) length which are chosen to include the complete  $MF(t)$  pattern of an arrival.

The BPNN used in this approach has three layers. Its input layer has 60 nodes, giving a  $MF(t)$  segment with a fixed length of 60-samples. There are three nodes in its output layer to flag the result: output (1,0,0) for a noise burst, (0,1,0) for a  $P$ -arrival, and (0,0,1) for an  $S$ -arrival in training. Ten hidden nodes are chosen after a process of trial and error with different training runs.

In this procedure, only a small number of recordings from station DP are used to train the BPNN by using the generalized delta rule (Rumelhart, Hinton and Williams, 1986). As the BPNN performance depends on the training dataset, selecting this training dataset is crucial. A BPNN trained with incorrect or inconsistent data cannot be expected to give a correct answer for new data.  $P$ - and  $S$ -waves with similar  $MF(t)$  patterns should be avoided in the training dataset because the BPNN cannot distinguish between them and the training procedure might not converge. At the start of training, only three  $MF(t)$  segments are selected, representing each of the three categories: noise burst,  $P$ -arrival and  $S$ -arrival, with the desired output (1,0,0; 0,1,0; 0,0,1) respectively. After training, this BPNN is used to handle other data. Using manual analysis results, another three segments incorrectly identified by this BPNN are combined with the former training segments, to re-train this BPNN. This procedure is repeated until the performance of the trained BPNN cannot be improved by further additions to the training dataset. Figure 4 shows all training segments of  $MF(t)$  used in the present work. With the training dataset selected and by using the generalized delta rule (Rumelhart, Hinton, and Williams, 1986; Pao, 1989), the final procedure with nine groups from station DP took less than two minutes CPU time on a VAX4000.

As each  $MF(t)$  segment is fed into the trained BPNN, the BPNN outputs three values:  $o_1$ ,  $o_2$ , and  $o_3$ . For the training segments, the output should be the desired ones: (1,0,0) for a noise burst, (0,1,0) for a  $P$ -arrival, and (0,0,1) for a  $S$ -arrival. For non-training segments, the output ( $o_1$ ,  $o_2$ , and  $o_3$ ) is a measurement of similarity between a new segment and the group characteristics defined by the training segments. If a non-training segment is similar to a training segment, the BPNN output will close to a desired output of the training segment. In order to identify segment types, we only need to seek the maximum of the three outputs ( $o_1$ ,  $o_2$ ,  $o_3$ ). Normally, the trained BPNN should have a high output and two low outputs in its three output nodes. However, sometimes it has two or three high outputs or three low outputs. This means these segments are quite different from training segments. In these cases, we

still use the maximum of the three outputs to identify their wave types.

## 4. Performance

### 4.1. Station DP

For the data from station DP, 345 *P*-arrivals, 302 *S*-arrivals and 174 noise bursts are automatically picked by our previous BPNN arrival picker from 371 recordings. A BPNN for arrival identification is finally trained with nine groups of  $MF(t)$  segments from noise bursts, *P*- and *S*-arrivals at station DP (Figure 4). Figure 5 shows a successful example of this trained BPNN in application. Table 1(a) displays the overall performance of this trained BPNN. This BPNN has a better performance in identifying the *P*-arrivals (82.3%) than for *S*-arrivals (62.6%) and noise bursts(47.7%). Excessive noise contaminated 17 recordings. In these recordings, noise bursts are very similar to the coherent seismic arrivals and their  $MF(t)$  patterns are similar to the patterns of *P*- or *S*-arrivals, therefore the BPNN cannot be expected to identify them correctly. If such recordings are omitted, there are only 118 noise bursts picked. The BPNN then classifies 66 (56.0%) of them as noise bursts, 11(9.3%) as *P*-arrivals, and 41(34.5%) as *S*-arrivals.

### 4.2. Station AY

For the data from station AY, 274 *P*-arrivals, 240 *S*-arrivals and 28 noise bursts are automatically picked by our previous BPNN arrival picker from 391 recordings. The above BPNN is applied to these arrivals. Table 1(b) displays its overall performance. Unfortunately it only correctly identifies 42% of the *P*-arrivals, 38.8% of the *S*-arrivals and 32.1% of the noise bursts. This is due to the  $MF(t)$  patterns of seismic arrivals being different between the data from stations DP and AY, even for the same arrivals from the same event (Figure 2). Such differences make the BPNN fail to identify these correctly, as the output of the trained BPNN is the measurement of the similarity between the segment to be tested and the trained segments. For example, some recordings have *P*-arrivals with low to mid-level

values of  $MF(t)$ , and the  $S$ -arrivals with high level values of  $MF(t)$  pattern, which are quite opposite to the trained waves. To deal with the data from station AY, it is necessary to train this BPNN by using the data from station AY, so that it can remember the  $MF(t)$  patterns from station AY and use this to classify the new data correctly.

The best performance is obtained when the BPNN is trained for station AY by using the five groups shown on Figure 7. The training procedure took less than one minute time on a VAX4000 computer and Table 2(a) displayed its overall performance. This BPNN now has clearly a much better performance than previous one. It successes 76.6% for the  $P$ -arrivals and 60.4% for the  $S$ -arrivals. This BPNN is also used to test the data from station DP (Table 2(b)). Although it can identify 84.9% of the  $P$ -arrivals, the 36.1% success rate for the  $S$ -arrivals is too low. Note that the identification performance for the noise bursts may be statistically insignificant as only 28 noise bursts are tested.

#### **4.3. Effect of changing training dataset**

As the performance of a trained BPNN clearly depends on the training dataset, an investigation is of the sensitivity to the training dataset from Station DP. Table 3 shows a comparison of three BPNNs trained with different datasets. As the number of segments in the dataset increases,  $P$ -arrival identification improves, but the performance for  $S$ -arrival and noise burst identification becomes worse. This appears consistent with the observation that the  $MF(t)$  patterns for the  $P$ -arrivals are more typically alike, but  $MF(t)$  patterns for the  $S$ -arrivals are often quite different. In addition, it is noted that some  $P$ -arrivals,  $S$ -arrivals and noise bursts also have similar  $MF(t)$  patterns. If such a  $P$ -arrival pattern is used to train the BPNN, the network will classify all of them as  $P$ -arrivals irrespective of their actual state. In this approach, arrivals are classified according to the linearity pattern of their polarization as defined in the training dataset, with  $P$ -arrival,  $S$ -arrivals and noise bursts characterising high, mid and low levels respectively. It appears that other properties of the seismic arrivals such as the direction of polarization and frequency might be required to supplement this information.

#### **4.4. Effect of changing input nodes**



We also investigated the sensitivity to input segment length as this determines the BPNN structure. Input nodes between 50 and 70 nodes are tested, but with the same hidden nodes and output nodes. The training procedure is the identical: one group of training segments, increasing finally to nine groups. The training segments are different for these three BPNNs due to their different performance at every training stage, but all are from station DP. Table 4 shows the results for the three BPNNs. On balance, the BPNN with 60 input nodes has the optimum performance. This suggests that segments should include appropriate information from an arrival, otherwise too much or too little information will degrade the BPNN performance. This also reflects the general observation that BPNN architecture must be specifically tailored to the individual application.

#### **4.5 Comparison with other methods of identification**

As discussed in the introduction, only a few methods can be used to pick specific types of arrival for single station data. Table 5 gives a comparison of their principles and general performance against our method. Note that the methods used by Roberts et al (1989) and Cichowicz (1993) are not for arrival identification, they pick only one kind of specific waves. The method used by Tong (1995) can only give wave "features", not the wave type. Because articles found in literature tend to describe principles and show a few examples these cannot directly or wholly compared with our result which is applied to a specific data set. It is also difficult to obtain their programs to deal with the data sets used in this paper. Consequently, this table is not truly representative of the optimal forms of each method. However, it does appear that our method needs a minimal assumption and can identify both P- and S-arrivals. We do not suggest without further tests and development that our method should be favoured, but readers must evaluate advantages over other methods listed in Table 5, in conjunction with the demerits elaborated above.

#### **5. Conclusions**

In this paper, a BPNN approach has been developed to identify *P*- and *S*-arrival types from local

earthquake data using only the polarization state. The results show that a BPNN trained with a small subset of the data from station DP can correctly identify 82.3% of the  $P$ -arrivals and 62.6% of the  $S$ -arrivals from station DP, and another BPNN trained with data from station AY can correctly identify 76.6% of the  $P$ -arrivals and 60.5% of  $S$ -arrivals from station AY. This performance, combined with the advantage of not requiring programs to construct special variables and parameters with complicated mathematics, suggests that the BPNN is a natural choice for such applications. This method is adaptive, and does not need other preliminary assumptions except the onset time. The training dataset can be altered to enhance particular features of different datasets. Adding a new training dataset and retraining a BPNN is easy and quick. Although the training time in this approach is longer, once trained the BPNN is sufficiently quick to operate in most real-time applications.

The performance of the trained BPNN, however, has inherent limitations due to the complexity of the  $MF(t)$  patterns. The first limitation is that the training dataset and test data must be from the same station due to the inter-station complexity of  $MF(t)$  patterns. It means that the polarization information is dependent on each station site. The second limitation is that the BPNN's performance cannot be improved by simply adding more events to the training dataset, again due to the complex structure of  $MF(t)$  patterns in the chosen attribute space. This suggests that other information such as the direction of polarization and frequency information may be required. The third limitation is in finding an optimum architecture for a particular application because no theory is currently available to help tackle with this task. The BPNN's performance depends upon the training set, and its ability to interpret cannot lie too far outside its experience. This approach works well if testing segments are similar to the training segments, otherwise, it fails.

In fact, the above limitations are also due to the disadvantage of the supervised learning scheme being used to train the BPNN. Without training, a BPNN cannot learn new strategies for a particular situation that is not covered by the set of examples used to train the network (Haykin, 1994). However, this might be overcome by the use of an unsupervised learning scheme or other kind of neural network such as the ART2 (Carpenter and Grossberg, 1987).

## **Acknowledgements**

This research is sponsored by Global Seismology Research Group (GSRG), British Geological Survey (BGS) of the Natural Environment Research Council (NERC), and is published with the approval of the Director of the BGS (NERC). We thank Chris Browitt, David Booth, and John Lovell of BGS for supplying the earthquake data. Thanks are extended to the staff and student of GSRG for their support and encouragement with this work.

## Appendix A: Covariance matrix analysis and the degree of polarization.

The covariance matrix of 3-C recordings provided a useful measurement of the polarizations of the seismic signals (Samson, 1977; Cichowicz, Green and Brink, 1988; Cichowicz, 1993). The covariance matrix is defined as:

$$C = \begin{Bmatrix} c_{xx} & c_{xy} & c_{xz} \\ c_{yx} & c_{yy} & c_{yz} \\ c_{zx} & c_{zy} & c_{zz} \end{Bmatrix}; \quad (A.1)$$

$$c_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \quad (A.2)$$

where the covariance is measured for  $N$  samples:

where  $\bar{x}$  and  $\bar{y}$  are the average values of  $x$  and  $y$ , and  $N$  is the length of a window in which the covariances are calculated.  $N$  is usually determined from the predominant frequency of signals (Cichowicz, 1993). However, in the case of a seismic network, it is difficult to calculate  $N$  because the recorded events often have a large variation in frequency. Usually  $N$  has to be chosen after gaining some experience from real data.

The covariance matrix contains all the information needed to characterize the polarization state of a wave. It is a real symmetrical matrix which has three real eigenvalues and can be diagonalized to give the eigenvalues and eigenvectors for its principal axes. Some parameters can then be defined from the eigenvalues and eigenvectors to display the polarization state of the wave (Cichowicz, 1993). For the purpose of arrival picking and identification, parameters which are independent of the source orientation should be defined to extract the polarization properties.

One of such useful parameter is the *degree of polarization*  $F(t)$  defined from the eigenvalues (Samson, 1977, Cichowicz, 1993):

$$F(t) = \frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}{2(\lambda_1 + \lambda_2 + \lambda_3)^2} \quad (A.3)$$

where the  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  are eigenvalues of the covariance matrix of a moving window of width  $N$  samples. This equation can be written as:

$$F(t) = \frac{3 \text{tr}S^2 - (\text{tr}S)^2}{2(\text{tr}S)^2} \quad (A.4)$$

where  $\text{tr}S$ , defined as  $\lambda_1 + \lambda_2 + \lambda_3$ , is the trace of  $C$ , and  $\text{tr}S^2$  is defined as  $\lambda_1^2 + \lambda_2^2 + \lambda_3^2$ . This equation

shows that the function can be calculated without having to diagonalize the covariance matrix. Mathematically, the trace of a matrix is independent of the rotation of coordinate system, and hence is also independent of the source orientation. According to this definition, if only one eigenvalue is non-zero, then  $F=1$ , and the wave is linearly polarized; if all of the eigenvalues are equal, then  $F=0$ , and the wave can be considered as completely unpolarized or circularly polarized (Cichowicz, Green and Brink, 1988). Each wave has its own characteristic pattern with time, not just one particular value. The variation of this quantity along the seismogram may indicate the type of a wave. Thus  $F(t)$  enables us to study the evolution of the degree of polarization of a wave.

In order to calculate  $F(t)$ , the window length  $N$  must be determined according to the data features. Computing  $N$  is numerically difficult for real data but it might be visually determined by checking the plotting of  $F(t)$ . Figure 9 shows an example of the variation of  $F(t)$  with different window length  $N$ . Note that the time  $t$  corresponds to the last point of the window. The basic pattern of  $F(t)$  does not change as the window length varies from 5 to 15 samples. However, the longer window length gives a smoother  $F(t)$ . An optimum number ( $N=10$ ) is obtained for the data used in this study. Note that all 3-C recordings must have the same frequency bandwidth, the same scale, and the same noise level. If one of the 3-C recordings has a significantly different property from the others, then  $F(t)$  is highly biased, and may give rise to a misleading interpretation.

## References:

- Bache, T., Bratt, S. R., Wang, J., Fung, R. M., Kobryn, C., and Given, J. W., 1990, The intelligent monitoring system, *Bulletin of the Seismological Society of America*, **80**, 1833-1851
- Basham, P. W., and Ellis, R. M., 1969, The composition of P-code using magnetic tape seismograms, *Bulletin of the Seismological Society of America*, **59**, 473-486.
- Carpenter, G., and Grossberg, S., 1987, ART2: self-organization of stable category recognition codes for analog input patterns, *Applied Optics*, **26**, 4919-4930.
- Cichowicz, A., 1993, An automatic S-phase picker, *Bulletin of the Seismological Society of America*, **83**, 180-189.
- Cichowicz, A., Green, R. W., and Brink, A. van Z., 1988, Coda polarization properties of high-frequency microseismic events, *Bulletin of the Seismological Society of America*, **78**, 1297-1318.
- Dai, H. C., and MacBeth, C., 1995, Automatic picking of seismic arrivals in local earthquake data using an artificial neural network, *Geophysical Journal International*, **120**, 758-774
- Der, Z. A., Baumgardt, D. R., and Shumway, R. H., 1993, The nature of particle motion in regional seismograms and its utilization for phase identification, *Geophysical Journal International*, **115**, 1012-1024.
- Fausett, L., 1994, *Fundamentals of Neural Networks*, Prentice Hall, Englewood Cliffs, New Jersey.
- Haykin, S., 1994, *Neural Networks, a comprehensive Foundation*, Macmillan College Publishing Company, New York.
- Kvaerna, T., and Ringdal, F., 1992, Integrated array and three-component processing using a seismic microarray, *Bulletin of the Seismological Society of America*, **82**, 870-882.
- Lovell, J. H., 1989, *Source parameters of a microearthquake swarm in Turkey*, thesis for the degree of Master of Philosophy, University of Edinburgh.
- Mykkeltveit, S., and Bungum, H., 1984, Processing of regional seismic events using data from small-aperture arrays, *Bulletin of the Seismological Society of America*, **74**, 2313-2333.

- Pao, Y. H., 1989, *Adaptive pattern Recognition and neural networks*, Addison-Wesley Publishing Company, Inc. New York
- Roberts, R. G., and Christoffersson, A., 1990, Decomposition of complex single-station three-component seismograms, *Geophysical Journal International*, **103**, 55-74.
- Roberts, R. G., Christoffersson, A., and Cassidy, F., 1989, Real-time event detection, phase identification and source location estimation using single station three-component seismic data, *Geophysical Journal International*, **97**, 471-480.
- Rumelhart, D., Hinton, G. E., and Williams, R. J., 1986, Learning representations by back-propagating errors, *Nature*, **323**, 533-536.
- Samson, J. C., 1977, Matrix and Stokes vector representations of detectors for polarized waveforms: theory, with some applications to teleseismic waves, *Geophysical Journal of Royal Astronomical Society*, **51**, 583-603.
- Tong, C., 1995, Characterization of seismic phase - an automatic analyser for seismograms, *Geophysical Journal International*, **123**, 937-947.
- Tong, C., and Kennett, B. L. N., 1995, Towards the identification of later seismic phases, *Geophysical Journal International*, **123**, 548-958.

**FIGURE CAPTIONS:**

**Figure 1.** The degree of polarization (lower diagram) determined from a 3-C seismograms (upper diagrams). Three vertical lines indicate the arrivals onset times of the noise burst, P-arrival and S-arrival. The degree of polarization has a lower level for the noise burst, a higher level for the P-arrival and a mid range value for the S-arrival. Their modified DOPs are shown below them respectively (arrowed).

**Figure 2.** 3-C recordings and the degree of polarization  $F(t)$  of a local earthquake recorded on stations DP (above) and AY (below) respectively. The patterns of the degree of polarization are different for the same P-arrival and S-arrival from the two stations with their modified DOPs shown below (arrowed).

**Figure 3.** The flow chart of the approach of identifying arrival types using a BPNN.

**Figure 4.** Nine groups of MF(t) segments of noise bursts, P-arrivals and S-arrivals for training a BPNN for arrival identification. Arrows on segments indicate the pre-picked onset times for these arrivals which all lie at the 31st sample.

**Figure 5.** 3-C seismograms, the vector modulus and the degree of polarization of a local earthquake. The vertical lines indicate the arrival onsets of a noise burst, a P-arrival and a S-arrival. In this case, Their modified DOP are shown below them respectively (arrowed). The BPNN correctly identifies them with its output (1.1, 0.0, -0.1), (0.0, 1.0, -0.1) and (-0.1, 0.6, 1.1) respectively. Note that the output for the S-arrival differs from the training one.

**Figure 6.** Five groups of MF(t) segments for noise bursts, P-arrivals and S-arrivals used in training the BPNN. Arrows on segments indicate the pre-picked onset times for these arrivals and all lie at the 31st sample.



**Figure 7.** The 3-C recording of a local earthquake and three traces of the degree of polarization  $F(t)$  of 3-C recordings with different window length  $N$ .

**Table 1**

The performance of the trained BPNN for arrivals identification. This BPNN has 60 input nodes and is trained with nine groups of training segments from station DP. (a) Identifying results for the data from station DP. (b) Identifying results for the data from station AY.

(a)

	P-arrivals (345)	S-arrivals (302)	Noise (174)
NN identifying P	82.3% (284)	22.0% (67)	9.2% (16)
NN identifying S	10.4% (36)	62.6% (189)	43.0% (75)
NN identifying N	7.2% (25)	15.2% (46)	47.7% (83)

(b)

	P-arrivals (274)	S-arrivals (240)	Noise (28)
NN identifying P	42.0% (115)	48.8% (117)	42.9% (12)
NN identifying S	43.8% (120)	38.8% (93)	25.0% (7)
NN identifying N	14.2% (39)	12.5% (30)	32.1% (9)

**Table 2.**

The performance of the trained BPNN for arrival identification. This BPNN has 60 input nodes and is trained with five groups of training segments from station AY. (a) Identifying results for the data from station AY. (b) Identifying results for the data from station DP

(a)

	P-arrivals (274)	S-arrivals (240)	Noise (28)
NN identifying P	76.6% (210)	22.1% (53)	21.4% (6)
NN identifying S	12.8% (35)	60.4% (145)	53.6% (15)
NN identifying N	11.3% (31)	17.5% (42)	32.1% (9)

(b)

	P-arrivals (345)	S-arrivals (302)	Noise (174)
NN identifying P	84.9% (293)	45.4% (137)	30.0% (52)
NN identifying S	8.1% (28)	36.1% (109)	17.8% (31)
NN identifying N	7.0% (24)	18.5% (56)	52.3% (91)

**Table 5**

A summary comparison of selected identifying methods

Author	Assumption	Input data	Method	Output result	Event type	Performance or testing
Roberts et al. (1989)	None	3-C	Auto- and cross-correlations	Onset of P-wave and linearly polarized S-wave	Tele-events	two examples
Cichowicz (1993)	Onset and polari-zation of P-wave	3-C	Characteristic function > threshold	Onset of S-wave	Local events	Six examples (65-70%)
Tong (1995)	Onset of arrivals	1-C	Intelligent segmentation	Wave feature (not type)	Tele-events	Two events in two stations
Tong and Kennett (1995)	P-wave feature and onset of later arrivals	3-C	Energy analysis, Threshold	Types of later arrivals	Tele-events, regional	Three examples

This paper	Onset of arrivals	3-C	Degree of polarizarization, BPNN	Arrival type	events local events	80% for P, 61% for S (more than 600 events)
------------	-------------------	-----	-------------------------------------	--------------	------------------------	--

---

Station: DP  
Date: 1984-05-06  
Start-time: 12h16m08s  
Scale: 564

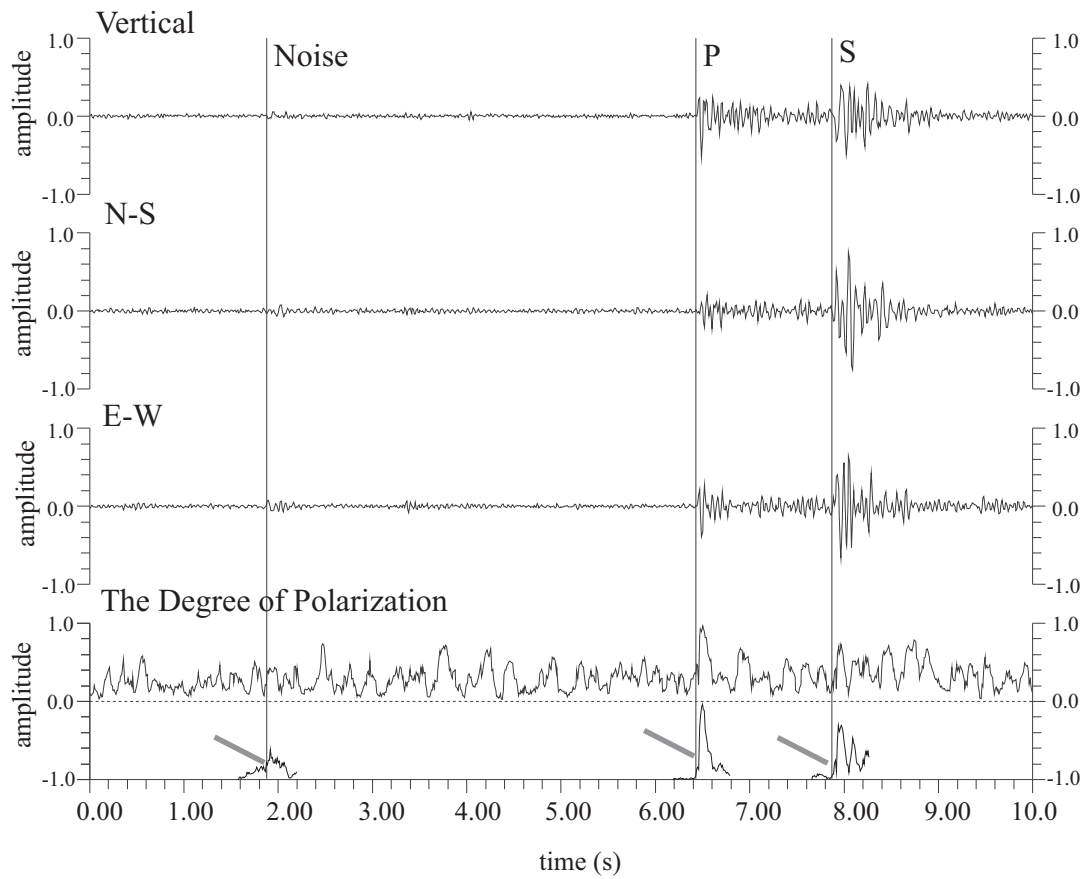
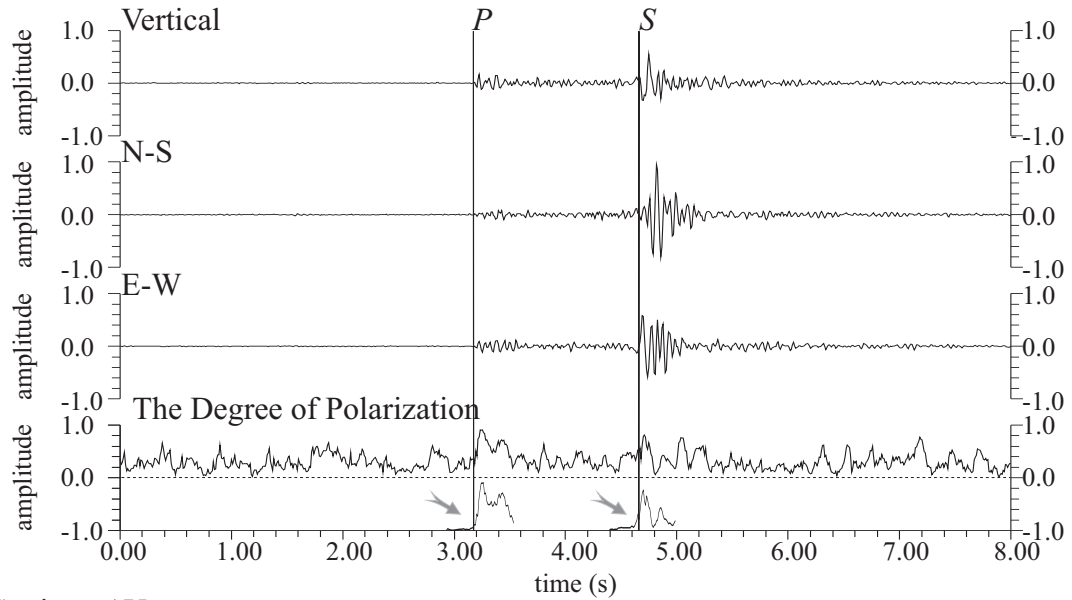


Figure 1 (Application of Back-propagation Neural Networks to Identification of seismic arrival types by Hengchang Dai and Colin MacBeth)

Station : DP  
Date: 1984-05-03  
Time: 16h25m26s  
Scale: 1386



Station : AY  
Date: 1984-05-03  
Time: 16h25m26s  
Scale: 379

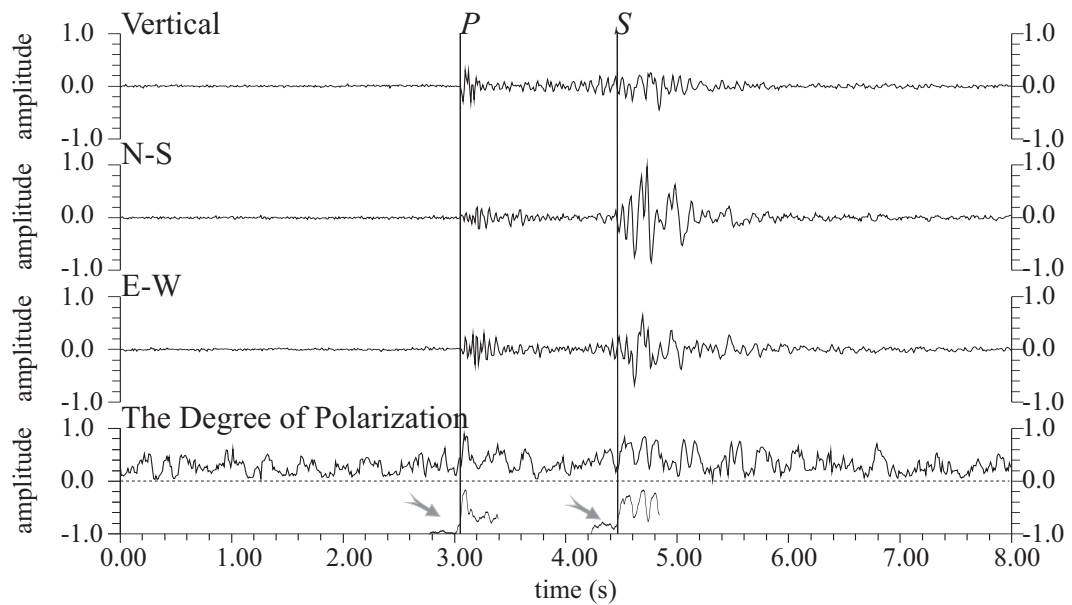


Figure 2. (Application of Back-propagation Neural Networks to Identification of seismic arrival types by Hengchang Dai and Colin MacBeth)

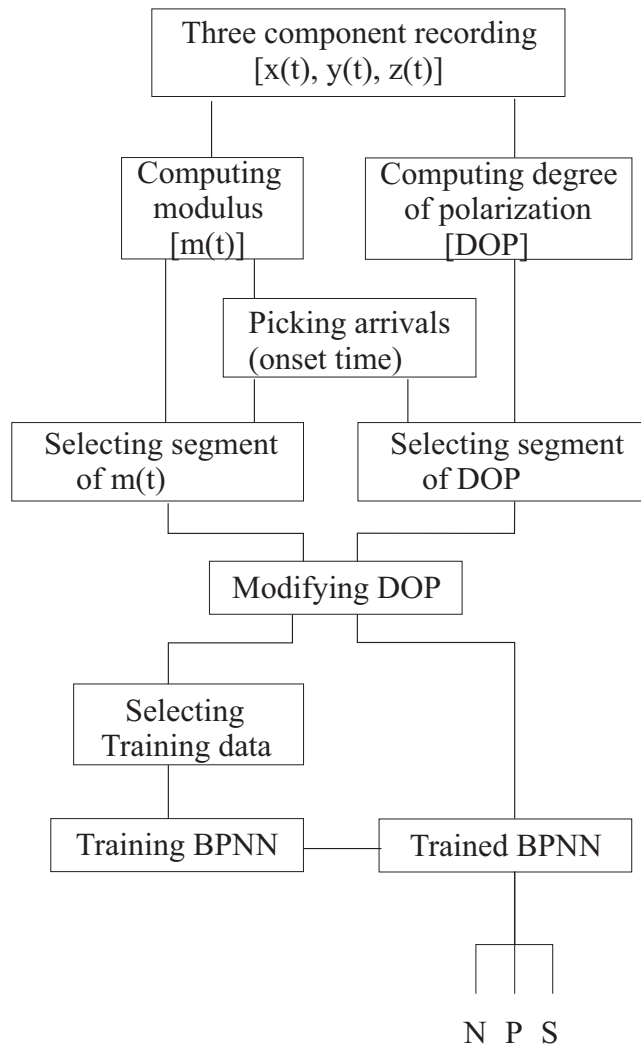


Figure 3 (Application of Back-propagation Neural Networks to Identification of seismic arrival types by Hengchang Dai and Colin MacBeth)

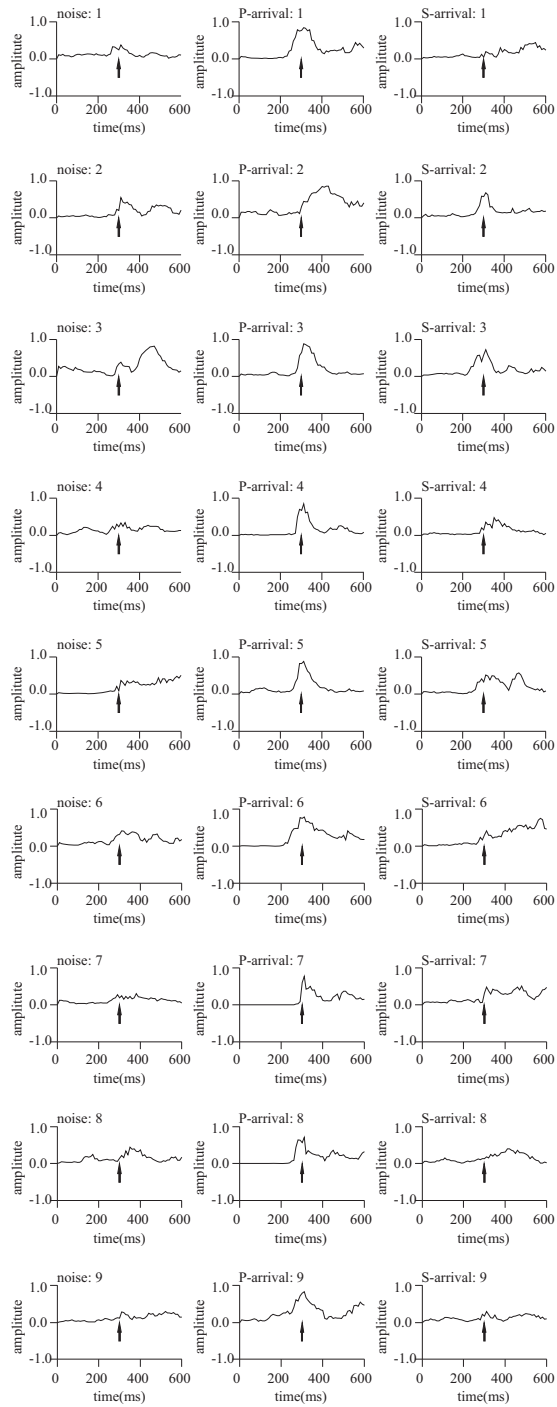


Figure 4 (Application of Back-propagation Neural Networks to Identification of seismic arrival types by Hengchang Dai and Colin MacBeth)

Station: DP  
Date: 1984-05-31  
Start-time: 09h23m36s  
Scale: 373

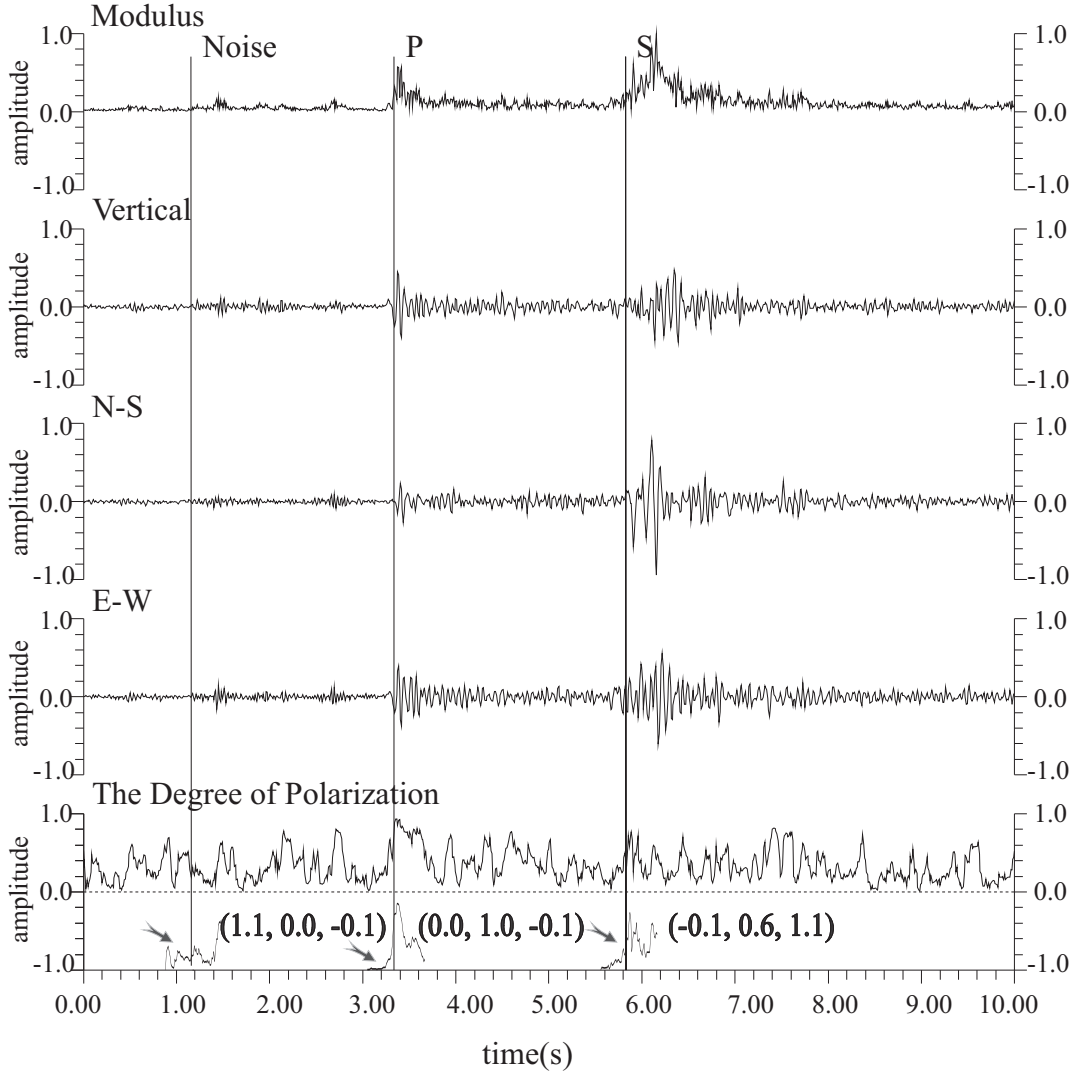


Figure 5 (Application of Back-propagation Neural Networks to Identification of seismic arrival types by Hengchang Dai and Colin MacBeth)



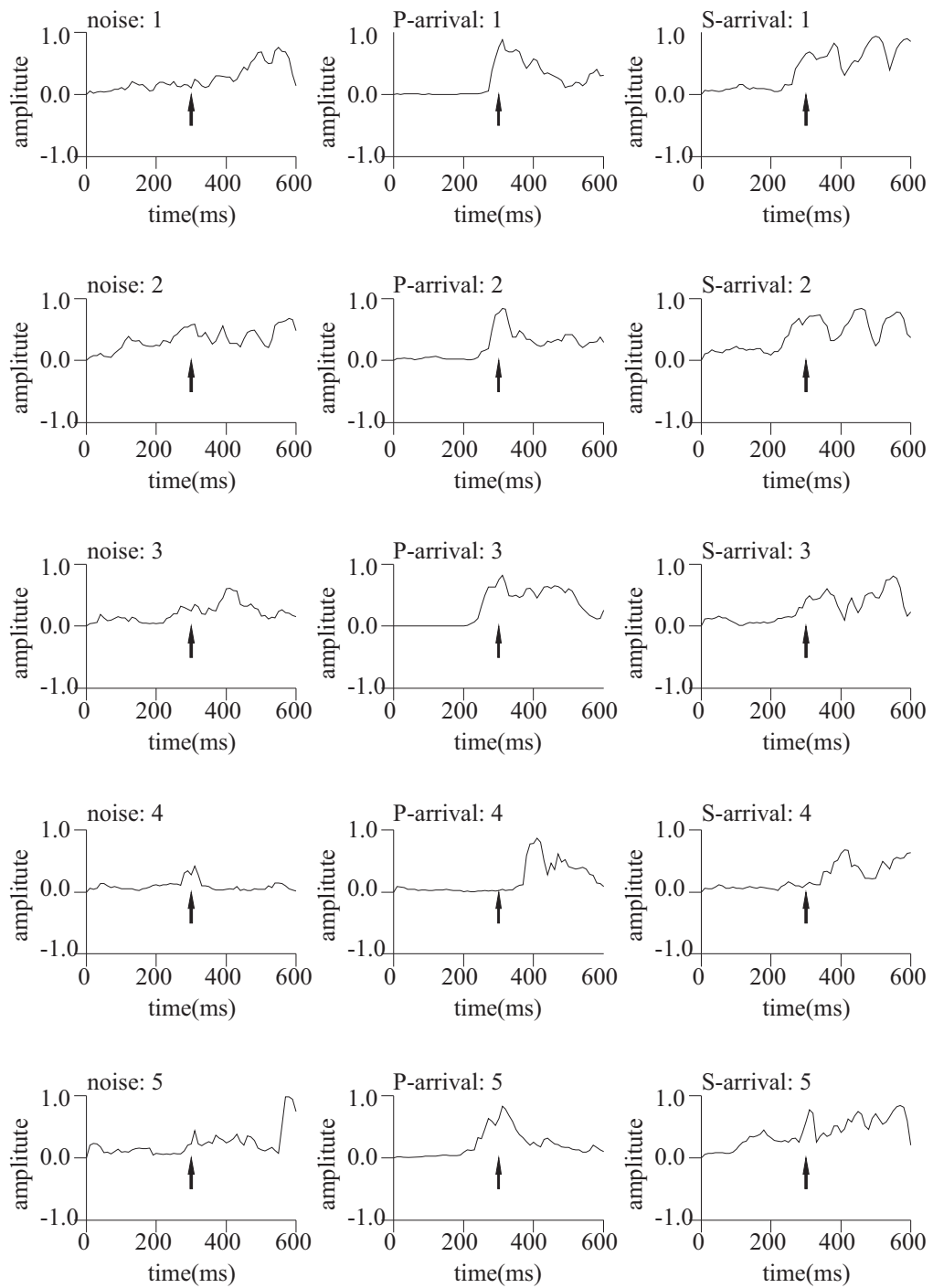


Figure 6 (Application of Back-propagation Neural Networks to Identification of seismic arrival types by Hengchang Dai and Colin MacBeth)

Station: DP  
Date: 1984-05-06  
Time: 12h17m27s  
Scale: 846

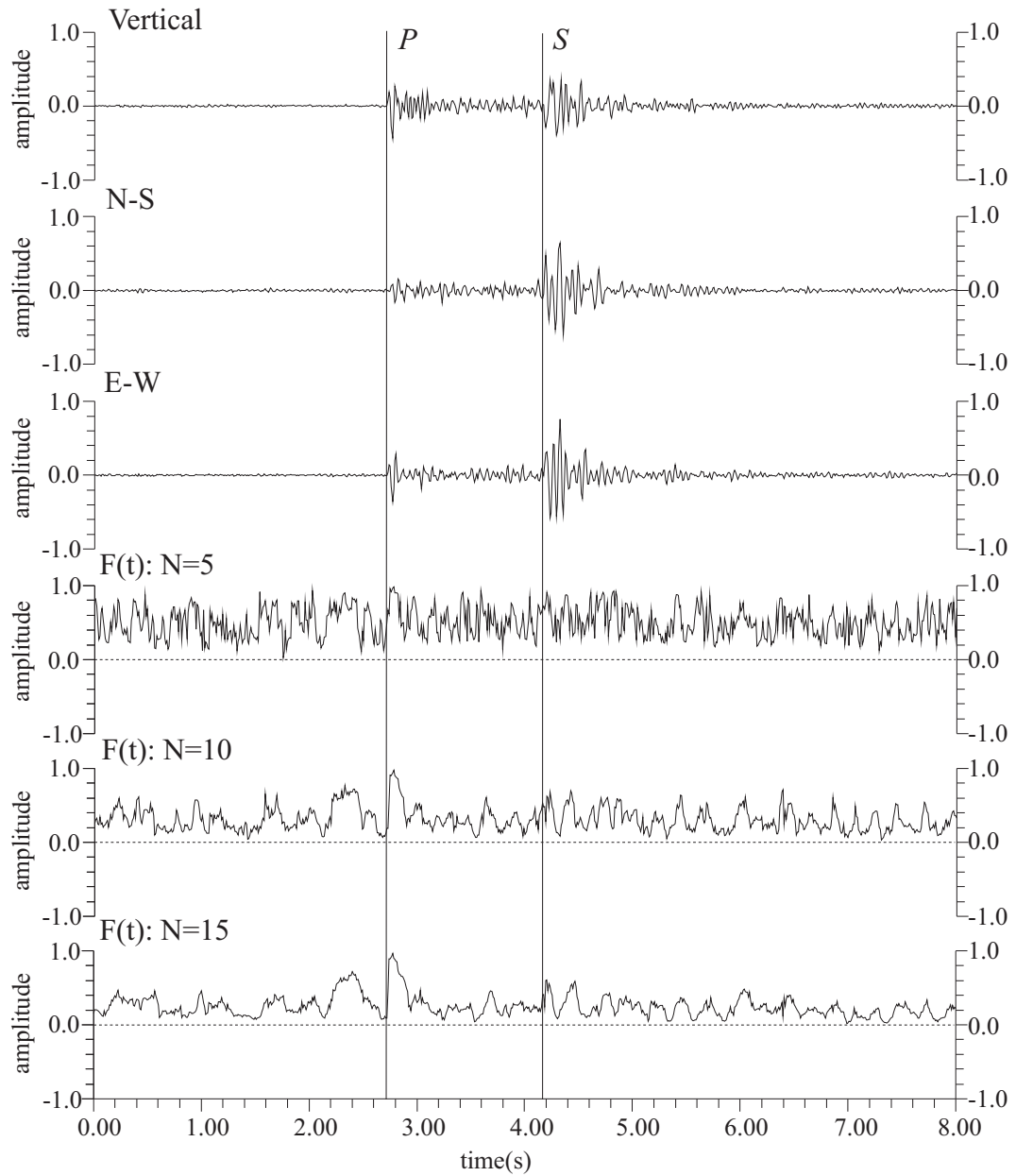


Figure 7 (Application of Back-propagation Neural Networks to Identification of seismic arrival types by Hengchang Dai and Colin MacBeth)