Multi-window algorithm for automatic phase pickings of P and S phases of local events in 3-C seismic recordings

Zuolin Chen and Robert Stewart, CREWES, University of Calgary*

Summary

A multi-window algorithm for automatic picking of the first arrivals of impulsive P and S phases of local seismic events in low *SNR* (signal to noise ratio) environments has been developed. This algorithm differs from the conventional *STA/LTA* method by employing both the instantaneous and the averaged absolute amplitude or/and power of traces in a series of time windows before and after each time point (sample) as the characteristic functions. When the instantaneous absolute value of a characteristic function exceeds an automatically adjusted threshold, ratios based on the averages of the windows over time samples provide parameters to differentiate an expected event from unwanted noise. Examination of the algorithm in various *SNR* environments by using synthetic and real data shows that the picking accuracy of impulsive first arrivals increases with the decline of the noise level, and can be less than 1-2 samples when *SNR* ratio is lower than 3. Considering the influence of angle of incidence of the ray path on *SNR*, both the component with highest *SNR* ratio and the square root of power of 3-C traces are recommended to be used in the event picking procedures. S phase picking is improved by rotating the 3-C traces to the local ray coordinates to remove the superimposition of P coda waves, and enhance the local *SNR*.

Introduction

Accurate onset time determination of P and S phases is of considerable importance for the automatic real-time location of local seismic events in large volumes of digital passive seismic data. Generally, picking of the first arrivals of seismic events is carried out in the time-domain (e.g., Allen, 1982; Der and Shumway, 1999). Among the time-domain approaches, the short-term average over long-term average ratio (*STA/LTA*) automatic picker is widely used. The absolute value, the power, or the envelope of the seismic trace is usually selected as the characteristic function for the calculation of *STA* and *LTA* (e.g., Kanasewich, 1981). The fundamental rationale of the algorithm detects when the *STA/LTA* ratio exceeds a pre-defined threshold. This approach is well suited for the arrival detection of seismic events, but has two inherent weaknesses of inaccurate arrival time detection due to the long delay associated with the length of the *STA* window, and inability to distinguish events from high-amplitude noise.

In this paper, we propose an improved *STA/LTA* algorithm which effectively picks first arrivals at various levels of noise environments, based on the instantaneous absolute values of amplitude or/and the square root of power of 3-C seismic traces and their averages over three moving time windows. A concept of waveform corrcetion is introduced to compensate for the time delay associated with the trigger threshold. This multi-window procedure resembles that of a human operator in detecting the first arrivals of seismic events and is comparable in accuracy to the latter. Similar processing is applied to the picking of the S phase by a manipulation to enhance the local *SNR*.

Methodology

The multi-window automatic phase picker operates in the time-domain. To pick the first arrival of a P phase, it includes procedures for defining time windows, standards, corresponding thresholds and waveform correction for the acquisition of a more accurate arrival time. After the picking of the P phase, rotation of the 3-C traces from original (N, E and Z) to local ray coordinates (L, T and Q) is carried out to remove the P coda waves, and enhance the local *SNR* of S phase in T and Q coordinates. Azimuth and angle of incidence in L direction are derived from the dominant eigenvector of a covariance matrix obtained from a time window enclosing the P phase. The remaining procedures are a repetition of the previous steps.

Windows and standards

The definitions of the multiple moving time windows are mainly tailored to the purpose of measuring the averages of the *SNR* within certain time extents before, just after and after a delay of an instantaneous time point (sample). The averages of absolute values of a characteristic function within *BTA* (Before Term Average), *ATA* (After Term Average) and *DTA* (Delayed Term Average) windows are respectively defined as follows:

$$\overline{BTA(t)} = \sum_{i=1}^{m} |u(t-i)| \Big/ m \quad (1)$$

$$\overline{ATA(t)} = \sum_{j=1}^{n} |u(t+j)| \Big/ n \quad (2)$$

$$\overline{DTA(t)} = \sum_{k=1}^{q} |u(t+j+d)| \Big/ q \quad (3)$$

Where, $u(t)$ represents the amplitude or the square root of the power of 3-C seismic traces at a time point t; m, n, and q represent the lengths of the windows in samples respectively; d is the time delay for a *DTA* window.

The instantaneous absolute value $|u(t)|$ is selected as the first standard $R_1(t)$ to judge the arrival of a possible high amplitude/power event. Two extraordinary standards $R_2(t)$ and $R_3(t)$ are introduced based on the ratios of averages of *ATA* and *DTA* over *BTA*, and take the forms

$$R_2(t) = \overline{ATA(t)} / \overline{BTA(t)} \quad (4)$$

$$R_3(t) = \overline{DTA(t)} / \overline{BTA(t)}. \quad (5)$$

The functions of $R_2(t)$ and $R_3(t)$ are mainly designed to discriminate between high-amplitude short-duration and long-duration noise, respectively. The definitions for the $R_2(t)$ and $R_3(t)$ are based on the following rationale:

When the first standard, $R_1(t)$, exceeded a pre-defined threshold $H_1(t)$ (see below) at a time point t, the second standard, $R_2(t)$, is used to separate high-amplitude but short-duration noise from a high-amplitude but long-duration event. Equation (4) shows that when the *ATA* window contains noise, and the length of the window is several times longer than the noise, the $R_2(t)$ value will only be weakly affected by the existence of the noise; however, when the *ATA* window is occupied by a part of an event, the $R_2(t)$ value will be much amplified. Based on this feature of $R_2(t)$, short duration noise will be removed. When the duration of high-amplitude noise is comparable to the length of the *ATA* window, $R_2(t)$ will lose its ability to distinguish an event from noise. This difficulty can be addressed by moving the *ATA* window backward a certain time interval as a *DTA* window. Yielding equation (5) an appropriate delay time will reduce the average of the *DTA* window and hence $R_3(t)$ to an expected level. In this way, false triggers caused by high-amplitude long-duration noise can be avoided.

Thresholds

Thresholds pertaining to the three standards can be pre-defined based on the expected *SNR*. An optimal first threshold, $H_1(t)$, should be larger than the fluctuations of most noise but lower than an expected event. This demand can be satisfied by measuring the mean (E_m) and standard deviation (E_{sd}) of the envelope ($E(t)$) of pre-existing noise within a *BTA* window. The envelope $E(t)$ can be acquired from the absolute value of the Hilbert transform of a seismic trace $u(t)$. In practice, $H_1(t)$ is often shifted several samples backwards to avoid the unnecessarily early rising of $H_1(t)$ caused by the first arrival. In this way, an instantaneous $H_1(t)$ is expressed as

$$H_1(t) = E_m(t-p) + \alpha E_{sd}(t-p). \quad (6)$$

Here, p is the number of shifted samples; α is a coefficient to adjust the height of the first threshold, and is often taken to be 3, whereby approximately 99% of the noise will not exceed $H_1(t)$ theoretically. From equation (6) it is obvious that $H_1(t)$ is automatically adjusted with the variance of the background noise.

Thresholds $H_2(t)$ *and* $H_3(t)$ corresponding to standards $R_2(t)$ and $R_3(t)$ are generally set to be ¾ level of the expected *SNR* of events. It is noted that extraordinary high or low specification of $H_2(t)$ *and* $H_3(t)$ will either lead to missing of events, or the triggering of unwanted noise.

Waveform correction

As $H_1(t)$ is defined higher than most pre-existing noise levels, and further the instantaneous absolute value at the trigger time point is higher than $H_1(t)$, the trigger time point must be somehow later than the real onset time according to the configuration of the first arrival of an event. This belated onset time can be compensated by introducing a concept of waveform correction. For an impulsive first arrival, this process is easily accomplished by using the height of the absolute value and the representative gradient at the trigger point.

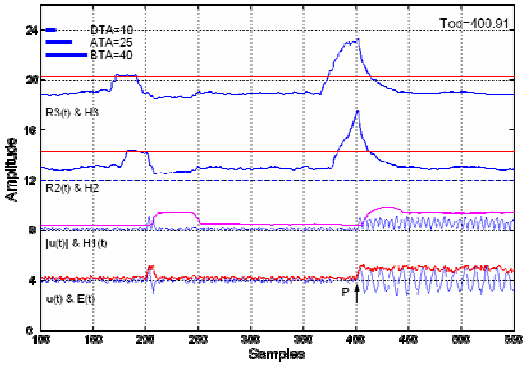

Figure 1: Illustration of the picking of the first arrival of an event in a 1-C synthetic seismogram. Diagrams from the bottom to the top are synthetic seismogram (blue) and envelope (red); absolute value of seismic trace $u(t)$ (blue) and automatically adjusted threshold $H_1(t)$ (magenta); second standard $R_2(t)$ (blue curve) and threshold $H_2(t)$ (red straight line); third standard $R_3(t)$ (blue curve) and threshold $H_3(t)$ (red straight line). The dash-dotted line in each diagram represents the zero reference levels. The lengths of moving time windows are shown in the top left corner. Expected *SNR* is set to be 3. The real onset time of the seismic event is set at time point 401. The corrected onset time (T_{oc}) is given in the top right corner.

Examinations on synthetic and real data

Picking of P onset time in 1-C synthetic data

The performance in accuracy and tolerance of our algorithm is firstly examined on a series of 1-C synthetic data with various *SNR* levels. Basically, the synthetic data is composed of three parts: random noise with various gains, a high-amplitude short-duration noise and an impulsive sine wave seismic event with its first arrival at a fixed time point. Random noise still superimposes with the event after its arrival.

Figure 1 gives an example to show the multiple time windows, standards and thresholds, and illustrates the picking accuracy of our algorithm on a synthetic seismic trace with a 20 Hz dominant frequency event. The maximum amplitudes of the event and a high-amplitude short-duration noise are assumed to be 1. The amplitude coefficient of random noise is assumed to be 0.25. It is

noted that the three standards of the event exceeded their corresponding pre-defined thresholds at sample 402, and the event is triggered there. The continuous procedure of waveform correction modifies the trigger point to the corrected onset time of 400.91, which is merely 0.09 samples earlier than the real onset time at time point 401.

Accuracy of our automatic P phase picker is also examined on some real local events recorded by a small six-station vertical-component seismic array deployed in southern Alberta, Canada (Bingham, 1996). Examinations show that the picking accuracy of real events with impulsive first arrivals can be stably picked as accurate as less than 1-2 samples compared to the pickings by a human operator.

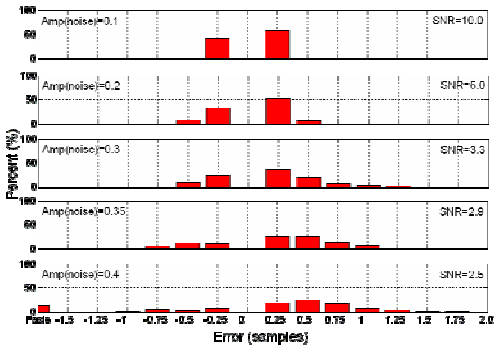

Figure 2: Error distribution of corrected onset times with various background noise levels. For each diagram, picking is repeated 100 times to get stable statistics. Histograms indicate the percentage of events with corrected onset time errors less than the indicated amounts in the horizontal axes. The *SNR* is estimated by equation $SNR = Amp_{event} / Amp_{noise,}$ where Amp_{event} and Amp_{noise} represent the maximum amplitudes of event and random noise respectively. The lengths of *BTA*, *ATA* and *DTA* windows are set to be 40, 30 and 30 samples respectively. The time delay of *DTA* window is set to be 10 samples. $H_2(t)$ and $H_3(t)$ are set to be 0.75 times the expected *SNR* of 2.

The performance in accuracy and tolerance is further examined by the synthetic data in the case of various background noise levels. As the noise component of the synthetic data is computer-generated randomly, the time series used in the examinations is slightly different each time. The examination of each noise levels is repeated 100 times to acquire a stable statistic of errors of detection between real and corrected onset times.

Figure 2 shows the statistical results on the accuracy and tolerance of our algorithm with the increase of the amplitude of random noise, where the maximum onset time error ranges from -1 to +1.25 samples. When the coefficient of random noise is increased to 0.4, false triggers due to the existence of high-amplitude noise began to occur.

The scatter of the corrected onset times increases with noise levels. This is considered to be caused by the disturbance of the waveform of first arrivals by the superimposed random noise. The disturbance of the waveform thus reduces the reliability of the two crucial factors of height and representative gradient adopted in the waveform correction at the trigger point.

Picking of P phase in 3-C synthetic data

Based on the results from 1-C synthetic data, it is obvious that the picking accuracy and tolerance vary with the *SNR* in a specified component. Assuming the noise to be random and have equal variance in each component of 3-C recordings, the amplitude and hence the *SNR* in each component depends on the azimuth and angle of incidence of the ray path. In more detail, the *SNR* of P phase in vertical component is determined by the angle of the incidence, and the *SNR* in the two horizontal components are related to both azimuth and angle of incidence. The relationships between the *SNR* in three orthogonal components and the angle of incidence are calculated and shown in Figure 3. It can be seen that a small angle of incidence (less than approximate 50°) causes highest *SNR* in the vertical component, while a large angle of incidence creates highest *SNR* in one of the horizontal components. It is noted that the square root of the power of 3-C traces possesses the second highest *SNR* in most cases. To insure that the picking of P phase is reliable, it is suggested to refer to both the component with the highest *SNR* and the square root of power of traces before the declaration of an event.

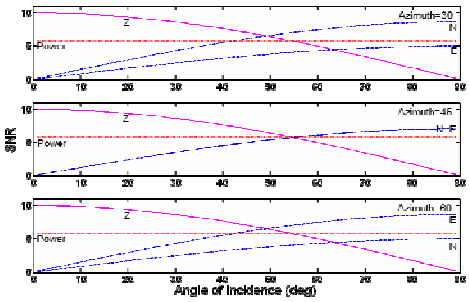

Figure 3: Relationships between *SNR* and angle of incidence of ray path. The vertical component, Z, is marked by a solid line (magenta); horizontal components; E and N are marked by dashed lines (blue); the square root of power of traces is marked by a dotted line (red).

Picking of S phase in 3-C synthetic data

Because the S phase is often superimposed by P coda waves, to enhance the S phase and suppress the P coda waves, we rotate the 3-C traces from the original coordinates (N, E and Z) into ray coordinates (L, T and Q). Rotating to align the P wave along the propagation direction (L) in the local ray coordinates is equivalent to the manipulation of a polarization filter which allows P wave pass in L direction, and S wave pass in T and Q directions.

Azimuth and angle of incidence for ray path rotation can be derived from the dominant eigenvector analysis method

using a covariance matrix obtained from a time window of the 3-C data containing the P first arrival (Flinn, 1965; Montalbetti and Kanasewich, 1970). The length of the time window is determined by the interval between the picking time and the first peak or trough of the P phase in a specified component. In addition, the 180° ambiguity is judged by the signs of the polarities of the P phase in three components.

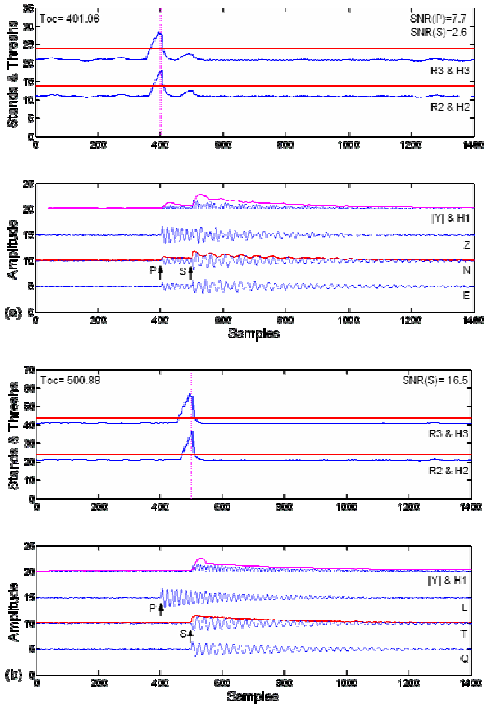

Figure 4: a) An example illustrates the failure of picking the S phase in a 3-C synthetic seismogram before ray path rotation. Component 'N' is used for the picking of the P and S phase, and the envelope of the trace is plotted together in a solid line (red). In the upper part of the figure, a dotted vertical line (magenta) marks the corrected P onset time (sample 401.05); the real onset time of P phase is at sample 401. The lengths of *BTA*, *ATA* and *DTA* windows are all set to be 40 samples; the time delay of the *DTA* window is set to be10 samples. H_2 *(t)* and H_3 *(t)* are set to be 0.75 times the expected *SNR* of 5.
b) Corresponding 3-C seismogram in local ray coordinates (L, T and Q) after ray path rotation. Markers are similar to those in a). The corrected S onset time is at sample 500.83, where the real onset time is at sample 501.

The picking precedures of P and S phases in a 3-C synthetic seismogram before and after ray path rotation are illustrated in Figure 4. Due to the low *SNR* of the S phase before rotation, the standard R_2 *(t)* and R_3 *(t)* at S onset time (sample 501) are lower than the pre-defined thresholds H_2 *(t)* and H_3 *(t)*. As a result, the S phase fails to be picked (Figure 4a). However, when the rotation procedure is carried out immediately after the picking of the P phase, the local *SNR* of the S phase is dramatically improved from 2.6 to 16.5. Neddless to say, the S phase is easily picked (Figure 4b).

Conclusions

We have developed an effective and accurate multi-window algorithm for automatically picking the onset times of P and S phase of impulsive local seismic events. The picking in accuracy of P phase is less than 1 sample in most cases, which is comparable to that achieved by a human operator. Picking of S phase is improved through the procedure of ray path rotation to enhance the local *SNR*.
The algorithm is examined to function reliably when the *SNR* is higher than 3.
The manipulation of parameters, such as lengths of time windows, time delay interval and thresholds, is brief to be mastered. After the limitation of expected *SNR*, duration of impulsive noise to be discarded and dominant frequency band of signal are indicated, all parameters can be defined directly or automatically.
Due to the efficiency and simplicity of the method, our algorithm is suitable and applicable in real-time automatic passive microseismic monitoring with only moderate computing capability.

References

Allen, R., 1982, Automatic phase-pickers: their present use and future prospects, Bull. Seism. Soc. Am., 72, S225–S242.

Bingham, D.K., 1996, Seismic monitoring of Turtle Mountain, Internal report, Alberta Environmental Protection, Government of Alberta.

Der, Z.A. and Shumway R.H., 1999, Phase onset time estimation at regional distance using the CUSUM algorithm, Phys. Earth planet. Inter., 113, 227-246.

Flinn, E.A., 1965, Signal analysis using rectilinearity and direction of particle motion, Proc. IEEE 53, 1874-1876.

Kanasewich, E.R, 1981, Time Sequence Analysis in Geophysics, Univ. of Alberta Press, Edmonton.

Montalbetti, J. and Kanasewich E., 1970, Enhancement of teleseismic body waves with a polarization filter.Geophys. J. R. Astr. Soc., 21, 119-129.

Acknowledgments

We thank Dr. R. Maier of CREWES, Univ. of Calgary for his helpful discussions.