Article
Identification of Similar Seismic Waves Using the Phase-Only Correlation Function and Wavelet Transform

Hirokazu Moriya

Graduate School of Engineering, Tohoku University, 6-6-04 Aramaki aza aoba, Aoba-ku, Sendai 980-8579, Japan; hirokazu.moriya.e1@tohoku.ac.jp

Abstract: Accurately determined acoustic emission (AE) locations provide significant information on fracture systems, such as the orientation of fractures in a geothermal reservoir. To determine the relative source locations among a group of seismic events, similar AE waveforms must be detected and the relative arrival times of the P and S waves must be determined. In this paper, a method to identify similar AE waveforms is proposed, in which wavelet transform scalograms are used to determine the phase-only correlation function. The proposed method was applied to arbitrarily selected seismic waveforms, and its feasibility was evaluated by comparing the results with those obtained when the phase-only correlation function was obtained by using Fourier transform results.

Keywords: phase-only correlation; wavelet transform; acoustic emission; multiplet; cluster analysis

1. Introduction

During acoustic emission (AE) measurements at geothermal fields, AE events with similar waveforms are often observed. An AE group with similar waveforms is called an “AE multiplet”. AE multiplets are considered to comprise AE events occurring at a fracture with the same characteristics. The source location of these seismic events can be determined with a high accuracy by using a cross-correlation analysis [1], which allows for the strike and dip of the fault plane to be estimated. For instance, when the distance from the source locations to the observatories is about 1 km, the relative position of the source locations can be detected with an accuracy of several tens of centimeters. It makes sense from the principle of measurement that determining the relative relationship between the source locations is more accurate than determining the absolute position of the source locations. By estimating the geometrical relationship among a group of microearthquakes with similar waveforms, the fracture orientation and the migration of microearthquake source locations during hydraulic fracturing and fluid injection can be detected [2]. If several similar waveforms are observed and their source locations are determined to the order of tens of centimeters, the fracture surface that has been shear-slipped due to the increase in pore pressure can be estimated from the strike and dip of the plane defined by the distribution of the source locations. The direction of the pressure propagation due to water injection can also be estimated from the direction of the movement of the source location. The multiplet is useful because it is able to estimate not only the fracture distribution in the reservoir, but also the repeating moment release on a fracture plane, which would be related to fluid flow in the fracture, etc. [3–7]. Conventionally, similar waveforms have been identified by finding the mutual cross-correlation coefficient or a coherence function. The development of an automated efficient searching method for similar waveforms is needed for the evaluation of a fracture system using microearthquake waveforms, because their visual identification and classification are time consuming. An automated efficient searching method is also important in detecting slow-slip seismic events, which have recently been of interest for understanding the mechanical behavior of faults.
Image matching and pattern recognition are important data processing tools in geophysical fields, and several image matching and pattern recognition methods have been proposed for the identification of similar seismic events, for the detection of low-frequency earthquakes within a tremor, seismic pattern recognition, and ground-penetrating radar applications [8–13]. The source location of a group of similar seismic events (a multiplet) can be determined with a high resolution by using cross-correlation or cross-spectrum analyses. These techniques are widely used to locate faults and to determine the orientation of rock fractures in geothermal and petroleum fields [14–16], and microearthquake signal processing for subsurface measurements is increasing in importance (e.g., [17]).

The phase-only correlation (POC) function is widely used in image data matching, such as for fingerprint recognition by automated teller machines (ATMs) [18]. The POC function can be used to recognize the similarity of two images with a high accuracy. In waveform data analysis, a method for detecting similar waveforms using the phase-only correlation function with a moving time window Fourier spectrum has been proposed [8,19]. This method is based on the new idea of transforming the waveform data to the time and frequency axes, and considering it as two-dimensional image data drawn on the time and frequency axes to calculate the phase-limited correlation function. Compared with conventional detection methods using correlation functions or a coherence function, this method evaluates the time variation of waveforms in both the time and frequency domains, which allow for small differences in the waveforms to be evaluated and the waveforms to be classified according to their similarity. With this method, however, the resolution of the spectrum in the time domain is limited by the length of the time window used for the Fourier transform; a comparatively longer time window length must be used to estimate the lower frequency components.

In this paper, a similar waveform detection method using the phase-only correlation function with wavelet transform, which enables spectral estimation with a relatively high time resolution, is investigated. Through the application to microseismic waveforms, it has been found that the proposed method can classify the waveforms more accurately and quantitatively than the FFT-based method previously proposed by the author. The proposed method, as an efficient automatic search method for similar waveforms, will contribute to the measurement of the geothermal reservoir structure by determining the source location of similar waveforms with a high accuracy.

2. Wavelet Transform

Wavelet transform, a time–frequency analysis method, has various geophysical applications [20–25]. In Fourier transform, the length of the analysis time window is fixed, regardless of whether high- or low-frequency components are being estimated. However, the length of the analysis time window used to estimate the spectrum should differ depending on the frequency component. In a moving time window Fourier transform, where the transform is performed while moving the analysis time window, the higher the frequency component, the lower the resolution in the time-axis direction. This is because the number of periods included in the time window increases when the frequency component is high. Different from Fourier transform, in wavelet transform, the length of the time window is changed during the calculation according to the frequency. In other words, the basis functions are scaled up or down so that the time resolution does not become lower at higher frequencies. Wavelet transform using a moving time window can thus more easily supplement the waveform changes in the time-axis direction than Fourier transform. Although a wavelet transform cannot overcome the uncertainty principle, it is well suited for the analysis of signals characterized by strong non-stationarity, such as microearthquake signals.

3. Phase-Only Correlation Function

Phase-only correlation of image data makes it possible to evaluate similarity. Phase-only correlation is defined by the 2D inverse discrete Fourier transform (IDFT) of the
combined phase $\hat{R}(k_1, k_2)$ obtained from the 2D discrete Fourier transform (DFT) of two sets of image data, $f(n_1, n_2)$ and $g(n_1, n_2)$, as described by the following equations (e.g., [17]).

\[
F(k_1, k_2) = \sum_{n_1=-M_1}^{M_1} \sum_{n_2=-M_2}^{M_2} f(n_1, n_2) W^{k_1 n_1}_{N_1} W^{k_2 n_2}_{N_2} = A_F(k_1, k_2) e^{i\theta_F(k_1, k_2)} \quad (1)
\]

\[
G(k_1, k_2) = \sum_{n_1=-M_1}^{M_1} \sum_{n_2=-M_2}^{M_2} g(n_1, n_2) W^{k_1 n_1}_{N_1} W^{k_2 n_2}_{N_2} = A_G(k_1, k_2) e^{i\theta_G(k_1, k_2)} \quad (2)
\]

where $k_1 = -M_1, \ldots, M_1, k_2 = -M_2, \ldots, M_2$, $W_{N_1} = e^{-j\frac{2\pi}{N_1}}$, and $W_{N_2} = e^{-j\frac{2\pi}{N_2}}$, and $N_1$ and $N_2$ denote the numbers of samples in the $n_1$ and $n_2$ directions, respectively. $A_F(k_1, k_2)$ and $A_G(k_1, k_2)$ denote the amplitudes of $F(k_1, k_2)$ and $G(k_1, k_2)$, respectively, and $\theta_F(k_1, k_2)$ and $\theta_G(k_1, k_2)$ are their respective phases. Then, the function $R(k_1, k_2)$ can be defined as follows:

\[
\hat{R}(k_1, k_2) = \frac{F(k_1, k_2) \overline{G}(k_1, k_2)}{|F(k_1, k_2)\overline{G}(k_1, k_2)|} = e^{i\Delta \theta(k_1, k_2)} \quad (3)
\]

where $\overline{G}(k_1, k_2)$ represents the complex conjugate of $G(k_1, k_2)$ and $\Delta \theta(k_1, k_2) = \theta_F(k_1, k_2) - \theta_G(k_1, k_2)$. The phase-only correlation function can then be represented by using inverse 2D IDFT, as follows

\[
\hat{r}(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1=-M_1}^{M_1} \sum_{k_2=-M_2}^{M_2} \hat{R}(k_1, k_2) W^{-k_1 n_1}_{N_1} W^{-k_2 n_2}_{N_2} \quad (4)
\]

where the value of $\hat{r}(n_1, n_2)$ has a range from 0 to 1. The peak value of $\hat{r}(n_1, n_2)$ provides a measure of the similarity of the two images, and the position of the peak shows the translational displacement between the two images.

### 4. Detection of Similar Seismic Events with the Phase-Only Correlation Function

The performance of the phase-only correlation method using a moving time window Fourier transform was evaluated previously [8]. In this study, the method with a wavelet transform was evaluated and then its performance was compared to that of the method using the Fourier transform.

Figure 1 shows a flow chart of the calculation method for the phase-only correlation using wavelet transform. AE waveform time series data were used in the analysis. In this method, wavelet transform was used for the spectrum estimation instead of Fourier transform. The scalogram, which is a transformation of the signal calculated by the wavelet transform, is regarded as 2D image data; thus, an image matching using the POC function is applicable. Two arbitrarily selected waveforms were first transformed to scalograms by transforming the time series data to time and frequency domain data, allowing the data to be regarded as 2D image data. Next, 2D DFT was applied to each of the two scalograms, and the complex vectors, one of which needed to be a conjugate complex number, obtained from each scalogram were multiplied together. Then, the resulting normalized cross-spectrum was inverted in the time domain by using 2D IDFT. The value of the peak of the phase-only correlation function $\hat{r}(n_1, n_2)$ represented the similarity of the two waveforms.

Figure 2 shows the group of waveforms used to evaluate the performance of the proposed method. These waveforms were produced by microearthquakes that occurred at the lower end of an active fault in Sendai City, Miyagi Prefecture, at a depth of about 11 km, and they were detected by a triaxial seismometer in a borehole. The distance from the hypocenters to the detector was around 15 km. The waveforms were those of the vertical component signal, and the sampling frequency was 100 Hz. These waveforms
of microearthquakes observed by a broad frequency band detector were used for the evaluation of the proposed method, because the P and S waves could be clearly read and it was easy to visually judge their similarity.

Figure 1. Flow chart of the calculation method for the phase-only correlation using wavelet transform of waveforms. When multiplying the results obtained by 2D DTF with each other, the phase difference can be obtained by setting one of them as a conjugate complex number.

Figure 2. The group of waveforms used to evaluate the performance of the proposed method.

Waveform numbers 1 to 5 were visually identified as a group of similar waveforms, whereas waveform numbers 6 to 10 were judged to be dissimilar waveforms. Two of the waveforms, which were processed by a band-pass filter with a bandwidth of 0.1 to 40 Hz to remove low-frequency fluctuations and power supply temperature extinction, used in the evaluation and the results of the wavelet transform are shown in Figures 3 and 4. A Morse wavelet with a symmetry parameter of 3 and a time–bandwidth product of 60 was
used as the analyzing wavelet [20]. The wavelet transform result is shown as frequency magnitude changes on the scalogram; on each scalogram, frequency increases of 10 to 30 Hz and two S-wave peaks can be seen. It can be also seen that the frequency value on the scalogram changed as the wave energy changed with the arrival of the P and S waves. As the two waveforms were similar, the results of the wavelet transforms were also similar. The phase-limited correlation function result calculated using the scalograms in Figures 3 and 4 shows a sharp peak with a value of 0.38 (Figure 5).

Figure 3. Vertical component of the particle velocity and the waveform scalogram, where waveform no.1 in Figure 2 is analyzed. The time-axis of the scalogram corresponds to the left edge of the moving time window used for the wavelet transform calculation.

Figure 4. Vertical component of the particle velocity and the waveform scalogram, where waveform no.2 of Figure 2 is analyzed. The time-axis for the scalogram corresponds to the left edge of the moving time window used for the wavelet transform calculation.
Figure 5. Phase-only correlation function calculated for the two similar waveforms (shown in Figures 3 and 4), where similarity was evaluated by visual inspection. The n1 and n2 axes correspond to the shift lags in the time- and frequency-axis directions, respectively.

Figure 6 shows the waveform of a microearthquake that does not resemble the two waveforms in the above evaluation and the wavelet transform results. This waveform has a low signal-to-noise ratio, and the P and S waves are not clear. The phase-only correlation function result calculated between the scalograms shown in Figures 3 and 6 (i.e., two waveforms visually judged to be dissimilar) shows several peaks with low values (Figure 7). This result indicates that the similarity between these two waveforms is lower compared with that between the waveforms shown in Figures 3 and 4.

Figure 6. Vertical component of the particle velocity and the waveform scalogram of a waveform visually judged to be dissimilar to the waveforms shown in Figures 3 and 4.
The wavelet transform does not compute the scalogram of the low-frequency components at the ends of the waveform in the time window. This region increased from the center of the time window to the end. If the energy of the waveform is concentrated in the low-frequency region, the scalogram at the lower frequency portion will be missing at both ends of the time window. In this sense, the sampling frequency and the frequency range of the dominant frequency components of the waveform are important. The influence of the sampling frequency on the estimation of the scalogram, as well as the detectability of this method, have not been investigated using real waveform data. In the presented paper, the main energy of the wave was distributed within 15-30 Hz under the sampling frequency of 100Hz. Therefore, the scalogram of that region could be estimated. In the analysis of this paper, for the part where the scalogram was not calculated, the value was set to zero and the POC was calculated.

5. Application to a Group of Waveforms

The proposed method was applied to the group of waveforms shown in Figure 2, where the seismic events were induced for several months at a depth of about 10km just below the Ohokura artificial lake. The analyses presented below were performed using MATLAB. The POC function was calculated for all combinations of the two waveforms. Figure 8 shows the matrix of the peak values calculated for each combination. The peak POC function value between waveform numbers 1 and 2 was 0.38, as described above. The peak values calculated between the combinations of waveform numbers 1 to 5 with waveform numbers 6 to 10 had relatively lower values, and the peak values of the combinations of waveform numbers 6 to 10 also had low values. This result shows that a group of similar waveforms could be extracted from arbitrarily mixed waveform data.
### Figure 8. Matrix of peak values calculated between all combinations of two waveforms among waveform numbers 1 to 10.

![Matrix of peak values](image)

### Figure 9. Matrix of peak values calculated between all combinations of two waveforms among waveform numbers 1 to 10, where FFT was used instead of wavelet transform.

![Matrix of peak values](image)

### Figure 10 shows the results when FFT is used instead of WT. Waveforms 1 to 5 were similar to each other. The average POC values of the similar waveforms were 0.41 for WT and 0.39 for FFT (the average of the values in the bold lines, respectively). On the other hand, the waveforms of nos. 6 to 10 were not similar to each other. The average value of POC between waveforms 1-5 and 6-10 was 0.10 for WT and 0.16 for FFT (the average of the values in the double line and numbers are shown in italic, respectively). These results show that using WT results in larger values of POC between similar waveforms than using FFT, and smaller values of POC when the waveforms are not similar. These results show that using WT resulted in larger values of POC between similar waveforms than using FFT, and smaller values of POC when the waveforms were not similar. This is because WT is highly time-resolved, so small changes in the waveform with respect to time are reflected in the scagram, and small differences between the waveforms are also reflected in the POC value. In this way, POC using WT can detect more similar waveforms and classify them quantitatively.

### Figure 11 shows the results when the method was applied to 30 waveforms with waveform numbers 1 to 10. Waveforms with waveform numbers 1 to 5 were similar and classified as 1 group. Waveforms with waveform numbers 6 to 10 were classified as another group. The results of the cluster analysis conducted by the method used in [8], where an agglomerative hierarchical clustering was introduced and the Ward method was used to calculate the distance between the pairs of observables. The distance here is represented by the maximum values of the POC function. In these results, groups of similar and dissimilar waveforms are distinguished. Mutual distances are smaller between
waveforms with larger peak POC function values, and waveforms with smaller mutual distances are grouped together. Figure 11 shows the results when the method was applied to 30 waveforms, including the 10 waveforms in Figure 2. It can be seen that similar waveforms with waveform numbers 1 to 5 were classified as one group. Thus, this method can detect similar waveforms even when the number of waveforms increases.

Figure 9. Matrix of peak values calculated between all combinations of two waveforms among waveform numbers 1 to 10, where FFT was used instead of wavelet transform.

Figure 10. Waveform classification results by cluster analysis.

Figure 11. Waveform classification results by cluster analysis, where the method was applied to 30 waveforms.

6. Conclusions

A method for the detection of similar waveforms was developed, in which the phase-only correlation function was determined by using scalograms obtained by wavelet transform. The proposed method was applied to a group of microearthquakes that included both similar and dissimilar waveforms. The result showed that the degree of similarity was expressed by the peak value of the phase-only correlation function. Thus, the waveforms could be classified into groups of similar or dissimilar waveforms, in the same way as in previous research, in which the phase-only correlation function was determined using Fourier transform results. Compared with the Fourier transform results, the scalograms
calculated by the wavelet transform had a higher time resolution in the time–frequency domain, allowing small waveform changes to be expressed. As a result, it has become possible to judge the similarity and dissimilarity of waveforms more accurately and quantitatively. In addition, it has also been shown that the waveforms can be automatically classified into similar waveform groups when combined with the agglomerative hierarchical cluster analysis method. Therefore, this method would enable better classification of waveforms of microearthquakes induced for geothermal reservoir evaluation, as well as for the detection of slow-slip seismic events using a reference waveform to find similar slow-slip events.

**Funding:** A part of this work was supported by JSPS KAKENHI Grant-in-Aid for Challenging Exploratory Research (19K22149).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data sharing not applicable.

**Acknowledgments:** I thank the National Research Institute for Earth Science and Disaster Prevention (NIED) for providing the earthquake waveform data, which were downloaded from the NIED website (http://www.hinet.bosai.go.jp/, accessed on 1 December 2017).

**Conflicts of Interest:** The author declares no conflict of interest.

**References**

1. Poupinet, G.; Ellsworth, W.L.; Frechet, J. Monitoring Velocity Variations in the Crust Using Earthquake Doublets: An Application to the Calaveras Fault, California. *J. Geophys. Res. Atmos.* **1984**, *89*, 5719–5731. [CrossRef]
2. Evans, K.; Moriya, H.; Niitsuma, H.; Jones, R.H. Microseismicity and permeability enhancement of hydrogeologic structures during massive fluid injections into granite at 3 km depth at the Soultz HDR site. *Geophys. J. Int.* **2005**, *160*, 389–412. [CrossRef]
3. Deichmann, N.; Garcia-Fernandez, M. Rupture geometry from high-precision relative hypocenter locations of microearthquake clusters. *Geophys. J. Int.* **1992**, *110*, 501–517. [CrossRef]
4. Dodge, D.A.; Beroza, G.C. Detailed observations of California foreshock sequences: Implications for the earthquake initiation process. *J. Geophys. Res.* **1996**, *101*, 22371–22392. [CrossRef]
5. Dodge, D.; Beroza, G.C.; Ellsworth, W.L. Foreshock sequence of the 1992 Landers, California, earthquake and its implications for earthquake nucleation. *J. Geophys. Res.* **1995**, *100*, 9865–9880. [CrossRef]
6. Waldhauser, F.; Ellsworth, W.L. A double-difference earthquake location algorithm: Method and application to the Northern Hayward Fault, California. *Bull. Seismol. Soc. Am.* **2000**, *90*, 1353–1368. [CrossRef]
7. Moriya, H.; Nakazato, K.; Niitsuma, H.; Baria, R. Detailed fracture system of the Soultz-sous-Forets HDR field evaluated using microseismic multiplet analysis. *Pure Appl. Geophys.* **2002**, *159*, 517–541. [CrossRef]
8. Moriya, H. Phase-only correlation of time-varying spectral representations of microseismic data for identification of similar seismic events. *Geophysics* **2011**, *76*, WC37–WC45. [CrossRef]
9. Brown, J.R.; Beroza, G.C.; Shelly, D.R. An autocorrelation method to detect low frequency earthquakes within tremor. *Geophys. Res. Lett.* **2008**, *35*, L16305. [CrossRef]
10. Castro de Matos, M.; Osorio, P.L.M.; Johann, P.R.S. Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. *Geophysics* **2007**, *72*, 9–21. [CrossRef]
11. Marroquin, I.D.; Brault, J.J.; Hart, B.S. A visual data-mining methodology for seismic facies analysis: Part 1—Testing and comparison with other unsupervised clustering methods. *Geophysics* **2009**, *74*, 1–11. [CrossRef]
12. Ehret, B. Pattern recognition of geophysical data. *Geoderma* **2010**, *160*, 111–125. [CrossRef]
13. Schaff, D.P. Semiempirical statistics of correlation-detector performance. *Bull. Seismol. Soc. Am.* **2008**, *98*, 1495–1507. [CrossRef]
14. Rowe, C.A.; Astar, R.C.; Borchers, B.; Young, C.J. An automatic, adaptive algorithm for refining phase picks in large seismic data sets. *Bull. Seismol. Soc. Am.* **2002**, *92*, 1660–1674. [CrossRef]
15. Moriya, H.; Niitsuma, H.; Baria, R. Multi-clustering analyses reveals structural details within the seismic cloud at the Soultz geothermal field, France. *Bull. Seismol. Soc. Am.* **2003**, *93*, 1606–1620. [CrossRef]
16. Arrowsmith, S.J.; Eisner, L. A technique for identifying microseismic multiplts and application to the Valhall field, North Sea. *Geophysics* **2006**, *71*, V31–V40. [CrossRef]
17. Mukuhira, Y.; Polianinkov, O.V.; Fehler, M.; Moriya, H. Low SNR microseismic detection using direct P-wave arrival polarization. *Bull. Seismol. Soc. Am.* **2020**, *110*, 3115–3129. [CrossRef]
18. Ito, K.; Nakajima, H.; Kobayashi, K.; Aoki, T.; Higuchi, T. A fingerprint matching algorithm using phase-only correlation. *IEICE Trans. Fundam.* **2004**, *E87-A*, 682–691.
19. Honsho, C.; Kido, M.; Ichikawa, M.; Ohashi, T.; Kawakami, T.; Fujimoto, H. Application of Phase-Only Correlation to Travel-Time Determination in GNSS-Acoustic Positioning. *Front. Earth Sci.* **2021**, *10*. 
20. Flandrin, P.; Goncalves, P. *From Wavelet to Time-Scale Energy Distributions, (in Recent Advances in Wavelet Analysis)*; Schumaker, L.L., Webb, G., Eds.; Academic Press: Boston, MA, USA, 1994; pp. 309–334.

21. Rioul, O.; Flandrin, P. Time-scale energy distributions, A general class extending wavelet transforms. *IEEE Trans. Signal Process.* **1992**, *40*, 1746–1757. [CrossRef]

22. Soma, N.; Niitsuma, N.; Baria, R. Reflection technique in time-frequency domain using multicomponent acoustic emission signals and application to geothermal reservoirs. *Geophysics* **2002**, *67*, 928–938. [CrossRef]

23. Soma, N.; Niitsuma, N.; Baria, R. Evaluation of subsurface structure at Soultz Hot Dry Rock site by the AE reflection method in time-frequency domain. *Pure Appl. Geophys.* **2002**, *159*, 543–562. [CrossRef]

24. Soma, N.; Niitsuma, N.; Baria, R. Reflection imaging of deep reservoir structure based on three-dimensional hodogram analysis of multicomponent microseismic waveforms. *J. Geophys. Res.* **2007**, *112*, B11303. [CrossRef]

25. Olhede, S.C.; Walden, A.T. Generalized morse wavelets. *IEEE Trans. Signal Process.* **2002**, *50*, 2661–2670. [CrossRef]